# CARDINALITY OF THE REALS

In these notes we show that $\mathbf{R}$, the set of real numbers, has the same cardinality as $\mathcal{P}(\mathbf{N})$, the power set of the set $\mathbf{N}$ of natural numbers.

Our argument will proceed in two stages (each with several steps): first, showing that the interval $[0, 1) \subset \mathbf{R}$ has the same cardinality as $\mathcal{P}(\mathbf{N})$, and second, showing that $[0, 1)$ has the same cardinality as the whole set $\mathbf{R}$. We assume that the "infinite decimal" model of $\mathbf{R}$ has been established, and that the uncountability of $\mathcal{P}(\mathbf{N})$ has been established.[1]

These notes were written for students in David Groisser's fall 2018 MAA 4211 class, so certain additional facts are assumed that, at the time of this writing, have already been seen by these students in class or in homework.

**Some notation and conventions used in these notes.**

- **(Convention on the meaning of "natural number")**

  $\mathbf{N}$ = the set of natural numbers $= \{1, 2, 3, \dots\}$ (i.e. 0 is not included in $\mathbf{N}$).[2]

- Given two sets $A$ and $B$, the notation "$A \sim B$" means that there is a bijection from $A$ to $B$.

- For $n \in \mathbf{N}$, define $J_n = \{m \in \mathbf{N} \mid m \leq n\} = \{1, \dots, n\} \subset \mathbf{N}$. Define $J_0 = \emptyset$.

- **(Convention on the meaning of "countable")**

  A set $S$ is *countable* if there exists a subset $A \subset \mathbf{N}$ such that $S \sim A$. Every finite set is countable. If $S$ is finite and nonempty, there exists a (necessarily unique) $n \in \mathbf{N}$ such that $S \sim J_n$. An infinite countable set is called *countably infinite*.[3]

- For a set $A$, the notation "$|A|$" is read "the cardinality of $A$". For sets $A$ and $B$, "$|A| = |B|$" is synonymous with "$A \sim B$".

[1] The latter is a special case of the fact for any set $X$, there never exists a surjective map from $X$ to $\mathcal{P}(X)$.

[2] This is the I use throughout my Advanced Calculus course, so for my own students, there is no necessity that I state this convention here. I'm stating it for other readers, and as a memory-aid for students in my class who are simultaneously taking another class in which the convention is that $\mathbf{N}$ includes 0.

[3] As with the convention for natural numbers, my own students already know that this is the convention is used throughout my Advanced Calculus class. There is another convention in which "countable" means "countably infinite", and in which a set that is either finite or countably infinite is called "at most countable".

- If $A$ is a finite set and $n \in \mathbf{N}$, the notation "$|A| = n$" means $|A| = |J_n|$ (equivalently, $A \sim J_n$ ). The notation "$|A| = 0$" means $A = \emptyset$.
- "Interval notation" has the customary meaning. In particular, $[0,1) = \{x \in \mathbf{R} \mid 0 \leq x < 1\}$.

**"Binary decimals"**

An "infinite binary decimal", or simply "binary decimal", is a formal expression

$$.a_1\, a_2\, a_3\, \ldots \tag{1}$$

where $a_n \in \{0,1\}$ for each $n \in [0,1]$.[4] More precisely, a binary decimal is simply the above choice of notation for a function $f : \mathbf{N} \to \{0,1\}$; the relation between the notation (1) and such a function $f$ is given by identifying $f(n)$ with the digit $a_n$ (for each $n \in \mathbf{N}$). Below, we will sometimes call such functions $f$ themselves "binary decimals".

