The bigness of things

Vaughn Climenhaga

University of Houston

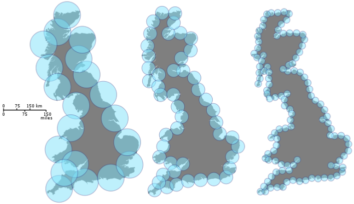


Image from Wikipedia

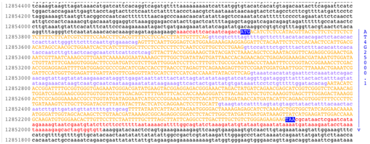


Image from Wikipedia

How big is it?

Meaning of “big” depends on what “it” is, and why we care.

How big is it?

Meaning of “big” depends on what “it” is, and why we care.

How big is . . .

a crowd of people?

How big is it?

Meaning of “big” depends on what “it” is, and why we care.

How big is . . .

a crowd of people? number weight

How big is it?

Meaning of “big” depends on what “it” is, and why we care.

How big is . . .

a crowd of people?	number	weight
a fish?		

How big is it?

Meaning of “big” depends on what “it” is, and why we care.

How big is . . .

a crowd of people?	number	weight
a fish?	length	weight

How big is it?

Meaning of "big" depends on what "it" is, and why we care.

How big is . . .

a crowd of people?	number	weight
a fish?	length	weight
a city?		

How big is it?

Meaning of “big” depends on what “it” is, and why we care.

How big is . . .

a crowd of people?	number	weight	
a fish?	length	weight	
a city?	# of people	diameter	area

How big is it?

Meaning of “big” depends on what “it” is, and why we care.

How big is ...

a crowd of people?	number	weight	
a fish?	length	weight	
a city?	# of people	diameter	area
a house?			

How big is it?

Meaning of “big” depends on what “it” is, and why we care.

How big is . . .

a crowd of people?	number	weight	
a fish?	length	weight	
a city?	# of people	diameter	area
a house?	# of bedrooms	area	volume

How big is it?

Meaning of "big" depends on what "it" is, and why we care.

How big is . . .

a crowd of people?	number	weight	
a fish?	length	weight	
a city?	# of people	diameter	area
a house?	# of bedrooms	area	volume
an assignment?			

How big is it?

Meaning of "big" depends on what "it" is, and why we care.

How big is . . .

a crowd of people?	number	weight	
a fish?	length	weight	
a city?	# of people	diameter	area
a house?	# of bedrooms	area	volume
an assignment?	# of problems	time	

How big is it?

Meaning of “big” depends on what “it” is, and why we care.

How big is . . .

a crowd of people?	number	weight	
a fish?	length	weight	
a city?	# of people	diameter	area
a house?	# of bedrooms	area	volume
an assignment?	# of problems	time	
a book?			

How big is it?

Meaning of “big” depends on what “it” is, and why we care.

How big is . . .

a crowd of people?	number	weight	
a fish?	length	weight	
a city?	# of people	diameter	area
a house?	# of bedrooms	area	volume
an assignment?	# of problems	time	
a book?	# of pages	information	

How big is it?

Meaning of “big” depends on what “it” is, and why we care.

How big is . . .

a crowd of people?	number	weight	
a fish?	length	weight	
a city?	# of people	diameter	area
a house?	# of bedrooms	area	volume
an assignment?	# of problems	time	
a book?	# of pages	information	
Facebook?			

How big is it?

Meaning of “big” depends on what “it” is, and why we care.

How big is . . .

a crowd of people?	number	weight	
a fish?	length	weight	
a city?	# of people	diameter	area
a house?	# of bedrooms	area	volume
an assignment?	# of problems	time	
a book?	# of pages	information	
Facebook?	# of users	data	

How big is it?

Meaning of "big" depends on what "it" is, and why we care.

How big is . . .

a crowd of people?	number	weight	
a fish?	length	weight	
a city?	# of people	diameter	area
a house?	# of bedrooms	area	volume
an assignment?	# of problems	time	
a book?	# of pages	information	
Facebook?	# of users	data	
the internet?			

How big is it?

