Trevor, Angel, and Michael

Measure Zero, the Cantor Set, and the Cantor Function

Measure Zero:

<u>Definition</u>: Let X be a subset of R, the real number line, X has <u>measure zero</u> if and only if $\forall\ \varepsilon > 0$ $\exists$ a set of open intervals, $\{I_1,\ldots,I_k\}$, $1 \le k \le \infty$,

such that (i) $X \subseteq \bigcup_{k=1}^{\infty} I_k$ and (ii) $\sum_{k=1}^{\infty} |I_k| \le \varepsilon$.

<u>Theorem 1</u>: If X is a finite set, X a subset of R, then X has measure zero.

Proof:

Let $X = \{x_1, \ldots, x_N\}$, $N \ge 1$

Given $\varepsilon > 0$ let $I_k = (x_k - \varepsilon/2N, x_k + \varepsilon/2N)$ be the kth interval.

$$\Rightarrow X \subseteq \bigcup_{k=1}^{N} I_k$$

and

$$\Rightarrow \sum_{k=1}^{N} |I_k| = \sum_{k=1}^{N} |x_k + \varepsilon/2N - x_k + \varepsilon/2N| = \sum_{k=1}^{N} |\varepsilon/N| = \sum_{k=1}^{N} \varepsilon/N = N\varepsilon/N = \varepsilon \le \varepsilon$$

Therefore if X is a finite subset of **R**, then X has measure zero.

<u>Theorem 2</u>: If X is a countable subset of R, then X has measure zero.

Proof:

Let $X = \{x_1, x_2, \ldots\}$, $x_i \in \mathbf{R}$

Given $\varepsilon > 0$ let $I_k = (x_k - \varepsilon/2^{n+1}, x_k + \varepsilon/2^{n+1})$ be the kth interval

$$\Rightarrow X \subseteq \bigcup_{k=1}^{\infty} I_k$$

and

$$\Rightarrow \sum_{k=1}^{\infty} |I_k| = \sum_{k=1}^{\infty} |x_k + \varepsilon/2^{k+1} - x_k + \varepsilon/2^{k+1}| = \sum_{k=1}^{\infty} |\varepsilon/2^k| = \varepsilon \sum_{k=1}^{\infty} 1/2^k = \varepsilon\left(\sum_{k=0}^{\infty} 1/2^k - 1\right) = \varepsilon(2-1) = \varepsilon \le \varepsilon$$

Therefore if X is a countable subset of **R**, then X has measure zero. A famous example of a set that is not countable but has measure zero is the Cantor Set, which is named after the German mathematician Georg Cantor (1845-1918).

Cantor Set:

We start with a closed interval from 0 to 1:

We remove the middle third of the interval, by removing an open interval in the middle of the line:

Now we remove the middle third of the remaining intervals:

And, we remove the middle third of those intervals:

We continue this process of removing the middle third of the remaining intervals. The points that are never removed from the interval [0,1] form the Cantor Set. There are an infinite number of points in the Cantor Set.

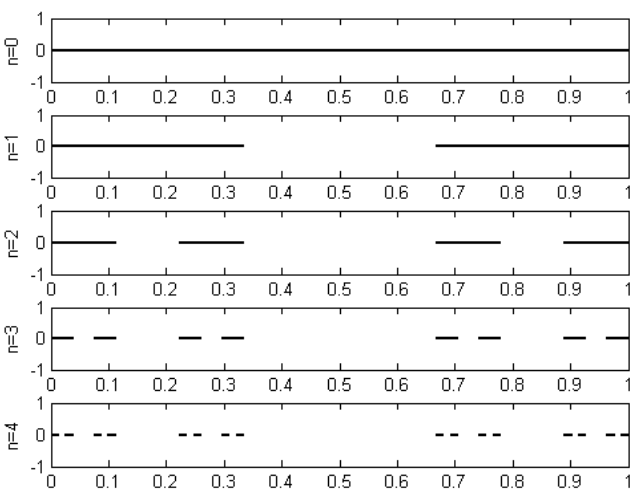

Figure 1: MatLab output for n=0,...,4

$C_0 = [0, 1]$

$\cup$

$C_1 = [0,\ 1/3\ ] \cup [2/3,\ 1]$

$\cup$

$C_2 = [0, 1/9] \cup [2/9, 3/9] \cup [6/9, 7/9] \cup [8/9, 9/9]$

$\cup$

$C_3 = \bigcup_{k=1}^{8} I_{3,k},$ *where* $I_{3,k}$ *are the subintervals that remain in* $[0, 1]$

$\cup$

$C_4 = \bigcup_{k=1}^{16} I_{4,k}$

$\cup$

.

