Massachusetts Institute of Technology | Handout 34
6.042J/18.062J: Mathematics for Computer Science | March 30, 1998 (typos 4/1/98)
Professor Albert R. Meyer

The Size of Sets

1 Sets of the Same Size: Bijections

Counting is a theme throughout discrete mathematics: how many leaves in a tree, minimal colorings of a graph, trees with a given set of vertices, five-card hands in a deck of fifty-two, consistent rankings of players in a tournament, stable marriages given boy's and girl's preferences, and so on.

A useful technique for determining the size of a set is to set up an exact correspondence between the elements of the set and the elements of some other set of known size. For example, for any finite set S with $n > 0$ elements, we set up a correspondence between subsets of S and binary words of length n. Since every computer scientist knows there are 2^n binary words of length n, she can conclude that there are also exactly 2^n subsets of a set of size n.

To specify the correspondence, suppose the distinct elements of S are $s_1, s_2, \ldots, s_n$. Let each length n binary word $b_1 b_2 \ldots b_n$ correspond to the subset $T(b_1 b_2 \ldots b_n) ::= \{s_i : b_i = 1\}$. So each length n binary word, b, determines a unique subset $T(b) \subseteq S$. It is easy to see that distinct binary words determine distinct subsets: if their ith bits differ, then one of the corresponding subsets contains s_i and the other doesn't — that is, T is an *injection*. Also, every subset corresponds to the word whose ith bit is 1 exactly when s_i is in the subset — that is, T is a *surjection*. So, the mapping $T : \{0,1\}^n \to P(S)$ is a *bijection*.

Remember "bijection" is defined as follows:

Definition 1.1 *Let f be a function with domain A and codomain B, that is, $f : A \to B$. Define f to be an* injection *(also called "one-to-one") iff*

$$f(a) = f(a') \implies a = a'$$

for all $a, a' \in A$. The range *of f is defined as*

$$\mathit{range}(f) ::= \{b \in B : b = f(a) \text{ for some } a \in A\}.$$

Define f to be a surjection *(also called "onto") iff $\mathit{range}(f) = B$. Define f to be a* bijection *iff f is both an injection and a surjection.*

Exact correspondences between infinite sets can be useful as well. For example, how does a single serial processor simulate a Java program which spawns a growing number of parallel threads? — by "dovetailing" the computation of the threads. Namely, at each interval of computation, the serial processor picks a single thread to simulate. Of course it won't do for the serial processor simply to simulate some fixed number of threads in a cycle, for then threads not in the cycle never get a turn. For a dovetailing strategy to work properly, every one of the growing number of threads must get a growing number of turns as time goes on.

More precisely, we require that for every pair of nonnegative integers (n, m), there is an number, $f(n, m)$ such that the nth thread gets its mth turn in the $f(n, m)$th time interval, and in every time interval exactly one thread gets a turn. So formally, a proper strategy is a bijection $f : \mathbb{N}^2 \to \mathbb{N}$. Since a thread can't have its $m + 1$st turn before it mth, the bijection f also has to be increasing in its second argument that is, $f(n, m_1) < f(n, m_2)$ whenever $m_1 < m_2$.

There are many ways to define such a bijection. For example, let $f(n, m)$ be the mth largest nonnegative integer which is divisible by 2^n and not by 2^{n+1} . A more uniform strategy arises by letting the processor simulate threads in a cycle, but to add another thread to the cycle each time around. The bijection determined by this strategy turns out to be a polynomial $f(n, m) ::= ((n + m)^2 + m - n)/2$.

So the correspondence above demonstrated

Theorem 1.2 *There is a bijection* $f : \mathbb{N}^2 \to \mathbb{N}$.

Exercise: Let $f(n, m) = ((n + m)^2 + m - n)/2$.

(a) List the eleven pairs of nonnegative integers (n, m) such that $f(n, m) \leq 10$.

(b) Describe in general how, given $k \in \mathbb{N}$, to find a pair $(n, m) \in \mathbb{N}^2$ such that $f(n, m) = k$.

(c) Verify that f is a bijection.

In these notes we'll demonstrate that using bijections to specify the size of infinite sets yields a sensible theory of infinite set sizes — yes there will be more than one size of infinity — but it's important to notice from the start that there will be some dramatic contrasts between size properties of finite and infinite sets.

For example, a proper subset of a finite set is always strictly smaller than the whole finite set. On the other hand, the set, $\mathbb{Z}$ of all integers is the "same size" as the set, $\mathbb{N}$, of all natural numbers, even though $\mathbb{N}$ is a proper subset of $\mathbb{Z}$. Namely, there is a bijection betweeen the natural numbers and all the integers. For example, the function $f : \mathbb{N} \to \mathbb{Z}$ such that $f(n) ::= -n/2$ when n is even, and $f(n) = (1 + n)/2$ when n is odd, is a bijection. In fact we shall show below that this behavior determines infiniteness: a set is infinite iff it has an bijection into proper subset of itself.

While it's pretty clear what is meant by "the number of elements" in a finite set, it isn't so clear for infinite sets. In fact, the definition of "finite" depends on being able to count the number of elements in a set in the first place. So to avoid circularity, and especially to be sure we're on firm ground when we handle infinite sets as well as finite ones, let's be very careful in our definitions about the size of sets. Before we define precisely what the size of a set may be, let's define the more fundamental notion of "same size" we've discussed above.

Definition 1.3 *We write $A \operatorname{inj} B$ to indicate that there is an injection from A to B. Similarly, we write $A \operatorname{sur} B$ iff there is a surjection from A to B. Finally, we write $A \operatorname{bij} B$ to indicate that there is a bijection from A to B.*

Two sets A, B are said to be of the same cardinality *iff $A \operatorname{bij} B$.*

2 Finite Sets

One way to demonstrate that a set is finite, and in fact has n elements, is to find a bijection between it and some standard set of n elements. We'll use the set $\{0, 1, \ldots, n-1\}$, indicated by the notation[1] "$[0, n)$," as the standard set of size $n \in \mathbb{N}$.

Definition 2.1 *For $n \in \mathbb{N}$, define $[0, n) ::= \{i \in \mathbb{N} : i < n\}$.*

So by definition, $[0, 0) = \emptyset$.

Definition 2.2 *A set, S, is said to be* finite *iff $S \operatorname{bij} [0, n)$ for some $n \in \mathbb{N}$. If A is a finite set, the* cardinality*, $|A|$, of A, is the least $n \in \mathbb{N}$ such that $A \operatorname{bij} [0, n)$.*

So if A is finite, then $A \operatorname{bij} [0, |A|)$ by definition of $|A|$.

As a initial, simple "sanity check" of our technical definition of cardinality, let's verify that a set of size zero is empty:

Lemma 2.3 $|A| = 0 \iff A = \emptyset$.

Proof. ($\Rightarrow$) By definition, if $|A| = 0$, then there is a bijection $f : A \to [0, 0)$. But $[0, 0) = \emptyset$, and the only function[2] from a set into the empty set is the function with empty domain, that is, $A = \emptyset$.

[1] This notation follows a common convention that when an interval is denoted by its endpoints, square left bracket "[" indicates that the smaller endpoint is included in the interval, and ordinary right parenthesis ")" indicates that the larger endpoint is not included in the interval.