Meaning of “big” depends on what “it” is, and why we care.

How big is . . .

a crowd of people?	number	weight	
a fish?	length	weight	
a city?	# of people	diameter	area
a house?	# of bedrooms	area	volume
an assignment?	# of problems	time	
a book?	# of pages	information	
Facebook?	# of users	data	
the internet?	# of websites	data	useful data

Various notions of “bigness”

Concrete, familiar meanings of “big” from the previous slide:

0. cardinality
1. length
2. area
3. volume

Various notions of "bigness"

Concrete, familiar meanings of "big" from the previous slide:

0. cardinality
1. length
2. area
3. volume

or "weighted" versions:

$$\text{weight} = \int \text{density}\, d(\text{volume})$$

Various notions of “bigness”

Concrete, familiar meanings of “big” from the previous slide:

0. cardinality
1. length
2. area
3. volume

or “weighted” versions:

$$\text{weight} = \int \text{density}\, d(\text{volume})$$

More abstract meanings: “amount of data”?

- We are used to thinking of kB, MB, GB, TB, etc.
- But a 500 GB hard drive where every bit is set to ‘0’ doesn’t have much data on it. . .

Subsets of $\mathbb{R}^3$

Focus on familiar meanings for now. Consider some subsets of $\mathbb{R}^3$.

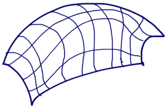
(a) finite set

0-dimensional

(b) curve

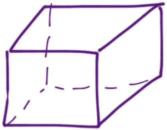
1-dimensional

(c) surface



2-dimensional

(d) open region



3-dimensional

Subsets of $\mathbb{R}^3$

Focus on familiar meanings for now. Consider some subsets of $\mathbb{R}^3$.

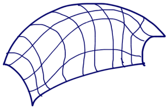
(a) finite set

0-dimensional

(b) curve

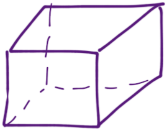
1-dimensional

(c) surface



2-dimensional

(d) open region



3-dimensional

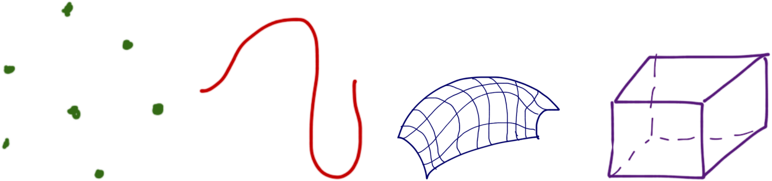
Cardinality:

- (a): Good way to measure how big a finite set is
- (b)–(d) have infinite cardinality

Subsets of $\mathbb{R}^3$

Focus on familiar meanings for now. Consider some subsets of $\mathbb{R}^3$.

(a) finite set (b) curve (c) surface (d) open region



0-dimensional 1-dimensional 2-dimensional 3-dimensional

Length:

- (a) has zero length. (Cover each point with tiny intervals)
- (b): Good way to measure how big a curve is
- (c)–(d) have infinite length: no curve of finite length can cover

Subsets of $\mathbb{R}^3$

Focus on familiar meanings for now. Consider some subsets of $\mathbb{R}^3$.

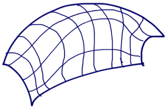
(a) finite set

0-dimensional

(b) curve

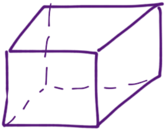
1-dimensional

(c) surface



2-dimensional

(d) open region



3-dimensional

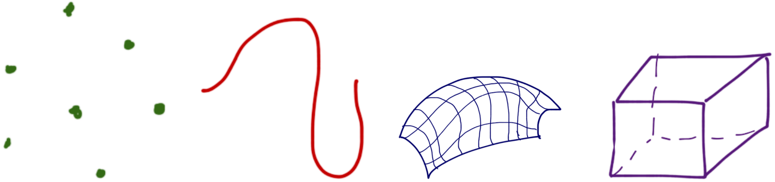
Area:

- (a)–(b) have zero area. (Cover with tiny discs)
- (c): Good way to measure how big a surface is
- (d) has infinite area

Subsets of $\mathbb{R}^3$

Focus on familiar meanings for now. Consider some subsets of $\mathbb{R}^3$.