.

.

$\cup$

$$C_N = \bigcup_{k=1}^{2^N} I_{N,k}$$

We do this infinitely many times, thus let N-->∞, and name the

$Cantor\ Set\ C_\infty = \bigcap_{n=0}^{\infty} C_n$: the Cantor set is the intersection of the sets from all of the iterations.

Measure of the Cantor Set:

In **R,** we can use the Euclidean distance to measure the length of the interval from 0 to 1, which has the length 1.

Now let's look at the length of what we have remaining in [0,1] at each step:

Step n	(k) Number of subintervals remaining	Length of one subinterval $\mid I_{n,k} \mid$	total length $\sum_{k=1}^{2^n} \mid I_{n,k} \mid$
0	1	1	1
1	2	1/3	2/3
2	4	1/9	4/9
3	8	1/27	8/27
4	16	1/81	16/81
N	2^N	$1/3^N$	$(2/3)^N$
$\lim_{N\to\infty}$	$\lim_{N\to\infty} 2^N \to \infty$	$\lim_{N\to\infty} 1/3^N \to 0$	$\lim_{N\to\infty} (2/3)^N \to 0$

CLAIM: The Cantor Set has measure zero.

PROOF: Given $\varepsilon > 0$ (arbitrarily small), choose N_* = ln(ε /3)/ln(2/3), such that $\forall$ N $\geq N_*$, at the Nth step we have the closed intervals $I_{N,k} = [a_k, b_k]$, k=1,..., 2^N where $|I_{N,k}| = |b_k - a_k|$ and $\sum_{k=1}^{2^N} \mid I_{N,k} \mid$ = $(2/3)^N$.

Let $J_{N,k} = (a_k - \frac{(2/3)^N}{2^N}, b_k + \frac{(2/3)^N}{2^N})$ be the open interval such that $I_{N,k} \subset J_{N,k}$, $\forall k$, and such that $|J_{N,k}| = |b_k + \frac{(2/3)^N}{2^N} - a_k + \frac{(2/3)^N}{2^N}| = |b_k - a_k + \frac{(2/3)^N}{2^{N-1}}|$.

Since $C_\infty \subset \bigcup_{k=1}^{2^N} I_{N,k}$ it follows that $C_\infty \subset \bigcup_{k=1}^{2^N} J_{N,k}$ (satisfying (i)).

Now let's consider $\sum_{k=1}^{2^N} \mid J_{N,k} \mid = \sum_{k=1}^{2^N} |b_k - a_k + \frac{(2/3)^N}{2^{N-1}}| \leq \sum_{k=1}^{2^N} \mid I_{N,k} \mid + \sum_{k=1}^{2^N} \frac{(2/3)^N}{2^{N-1}} =$

$= (2/3)^N + 2^N \frac{(2/3)^N}{2^{N-1}} = 3(2/3)^N = 3e^{N ln(2/3)} = 3e^{ln(\varepsilon/3)} = 3(\varepsilon/3) = \varepsilon$ (satisfying (ii)).

Therefore, by the definition, the Cantor Set has measure zero. ∎

By Corollary 1.5.9 in Book II (Tao) proved in Joclyn and Michael's presentation, since the sets $C_0, C_1, C_2, \ldots$ are a sequence of nonempty compact subsets of **R** such that $C_0 \supset C_1 \supset C_2 \supset \ldots$, then the intersection $\bigcap_{n=0}^{\infty} C_n$ is non-empty. That is, the Cantor Set is a non-empty set.

