The Size of the Cantor Set

Washington University Math Circle

November 6, 2016

In mathematics, a *set* is a collection of things called *elements*. For example, $\{1, 2, 3, 4\}$, {a,b,c,...,z}, and $\{\text{cat}, \text{dog}, \text{chicken}\}$ are sets. Most often, mathematicians are interested in sets whose elements are numbers. Some important examples of such sets are below:

1. The *natural numbers*, denoted $\mathbb{N}$, is the set $\{1, 2, 3, \ldots\}$.

2. The *integers*, denoted $\mathbb{Z}$, is the set $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$.

3. The *rational numbers*, denoted $\mathbb{Q}$, is the set of fractions. Specifically, $\mathbb{Q} = \{\frac{a}{b}\}$ where a and b are integers and $b \neq 0$.

4. The *irrational numbers* is the set of real numbers that are not rational, that is, the set of real numbers that cannot be expressed as a fraction. Some examples of irrational numbers are $\sqrt{2}, 17^{1/3}$, and π.

5. The *real numbers*, denoted $\mathbb{R}$, is the set of all numbers that can be written as a decimal. This set can also be written as $(-\infty, \infty)$ and is the set of all numbers that are either rational or irrational.

There are many ways to think about the "size" of a set. In this workshop, we will discuss two of these ways – cardinality and measure. We can think of a set's cardinality as the number of elements it contains and we can think of a set's measure as its length. We will discuss both of these notions in detail, then determine the cardinality and measure of some particularly interesting sets.

Cardinality

If A is a set, we denote its cardinality by $|A|$. For finite sets, the notion of cardinality is easy to understand – it is just the number of elements in the set. For example $|\{1, 2, 3, 4\}| = 4$, $|\{a, b, c, \ldots, z\}| = 26$, and $|\{\text{cat}, \text{dog}, \text{chicken}\}| = 3$. For infinite sets, we need to be a bit more clever to understand cardinality.

Two sets A and B have the same cardinality (written $|A| = |B|$) if there is a way to match up their elements with a *bijective correspondence*. In other words, we are able to pair the elements of A with the elements of B in a way where *every* element of A is paired with *exactly one* element of B. Here are a couple examples:

1. $|\{1, 2, 3\}| = |\{\text{cat}, \text{dog}, \text{chicken}\}|$ since there is the bijective correspondence $(1, \text{cat})$, $(2, \text{dog})$, $(3, \text{chicken})$.

2. $|\{a, b, c, \ldots, z\}| = |\{$ways to roll a pair of 6-sided dice such that the sum of the dice values is greater than 5$\}|$

Exercise: Find a bijective correspondence between the two sets in example 2 above.

Exercise: Think of two sets that have the same cardinality and write down a bijective correspondence between them.

As can be expected, finding bijective correspondences can be more difficult when the sets are infinite. The first three sets discussed in this workshop were $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$. Upon first consideration, we might expect that the cardinality of $\mathbb{N}$ is smaller than the cardinality of $\mathbb{Z}$ since every element in $\mathbb{N}$ is also in $\mathbb{Z}$, but there are elements of $\mathbb{Z}$ that are not elements of $\mathbb{N}$ (such as $0, -1, -2, \ldots$). For the same reasons, we might expect that the cardinality of $\mathbb{Z}$ is smaller than the cardinality of $\mathbb{Q}$. However, this is not the case. We will now go through an activity that shows $|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}|$.

Hilbert's Hotel: A famous mathematician named David Hilbert owns a hotel that has an infinite number of rooms that are labeled by the numbers $1, 2, 3, \ldots$. This hotel is very popular, and so every room is occupied.

Exercise: If a new guest arrives and wishes to be accommodated in the hotel, how can we accommodate this guest without forcing anyone else to leave the hotel?

Exercise: How can we accommodate a group of infinitely many new guests (Guest 1, Guest 2, ...) in the hotel without forcing anyone else to leave the hotel?

Exercise: How can we accommodate infinitely many groups (Group 1, Group 2, ...) each with infinitely many guests (for Group 1: Guest (1,1), Guest (1,2), Guest (1,3), ..., for Group 2: Guest (2,1), Guest (2,2), Guest (2,3), ..., and so on) in the hotel without forcing anyone to leave?

