Understanding Analysis, 2nd Edition: Stephen Abbott
Chapter 1 - Section 1.6 - Cantor's Theorem

1. Show that $(0,1)$ is uncountable if and only if R is uncountable.

Solution: First we prove the necessity: if $(0,1)$ is uncountable then R is uncountable. This follows immediately from the fact that $(0,1)$ is a subset of R. A superset of an uncountable set must be uncountable as well.

Now the sufficiency: if R is uncountable then $(0,1)$ is uncountable. I initially wanted to prove this by contrapositive: if $(0,1)$ is countable then R is countable. But this is tricky. The fact that $(0,1)$ is a subset of R does not imply R is countable. Then I tried using the fact that the union of (infinitely many) countable sets is countable, by merging infinitely many intervals. But this requires proving that the intervals other than $(0,1)$ are also countable.

The idea hit me to produce a bijection between $(0,1)$ and R instead, and indeed such a bijection is possible by Exercise 1.5.4. For example, the tangent function maps $(\frac{-\pi}{2}, \frac{\pi}{2})$ to R. We need just need to map $(0,1)$ to $(\frac{-\pi}{2}, \frac{\pi}{2})$ and compose it with tan. Find the line equation joining $(0,1)$ to $(\frac{-\pi}{2}, \frac{\pi}{2})$ and we get $\pi x - \frac{\pi}{2}$, therefore $f(x) = tan(\pi x - \frac{\pi}{2})$ is a bijection between $(0,1)$ and R. This proves both necessity and sufficiency at once.

2. (a) Explain why the real number $x = .b_1b_2b_3...$ cannot be $f(1)$.

   (b) Now explain why $x \neq f(2)$, and in general why $x \neq f(n)$ for any $n \in N$.

   (c) Point out the contradiction that arises from these observations and conclude that $(0, 1)$ is uncountable.

Solution:

(a) Because b_1 is already different from a_{11}.

(b) In general, $b_n \neq a_{nn}$, so the decimal expansion differs in at least one digit.

(c) If $(0, 1)$ were countable, then $x = f(n)$ for an n, since x is defined to be a real number $\in (0, 1)$. But from part (b), $x \neq f(n)$ for all n, a contradiction. Therefore $(0, 1)$ must be uncountable.

3. Supply rebuttals to the following complaints about the proof of Theorem 1.6.1 (the open interval $(0, 1)$ is uncountable).

   (a) Every rational number has a decimal expansion, so we could apply the same argument to show that the set of rational numbers between 0 and 1 is uncountable. However, because we know that every subset of Q must be uncountable, the proof of Theorem 1.6.1 must be flawed.

   (b) Some numbers have *two* different decimal representations. Specifically, any decimal expansion that terminates can also be written with repeating 9's. For instance, 1/2 can be written as 0.5 or as 0.49999.... Doesn't this cause some problems?

Solution:

(a) Every rational number has a decimal expansion, but the converse is not true; not every decimal expansion represents a rational number. In the proof of Theorem 1.6.1, x may be irrational, so it can't be used to conclude that Q is uncountable.

(b) First, it's worth noting that the *only way* a number can have two different decimal expansion is the one the problem statement says; one expansion that terminates, and another with repeating 9's. I claim this without proof, as the book also intentionally uses decimal representations without formal definitions.

The proof of Theorem 1.6.1 works fine because it uses only the digits 2 and 3, so the counterexample it constructs is a number with a unique decimal expansion. It cannot be the case that x is just another representation of the decimal expansion of a real number $f(n)$.

4. Let S be the set consisting of all sequences of 0's and 1's. Observe that S is not a particular sequence, but rather a large set whose elements are sequences, namely,

$$S = \{(a_1, a_2, a_3, ...) : a_n = 0 \text{ or } 1\}.$$

As an example, the sequence $(1, 0, 1, 0, 1, 0, ...)$ is an element of S, as is the sequence $(1, 1, 1, 1, 1, ...)$. Give a rigorous argument showing that S is uncountable.