The endpoints of each $I_{N,k}$ are in the Cantor Set C_∞:

$$0, 1, \tfrac{1}{3}, \tfrac{2}{3}, \tfrac{1}{9}, \tfrac{2}{9}, \ldots \in C_\infty.$$

All of the points in C_∞ are of the form $a/3^n$ for some integers a and n. In fact, each point x in C_∞ can be written $\sum_{n=1}^{\infty} 2a_n/3^n$, where each a_n is either 0 or 1.

For example:

0 = 0 (base 3)

1 = 1.0 (base 3) = 0.2222… (base 3) = $\sum_{n=1}^{\infty} 2/3^n = 2(\frac{1/3}{1-1/3}) = 2(1/2) = 1$

$\frac{1}{3}$ = 0.1 (base 3) = 0.0222… (base 3) = $\sum_{n=2}^{\infty} 2/3^n = 2(\frac{1/9}{1-1/3}) = 2(1/6) = 1/3$

$\frac{2}{3}$ = 0.2 (base 3)

$\frac{1}{9}$ = 0.01 (base 3) = 0.0022… (base 3) = $\sum_{n=3}^{\infty} 2/3^n = 2(\frac{1/27}{1-1/3}) = 2(1/18) = 1/9$

$\frac{2}{9}$ = 0.02 (base 3)

⋮

Thus the Cantor Set consists of the points in [0,1] having some base 3 decimal representation that does not include the digit 1 (Carothers, ch 2).

To see that the Cantor Set is uncountable, we will use Cantor's diagonal argument (en.wikipedia.org). First let $x_i \in C_\infty$, for i = 1,2,3,...

$x_i = 0.a_1,a_2,a_3,\ldots$ where a_n=0,2.

Now let's try to enumerate:

x_1 = 0.00000000000...
x_2 = 0.22222222222...
x_3 = 0.02020202020...
x_4 = 0.20202020202...
x_5 = 0.22020220202...
x_6 = 0.00220220220...
x_7 = 0.20002000200...
x_8 = 0.00220022002...
x_9 = 0.22002200220...
x_{10} = 0.22022200202...
x_{11} = 0.22020200200...
⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋱

By construction x differs from each x_n, since their n^{th} digits differ (highlighted).

x = 0.20222020022...

Hence, *x* cannot occur in the enumeration. Consequently, it is impossible to enumerate the Cantor Set points; they are uncountable.

We can now see that **the Cantor Set is not countable but has measure zero**, that is, it is simultaneously large and meager.

Cantor Set 2D example:

The Sierpinski Carpet is a Cantor Set in 2D. The square [0,1]x[0,1] is viewed as having 9 sub-squares. We remove the middle most square (open set). To recurse, we remove the middle square of each of the remaining 8 sub-squares. And, repeat. (See Figure 2.)

Area:

We can use the Euclidean distance to see the area of the square from [0,1]x[0,1] is 1. This time we will measure the area of the original square (Area=1) and subtract the area of the middle square.

Now let's look at the area of what we have removed:

Step n	Number of sub-squares removed	Length of one sub-square
1	1	$(1/3)^2$
2	8	$(1/9)^2$
3	8^2	$(1/27)^2$
4	8^3	$(1/81)^2$
i	8^{i-1}	$(1/3^i)^2$

To find the total area of the removed sub-squares we add up the number of sub-squares removed per step multiplied by the area of one such sub-square:

$\sum_{i=1}^{\infty} 8^{i-1} * (\frac{1}{3^i})^2 = \sum_{i=1}^{\infty} \frac{8^{i-1}}{3^{2i}} = \frac{1}{9}\sum_{j=0}^{\infty} (\frac{8}{9})^j$; we have a convergent geometric series.

$\frac{1}{9}\sum_{j=0}^{\infty} (\frac{8}{9})^j = \frac{1}{9}(\frac{1}{1-\frac{8}{9}}) = \frac{\frac{1}{9}}{\frac{1}{9}} = 1$; we can see that the sum of the areas of the sub-squares that we removed **total to area 1**, the area of the original square. Therefore, the area of the Sierpinski Carpet is zero. This Cantor Set in 2D (which has infinitely many points) is another example of a set with measure zero.