(a) finite set (b) curve (c) surface (d) open region



0-dimensional 1-dimensional 2-dimensional 3-dimensional

Volume:

- (a)–(c) have zero volume
- (d): Good way to measure how big an open region is

Subsets of $\mathbb{R}^3$

Focus on familiar meanings for now. Consider some subsets of $\mathbb{R}^3$.

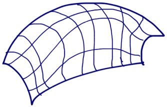
(a) finite set

0-dimensional

(b) curve

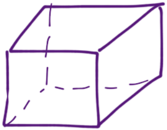
1-dimensional

(c) surface



2-dimensional

(d) open region



3-dimensional

Moral: To say how "big" a thing is, need to know its dimension.

Subsets of $\mathbb{R}^3$

Focus on familiar meanings for now. Consider some subsets of $\mathbb{R}^3$.

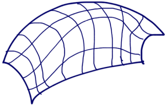
(a) finite set

0-dimensional

(b) curve

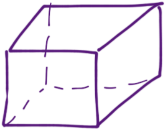
1-dimensional

(c) surface



2-dimensional

(d) open region



3-dimensional

Moral: To say how "big" a thing is, need to know its dimension.

- Dimension itself is a notion of bigness
- What is "dimension"? Seems to be which measure we use...

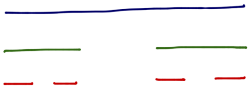
Example 1: A Cantor set

Consider the sets

$C_0 = [0, 1]$

$C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$

$C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$

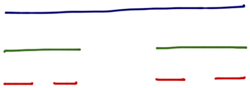


- C_n is disjoint union of 2^n intervals of length 3^{-n}
- Get C_{n+1} from C_n by removing middle third of each interval

Example 1: A Cantor set

Consider the sets

$C_0 = [0,1]$
$C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$
$C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$



- C_n is disjoint union of 2^n intervals of length 3^{-n}
- Get C_{n+1} from C_n by removing middle third of each interval
- The middle-third Cantor set is $C = \bigcap_{n \geq 0} C_n$.

Fact 1: C is infinite.

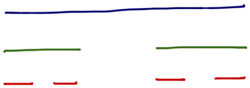
Example 1: A Cantor set

Consider the sets

$C_0 = [0, 1]$

$C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$

$C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$



- C_n is disjoint union of 2^n intervals of length 3^{-n}
- Get C_{n+1} from C_n by removing middle third of each interval
- The middle-third Cantor set is $C = \bigcap_{n \geq 0} C_n$.

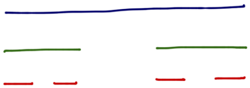
Fact 1: C is infinite.

- In fact, C is uncountable. (Bijection between $\{0, 1\}^{\mathbb{N}}$ and C)

Example 1: A Cantor set

Consider the sets

$C_0 = [0,1]$
$C_1 = [0,\frac{1}{3}] \cup [\frac{2}{3},1]$
$C_2 = [0,\frac{1}{9}] \cup [\frac{2}{9},\frac{1}{3}] \cup [\frac{2}{3},\frac{7}{9}] \cup [\frac{8}{9},1]$



- C_n is disjoint union of 2^n intervals of length 3^{-n}
- Get C_{n+1} from C_n by removing middle third of each interval
- The middle-third Cantor set is $C = \bigcap_{n\geq 0} C_n$.

Fact 1: C is infinite.

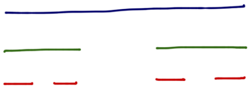
- In fact, C is uncountable. (Bijection between $\{0,1\}^{\mathbb{N}}$ and C)

Fact 2: C has zero length. (Length of C_n is $2^n 3^{-n} \to 0$)

Example 1: A Cantor set

Consider the sets

$C_0 = [0, 1]$
$C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$
$C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$



- C_n is disjoint union of 2^n intervals of length 3^{-n}
- Get C_{n+1} from C_n by removing middle third of each interval
- The middle-third Cantor set is $C = \bigcap_{n \geq 0} C_n$.