Solution: The same diagonal argument works. We just replace the decimal digits 0-9 with 0 and 1. Consider the following table like in the book

$$\begin{array}{c|l} \mathbb{N} & (0,1) \\ \hline 1 & f(1) = \ .\,\mathbf{a_{11}}\, a_{12}\, a_{13}\, a_{14}\, a_{15}\, a_{16}\, \cdots \\ 2 & f(2) = \ .\,a_{21}\, \mathbf{a_{22}}\, a_{23}\, a_{24}\, a_{25}\, a_{26}\, \cdots \\ 3 & f(3) = \ .\,a_{31}\, a_{32}\, \mathbf{a_{33}}\, a_{34}\, a_{35}\, a_{36}\, \cdots \\ 4 & f(4) = \ .\,a_{41}\, a_{42}\, a_{43}\, \mathbf{a_{44}}\, a_{45}\, a_{46}\, \cdots \\ 5 & f(5) = \ .\,a_{51}\, a_{52}\, a_{53}\, a_{54}\, \mathbf{a_{55}}\, a_{56}\, \cdots \\ 6 & f(6) = \ .\,a_{61}\, a_{62}\, a_{63}\, a_{64}\, a_{65}\, \mathbf{a_{66}}\, \cdots \\ \vdots & \vdots \end{array}$$

where $a_{mn} \in \{0, 1\}$. Construct a sequence b from the diagonal such that $b_n = 1$ if $a_{nn} = 0$ else $b_n = 0$. Exactly the same proof as Exercise 2 works here.

5. (a) Let $A = \{a, b, c\}$. List the eight elements of $P(A)$. (Do not forget that $\emptyset$ is considered a subset of every set).

   (b) If A is finite with n elements, show that $P(A)$ has 2^n elements.

Solution:

(a) $\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}$.

(b) For a particular subset, an element is either chosen or not chosen. So there are two choices for each of the n elements, hence 2^n.

6. (a) Using the particular set $A = \{a, b, c\}$, exhibit two different 1-1 mappings from A to $P(A)$.

   (b) Letting $C = \{1, 2, 3, 4\}$, produce an example of a 1-1 map $g : C \to P(C)$.

   (c) Explain why in parts (a) and (b), it's impossible to construct mappings that are *onto*.

Solution: Note that the elements of $P(A)$ are *sets*. So for example in part (a), $f(b) = b$ is not a valid mapping. It should instead be $f(b) = \{b\}$.

(a) $f(a) = \{a\}, f(b) = \{b\}, f(c) = \{c\}$ or $f(a) = \{b\}, f(b) = \{c\}, f(c) = \{a\}$.

(b) $g(1) = \{1\}, g(2) = \{2\}, g(3) = \{3\}, g(4) = \{4\}$.

(c) Because $|P(A)| > |A|$, at least for any nonempty finite set A. It's also true for infinite sets, but that's the next exercise.

7. Return to the particular functions constructed in Exercise 1.6.6 and construct the subset B that results using the preceding rule (see the book). In each case, note that B is not in the range of the function used.

Solution: We'll just do it for part (a) of Exercise 1.6.6. When the mapping is

$$f(a) = f(a) = \{a\}, f(b) = \{b\}, f(c) = \{c\}$$

then $B = \emptyset$, and this is outside the range of f.

When the mapping is

$$f(a) = \{b\}, f(b) = \{c\}, f(c) = \{a\}$$

then $B = \{a, b, c\}$, and this is again outside the range of f.

Also note that this is pretty much the diagonal argument. For a set $A = \{a_1, a_2, a_3, ...\}$ and say, $f(a_1) = \{a_1, a_5\}$, then draw the table with $f(a_1) = \{a_1, a_2', a_3', a_4', a_5, a_6', ...\}$ where an accent means the element is not part of $f(a_1)$, but we list it anyway so the table stays rectangular. Now go through the diagonal and construct B by taking accented elements.