[2] Remember, a function $f : A \to B$ is, by definition, *total.* That is, $f(a)$ is defined and $f(a) \in B$ for *all* $a \in A$.

($\Leftarrow$) The "empty" function from $\emptyset \to \emptyset$ satisfies the definition of bijection. Since $\emptyset = [0, 0)$, we have a bijection from the empty set to $[0, 0)$. So according to the definition of cardinality, $|\emptyset| = 0$. ■

Since we are using the set of nonnegative integers less than n as the standard in the definition of cardinality n, we better be sure that in fact this set has cardinality n. This property seems obvious, and it certainly is utterly familiar, but it does does not quite follow immediately.

The definition says that the cardinality of the set of nonnegative integers less than n is the *least* $m \in \mathbb{N}$ such that there is a bijection from this set to $[0, m)$. This obviously implies that $|[0, n)| \leq n$. But is it possible that $|[0, n)| < n$? In other words, is it possible that there is a bijection between the nonnegative integers less than n and those less than m for some $m < n$? Well of course not! In fact, the following stronger fact indeed can be proved:

Lemma 2.4 $[0, m) \operatorname{inj} [0, n)$ *iff* $m \leq n$.

The proof is by induction on n and appears at the beginning of Appendix B.

The remaining basic properties of finite cardinality are summarized in:

Theorem 2.5 *If B is a finite set, then*

$$A \operatorname{inj} B \iff A \text{ is finite and } |A| \leq |B|.$$

In particular,

$$A \operatorname{bij} B \iff A \text{ is finite and } |A| = |B|.$$

Corollary 2.6 *Any subset of a finite set is finite.*

Proof. If $A \subseteq B$, the "inclusion" function $\mathrm{id}_{AB} : A \to B$ where $\mathrm{id}_{AB}(a) ::= a$, is an injection, so $A \operatorname{inj} B$. So if B is finite, then by Theorem 2.5, A is also finite. ■

The statements in Lemma 2.4 and Theorem 2.5 are pretty obvious, and it's hard to imagine that anyone would be upset if we simply took them as axioms from the start. But these properties are in fact logical consequences of the definition of cardinality, injections and bijections. We've placed their proofs in Appendix B to illustrate this kind of reasoning from first principles. The proofs break down into a series of elementary Lemmas, but actually take several pages to work out. It isn't necessary to look at Appendix B to understand the rest of these notes.

We assume below that the reader knows the basic properties of bijections, injections, and surjections. The main properties used are:

- Composition of functions is an associative operation:

$$f \circ (g \circ h) = (f \circ g) \circ h.$$

- The relation, bij, is an equivalence relation on sets.
- The relations inj and sur are transitive.
- $A \operatorname{inj} B$ iff $B \operatorname{sur} A$.

Appendix A provides a review of the definitions and proofs of these properties.

3 Inclusion-Exclusion

Now we can exhibit yet another proof of the principle of Inclusion-Exclusion. This one is kind of elegant.

Lemma 3.1 *If A and B are disjoint finite sets, then $|A \cup B| = |A| + |B|$.*

Proof. Since A, B are finite, there are bijections $f : A \to [0, |A|)$ and $g : B \to [0, |B|)$. Since A and B are disjoint, we can define a function h on $A \cup B$ where $h(a) ::= f(a)$ for $a \in A$, and $h(b) ::= g(b) + |A|$ for $b \in B$. Then $h : A \cup B \to [0, |A| + |B|)$ is a bijection, so $|A \cup B| = |A| + |B|$ by Theorem 2.5. ∎

Lemma 3.2 *For any set, A, and finite set, B,*

$$|B - A| = |B| - |A \cap B|.$$

Proof. For any sets A, B, we have $B = (B - A) \cup (A \cap B)$. Moreover, $(B - A)$ and $A \cap B$ are disjoint subsets of B. Now if B is finite, then by Corollary 2.6, so are $(B - A)$ and $A \cap B$. Hence, by the previous Lemma 3.1,

$$|B| = |B - A| + |A \cap B|.$$

∎

Theorem 3.3 *(Inclusion-Exclusion) For finite sets A, B,*

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Proof. For any sets A, B, we have $A \cup B = A \cup (B - A)$; also A and $B - A$ are disjoint. So for finite A, B, the two previous Lemmas 3.1 and 3.2 imply

$$|A \cup B| = |A| + |B - A| = |A| + (|B| - |A \cap B|).$$

■

Notation: We write $A \subset B$ to indicate that A is a proper subset of B, that is, $(A \subseteq B)$ and $A \neq B$.

Corollary 3.4 *If B is a finite set and $A \subset B$, then A is finite and $|A| < |B|$.*

Proof. Since B is finite, so are A and $B - A$ by Corollary 2.6. Since $A \subset B$, we have $B - A \neq \emptyset$, so $|B - A| > 0$ by Lemma 2.3. Also, $A \cap B = A$. Now by Lemma 3.2,

$$0 < |B - A| = |B| - |A \cap B| = |B| - |A|.$$

■

4 Infinite Sets

The most familiar example of an infinite set is probably the nonnegative integers, $\mathbb{N}$. Let's check that $\mathbb{N}$ is infinite according to our definitions. Actually, we'll prove something slightly stronger: a set is infinite iff there is an injection of the nonnegative integers into the set. This means that $\mathbb{N}$ is the "smallest" possible infinite set.

Definition 4.1 *A set is said to be* infinite *iff it is not finite.*

Theorem 4.2 *A set A is infinite set iff $\mathbb{N}\,\mathrm{inj}\,A$.*

Proof. ($\Rightarrow$) We define a sequence of distinct elements $a_1, a_2, \ldots, a_n, \ldots$ recursively. Since A is infinite, it cannot be empty, so there is an element in A. Let a_1 be this element. To define a_{n+1}, let $f : [0, n) \to A$ be defined by the rule $f(k) ::= a_{k+1}$ for $0 \leq k < n$. Since the a_i for $i \leq n$ are distinct, we know that f is an injection. Now if $\mathrm{range}(f) = A$, then f would be a bijection, contradicting the fact that A is infinite. So there is an element in $A - \mathrm{range}(f)$; let a_{n+1} be this element.

Now define $g : \mathbb{N} \to A$ by the rule that $g(m) ::= a_{m+1}$. Since the elements of the sequence of a_i's are distinct, g is an injection. Hence, $\mathbb{N}\,\mathrm{inj}\,A$.

($\Leftarrow$) Suppose for contradiction that A is finite, so $A\,\mathrm{inj}\,[0, |A|)$. But $\mathbb{N}\,\mathrm{inj}\,A$, hence $\mathbb{N}\,\mathrm{inj}\,[0, |A|)$ because inj is transitive (*cf.* Appendix A.) That is, there is an injection

$f : \mathbb{N} \to [0, |A|)$. Now the restriction of f to the domain $[0, |A|+1) \subset \mathbb{N}$ is an injection from $[0, |A|+1)$ to $[0, |A|)$, contradicting Lemma 2.4. ■

It's important to notice that we have only said that there is an *inj*ection from the nonnegative integers into any infinite set, *not* that there is a *bij*ection. In the 1870's, the mathematician Georg Cantor made the dramatic discovery that not every infinite set can be put in exact correspondence with the nonnegative integers — there may not be a bijection. We'll prove this in Section 8.

Here is another characterization of infiniteness we mentioned earlier.