The second exercise above exhibits a bijective correspondence between $\mathbb{N}$ and $\mathbb{Z}$, showing $|\mathbb{N}| = |\mathbb{Z}|$. Meanwhile, the third exercise above exhibits a bijective correspondence between $\mathbb{N}$ and $\mathbb{Q}$, showing $|\mathbb{N}| = |\mathbb{Q}|$. Thus $|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}|$. It turns out (see challenge exercises on the last page) that $|\mathbb{N}| \neq |\mathbb{R}|$. It is another fact that $|[0,1]| = |\mathbb{R}|$. Sets that are either finite of have the same cardinality as $\mathbb{N}$ are called *countable* sets, and all other sets are called *uncountable*.

Measure

If A is a set of real numbers, we denote its measure by $m(A)$. The measure of a set of real numbers can be thought of as the length of the set. For this reason, the measure of any set must be at least 0. The open interval (a, b) (the set of real numbers that are greater than a and less than b) has measure $b - a$. We also define the measure of the closed interval $[a, b]$ (the set of real numbers that are greater than *or equal to* a and less than *or equal to* b) to be $b - a$.

We can use the following property to understand the measure of more complicated subsets of real numbers: if A and B are two sets of real numbers that share no common elements, then the measure of the set of all elements that are in A or in B (written $A \cup B$) is equal to the measure of A plus the measure of B. In mathematical notation, this is written $m(A \cup B) = m(A) + m(B)$.

Exercise:

1. Find the measure of $(-2, 5)$.

2. Find the measure of $(1, 3) \cup (2, 6)$.

3. Find the measure of $\{3\}$.

4. Find the measure of $\{1, 2, 3, \ldots, n\}$ where n is any positive integer.

Our goal is to discuss the cardinality and measure of a couple interesting sets. In order to define these sets, we must first discuss how to represent real numbers in different base systems.

Real Numbers in Different Bases

We usually represent numbers using the ten digits $0, 1, 2, \ldots, 9$. This is a *base 10* number system since we use ten digits. The base 10 system is commonly called the *decimal system* (the prefix "dec-" means 10). In this system we represent a real number by the coefficients when expressing the number as a sum of powers of 10. Written mathematically, this means that to represent a number x in the decimal system, we write

$$x = a_n * 10^n + a_{n-1} * 10^{n-1} + \cdots + a_1 * 10^1 + a_0 * 10^0 + a_{-1} * 10^{-1} + a_{-2} * 10^{-2} \cdots ,$$

then represent x as $(a_n a_{n-1} \ldots a_1 a_0 . a_{-1} a_{-2} \ldots)$. For example,

$$352.84 = 3 * 10^2 + 5 * 10^1 + 2 * 10^0 + 8 * 10^{-1} + 4 * 10^{-2}.$$

It is possible to represent real numbers using a different number of possible digits. For example, a number system that only has two digits is called the *binary* system and a number system that has 16 digits is called the *hexadecimal system.* For any positive integer b, we can describe a real number x in the base b number system in the following way: write

$$x = a_n * b^n + a_{n-1} * b^{n-1} + \cdots + a_1 * b^1 + a_0 * b^0 + a_{-1} * b^{-1} + a_{-2} * b^{-2} \cdots ,$$

then represent x as $(a_n a_{n-1} \ldots a_1 a_0 . a_{-1} a_{-2} \ldots)_b$. For example, to express 27.25 in binary, write

$$27.25 = 1 * 2^4 + 1 * 2^3 + 0 * 2^2 + 1 * 2^1 + 1 * 2^0 + 0 * 2^{-1} + 1 * 2^{-2},$$

so $27.25 = (11011.01)_2$.

Exercise: Express $99.\overline{333}\ldots$ in base 3.

Exercise: Express $(1302)_4$ (given in base 4) in base 7.

The Cantor Set

The Cantor set is a famous set of real numbers named after the mathematician Georg Cantor. We denote the Cantor set by $\mathcal{C}$ and construct it in the following way:

Exercise: Draw a picture for each step below to visualize the Cantor set.

1. Start with the closed interval $[0, 1]$.

2. Remove the middle third of this interval $\left(\frac{1}{3}, \frac{2}{3}\right)$. You should be left with $\left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$.

3. Remove the middle thirds of the remaining two intervals. You should be left with $\left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{3}{9}\right] \cup \left[\frac{6}{9}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right]$.

4. Remove the middle thirds of the remaining four intervals.

5. Continue this process infinitely many times. The points that remain are the elements of the Cantor set.

Cardinality of the Cantor Set

Below is a picture of the construction of the Cantor set.