Too lazy to draw the table here. Just see the illustration in Wikipedia's page about Cantor's diagonal argument.

8. (a) First, show that the case $a' \in B$ leads to a contradiction.

   (b) Now, finish the argument by showing that the case $a' \notin B$ is equally unacceptable.

Solution:

(a) Our assumptions are $B = f(a')$ and $a' \in B$.

From the second assumption, this must mean that $a' \notin f(a')$, because that's how we constructed B. But if we have $a' \in B$ and $a' \notin f(a')$, then $B \neq f(a')$, contradicting our first assumption.

(b) Our assumptions are $B = f(a')$ and $a' \notin B$.

From the second assumption, this must mean that $a' \in f(a')$, because that's how we constructed B. But if we have $a' \notin B$ and $a' \in f(a')$, then $B \neq f(a')$, contradicting our first assumption.

9. Using the various tools and techniques developed in the last two sections (including the exercises from Section 1.5), give a compelling argument showing that $P(N) \sim R$.

Solution:

——————————— INCORRECT ———————————

We'll first produce the bijection $P(N) \sim (0, 1)$. Then by Exercise 1, $(0, 1) \sim R$.

First, we produce an injective function from $P(N)$ to $(0,1)$. We simply map an element $a \in P(N)$ to the decimal expansion of an element $f(a) \in (0,1)$. For example, if $a = \{1,2,3,4,11,24\}$ then $f(a) = 0.12341124$.

Wait no, this is not injective for at least two reasons. $f(\{1,11\}) = 0.111$ and $f(\{111\}) = 0.111$. Two different inputs, same output. Also $f(\{1,2,3\}) = 0.123$. But we can't say $f(\{3,2,1\}) = 0.321$ because $\{1,2,3\}$ and $\{3,2,1\}$ are the same set! So we can't produce 0.321.

Okay, I gave up and looked up the solution :(

The first thing is to forget about constructing a "nice", direct bijection between $P(N)$ and R. An MSE answer says we can't hope for a nice bijection because some obscure topological reasons that I don't understand yet so I'll just put the link here (see the comment by user Noah Schweber).

Now, the trick (and from some internet browsing, this seems like a standard, well-known trick) is to first produce a bijection from $P(N)$ to the set S of infinite binary sequences, as defined in Exercise 1.6.4. Then we'll produce a bijection between S and $[0,1)$, which will complete the proof because $[0,1) \sim R$ by Exercise 1.5.4(c).

The bijection between $P(N)$ and S is straightforward. For $s \in S$, $f(s) = \{n : s_n = 1\}$. For example, $s = 10100111000..$ then $f(s) = \{1,3,6,7,8\}$.

Next, the bijection $S \sim [0,1)$. The natural bijection here would be binary representation. Every real number has a binary representation, which we will accept and use without formal definition (just like how the book treats decimal representation). It's then easy to define the map $f : [0,1) \to S$ where $f(x)$ is the binary representation of x. For example, $f(1/4) = 0.01000....$

f is *almost* injective. Almost, because we need to deal with non-unique binary representations. For example, $f(1/4)$ can be either $0.01000...$ or $0.0011111....$ How do we handle this? There are two ways.

Approach #1: Schroder-Bernstein Theorem.

This problem showcases the usefulness of the Schroder-Bernstein Theorem. Why do we need a theorem for something as "obvious" as a bijection?

For numbers with two binary representations, we pick the version with infinite zeroes. Now f is injective from $[0,1)$ to S, but it's not surjective, so we can't just invert f to create a bijection.

This is where the Schroder-Bernstein Theorem comes into play. When it's difficult to use only a single function (and its inverse) to create a bijection, the Theorem allows us to use a **completely different function** to create an injection from S to $[0, 1)$, and it guarantees that there is a bijection.