Modifying some notation used in earlier homework, let $\underline{\mathbf{2}}^{\mathbf{N}}$ denote the set of all functions from $\mathbf{N}$ to $\{0,1\}$ (hence the set of all binary decimals). In homework it was proven that

$$\left|\underline{\mathbf{2}}^{\mathbf{N}}\right| = \left|\mathcal{P}(\mathbf{N})\right|. \tag{2}$$

Analogously to what we did for base-10 decimals, call a binary decimal *normalized* if it has no infinite strings of 1's. The set of functions from $\mathbf{N}$ to $\{0,1\}$ corresponding to normalized binary decimals is therefore

$$\text{NBD} \;:=\; \{f : \mathbf{N} \to \{0,1\} \mid \nexists N \in \mathbf{N} \text{ such that for all } n \geq N \text{ we have } f(n) = 1\}\,.$$

The complement of NBD in $\underline{\mathbf{2}}^{\mathbf{N}}$ is thus the set

$$\begin{aligned} Z \;&:=\; \underline{\mathbf{2}}^{\mathbf{N}} \setminus \text{NBD} && (3)\\ &=\; \{f : \mathbf{N} \to \{0,1\} \mid \exists N \in \mathbf{N} \text{ such that for all } n \geq N \text{ we have } f(n) = 1\}\,. && (4) \end{aligned}$$

We can map the set of binary decimals (equivalently, $\underline{\mathbf{2}}^{\mathbf{N}}$) to the interval $[0,1] \subset \mathbf{R}$ exactly the way we did for base-10 infinite decimals: for each $f \in \underline{\mathbf{2}}^{\mathbf{N}}$ and $N \in \mathbf{N}$, we define $s_N(f) = \sum_{n=1}^{N} \frac{f(n)}{2^n}$, and define the real number $r(f)$ by $r(f) = \text{l.u.b.}(\{S_N(f) \mid N \in \mathbf{N}\})$. Essentially the same proof used for base-10 decimals shows that the map $f \mapsto r(f)$ restricts to a bijection from NBD to $[0,1)$.

This establishes that

$$\left|\text{NBD}\right| = \left|[0,1)\right|. \tag{5}$$

To complete stage-one of our argument, we will show that $\underline{\mathbf{2}}^{\mathbf{N}}$ has the same cardinality as its proper subset NBD. For this we will need some lemmas.

[4]The term "binary decimal" is, of course, an oxymoron, but I am not aware of any simple, generally accepted terminology for this object. A binary decimal is simply the base-2 analog of a base-10 (infinite) decimal with all the digits to the right of the decimal point. In some uses, the term "binary decimal" allows for an integer $a_0$ to be included to the left of the decimal point, but in these notes we won't do that.

**Lemma 1.1** *Let $X$ be an infinite set, $S \subset X$ a countable subset. Then $X$ has a countably infinite subset containing $S$.*

**Proof**: Left to reader.[5] ■

**Lemma 1.2** *Let $S$ be a finite nonempty subset of $\mathbf{N}$, and let $n = |S|$. Then there exists a bijection $\psi : \mathbf{N} \to \mathbf{N}$ such that $\psi(S) = J_n$.*

**Proof**: Left to reader. ■

**Lemma 1.3** *Let $X$ be an uncountable set, $S \subset X$ a countable subset, and $A$ a nonempty countable set. Then*

*(a) $X \setminus S$ has the same cardinality as $X$.*

*(b) $X \bigcup A$ has the same cardinality as $X$.*

**Proof**: (a) If $S$ is empty there is nothing to show, so assume $S \neq \emptyset$.

The set $X \setminus S$ is uncountable, since otherwise $X$ would be the union of two countable sets ($S$ and $X \setminus S$), hence would be countable. In particular, $X \setminus S$ is infinite, hence (by Lemma 1.1) contains a countably infinite subset $S_1$. Then $\tilde{S} := S \bigcup S_1$ is countably infinite: it is countable since it is the union of two countable sets, and infinite since it contains the infinite set $S_1$. Let $\varphi : \tilde{S} \to S_1$ be a bijection; such $\varphi$ exists since all countably infinite sets have the same cardinality.

Define $\tilde{f} : X \to X$ by

$$\tilde{f}(x) = \begin{cases} x & \text{if } x \notin \tilde{S}, \\ \varphi(x) & \text{if } x \in \tilde{S}. \end{cases}$$

It is easily seen that $\tilde{f}$ is injective and that its range is $(X \setminus \tilde{S}) \bigcup S_1 = X \setminus S$. Defining $f : X \to X \setminus S$ by $f(x) = \tilde{f}(x)$ for all $x \in X$, it follows that $f$ is a bijection. Hence $|X| = |X \setminus S|$.