Theorem 4.3 *A set is infinite iff there is an injection from the set to a proper subset of itself.*

Proof. ($\Leftarrow$) We have an injection $f : A \to A$ with at least one element $a \in A - \text{range}(f)$. Define $g : \mathbb{N} \to A$ recursively: $g(0) ::= a$, and $g(n+1) ::= f(g(n))$. We claim g is an injection, which implies that A is infinite by the previous Theorem 4.2.

Suppose to the contrary that g is not an injection. Then there must be a least $j_0 \in \mathbb{N}$ such that $g(j_0) = g(j_1)$ for some $j_1 > j_0$. Note that $g(j) \in \text{range}(f)$ for $j > 0$ by definition. Since $g(0) = a \notin \text{range}(f)$, we know in particular that $g(0) \neq g(j)$ for all $j > 0$. So $j_1 > j_0 > 0$. Now we have $f(g(j_0 - 1)) = g(j_0) = g(j_1) = f(g(j_1 - 1))$, and since f is an injection, we conclude that $g(j_0 - 1) = g(j_1 - 1)$, contradicting the minimality of j_0. ■

Exercise. Prove the left-to-right direction of Theorem 4.3. *Hint*: Corollary 3.4.

Negating both sides of the "iff" in Theorem 4.3 gives an immediate, but informative Corollary:

Corollary 4.4 *A set is finite iff every injection from the set to itself is also a bijection.*

5 Schröder-Bernstein Theorem

If the sizes of two finite sets are each less than or equal to the other, then the sets are the same size. In other words, the "smaller than" relation is *symmetric*, at least on finite sets. We wouldn't be happy with our interpretation of "smaller than" unless it was symmetric on infinite sets as well as finite ones, and it is.

We've been interpreting "A is smaller than B" as $A \,\text{inj}\, B$. So we can formulate the Symmetry property precisely as:

Theorem 5.1 *(Schröder-Bernstein) [$A \,\text{inj}\, B$ and $B \,\text{inj}\, A$] iff $A \,\text{bij}\, B$.*

Exercise. Show that the Schröder-Bernstein Theorem for finite A, B follows directly from Theorem 2.5.

The direct proof of the previous exercise depends heavily on properties of finite sets, so it will not extend to infinite sets. Instead, we present the original elegant proof, which works equally well for finite and infinite sets.

Proof. ($\Leftarrow$) If $A \operatorname{bij} B$ then $A \operatorname{inj} B$ and $A \operatorname{sur} B$ by definition. But $A \operatorname{sur} B$ iff $B \operatorname{inj} A$.

($\Rightarrow$) Suppose there is an injection $f : A \to B$ and an injection $g : B \to A$. We show how to piece together portions of f and g into a bijection $h : A \to B$.

We'll assume for convenience that A and B are disjoint (see the Exercise below). Now we treat the elements of A and B as the vertices of a bipartite digraph. (The graphs we've considered up to now have been finite. This one will be infinite whenever A or B is infinite.) The edges are

$$E ::= \{(a, f(a)) : a \in A\} \cup \{(b, g(b)) : b \in B\}.$$

The graph is illustrated in Figure 1. Two connected elements are pictured in Figure 1 as having a zig-zag path between them.

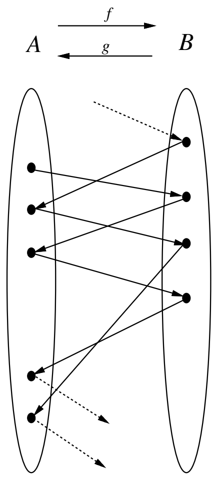

Figure 1: *Sets A and B with injections $f : A \to B$ and $g : B \to A$.*

Notice that because f and g are functions, each vertex has out-degree one. This implies there is a unique infinite path starting at each vertex. Also, a vertex has in-degree one iff it is in the range of f or g; otherwise it has in-degree zero. This implies there is a unique maximal path, which may be finite or infinite, into each vertex. So we can partition the vertices into four kinds of components, with no edges between distinct components:

1. a one-way infinite simple path starting at a vertex in A,

2. a one-way infinite simple path starting at a vertex in B,
3. a simple cycle,
4. a simple path which is infinite in both directions (all its vertices are degree two).

These four possibilities are illustrated in Figure 2.

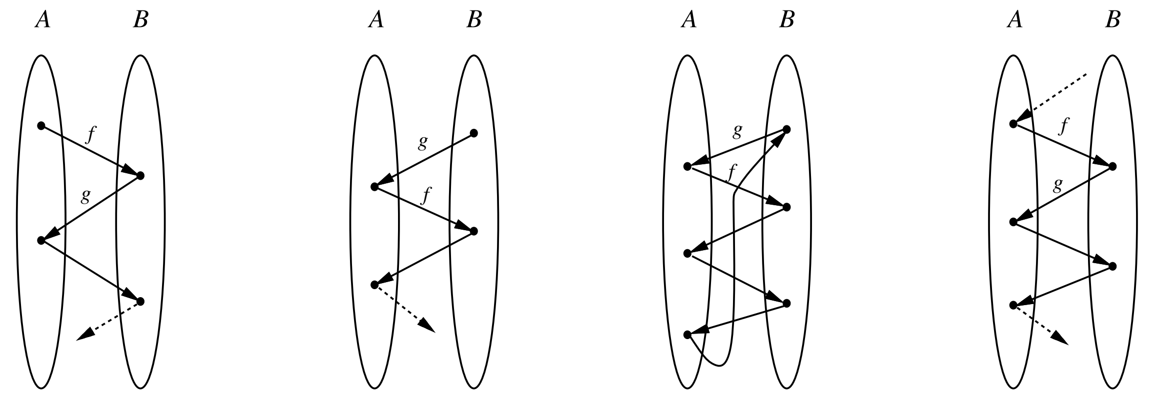

Figure 2: *The four kinds of connected components.*

In case 1 of a one-way path starting in A, the function f defines a bijection from the A-vertices on the path to the B-vertices on the path. Similarly, in case 2 of a one-way path starting in B, the function g defines a bijection from the B-vertices to the A-vertices on the path. In cases 3 and 4 of a finite cycle or two-way infinite path, both f and g define bijections between the A-vertices and the B-vertices on the path.

So we simply define h component by component. Except for case 2, the function f defines on bijection from A-vertices to B-vertices in a component, so define $h(a) ::= f(a)$ when a is not in a component with a start vertex in B. In case 2, g defines a bijection from the B-vertices to the A-vertices on the path, so its inverse g^{-1} is a uniquely defined bijection from the A vertices to the B-vertices. So we simply define $h(a) ::= g^{-1}(a)$ when a is in a component with a start vertex in B.

Since h is a bijection from the A-vertices to the B-vertices in each component, and the components are disjoint, h is the desired bijection from A from to B. ■

Exercise: Prove that for any sets A, B, there are disjoint sets A', B' such that A' bij A and B' bij B.