A more formal way to describe the Cantor set uses the base 3 number system. In this way, the Cantor set is defined as

$$\mathcal{C} = \{x : 0 \le x \le 1 \text{ and the base 3 representation of } x \text{ does not contain the digit } 1\}.$$

In other words, the Cantor set is the set of numbers x in the interval $[0, 1]$ where when we write $x = (0.a_{-1}a_{-2}\ldots)_3$, all of the digits $a_{-1}, a_{-2}, \ldots$ are either 0 or 2. For example, $\frac{8}{27} = (0.022)_3$ is in the Cantor set, but $\frac{5}{27} = (0.012)_3$ is not.

We will next determine the cardinality of the Cantor set. It turns out that the Cantor set is uncountable. In particular, $|\mathcal{C}| = |[0, 1]|$. So in the sense of cardinality, the Cantor set and the closed interval $[0, 1]$ have the same size.

Exercise: Use the definition of the Cantor set involving base 3 representations of numbers to show $|\mathbb{N}| \neq |\mathcal{C}|$.

1. Try to list out all of the elements of $\mathcal{C}$ in a bijective correspondence with $\mathbb{N}$. For example, write $c_1 = (0.a_{1,-1}a_{1,-2}\ldots)_3$, $c_2 = (0.a_{2,-1}a_{2,-2}\ldots)_3$, $\ldots$ where the numbers c_i represent all of the elements of the Cantor set.

2. Find an element of the Cantor set that is not in your list above by considering the digits $a_{1,-1}, a_{2,-2}, a_{3,-3}, \ldots$.

The exercise on the previous page shows that there cannot be a bijective correspondence between $\mathbb{N}$ and $\mathcal{C}$. This means that $\mathcal{C}$ is uncountable. Since every element of $\mathcal{C}$ is also an element of $[0,1]$, $|\mathcal{C}| = |[0,1]|$. So in the sense of cardinality, the Cantor set and $[0,1]$ have the same size.

We will next discuss the measure of the Cantor set. We will determine that the measure of the Cantor set is 0, although the measure of $[0,1]$ is 1. Therefore, in the sense of measure, the Cantor set is much smaller than $[0,1]$. The fact that the Cantor set is either the same size as $[0,1]$ or smaller than $[0,1]$ depending on which type of “size” we look at is one reason mathematicians are fascinated with this set.

Measure of the Cantor Set

Exercise: Refer to the construction of the Cantor set on page 6. We will call the closed interval $[0,1]$ the 0^{th} iteration of the Cantor set, the set after the first middle third is removed the 1^{st} iteration of the Cantor set, the set after the next two middle thirds are removed the 2^{nd} iteration of the Cantor set, and so on.

1. Find the measure of the first four iterations of the Cantor set.

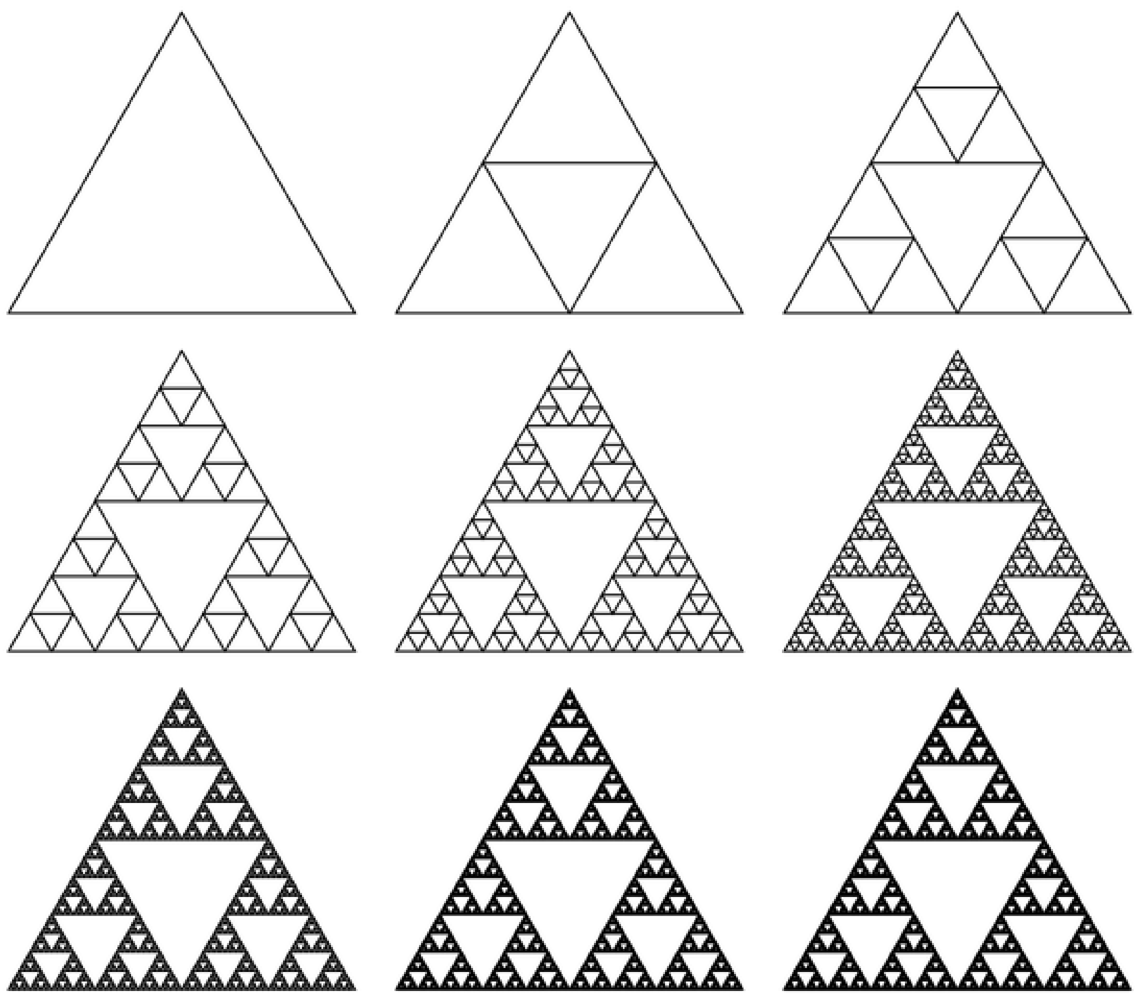
2. What is the measure of the n^{th} iteration of the Cantor set?