(b) Let $Y = X \bigcup A$. Then $X = Y \setminus A'$, where $A' = A \setminus X = A \setminus (A \bigcap X)$. Since every subset of a countable set is countable, $A'$ is countable. It is easily shown that every set containing an uncountable subset is itself uncountable; hence $Y$ is uncountable. Since the set $X$ used in the proof of (a) was an arbitrary uncountable set, it follows that $|Y \setminus A'| = |Y|$; i.e. $|X| = |X \bigcup A|$. ■

[5] Dr. Groisser's MAA 4211 students have done the following homework exercise: Prove that every infinite set has a countably infinite subset. The proof of Lemma 1.1 is just a minor modification of that homework exercise.

**Remark 1.4** In the only uses we will make of Lemma 1.3, the countable set $S$ or $A$ will either have just one element or will be countably infinite.

Now consider the set $Z$ defined in equation (3). From equation (4), we have

$$\begin{aligned} Z &= \bigcup_{N\in\mathbf{N}} U_N, \\ \text{where } U_N &:= \{f : \mathbf{N} \to \{0,1\} \mid f(n) = 1 \text{ for all } n \geq N\}. \end{aligned}$$

For each $N \in \mathbf{N}$, a function $f \in U_N$ is completely determined by the values $f(1), f(2), \ldots, f(N-1)$; more precisely by the function $f\big|_{J_{N-1}}$ (the restriction of $f$ to the set $J_{N-1}$, which is the empty set if $N = 1$). The map $f \mapsto f\big|_{J_{N-1}}$ is easily seen to yield a bijection from $U_N$ to the finite set $\underline{\mathbf{2}}^{J_{N-1}} :=$ {all functions from $J_{N-1}$ to $\{0,1\}$}. Hence $Z$ is a countable union of finite sets (hence a countable union of countable sets), so by an earlier homework problem, $Z$ is countable.

**Corollary 1.5** *The sets* $\underline{\mathbf{2}}^{\mathbf{N}}$ *and* NBD *have the same cardinality.*

**Proof**: Since $\mathcal{P}(\mathbf{N})$ is uncountable and $\underline{\mathbf{2}}^{\mathbf{N}} \sim \mathcal{P}(\mathbf{N})$, the set $\underline{\mathbf{2}}^{\mathbf{N}}$ is uncountable. We have just seen that the subset $Z \subset \underline{\mathbf{2}}^{\mathbf{N}}$ is countable. Since NBD $= \underline{\mathbf{2}}^{\mathbf{N}} \setminus Z$, Lemma 1.3 shows that NBD $\sim \underline{\mathbf{2}}^{\mathbf{N}}$. ■

**Corollary 1.6** *The interval* $[0,1) \subset \mathbf{R}$ *has the same cardinality as* $\mathcal{P}(\mathbf{N})$.

**Proof**: Combining results from above, we have $\mathcal{P}(\mathbf{N}) \sim \underline{\mathbf{2}}^{N} \sim$ NBD $\sim [0,1)$. ■

Stage 1 of our argument is now complete. For Stage 2 we have several options, depending on how much we are willing to take for granted. The simplest options require the use of facts we have not yet proven in MAA 4211 as of this writing, but will prove later in MAA 4211–4212, such as the Intermediate Value Theorem and the existence and properties of the basic trigonometric functions. We have not yet even *defined* "continuous function", let alone proved anything about continuous real-valued functions (e.g. the Intermediate Value Theorem); we will do this later in MAA 4211. The trigonometric functions are developed in MAA 4212.

Below are four ways of showing that $\mathbf{R} \sim [0,1)$, using progressively less knowledge. The last two ways use nothing that we haven't proven yet in MAA 4211. The first two rely on facts we have not yet proven, but I'm including them since these arguments—especially the first—are the ones that one mathematician would usually give to another; once you have the background for them it would be silly to hunt for a proof of "$\mathbf{R} \sim [0,1)$" that deliberately avoids using these facts.