6 Countable Sets

The Schröder-Bernstein Theorem makes it much easier to prove that there are bijections between sets. For example, let $\mathbb{Q}$ denote the rational numbers. Define $h : \mathbb{Z}^2 \to \mathbb{Q}$ so that $h(m, n) ::= m/n$ when $n \neq 0$, and $h(m, 0) ::= 1$. Clearly h is a surjection, so $\mathbb{Z}^2 \operatorname{sur} \mathbb{Q}$, and hence

$$\mathbb{Q} \operatorname{inj} \mathbb{Z}^2.$$

Conversely, we know $\mathbb{N} \operatorname{bij} \mathbb{Z}$, which implies

$$\mathbb{Z}^2 \operatorname{bij} \mathbb{N}^2.$$

Also, we showed at the outset in Theorem 1.2 that

$$\mathbb{N}^2 \operatorname{bij} \mathbb{N},$$

so altogether we conclude

$$\mathbb{Q} \operatorname{inj} \mathbb{N}.$$

On the other hand, $\mathbb{N} \subset \mathbb{Q}$, so $\mathbb{N} \operatorname{inj} \mathbb{Q}$. Therefore, by Schröder-Bernstein

Theorem 6.1 $\mathbb{N} \operatorname{bij} \mathbb{Q}$.

The rationals are densely spread over the entire real line, and so appear more numerous than the nonnegative integers, which are sparsely spread over half the line. Nevertheless, we have deduced, indirectly but undeniably, that there is an exact correspondence between these sets, so the rationals are not in fact more numerous.

Definition 6.2 *A set A is* countably infinite *iff* $\mathbb{N} \operatorname{bij} A$. *It is* countable *iff it is finite or countably infinite.*

For example, the results above show that $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{Z}^2, \mathbb{N}^2$ are all countably infinite.

The reason for the name is that we can "count" one after another through the elements of a countable set so that every element of the set is eventually counted. That is, if f is a bijection from the nonnegative integers to a set, we can count each of the elements of the set in the order $f(0), f(1), \ldots$. Other examples of countably infinite sets are

$$\begin{aligned} \text{even numbers} &= \{0, 2, 4, 6, \ldots\} \\ \text{powers of two} &= \{1, 2, 4, 8, \ldots\} \\ \text{primes} &= \{2, 3, 5, 7, 11, \ldots\} \end{aligned}$$

The bijection we described above between the nonnegative integers and the integers, namely, $f(n) ::= -n/2$ when n is even, and $f(n) = (1+n)/2$ when n is odd, corresponds to counting the integers in the order

$$0, 1, -1, 2, -2, 3, -3, 4, -4, \ldots.$$

This is the typical way to prove that a set is countable — just list the elements in the order they should be counted in. This implicitly defines a bijection with $\mathbb{N}$. Note that we only have to prove that a bijection f exists; it needn't have a simple description like the f above. It needn't even be a function that we can actually compute.

This method of proving a set countable by listing elements has one potential pitfall. As an example, the following listing of the integers is incorrect:

$$0, -1, -2, -3, \ldots, 1, 2, 3, \ldots.$$

The problem is that the "…" in the middle of the list represents the infinite list of negative integers from -4 on down. As a result, we never actually finish counting down, so we never get around to starting a count of the positive integers. More precisely, if $g(n)$ is defined to be the nth number in this list, then $g(n)$ is not positive for any $n \in \mathbb{N}$. This means that the function g defined is not a bijection to *all* the integers, only to the nonpositive ones.

This counting approach offers a direct way, for example, to prove that the positive rationals are countable. As a first attempt, we might try to count the positive rational numbers in the following order:

$$\frac{1}{1}, \frac{2}{1}, \frac{3}{1}, \ldots, \frac{1}{2}, \frac{2}{2}, \frac{3}{2}, \ldots, \frac{1}{3}, \frac{2}{3}, \frac{3}{3}, \ldots$$

A fixable problem with this list is that each rational is counted many times. We need to omit from the list all numbers that have already appeared. For example, 2/2 and 3/3 should be omitted, since the rational number 1 already appeared as 1/1.

The killer problem is that we used "…" again in the middle of the list to represent an infinite number of elements. This means that we will never reach, say, 1/2; that is, the nth element of the list above is the integer n. There is no $n \in \mathbb{N}$ such that 1/2 — or any other noninteger rational number for that matter — appears as the nth element. So we have not counted all the positive rationals and have not implicitly defined the required bijection. We need to list the rational numbers in a completely different order to make the proof work.

Having recognized the problem, there is a simple solution: list the positive rational numbers in the order indicated in Figure 3, omitting numbers that have already appeared in the list. This proves that the positive rationals are countable.

The ordering technique used in the proof and indicated in Figure 3 is just the kind of dovetailing described earlier for simulation of Java threads. We can also use dovetailing to prove that the product of any two countable sets is countable:

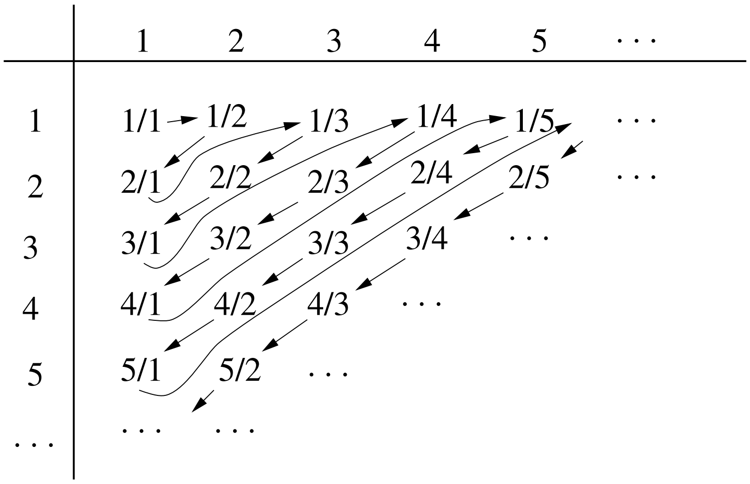

Figure 3: *A dovetailing order to count the positive rationals.*

Theorem 6.3 *If A and B are countable, then $A \times B$ is countable.*

Proof. Since A and B are countable, we can list the elements of A as $\{a_1, a_2, a_3, \ldots\}$ and the elements of B as $\{b_1, b_2, b_3, \ldots\}$. We can obtain a listing of $A \times B$ by dovetailing:

$$
\begin{array}{l}
(a_1, b_1), \\
(a_1, b_2), (a_2, b_1), \\
(a_1, b_3), (a_2, b_2), (a_3, b_1), \\
(a_1, b_4), (a_2, b_3), (a_3, b_2), (a_4, b_1), \\
\quad \ldots \qquad \ldots
\end{array}
$$

If either or both sets A and B are finite, omit from the listing all pairs containing non-existent elements a_i or b_j. ∎

Note that the ordering of elements in this proof corresponds to the ordering in Figure 3. In fact, the way we showed that the positive rationals are countable was a special case of the above proof.

Exercise. Prove that A is countable iff $A \operatorname{inj} \mathbb{N}$.

Exercise. Prove that if $A_0, A_1, \ldots$ is a sequence of countable sets, then so is $\cup_{i=0}^{\infty} A_i$.

Exercise. Suppose $A_0 \operatorname{bij} A_1$ and $B_0 \operatorname{bij} B_1$.

(a) Give a counterexample to the claim that

$$A_0 \cup B_0 \ \operatorname{bij} \ A_1 \cup B_1.$$

(b) Show that the claim holds if at least one of A_0 and B_0 is infinite.