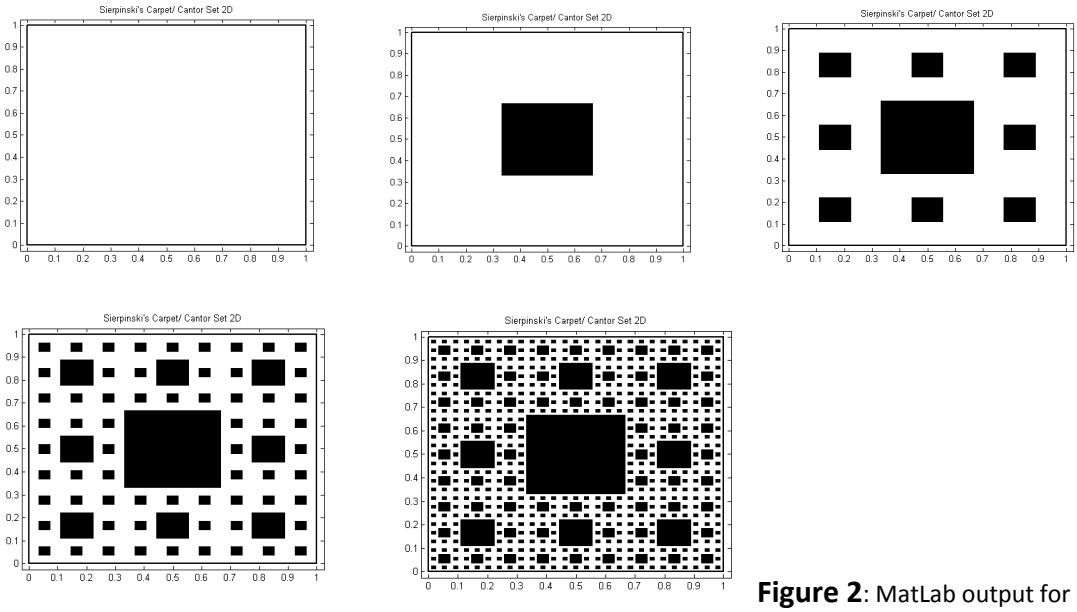

Figure 2: MatLab output for n=0,...,4

In the long run the white part of the square will be so thin that it will not take up any area. There will be individual white strings in the carpet, but their area will equal 0. A line in 2D has no area, as a points in 1D had no length.

The Sierpinski Triangle is another 2D version of a set with measure zero.

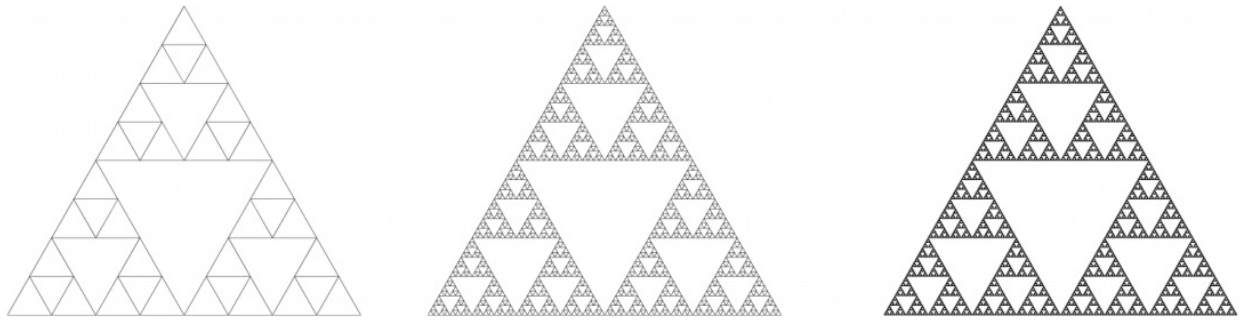

Figure 3: n=3,6,9 <http://www.intechopen.com/source/html/39360/media/image1.jpeg>

Cantor Function:

The Cantor function is a function which is continuous everywhere, differentiable almost everywhere, with a derivative of 0 where it exists, and increases from 0 to 1 on the interval [0,1]. Its construction and a proof of its continuity is provided below.