So let's define $g : S \to [0, 1)$ where $g(x) = (0.x)_{10}$. That is, we treat the binary string as its decimal version. For example, $g(1011000...)$ is the *decimal* number $(0.1011000...)_{10}$. As the decimal only involves 0s and 1s, it won't encounter non-unique decimal issues, which only involves infinite 9s. Therefore, g is injective. It's clearly not surjective, as it misses any real numbers in $[0, 1)$ that contains digits other than $\{0, 1\}$.

Now that we have an injection from both $S \to [0, 1)$ and $[0, 1) \to S$, by Schroder-Bernstein there is a bijection. Cool!

Unfinished Approach #2: $|S| = |S'|$

As in approach 1, for numbers with non-unique binaries, we pick the version with infinite zeroes. We're also going to give this new set a name:

$$S' = \{\text{binary sequences that don't end in trailing 1s}\}$$

Then there's a bijection between S' and $[0, 1)$ (but not S and $[0, 1)$).

Let's review our sequence of bijections. We have $|P(N)| = |S|$ and $|S'| = |[0, 1)| = |R|$. If we can prove that $|S| = |S'|$, then we are done.

First, note that the subset we removed to create S' is *countable*. This is because the set of sequences that *do* end in trailing ones is specified by the prefix before the trailing ones, so we can enumerate them like so:

$$\{01111.., (0)0111.., (1)0111..., (00)0111..., (01)0111..., (10)0111..., \}$$

More precisely, the numbers in the parentheses forms the set of all finite binary strings $\cup_{n=1}^{\infty}\{0, 1\}^n$. As per Theorem 1.5.8(ii), this set is countable.

Now, to prove $|S| = |S'|$, we'll prove the following more general lemma instead: removing a countable subset of an uncountable set does not change its cardinality. Note the statement of the lemma carefully. It asserts that not only the uncountable set stays uncountable, it retains its cardinality, which is a stronger result.

Let A be an uncountable set and C be a countable subset of A and $B = A \setminus C$. We want to prove $|A| = |A \setminus C|$. We use the following two facts:

- An uncountable subset stays uncountable after removing a countable subset. In particular, $A \setminus C$ is still uncountable.
- An uncountable set always contains a countable subset.

Now let D be a countable subset of $A \setminus C$.

——————————— INCORRECT ———————————

Define $f : A \to A \setminus C$:

$$f(x) = \begin{cases} d_i & x = c_i \in C \\ x & \text{otherwise} \end{cases}$$

The other direction $g : A \setminus C \to A$ is similar: $g(x) = c_i$ if $x = d_i \in D$ and x otherwise.

It should be clear that this is a bijection. Any element in $c_i \in C$ will be mapped to $d_i = D$ (and vice versa). Any element outside both will be mapped to itself, and by definition will not map to an element in either C or D.

I'm such a fucking idiot. Of course this doesn't work. $f(c_i) = f(d_i) = d_i$.

Instead, we'll use Schroder-Bernstein again. We need to produce injection between A and $A \setminus C$ in both directions. The injection $A \setminus C \to A$ is trivial: just use the identity function.

For the other direction, first note that there is a bijection between C and D, since they're both countable. Let $g : C \to D$ be the bijective function. Now, define $f : A \to A \setminus C$:

$$f(x) = \begin{cases} g(c_i) & x = c_i \in C \\ g'(d_i) & x = d_i \in D \\ x & \text{otherwise} \end{cases}$$

This is an injective function from A to $A \setminus C$. Elements from C and D just use the existing bijection, so they're injective. As for $x \notin C \cup D$, it maps to itself which means it's guaranteed not to coincide with either C or D.

Now that we have an injective function in both directions, by Schroder-Bernstein we can conclude $|A| = |A \setminus C|$.

Why half-open interval [0, 1)

$P(N)$ includes the empty set and this maps to the binary 000.... Including zero makes it easier to map to this binary.

If we include one, when constructing the mapping between S and $[0, 1)$, should the binary 000... map to the real number 0.000.. or 1.000...?