7 The Continuum

The real numbers, $\mathbb{R}$, are sometimes, dramatically, referred to as the *continuum*. We'll examine how they compare in size to some other sets.

Every real number is the least upper bound of the set of rationals less than it. That is, $\text{lub} : P(\mathbb{Q}) \to \mathbb{R}$ is a surjection, and therefore $\mathbb{R} \,\text{inj}\, P(\mathbb{Q})$. But we showed in Section 6 that $\mathbb{N} \,\text{bij}\, \mathbb{Q}$, which implies $P(\mathbb{N}) \,\text{bij}\, P(\mathbb{Q})$, and altogether we conclude

$$\mathbb{R} \,\text{inj}\, P(\mathbb{N}).$$

On the other hand, it's easy to find a surjection from the reals to the sets of nonnegative integers. For example, let 1^n denote a sequence of n consecutive 1's. Then the function $f : \mathbb{R} \to P(\mathbb{N})$ where

$$f(r) ::= \{n \in \mathbb{N} : \text{ the digits } 21^n2 \text{ occur in the infinite decimal expansion of } r\}$$

is a surjection. So $\mathbb{R} \,\text{sur}\, P(\mathbb{N})$, implying

$$P(\mathbb{N}) \,\text{inj}\, \mathbb{R}.$$

Therefore, by Schröder-Bernstein

Theorem 7.1 $\mathbb{R} \,\text{bij}\, P(\mathbb{N})$.

To appreciate the value of the Schröder-Bernstein Theorem, the reader can try directly contructing a bijection between the reals and $P(\mathbb{N})$.

A genuinely remarkable consequence of our definitions is that the real line is the same size as the two-dimensional plane:

Theorem 7.2 $\mathbb{R}^2 \,\text{bij}\, \mathbb{R}$

Proof. By the previous Theorem 7.1, all we need to show is that

$$P(\mathbb{N}) \,\text{bij}\, P(\mathbb{N})^2.$$

But that's easy: map $A \subseteq \mathbb{N}$ to the pair of sets

$$(\{n : 2n \in A\}, \{n : 2n + 1 \in A\}).$$

■

This Theorem stands in dramatic contrast to a geometric picture of the infinitely thin line with zero area versus the plane with its infinite area. Perhaps this violation of geometric intuition[3] accounts for continued submissions to mathematics journals of crank proofs that there are more points in two dimensions than in one (*cf.* Hodges, Wilfred, "An Editor Recalls Some Hopeless Papers," *Bulletin of Symbolic Logic*, March, 1998, v4,#1).

[3]Geometric intuition can be substantially restored: injections from the plane into the line have to be weird — they can't be even piecewise continuous, for example.

8 Diagonal Arguments

There would be little need for intuition and definitions about countable sets if all sets were countable. Many important sets turn out not to be countable. Sets that are not countable are called *uncountable*.

Theorem 8.1 *The set of real numbers is uncountable.*

This theorem says that there is no way to make a list of all the real numbers. Naturally, one is immediately tempted to try anyway:

$$0,\ 1,\ 0.5,\ 0.75,\ 0.1,\ 0.05,\ 0.001,\ \ldots$$

But wait a minute! All we're doing is listing the reals with finite decimal expansion. That doesn't even include all the rational numbers. Where are we going to stick in $\sqrt{2}, \pi, pi^{-173}, \pi^e$, all other numbers with nonrepeating expansions like

$$0.1010010001000010\ldots,$$

etc.? There just isn't any way to do it.

We'll first work through the proof informally, and then write it up. The argument is by contradiction; we assume for the purpose of contradiction that $\mathbb{R}$ is countable. This means that there exists a bijection $f : \mathbb{N} \to \mathbb{R}$. We can then make a list of all the real numbers that looks something like the one shown below.

```
f(0) =  76.5430000000...
f(1) = 163.7777777777...
f(2) =   2.7878787878...
f(3) =  -1.1121231234...
f(4) =   3.1415926535...
f(5) = -15.3727323928...
      ...etc...
```

The plan is to reach a contradiction by constructing a real number that appears nowhere in the list; this will prove that f is not surjective and contradict the assumption that f is a bijection.

Consider the number defined by the digits on the diagonal starting at the decimal point in the above table, $0.577192\ldots$. Here is the same table with extra spaces inserted to highlight this diagonal sequence:

```
f(0) =  76 .  5 430000000...
f(1) = 163 . 7 7 77777777...
f(2) =   2 . 78 7 8787878...
f(3) =  -1 . 112 1 231234...
f(4) =   3 . 1415 9 26535...
f(5) = -15 . 37273 2 3928...
        ...etc...
```

Now suppose that we add 5 to every digit, working modulo 10. This gives $0.022647\ldots$; call this value r. This quantity r is totally bizarre, but it is a well-defined real number, since we have specified all the digits.

By our construction r differs from $f(0)$ in the first position to the right of the decimal. Similarly, $f(1)$ differs from r in the second decimal position, $f(2)$ differs in the third position, and so forth. In general, $f(n)$ differs from r in the $(n+1)$-st position. So for all n, $f(n)$ is not equal to r. This implies that the real number r does not appear anywhere in our list!

Actually there is an irritating hole in this argument. We have shown that r has a decimal expansion that differs from the decimal expansion of $f(n)$. But real numbers can have more than one decimal expansion: for example, $1 = 0.999\ldots$. So the fact that some expansion of r differs from some expansion of $f(n)$ does not guarantee $r \neq f(n)$! It might actually happen that $r = f(n)$ had we allowed in the construction that the nth digit of r differed from the nth digit of $f(n)$ by 1 or 9. But we were careful not to allow this. Our construction ensures that r actually differs from $f(n)$ by at least $4 \cdot 10^{-(n+1)}$. For example, $1.0002999\ldots$ differs from $1.0007000\ldots$ by $4 \cdot 10^{-4}$.

Now since the reals are not countable, it is not true that $\mathbb{R} \operatorname{inj} \mathbb{N}$, but since the nonnegative integers are a subset of the reals, we have $\mathbb{N} \operatorname{inj} \mathbb{R}$. Now it makes sense to say that the set of nonnegative integers is strictly smaller than the set of reals.

Definition 8.2 *A set A is* strictly smaller *than a set B, in symbols,*

$$A \ll B,$$

iff $A \operatorname{inj} B$ and not $B \operatorname{inj} A$

Here is how the argument above might be written as a formal proof without mentioning the table and its diagonal:

Proof. We must show that there is no bijection $f : \mathbb{N} \to \mathbb{R}$. The proof is by contradiction. Assume for the purpose of contradiction that such a bijection f exists. Let r be the real number with decimal expansion $0.r_0r_1r_2r_3\ldots$ where r_i is 5 + the $(i+1)$-st decimal digit of $f(i)$, taken modulo 10. For all i, $f(i) \neq r$, since the $(i+1)$-st digits of $f(i)$ and r disagree by 5 and so $|r - f(i)| \geq 4 \cdot 10^{-(i+1)}$. This implies that f is not surjective, contradicting the assumption that f is a bijection. ■

This is called a "diagonalization argument", because of the construction based on diagonal entries. Diagonalization arguments are used all the time to prove that there is no bijection between two infinite sets. The generic idea is to assume for the purpose of contradiction that a bijection f exists. Then an element r of the range is constructed that differs a little bit from $f(\alpha)$ for every α in the domain. (In this case, for example, the real number r differs from $f(k)$ in the $(k+1)$-st decimal position for every natural k.) The existence of such an r shows that f is not surjective, contradicting the assumption that f is a bijection. We conclude that no bijection between the two sets exists.