Construction of the Cantor Function:

To construct the Cantor function, we create a uniformly Cauchy sequence of continuous increasing functions, where F_m is a piecewise function which is always 0 at 0 and 1 at 1, linear on the cantor set, and has values of $F_m(x) = (2k-1)/2^n,\ n \leq m$, where $x \in I_{n,k}$ and $I_{n,k}$ is the kth interval removed from [0,1] to create C_n. A visual of the first 3 iterations is shown below, as well as another visual of the Cantor function, also known as the Cantor-Lebesgue function, or the devil's staircase.

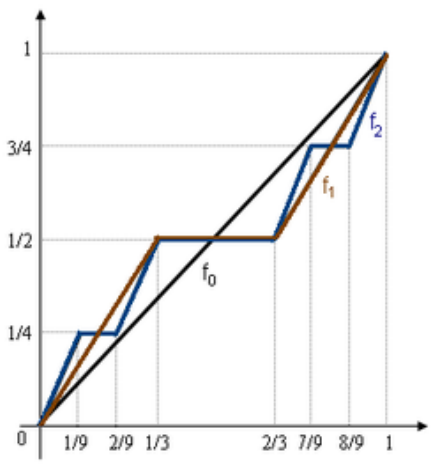

Figure 4: The first three Cantor function iterations.

<http://upload.wikimedia.org/wikipedia/commons/thumb/e/e1/Cantor_function_sequence.png/250px-Cantor_function_sequence.png>

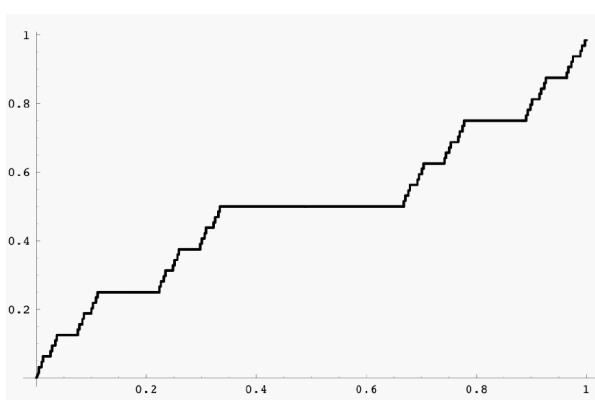

Figure 5: The Cantor function.

<http://www.math.harvard.edu/~ctm/home/text/class/harvard/114/07/html/cantor.gif>

Proof that F_m is Cauchy: adapted from <http://www.math.wustl.edu/~luthy/2013_F/cantor.pdf>

By construction, for $|F_{n+1}(x) - F_n(x)|$, it suffices to, without loss of generality, consider only the interval $[0, \frac{1}{3^n}]$. This is caused by the symmetry of F_n; where F_n is larger than F_{n+1} on the first interval, it is less than F_{n+1} by the same amount on the second interval, and this pattern is repeated. This may be seen in Figure 4.

On $[0, \frac{1}{3^n}]$, $F_n(x) = \frac{3^n}{2^n}x$, and $F_{n+1}(x)$ on $[0, \frac{1}{3^n}]$ is given by

$$F_{n+1}(x) = \begin{cases} \frac{3^{n+1}}{2^{n+1}}x & if\ x \in [0, \frac{1}{3^{n+1}}] \\ \frac{1}{2^{n+1}} & if\ x \in [\frac{1}{3^{n+1}}, \frac{2}{3^{n+1}}] \\ \frac{-1}{2^{n+1}} + \frac{3^{n+1}}{2^{n+1}}x & if\ x \in [\frac{2}{3^{n+1}}, \frac{3}{3^{n+1}}] \end{cases}$$

So $|F_{n+1}(x) - F_n(x)| \leq \frac{1}{2^{n+1}}$ for all $x \in [0, \frac{1}{3^n}]$, as the difference on each sub-interval obeys such.

So $\sup_{x \in [0,1]} |F_{n+1}(x) - F_n(x)| \leq \frac{1}{2^{n+1}}$.