**Option 1: Assume complete knowledge of the function** $\tan : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbf{R}$**.**

Assuming the usual properties of the tangent function, the map from the open interval $(0,1)$ to $\mathbf{R}$ given by $x \mapsto \tan(\pi(x-\frac{1}{2}))$ is a bijection. Hence

$$\big|(0,1)\big| = \big|\mathbf{R}\big|. \tag{6}$$

As seen earlier in these notes, the interval $[0,1)$ is uncountable, so by Lemma 1.3, $[0,1)\setminus\{0\} \sim [0,1)$. Hence

$$\big|\mathcal{P}(\mathbf{N})\big| = \big|[0,1)\big| = \big|[0,1)\setminus\{0\}\big| = \big|(0,1)\big| = \big|\mathbf{R}\big|. \tag{7}$$

■

**<u>Option 2: Assume the Intermediate Value Theorem and the continuity of rational functions.</u>**

Define $f:(0,1)\to\mathbf{R}$ by $f(x)=\frac{1}{1-x}-\frac{1}{x}$. By what we are allowing ourselves to assume in "Option 2", this function is continuous. Each of the functions $x\mapsto\frac{1}{1-x}$ and $x\mapsto -\frac{1}{x}$ is strictly increasing[6], so their sum is strictly increasing. It is easily seen that any strictly increasing function from an interval to $\mathbf{R}$ is injective. The function $f$ achieves arbitrarily large values (given any $y_0\in\mathbf{R}$ we have $f(1-\epsilon)>y_0$ for $\epsilon>0$ sufficiently small) and arbitrarily small values (given any $y_0\in\mathbf{R}$ we have $f(\epsilon)<y_0$ for $\epsilon>0$ sufficiently small). Using the Intermediate Value Theorem, it follows that $f$ achieves every value in $\mathbf{R}$—i.e. that $f$ is surjective—and hence that $f$ is bijective. Equations (6) and (7) then follow just as in Option 1. ■

**<u>Option 3: Assume nothing we haven't proven to date, but still exploit the function in Option 2.</u>**

To simplify some algebra, instead of literally using the function $f$ in Option 2, we will use the function $f:(-1,1)\to\mathbf{R}$ defined by $f(x)=\frac{1}{1-x}-\frac{1}{1+x}=\frac{2x}{1-x^2}$. In class, for any $a,b\in\mathbf{R}$ with $a<b$, we exhibited a bijection from $(0,1)$ to $(a,b)$. Hence if we show that $(-1,1)\sim\mathbf{R}$, it will follow that $(0,1)\sim\mathbf{R}$.

Define $\tilde{g}:\mathbf{R}\to\mathbf{R}$ by $\tilde{g}(y)=\frac{y}{1+\sqrt{1+y^2}}$. (Here we are using the fact, recently proven in class, that every positive real number has a unique positive square root, in order to define "$\sqrt{1+y^2}$".) It is easily seen that $|\tilde{g}(y)|<1$ (and thus $\tilde{g}(y)\in(-1,1)$) for all $y\in\mathbf{R}$. Define $g:\mathbf{R}\to(-1,1)$ by $g(y)=\tilde{g}(y)$. As the reader may check by brute-force substitution, $g(f(x))=x$ for all $x\in(-1,1)$, and $f(g(y))=y$ for all $y\in\mathbf{R}$.[7] Hence $f$ is a bijection.

[6] A real-valued function $g$ defined on an interval $I\subset\mathbf{R}$ is *strictly increasing* if for all $x_1,x_2\in I$ with $x_1<x_2$, we have $g(x_1)<g(x_2)$

[7] The formula for $\tilde{g}$ was not pulled from thin air, of course. One can produce this formula by writing $y=\frac{2x}{1-x^2}$ and solving for $x$ in terms of $y$. For $y\neq 0$ we must solve a quadratic equation,