Diagonalization arguments are important in logic and computer science. For example, this type of argument is used to show that there are some problems that can not be solved by a computer no matter how much time you give them. A similar method is used in logic to prove Gödel's Theorem, which says that there are some true theorems for which no proof exists!

The distinction between countable and uncountable sets is so important that it distinguishes entire fields of mathematics. Discrete math is the study of countable sets; calculus, analysis, and continuous math all study uncountable sets. In this course, we will stick to discrete mathematics after today; all sets will be countable.

A Paradox

Here is a set that appears to be both countable and uncountable! This is impossible, so one argument must be incorrect, but which one?

Define S_n to be the set of n-digit decimal numbers. For example:

$$\begin{aligned} S_1 &= \{.0, .1, .2, \ldots, .9\} \\ S_2 &= \{.00, .01, .02, \ldots, .99\} \end{aligned}$$

Now let $S = \bigcup_{n=1}^{\infty} S_n$. Is S countable?

Argument that S is uncountable: We will prove that S is uncountable by the same diagonalization argument that we used to prove that the set of real numbers is uncountable. Assume for the purpose of contradiction that we can list the elements of S in some order $s_0, s_1, s_2, \ldots$. Define $r = .r_0 r_1 r_2 \ldots$ where r_i is 5 + the $(i+1)$-st decimal digit of s_i taken modulo 10. Since S contains decimal numbers of every length, r must be in S. However, r differs from every element of S by at least one digit. Therefore, r is not in S, and we have a contradiction. ■

Argument that S is countable: We can list the elements of S by listing the elements of S_1, followed by S_2, followed by S_3, etc. Since each of these sets is finite, every set is fully listed eventually. ■

The first argument is incorrect; the false statement in the first argument is, "Since S contains decimal numbers of every length, r must be in S." The set S does contain decimal numbers of every *finite* length, but no numbers with an infinite number of non-zero digits to the right of the decimal. For example, the number $1/9 = 0.11111\ldots$ is not contained in any set S_i and therefore is not contained in the union S. As a result, there are many real numbers not in S. For this reason, the statement, "r must be in S" is false. This eliminates the contradiction from the proof by contradiction!

The second argument is correct; the union of a countable number of finite sets is always countable. In fact, in an earlier exercise we already observed that a union of a countable number of countable sets is always countable.

There is a nonpictorial version of the diagonal argument which is both simpler and more general, although maybe not as intuitive.

Lemma 8.3 *For any set S and function $f : S \to P(S)$, let*

$$\mathit{Diag}(f) ::= \{s \in S : s \notin f(s)\}.$$

Then $\mathit{Diag}(f) \notin \mathit{range}(f)$.

Proof. By contradiction. Suppose that $\mathrm{Diag}(f) \in \mathrm{range}(f)$, i.e., $\mathrm{Diag}(f) = f(s_0)$ for some $s_0 \in S$. Then by definition of $\mathrm{Diag}(f)$,

$$s \in f(s_0) \iff s \in \mathrm{Diag}(f) \iff s \notin f(s)$$

for all $s \in S$. So letting $s = s_0$, we conclude

$$s_0 \in f(s_0) \iff s_0 \notin f(s_0),$$

a contradiction. ■

Theorem 8.4 *(Cantor) For every set S,*

$$S \ll P(S).$$

Proof. The singleton function $g : S \to P(S)$ where $g(s) ::= \{s\}$, is an injection, so $S \operatorname{inj} P(S)$. By Lemma 8.3, it is not the case that $S \operatorname{sur} P(S)$, so it must not be the case that $P(S) \operatorname{inj} S$. ∎

So Cantor's Theorem implies that we can keep finding bigger infinities:

$$\mathbb{N} \ll P(\mathbb{N}) \ll P(P(\mathbb{N})) \ll P(P(P(\mathbb{N}))) \ll \ldots.$$

But there are more: let $Q ::= \cup_{i=0}^{\infty} P^i(\mathbb{N})$, then every set $P^i(\mathbb{N}) \subset Q \ll P(Q)$, so $P(Q)$ is strictly larger than every set in the sequence!

9 Set Theory

Another important property we would expect of a "smaller than" relation is Totality, which we can state precisely as

Claim 9.1 *(Totality) For any sets A and B, either $A \operatorname{inj} B$ or $B \operatorname{inj} A$.*

This claim follows easily from Theorem 4.2 if at least one of A or B is countable. However, it is far from obvious in general. In fact the only way to prove it requires sophisticated principles of the theory of infinite sets such as the Well-ordering Axiom. The Well-ordering Axiom says that, not only $\mathbb{N}$, but *every* set can be ordered in a way which ensures that every nonempty subset of the set has a least element.

There is another question about set size which was raised by Cantor when he first developed his theory of cardinality. By Cantor's Theorem, we know $P(\mathbb{N})$ is strictly larger than the smallest infinite set, $\mathbb{N}$. Is $P(\mathbb{N})$ the next smallest set, or could there be some set of size strictly between $\mathbb{N}$ and $P(\mathbb{N})$? Cantor thought not, but was unable to prove it. Using the fact that $P(\mathbb{N}) \operatorname{bij} \mathbb{R}$, he formulated the question as:

The Continuum Hypothesis. For every set S,

$$S \operatorname{inj} \mathbb{R} \implies [S \text{ is countable or } S \text{ bij } \mathbb{R}].$$

It took almost a century to discover the remarkable result that there is no way to prove or disprove the Totality Claim or the Continuum Hypothesis. It is consistent with the other more basic axioms of Set Theory — known as Zermelo-Frankel Set Theory, or *ZF* for short — to assume the Well-ordering Axiom and hence Totality, and in practice essentially all mathematicians do accept Well-ordering as a fundamental axiom. But in 1962, Paul Cohen astonished the mathematical world by showing how to construct a mathematical model satisfying ZF but not the Well-ordering axiom! In other words, the truth of the Well-ordering Axiom is *independent* of the other basic axioms of ZF.

The Continuum Hypothesis is also independent of ZF and Well-ordering together. In the 1930's, Kurt Gödel showed that it is consistent with the other axioms to assume the Continuum Hypothesis is true. On the other hand, Cohen's methods applied as well to the Continuum Hypothesis. Cohen showed there is a model satisfying ZF *and* the Well-ordering Axiom, but not the Continuum Hypothesis.

These independence results are usually studied in graduate level Set Theory courses. But aside from nicely completing the "smaller than" story, the Totality Claim and the Continuum Hypothesis play no role in discrete Mathematics or the rest of 6.042, so we needn't pursue this part of the theory of infinite sets.

A Bijections, Injections, Surjections

In the arguments above we used various elementary properties of bijections, injections, and surjections. In this section we review and prove these basic facts.