Fact 1: C is infinite.

- In fact, C is uncountable. (Bijection between $\{0, 1\}^{\mathbb{N}}$ and C)

Fact 2: C has zero length. (Length of C_n is $2^n 3^{-n} \to 0$)

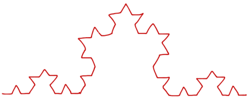
What is the dimension of C? Between 0 and 1.

Example 2: The Koch curve

Consider the curves

$K_0 =$ 

$K_1 =$ 

$K_2 =$ 

$\cdots$

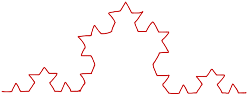
- K_n has 4^n line segments of length 3^{-n}
- Get K_{n+1} from K_n by replacing each line segment with a scaled-down copy of K_0

Example 2: The Koch curve

Consider the curves

$K_0 =$ 

$K_1 =$ 

$K_2 =$ 

$\ldots$

- K_n has 4^n line segments of length 3^{-n}
- Get K_{n+1} from K_n by replacing each line segment with a scaled-down copy of K_0
- The Koch curve is $K = \lim_{n\to\infty} K_n$

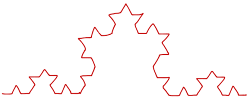
Fact 1: K has infinite length. (Length of K_n is $4^n 3^{-n}$)

Example 2: The Koch curve

Consider the curves

$K_0 =$ 

$K_1 =$ 

$K_2 =$ 

$\cdots$

- K_n has 4^n line segments of length 3^{-n}
- Get K_{n+1} from K_n by replacing each line segment with a scaled-down copy of K_0
- The Koch curve is $K = \lim_{n\to\infty} K_n$

Fact 1: K has infinite length. (Length of K_n is $4^n 3^{-n}$)

Fact 2: K has zero area. (Exercise – cover it with small rectangles)

What is dimension?

Algebraic idea: # of parameters/coordinates. (Always an integer!)

What is dimension?

Algebraic idea: # of parameters/coordinates. (Always an integer!)

More geometric idea: dimension is a scaling exponent.

Given $\lambda > 0$ and $E \subset \mathbb{R}^3$, let $\lambda E = \{\lambda \mathbf{x} \mid \mathbf{x} \in E\}$

- $\mathrm{volume}(\lambda E) = \lambda^3 \cdot \mathrm{volume}(E)$
- $\mathrm{area}(\lambda E) = \lambda^2 \cdot \mathrm{area}(E)$
- $\mathrm{length}(\lambda E) = \lambda^1 \cdot \mathrm{length}(E)$
- $\mathrm{cardinality}(\lambda E) = \lambda^0 \cdot \mathrm{cardinality}(E)$

What is dimension?

Algebraic idea: # of parameters/coordinates. (Always an integer!)

More geometric idea: dimension is a scaling exponent.

Given $\lambda > 0$ and $E \subset \mathbb{R}^3$, let $\lambda E = \{\lambda \mathbf{x} \mid \mathbf{x} \in E\}$

- $\mathrm{volume}(\lambda E) = \lambda^3 \cdot \mathrm{volume}(E)$
- $\mathrm{area}(\lambda E) = \lambda^2 \cdot \mathrm{area}(E)$
- $\mathrm{length}(\lambda E) = \lambda^1 \cdot \mathrm{length}(E)$
- $\mathrm{cardinality}(\lambda E) = \lambda^0 \cdot \mathrm{cardinality}(E)$

"Correct" thing to do now is find for each $\alpha > 0$ a measure

$$\mu_\alpha \colon \{\text{subsets of } \mathbb{R}^3\} \to [0, \infty] \text{ such that } \mu_\alpha(\lambda E) = \lambda^\alpha \mu(E)$$

This is α-dimensional Hausdorff measure, but requires technicalities

Self-similarity

Previous slide highlighted self-similarity of measures.

Think about self-similarity of sets. Scale a set E by a factor of $\frac{1}{2}$. How many copies needed to recover original shape?

Self-similarity

Previous slide highlighted self-similarity of measures.