Thus

$$\big|(0,1)\big| = \big|(-1,1)\big| = \big|\mathbf{R}\big|, \tag{8}$$

and the same argument as in Option 1 yields the equality (7). ■

**Option 4: Proceed directly from simple consequences of the Least Upper Bound property.**

Using the Least Upper Bound property, we proved in class that for every real number $y$, there is a unique integer $n$ such that $n \le y < n+1$. Equivalently, for each $y \in \mathbf{R}$ there is a unique pair $(n,x) \in \mathbf{Z} \times [0,1)$ such that $y = n+x$. This yields a map $\mathbf{R} \to \mathbf{N} \times [0,1)$. Clearly this map is bijective (its inverse is the map $(n,x) \mapsto n+x$). Hence

$$\big|\mathbf{R}\big| = \big|\mathbf{N} \times [0,1)\big|. \tag{9}$$

For any sets $A, B, C$, if $f : B \to C$ is a bijection then the map $A \times B \to A \times C$ defined by $(a,b) \mapsto (a, f(b))$ is also a bijection. Since we have already established that $[0,1) \sim \mathcal{P}(\mathbf{N})$, we therefore have $\mathbf{N} \times [0,1) \sim \mathbf{N} \times \mathcal{P}(\mathbf{N})$. Thus it suffices to show that $\mathbf{N} \times \mathcal{P}(\mathbf{N}) \sim \mathcal{P}(\mathbf{N})$.

For each $n \in \mathbf{N}$ define $\varphi_n : \mathbf{N} \to \mathbf{N}$ by $\varphi_n(m) = n + m$. Then define $h : \mathbf{N} \times \mathcal{P}(\mathbf{N}) \to \mathcal{P}(\mathbf{N}) \setminus \{\emptyset\}$ by

$$h(n,S) = \{n\} \bigcup \varphi_n(S). \tag{10}$$

Note that since every element of $\mathbf{N}$ is positive, $\varphi_n(S)$ never contains $n$ itself, and the minimal element of $h(n,S)$ is always $n$. Thus if $h(n_1,S_1) = h(n_2,S_2)$, we have $n_1 = n_2$, $\varphi_{n_1}(S_1) = \varphi_{n_2}(S_2) = \varphi_{n_1}(S_2)$, and therefore $S_1 = S_2$ as well. Hence $h$ is injective. If $S \in \mathcal{P}(\mathbf{N}) \setminus \{\emptyset\}$, then $S$ has a minimal element, so we may define $S' = \{m - \min(S) \mid m \in S \text{ and } m > \min(S)\}$. Then $h(\min(S), S') = S$. Hence $h$ is surjective as well as injective.

Thus $\mathbf{N} \times \mathcal{P}(\mathbf{N}) \sim \mathcal{P}(\mathbf{N}) \setminus \{\emptyset\}$. By Lemma 1.3, deleting the single element $\{\emptyset\}$ from $\mathcal{P}(\mathbf{N})$ does not change the cardinality: $\mathcal{P}(\mathbf{N}) \setminus \{\emptyset\} \sim \mathcal{P}(\mathbf{N})$. Hence $\mathbf{N} \times \mathcal{P}(\mathbf{N}) \sim \mathcal{P}(\mathbf{N})$. Combining this with equation (9), we have

$$\big|\mathbf{R}\big| = \big|\mathbf{N} \times [0,1)\big| = \big|\mathbf{N} \times \mathcal{P}(\mathbf{N})\big| = \big|\mathcal{P}(\mathbf{N})\big|.$$

■

---

and figure out which of the two roots lies in $(-1,1)$; the answer is $\frac{-1+\sqrt{1+y^2}}{y}$. We could leave the definition of $\tilde{g}(y)$ in this form for $y \neq 0$ and separately define $\tilde{g}(0) = 0$. However, multiplying numerator and denominator of our "$y \neq 0$" formula by $1 + \sqrt{1+y^2}$ and doing some cancellation, we arrive at $\tilde{g}(y) = \frac{y}{1+\sqrt{1+y^2}}$, which very conveniently gives the right value of $\tilde{g}(0)$ as well, and "more obviously" lies in $(-1,1)$ than does $\frac{-1+\sqrt{1+y^2}}{y}$.