To begin, let's review some further definitions:

Definition A.1 *If $f' : B \to C$, the* composition *of f' and f, written $f' \circ f$ is the function $e : A \to C$ defined by the rule that $e(a) ::= f'(f(a))$.*

For any set S, the identity function *$id_S : S \to S$ is defined by the property that $id_S(s) = s$ for all $s \in S$.*

Let $g, h : B \to A$. Define g to be a left inverse *of f iff*

$$g \circ f = id_A.$$

Similarly, h is a right inverse *of f iff*

$$f \circ h = id_B.$$

Lemma A.2 *Let A be a nonempty set. A function $f : A \to B$ is an injection iff it has a left inverse.*

Proof. ($\Leftarrow$). Suppose g is a left inverse of f. Now if $f(a_1) = f(a_2)$, then

$$a_1 = \mathrm{id}_A(a_1) = (g \circ f)(a_1) = (g(f(a_1)) = g(f(a_2)) = (g \circ f)(a_2) = \mathrm{id}_A(a2) = a_2$$

so f is an injection.

($\Rightarrow$) Suppose $f : A \to B$ is an injection. For every $b \in \mathrm{range}(f)$ there is by definition some $a \in A$ such that $f(a) = b$. Since f is an injection, this a is unique, so for $b \in \mathrm{range}(f)$, we can define $g(b)$ to be this unique a. Since A is nonempty, there is some $a_0 \in A$. For $b \in B - \mathrm{range}(f)$, define $g(b) = a_0$. So $g : B \to A$ and $g(f(a)) = a$ by definition of g. That is, g is a left inverse of f. ■

Lemma A.3 *A function $f : A \to B$ is an surjection iff it has a right inverse.*

Proof. ($\Leftarrow$). Suppose $h : B \to A$ is a right inverse of f. So for $b \in B$, $f(h(b)) = b$. Hence $b = f(a)$ where $a = h(b)$. Thus $b \in \text{range}(f)$. So $\text{range}(f) = B$, that is, f is a surjection.

($\Rightarrow$) For each $b \in B$, define $g(b)$ to be some a such that $f(a) = b$. Since f is a surjection, for each $b \in B$ there always is at least one such a. So $g : B \to A$ and $f(g(b)) = b$ for all $b \in B$. That is, g is a right inverse of f. ∎

Corollary A.4

$$A \operatorname{inj} B \iff B \operatorname{sur} A.$$

Proof. Notice that the definitions of left and right inverse imply that g is a left inverse of f iff f is a right inverse of g. So

$$\begin{aligned}
A \operatorname{inj} B &\iff \exists f : A \to B\,[f \text{ is an injection }] \\
&\iff \text{(by Lemma A.2) } \exists f : A \to B, g : B \to A\,[g \text{ is a left inverse of } f] \\
&\iff \exists f : A \to B, g : B \to A\,[f \text{ is a right inverse of } g] \\
&\iff \text{(by Lemma A.3) } \exists g : B \to A\,[g \text{ is a surjection}]
\end{aligned}$$

∎

In fact, we can say something stronger

Corollary A.5 *A function is a bijection iff it has a right inverse and a left inverse. If a function f has both right inverse and a left inverse, then they are equal.*

Proof. That a function is a bijection iff it has a right inverse and a left inverse follows immediately from the definition of bijection and the two previous Lemmas A.2 and A.3.

Now suppose $g : B \to A$ is a left inverse of f and $h : B \to A$ is a right inverse. Then

$$g(b) = g(\text{id}_B(b)) = g((f \circ h)(b)) = g(f(h(b))) = (g \circ f)(h(b)) = (\text{id}_A)(h(b)) = h(b),$$

so $g = h$. ∎

For a bijection $f : A \to B$, we conclude from Corollary A.5 that all right inverses and left inverses of f are equal. We write $f^{-1} : B \to A$ to denote this unique function.

Theorem A.6 *For any bijection $f : A \to B$, the function $f^{-1} : B \to A$ is also a bijection, and $(f^{-1})^{-1} = f$.*

Proof. We noted that the definitions of left and right inverse imply that g is a left inverse of f iff f is a right inverse of g. Since f^{-1} is a left and right inverse of f, it follows that f is a right and left inverse of f^{-1}, so by Lemma A.5, f^{-1} is a bijection and $(f^{-1})^{-1} = f$. ∎

Theorem A.6 immediately implies that $A \operatorname{bij} B \iff B \operatorname{bij} A$, that is, bij is *symmetric.* Obviously $A \operatorname{bij} A$, that is, bij is also *reflexive.* So to show that bij is an equivalence relation, we just have to verify that it is *transitive.* We'll actually prove something somewhat stronger, namely that both inj and sur are transitive relations.

Lemma A.7 *The composition of two injections is an injection.*

Proof. Suppose $f : A \to B$ and $f' : B \to C$ are injections. Let $e = f' \circ f$. Suppose $e(a_1) = e(a_2)$. We want to show that $a_1 = a_2$.

Let $b_i = f(a_i)$. So

$$f'(b_1) = f'(f(a_1)) = e(a_1) = e(a_2) = f'(f(a_2)) = f'(b_2).$$

Since f' is an injection, we conclude that $b_1 = b_2$. So $f(a_1) = f(a_2)$, and since f is an injection, $a_1 = a_2$. ∎

Corollary A.8 *The relation* inj *is transitive.*

Proof. Suppose $A \operatorname{inj} B$ and $B \operatorname{inj} C$. So there is an injection $f : A \to B$ and an injection $f' : B \to C$. Then $(f' \circ f) : A \to C$ is an injection by Lemma A.7. Hence $A \operatorname{inj} C$. ∎

Lemma A.9 *The composition of two surjections is a surjection.*

Proof. We could give a proof directly from the definitions as in the previous Lemma, but let's use another argument based on inverses.

Suppose $f : A \to B$ and $f' : B \to C$ are surjections. We want to show that $(f' \circ f)$ is a surjection. By Lemma A.3, we need only find a right inverse for $(f' \circ f)$. We'll do this as follows. Let $g : B \to A$ be a right inverse of f, which exists by Lemma A.3. Likewise, let $g' : C \to B$ be a right inverse of f'. Let $e' = g \circ g'$. We claim that e' is the desired right inverse of $(f' \circ f)$. This follows because

$$(f' \circ f) \circ e = (f' \circ (f \circ e)) = (f' \circ (f \circ (g \circ g'))) = (f' \circ ((f \circ g) \circ g')) = (f' \circ (\mathrm{id}_B \circ g')) = f' \circ g' = \mathrm{id}_C.$$

∎

Lemma A.9 implies that the relation sur is transitive. We have finally completed the proof of

Theorem A.10 *The relation* bij *is an equivalence relation on sets.*

Exercise. Suppose $A_0 \operatorname{bij} A_1$ and $B_0 \operatorname{bij} B_1$.

(a) Prove that $A_0 \times B_0$ bij $A_1 \times B_1$, $P(A_0) \operatorname{bij} P(A_1)$.

(b) Give a counterexample to the claim that $A_0 \cap B_0$ inj $A_1 \cap B_1$.

B Finite Cardinality

In this section we prove the basic properties of finite cardinality, starting from the definitions of $|A|$ and the basic properties of bijections, injections, and surjections in section A.

We begin with the key Lemma which justifying the use of the sets $[0, n)$ as standard finite sets.

Lemma B.1 $[0, m) \operatorname{inj} [0, n)$ *iff* $m \leq n$.

Proof. ($\Rightarrow$) If $m \leq n$, then the "inclusion function" mapping $m \in [0, m)$ to itself, is an injection from $[0, m)$ to $[0, n)$.