Think about self-similarity of sets. Scale a set E by a factor of $\frac{1}{2}$. How many copies needed to recover original shape?

$E = [0, 1]$  $2 = 2^1$ copies

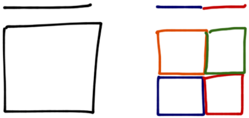
Self-similarity

Previous slide highlighted self-similarity of measures.

Think about self-similarity of sets. Scale a set E by a factor of $\frac{1}{2}$. How many copies needed to recover original shape?

$E = [0,1]$

$E = [0,1]^2$



$2 = 2^1$ copies

$4 = 2^2$ copies

Self-similarity

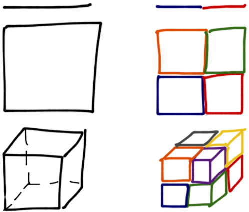
Previous slide highlighted self-similarity of measures.

Think about self-similarity of sets. Scale a set E by a factor of $\frac{1}{2}$. How many copies needed to recover original shape?

$E = [0,1]$ — $2 = 2^1$ copies

$E = [0,1]^2$ — $4 = 2^2$ copies

$E = [0,1]^3$ — $8 = 2^3$ copies



Moral: If E is a union of n copies of λE, then E is self-similar, and the dimension of E is α, where $n = \lambda^{-\alpha}$.

Self-similarity

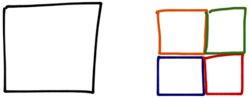
Previous slide highlighted self-similarity of measures.

Think about self-similarity of sets. Scale a set E by a factor of $\frac{1}{2}$. How many copies needed to recover original shape?

$E = [0, 1]$

$2 = 2^1$ copies

$E = [0, 1]^2$



$4 = 2^2$ copies

$E = [0, 1]^3$

$8 = 2^3$ copies

Moral: If E is a union of n copies of λE, then E is self-similar, and the dimension of E is α, where $n = \lambda^{-\alpha}$.

Solve this to write $\dim E = \alpha = \frac{\log n}{-\log \lambda}$.

Examples

Apply the formula $\dim E = \frac{\log n}{-\log \lambda}$ to some examples.

E	λ	n	$\dim E$
interval	$\frac{1}{2}$	2	$\frac{\log 2}{\log 2} = 1$
square	$\frac{1}{2}$	4	$\frac{\log 4}{\log 2} = 2$
cube	$\frac{1}{2}$	8	$\frac{\log 8}{\log 2} = 3$
Cantor set	$\frac{1}{3}$	2	$\frac{\log 2}{\log 3} \in (0, 1)$
Koch curve	$\frac{1}{3}$	4	$\frac{\log 4}{\log 3} \in (1, 2)$

Examples

Apply the formula $\dim E = \frac{\log n}{-\log \lambda}$ to some examples.

E	λ	n	$\dim E$
interval	$\frac{1}{2}$	2	$\frac{\log 2}{\log 2} = 1$
square	$\frac{1}{2}$	4	$\frac{\log 4}{\log 2} = 2$
cube	$\frac{1}{2}$	8	$\frac{\log 8}{\log 2} = 3$
Cantor set	$\frac{1}{3}$	2	$\frac{\log 2}{\log 3} \in (0, 1)$
Koch curve	$\frac{1}{3}$	4	$\frac{\log 4}{\log 3} \in (1, 2)$

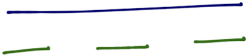
May consider other Cantor sets: 

- scale by $\frac{1}{5}$, use 3 copies to build: $\dim = \frac{\log 3}{\log 5}$

Examples

Apply the formula $\dim E = \frac{\log n}{-\log \lambda}$ to some examples.

E	λ	n	$\dim E$
interval	$\frac{1}{2}$	2	$\frac{\log 2}{\log 2} = 1$
square	$\frac{1}{2}$	4	$\frac{\log 4}{\log 2} = 2$
cube	$\frac{1}{2}$	8	$\frac{\log 8}{\log 2} = 3$
Cantor set	$\frac{1}{3}$	2	$\frac{\log 2}{\log 3} \in (0,1)$
Koch curve	$\frac{1}{3}$	4	$\frac{\log 4}{\log 3} \in (1,2)$

May consider other Cantor sets: 

- scale by $\frac{1}{5}$, use 3 copies to build: $\dim = \frac{\log 3}{\log 5}$

What about something like 

?