3. What happens to the formula for the measure of the n^{th} iteration of the Cantor set as the value for n gets bigger and bigger?

4. Determine the measure of the Cantor set.

Area of the Sierpinski Triangle

The Sierpinski triangle is a two-dimensional analogue of the Cantor set. Its construction is illustrated below. Begin with an equilateral triangle with side lengths 1, remove the middle equilateral triangle to get the first iteration, remove the middle equilateral triangles from the remaining three triangles to get the second iteration, and so on. Like with the Cantor set, we will show that the measure of the Sierpinski triangle is 0. In two dimensions, measure can be thought of as the area of an object, so we are in fact showing that the area of the Sierpinski triangle is 0.



Exercise: Refer to the construction of the Sierpinski triangle on the previous page. We will call the equilateral triangle with side lengths 1 the 0^{th} iteration of the Sierpinski triangle, the object after the first middle triangle is removed the 1^{st} iteration of the Sierpinski triangle, the object after the next three middle triangles are removed the 2^{nd} iteration of the Sierpinski triangle, and so on.

1. Find the area of the first four iterations of the Sierpinski triangle.

2. What is the area of the n^{th} iteration of the Sierpinski triangle?

3. What happens to the formula for the area of the n^{th} iteration of the Sierpinski triangle as the value for n gets bigger and bigger?

4. Determine the area of the Sierpinski triangle.

Challenge Exercises

Exercise 1: Explain why $\sqrt{2}$ is irrational. (*Hint:* Assume that $\sqrt{2} = \frac{a}{b}$ where a and b are integers and $b \neq 0$, then logically conclude something you know to be false.)

Exercise 2: A set A is a *subset* of another set B (written $A \subseteq B$) if every element of A is also an element of B. For example, $\{\text{cat}, \text{chicken}\} \subseteq \{\text{cat}, \text{dog}, \text{chicken}\}$ and $\mathbb{N} \subseteq \mathbb{Z}$. The *power set* of a set A (written $\mathcal{P}(A)$) is the collection of all subsets of A. For example, $\mathcal{P}(\{\text{cat}, \text{chicken}\}) = \{\{\}, \{\text{cat}\}, \{\text{chicken}\}, \{\text{cat}, \text{chicken}\}\}$.

What is $|\mathcal{P}(\{1, 2, \ldots, n\})|$ (the cardinality of the power set of $\{1, 2, \ldots, n\}$)? First try for $n = 1$, $n = 2$, and $n = 3$, then find a formula in general.

Exercise 3: Explain why $b + b^2 + b^3 + \cdots + b^n = b^{n+1} - b$ for any positive integers b and n.