So for all k > 0 and x ∈ [0 ,1],

$$\begin{aligned} |F_{n+k}(x) - F_n(x)| &\leq |F_{n+k}(x) - F_{n+k-1}(x)| + |F_{n+k-1}(x) - F_{n+k-2}(x)| \\ &\leq 2^{-(n+k+1)} + |F_{n+k-1}1(x) - F_n(x)| \\ &\leq 2^{-(n+k+1)} + |F_{n+k-1}1(x) - F_{n+k-2}(x)| + |F_{n+k-2}(x) - F_n(x)| \\ &\leq 2^{-(n+k+1)} + 2^{-(n+k)} + |F_{n+k-2}(x) - F_n(x)| \\ &\leq \dots \\ &\leq \sum_{j=n-1}^{n+k+1} 2^{-j} \\ &\leq \sum_{j=n-1}^{\infty} 2^{-j} \\ &= 2^{-n} \end{aligned}$$

Thus $|F_m(x) - F_n(x)| \leq 2^{-min\{m,n\}}$.

So $\forall \varepsilon\ \exists M\ s.t. \forall n, m \geq M, |F_m(x) - F_n(x)| \leq \varepsilon$, namely $M = \log_2 \varepsilon$.

So F_m is Cauchy, and the Cantor-Lebesgue function is the limit of a convergent sequence of functions.

Proof of Continuity:

Proof of continuity of Cantor – Lebesgue function:

By cases:

1) Let $x, x' \in C$.

Then x may be written as $\sum_{n=1}^{\infty} \frac{2a_n}{3^n}$, where $a_n = 0,1$

and x' may be written as $\sum_{n=1}^{\infty} \frac{2b_n}{3^n}$, where $b_n = 0,1$.

Let $|f(x) - f(x')| < \varepsilon$. Then $\left|\sum_{n=1}^{\infty} \frac{a_n}{2^n} - \sum_{n=1}^{\infty} \frac{b_n}{2^n}\right| < \varepsilon$, where $a_n, b_n = 0,1$.

So $\exists_{M>0}$ such that $\forall_{n \leq M}\ a_n = b_n$, and $\left|\sum_{n=M}^{\infty} \frac{a_n}{2^n} - \sum_{n=M}^{\infty} \frac{b_n}{2^n}\right|$

So $\left|\sum_{n=M}^{\infty} \frac{a_n - b_n}{2^n}\right| \leq \sum_{n=M}^{\infty} \left|\frac{1}{2^n}\right| = \frac{1}{2^{M-1}} < \varepsilon$, so $1 - M < \log_2 \varepsilon$.

Which is equivalent to $|x - x'| = \left|\sum_{n=M}^{\infty} \frac{2a_n}{3^n} - \sum_{n=M}^{\infty} \frac{2b_n}{3^n}\right| \leq \sum_{n=M}^{\infty} \left|\frac{2}{3^n}\right| = \frac{1}{3^{M-1}} = 3^{1-M} < 3^{\log_2 \varepsilon} = \delta$

So $\forall_{x,x' \in C,\ \varepsilon>0}\ \exists_{\delta}$ such that if $|x - x'| < \delta$, then $|f(x) - f(x')| < \varepsilon$, namely $\delta = 3^{\log_2 \varepsilon}$

So f continuous on C.

2) Let $x \in [0,1] \backslash C$. Then $x \in I = (c, d)$and $f(x) = f(x')\ \forall_{x,x' \in (c,d)}$ by construction.

So f is continuous on $[0,1] \backslash C$.

Also,

$f(c) = f(d) = f(x) \forall_{x \in (c,d)}$ by construction, so the endpoints match up and f is continuous on $[0,1]$.

Sources:

1. Carothers, N. L. (2000). *Real analysis*. Cambridge [UK]: Cambridge University Press.
2. Pereyra, M. C., & Ward, L. A. (2012). *Harmonic analysis: From fourier to wavelets*. Providence, R.I: American Mathematical Society.
3. Stein, E. M., & Shakarchi, R. (2005). *Real analysis: Measure theory, integration, and Hilbert spaces*. Princeton, N.J: Princeton University Press.
4. <http://en.wikipedia.org/wiki/Cantor%27s_diagonal_argument>
5. <http://www.math.wustl.edu/~luthy/2013_F/cantor.pdf>