Dimension as a growth rate

Alternate way to derive dimension of our examples:

1. Given $r > 0$, break set into pieces of diameter $\leq r$
2. $N(r) =$ number of such pieces

Observe that $N(r) \approx r^{-\dim}$ $\left\{ \right.$

- interval: $N(r) = r^{-1}$
- square: $N(r) \approx r^{-2}$
- cube: $N(r) \approx r^{-3}$

Dimension as a growth rate

Alternate way to derive dimension of our examples:

1. Given $r > 0$, break set into pieces of diameter $\leq r$
2. $N(r) =$ number of such pieces

Observe that $N(r) \approx r^{-\dim}$

- interval: $N(r) = r^{-1}$
- square: $N(r) \approx r^{-2}$
- cube: $N(r) \approx r^{-3}$

Conclusion: $\dim = \lim_{r \to 0} \frac{\log N(r)}{-\log r}$ (growth rate of $N(r)$)

- Cantor set: $N(3^{-n}) = 2^n$, so $\frac{\log N(3^{-n})}{-\log(3^{-n})} = \frac{\log 2}{\log 3}$
- Koch curve: $N(3^{-n}) = 4^n$, so $\frac{\log N(3^{-n})}{-\log(3^{-n})} = \frac{\log 4}{\log 3}$

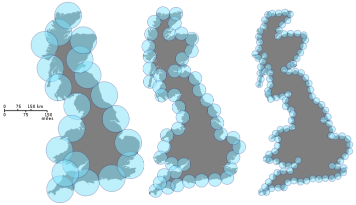
More general examples

Coastline of Britain

- r = size of ruler
- $rN(r)$ = measured length

$N(r) \approx r^{-1.25}$

Measured length $\approx r^{-.25} \to \infty$



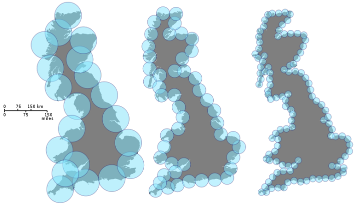
More general examples

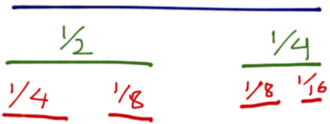
Coastline of Britain

- r = size of ruler
- $rN(r)$ = measured length

$N(r) \approx r^{-1.25}$

Measured length $\approx r^{-.25} \to \infty$





Can show that $N(2^{-(k+2)}) = F_k$, the kth Fibonacci number

Use fact that $F_k \approx \left(\frac{1+\sqrt{5}}{2}\right)^k$ to deduce that $\dim = \frac{\log(1+\sqrt{5})}{\log 2} + 1$

Bernoulli processes

Consider the following two stochastic processes:

1. Flip a coin repeatedly, write down outcome (H or T)
2. Roll a die repeatedly, write down the number from 1 to 6

Which one is "bigger"? The second one, but why?

Bernoulli processes

Consider the following two stochastic processes:

1. Flip a coin repeatedly, write down outcome (H or T)
2. Roll a die repeatedly, write down the number from 1 to 6

Which one is "bigger"? The second one, but why?

- "Bigness" = amount of information to record after each iteration of the experiment
- Information measured in number of bits
- h bits can store 2^h possible sequences

Bernoulli processes

Consider the following two stochastic processes:

1. Flip a coin repeatedly, write down outcome (H or T)
2. Roll a die repeatedly, write down the number from 1 to 6

Which one is "bigger"? The second one, but why?

- "Bigness" $=$ amount of information to record after each iteration of the experiment
- Information measured in number of bits
- h bits can store 2^h possible sequences

For n possible outcomes, need $2^h = n$, so $h = \log_2 n$.

- First process: $h = \log_2 2 = 1$
- Second process: $h = \log_2 6 \in (2, 3)$

$\boxed{h = \text{entropy}}$

Unequal probabilities

What if I use a weighted coin? Say $\mathbb{P}(H) = \frac{1}{3}$ and $\mathbb{P}(T) = \frac{2}{3}$.