($\Leftarrow$) By induction on n.

(Base case $n = 0$). Since $n = 0$, we know $[0, n) = \emptyset$. So if $[0, m) \operatorname{inj} [0, n)$, then $[0, m)$ must also be empty. That is, $m = 0 \leq n$.

(Induction). Assume that $n \geq 0$ and that

$$\forall m \in \mathbb{N} : [0, m) \operatorname{inj} [0, n) \implies m \leq n.$$

Suppose $f : [0, m) \to [0, n+1)$ is an injection. We must show that $m \leq n+1$. If $m = 0$, we are done, so we can assume $m > 0$. Now there are three cases:

1. Suppose $n \notin \operatorname{range}(f)$. Then f can also be regarded as a function from $[0, m)$ to $[0, n)$, and so $[0, m) \operatorname{inj} [0, n)$. Then by induction, $m \leq n$, so $m \leq n + 1$ as required.

2. Suppose $f(m-1) = n$. Since f is a bijection to $[0, n+1)$, we know that $f(k) < n$ for $0 \leq k < m - 1$. Now let g be the restriction of f to $[0, m-1)$, that is $g(k) ::= f(k)$ for $0 \leq k < m - 1$. The domain of g is the domain of f except for $m - 1$. Since f is an injection, the range of g is the range of f except for $f(m-1) = n$. So $g : [0, m-1) \to [0, n)$, and g is an injection since f is. Hence, $[0, m-1) \operatorname{inj} [0, n)$. By induction, we have that $m - 1 \leq n$, which implies $m \leq n + 1$.

3. Finally, supppose $n \in \text{range}(f)$ and $f(m-1) \neq n$. Since $n \in \text{range}(f)$, there is a nonnegative integer $j_0 < m-1$ such that $f(j_0) = n$. Now let g be defined so that it is the same as f except that the values at j_0 and $m-1$ are switched. That is $g : [0, m) \to [0, n+1)$ is defined by the rule that $g(j) ::= f(m-1)$, $g(m-1) ::= f(j_0) = n$, and $g(i) ::= f(i)$ for $i \neq j_0, m-1$. Since f is an injection, so is g. But $g(m-1) = n$, so our argument in the previous case applies to g, allowing us again to conclude that $m \leq n+1$.

■

Now we can formally conclude that there really are exactly n nonnegative integers less than n, namely, the set $\{0, 1, \ldots, n-1\}$ has cardinality n:

Corollary B.2 *For all* $n \in \mathbb{N}$, $|[0, n)| = n$.

Proof. Suppose $|[0, n)| = m$, so $[0, n)$ bij $[0, m)$ by definition, and therefore $n \leq m$ by the previous Lemma B.1. But we've already observed that $m = |[0, n)| \leq n$. Therefore, $m = n$. ■

Lemma B.3 *If* A inj $[0, n)$, *then* A *is finite and* $|A| \leq n$.

Proof. By induction on n.

(Base case $n = 0$). We have that $[0, n) = \emptyset$, so A must be empty and hence $|A| = 0$ by Lemma 2.3.

(Induction). Assume that $n \geq 0$ and B inj $[0, n)$ implies B is finite and $|B| \leq n$, for all sets B.

Now suppose $f : A \to [0, n+1)$ is an injection. We must show that A is finite and $|A| \leq n+1$. There are two cases:

1. Suppose $n \notin \text{range}(f)$. Then f can also be regarded as a function from A to $[0, n)$, and so A inj $[0, n)$. Then by induction, A is finite and $|A| \leq n \leq n+1$ as required.

2. Otherwise, $f(a) = n$ for some $a \in A$. Let $B ::= A - \{a\}$. Since f is an injection, we know that $f(b) < n$ for $b \in B$. Now let g be the restriction of f to B, that is $g(b) = f(b)$ for $b \in B$. Since f is an injection, the range of g is the range of f except for $f(a) = n$. So $g : B \to [0, n)$, and g is an injection since f is. Hence, B inj $[0, n)$.

   By induction, we conclude that B is finite and $|B| \leq n$. That is, there is a bijection $h' : B \to [0, m)$ for some $m \leq n$. Now define a function h on A by the rule that $h(b) ::= h'(b)$ for $b \in B$ and $h(a) ::= m+1$. Then $h : A \to [0, m+1)$ is a bijection. So A is finite and $|A| \leq m+1$ by definition of $|A|$. But $m+1 \leq n+1$, so $|A| \leq n+1$ as required.

■

Corollary B.4 *If B is finite and $A \operatorname{inj} B$, then A is finite. In particular, if B is finite and $A \subseteq B$, then A is finite.*

Proof. If B is finite, then $B \operatorname{bij} [0, |B|)$ by definition of $|B|$. If also $A \operatorname{inj} B$, then since inj is transitive, we conclude that $A \operatorname{inj} [0, n)$, so A is finite by Lemma B.3.

If $A \subseteq B$, then the "inclusion" function $\mathrm{id}_{AB} : A \to B$ defined by the rule $\mathrm{id}_{AB}(a) ::= a$ is an injection, so $A \operatorname{inj} B$, so the second statement in the Corollary follows from the first. ■

Corollary B.5

$$A \operatorname{bij} [0, n) \iff A \textit{ is finite and } |A| = n.$$

Proof. ($\Leftarrow$) By definition of $|A|$.

($\Rightarrow$) If $A \operatorname{bij} [0, n)$, then A is finite and $A \operatorname{bij} [0, |A|)$ by definition. Since bij is an equivalence, $[0, n) \operatorname{bij} [0, |A|)$, so $m = n$ by Lemma B.1. ■

We still need another "sanity check" of our definitions: we have defined "same size" separately from "the size" of finite sets. We better be sure the definitions agree:

Theorem B.6 *If B is a finite set, then*

$$[A \textit{ is finite and } |A| \leq |B|] \iff A \operatorname{inj} B.$$

Proof. Since B is finite, if $A \operatorname{inj} B$, then A is finite by we have by Corollary B.4. So we may assume in the proof that A is finite.

By definition of cardinality, $A \operatorname{bij} [0, |A|)$ and $B \operatorname{bij} [0, |B|)$. So since bij is an equivalence, bij implies inj, and inj is transitive, we have that

$$A \operatorname{inj} B \iff [0, |A|) \operatorname{inj} [0, |B|).$$

But by Lemma B.1,

$$[0, |A|) \operatorname{inj} [0, |B|) \iff |A| \leq |B|.$$

■

Corollary B.7 *If B is a finite set, then*

$$[A \textit{ is finite and } |A| = |B|] \iff A \operatorname{bij} B.$$

Proof. ($\Leftarrow$) $A \operatorname{bij} B$ implies that $A \operatorname{inj} B$ and $B \operatorname{inj} A$, so by Theorem B.6, A is finite, $|A| \leq |B|$ and $|B| \leq |A|$.

($\Rightarrow$) By definition, $A \operatorname{bij} [0, |A|)$ and $B \operatorname{bij} [0, |B|)$. If $|A| = |B|$, then obviously $[0, |A|) = [0, |B|)$, so A and B have a bijection to the same set, namely, $[0, |A|)$. So $A \operatorname{bij} B$, since bij is an equivalence. ■

So to summarize, we have found that the relative sizes of finite sets are completely captured by injections and bijections among them.