- More or less information? What's the entropy?

Unequal probabilities

What if I use a weighted coin? Say $\mathbb{P}(H) = \frac{1}{3}$ and $\mathbb{P}(T) = \frac{2}{3}$.

- More or less information? What's the entropy?

Think of extreme case: $\mathbb{P}(H) = \frac{1}{1000}$ and $\mathbb{P}(T) = \frac{999}{1000}$.

- The event $TTTTT$ doesn't carry as much information now
- Most events carry less information

Unequal probabilities

What if I use a weighted coin? Say $\mathbb{P}(H) = \frac{1}{3}$ and $\mathbb{P}(T) = \frac{2}{3}$.

- More or less information? What's the entropy?

Think of extreme case: $\mathbb{P}(H) = \frac{1}{1000}$ and $\mathbb{P}(T) = \frac{999}{1000}$.

- The event $TTTTT$ doesn't carry as much information now
- Most events carry less information

Definition: the information content of an event E is $-\log_2 \mathbb{P}(E)$

- Entropy = expected information content of each experiment

Unequal probabilities

What if I use a weighted coin? Say $\mathbb{P}(H) = \frac{1}{3}$ and $\mathbb{P}(T) = \frac{2}{3}$.

- More or less information? What's the entropy?

Think of extreme case: $\mathbb{P}(H) = \frac{1}{1000}$ and $\mathbb{P}(T) = \frac{999}{1000}$.

- The event $TTTTT$ doesn't carry as much information now
- Most events carry less information

Definition: the information content of an event E is $-\log_2 \mathbb{P}(E)$

- Entropy = expected information content of each experiment

Coin with weights $\frac{1}{3}$ and $\frac{2}{3}$:

- H carries information $\log_2(3)$, and T carries info $\log_2(\frac{3}{2})$

Entropy $= \frac{1}{3}\log_2(3) + \frac{2}{3}\log_2(\frac{3}{2}) = \log_2(3) - \frac{2}{3} < 1$ (log is concave)

Maximising entropy

Suppose I use a coin with weights p and q.

$$\boxed{\begin{array}{l} p, q \in [0,1] \\ p + q = 1 \end{array}}$$

- Information content of event H is $-\log_2 p$
- Information content of event T is $-\log_2 q$

Maximising entropy

Suppose I use a coin with weights p and q.

$$\boxed{\begin{array}{l} p, q \in [0,1] \\ p + q = 1 \end{array}}$$

- Information content of event H is $-\log_2 p$
- Information content of event T is $-\log_2 q$

$$\begin{aligned}
\text{entropy} &= \text{expected information content} \\
&= \mathbb{P}(H)(-\log_2 p) + \mathbb{P}(T)(-\log_2 q) \\
&= -p \log_2 p - q \log_2 q = p \log_2 \left(\frac{1}{p}\right) + q \log_2 \left(\frac{1}{q}\right)
\end{aligned}$$

Maximising entropy

$$\boxed{\begin{array}{l} p, q \in [0,1] \\ p+q=1 \end{array}}$$

Suppose I use a coin with weights p and q.

- Information content of event H is $-\log_2 p$
- Information content of event T is $-\log_2 q$

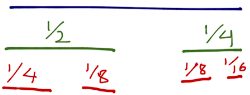
$$\begin{aligned}
\text{entropy} &= \text{expected information content} \\
&= \mathbb{P}(H)(-\log_2 p) + \mathbb{P}(T)(-\log_2 q) \\
&= -p\log_2 p - q\log_2 q = p\log_2\left(\frac{1}{p}\right) + q\log_2\left(\frac{1}{q}\right)
\end{aligned}$$

Because log is concave down we always have entropy ≤ 1

Strictly concave $\Rightarrow$ equality iff $p = q = \frac{1}{2}$

Relationship to dimension

Recall example of $(\frac{1}{2}, \frac{1}{4})$-Cantor set.

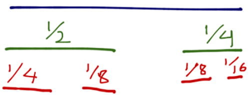


After n iterations, get a set C_n with 2^n intervals

- Length varies: left a times, right b times $\Rightarrow r = (\frac{1}{2})^a (\frac{1}{4})^b$
- $\#$ with this length is $\binom{n}{a}$
- If $a = pn$ and $b = qn$, then $\binom{n}{a} \approx e^{(-p \log p - q \log q)n}$

Relationship to dimension

Recall example of $(\frac{1}{2}, \frac{1}{4})$-Cantor set.



After n iterations, get a set C_n with 2^n intervals

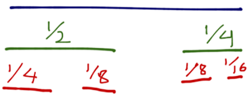
- Length varies: left a times, right b times $\Rightarrow r = (\frac{1}{2})^a(\frac{1}{4})^b$
- $\#$ with this length is $\binom{n}{a}$
- If $a = pn$ and $b = qn$, then $\binom{n}{a} \approx e^{(-p\log p - q\log q)n}$

Instead of covering all of C, just cover the part where $\frac{\#\text{left}}{\#\text{right}} \approx \frac{p}{q}$.

- $r = \left((\frac{1}{2})^p(\frac{1}{4})^q\right)^n \Rightarrow -\log r = n(p\log 2 + q\log 4)$

Relationship to dimension

Recall example of $(\frac{1}{2}, \frac{1}{4})$-Cantor set.



After n iterations, get a set C_n with 2^n intervals

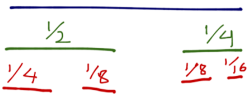
- Length varies: left a times, right b times $\Rightarrow r = (\frac{1}{2})^a(\frac{1}{4})^b$
- $\#$ with this length is $\binom{n}{a}$
- If $a = pn$ and $b = qn$, then $\binom{n}{a} \approx e^{(-p\log p - q\log q)n}$

Instead of covering all of C, just cover the part where $\frac{\#\text{left}}{\#\text{right}} \approx \frac{p}{q}$.

- $r = \left((\frac{1}{2})^p(\frac{1}{4})^q\right)^n \Rightarrow -\log r = n(p\log 2 + q\log 4)$
- $\log N(r) \geq n(-p\log p - q\log q)$

Relationship to dimension

Recall example of $(\frac{1}{2}, \frac{1}{4})$-Cantor set.



After n iterations, get a set C_n with 2^n intervals

- Length varies: left a times, right b times $\Rightarrow r = (\frac{1}{2})^a (\frac{1}{4})^b$
- $\#$ with this length is $\binom{n}{a}$
- If $a = pn$ and $b = qn$, then $\binom{n}{a} \approx e^{(-p\log p - q\log q)n}$

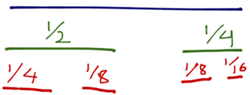
Instead of covering all of C, just cover the part where $\frac{\#\text{left}}{\#\text{right}} \approx \frac{p}{q}$.

- $r = \left((\frac{1}{2})^p (\frac{1}{4})^q\right)^n \Rightarrow -\log r = n(p\log 2 + q\log 4)$
- $\log N(r) \geq n(-p\log p - q\log q)$

$$\dim \approx \frac{\log N(r)}{-\log r} \geq \frac{-p\log p - q\log q}{p\log 2 + q\log 4} = \frac{\text{entropy}}{\text{average expansion}}$$

Relationship to dimension

Recall example of $(\frac{1}{2}, \frac{1}{4})$-Cantor set.



After n iterations, get a set C_n with 2^n intervals

- Length varies: left a times, right b times $\Rightarrow r = (\frac{1}{2})^a (\frac{1}{4})^b$
- $\#$ with this length is $\binom{n}{a}$
- If $a = pn$ and $b = qn$, then $\binom{n}{a} \approx e^{(-p \log p - q \log q)n}$

Instead of covering all of C, just cover the part where $\frac{\#\text{left}}{\#\text{right}} \approx \frac{p}{q}$.

- $r = \left((\frac{1}{2})^p (\frac{1}{4})^q\right)^n \Rightarrow -\log r = n(p \log 2 + q \log 4)$
- $\log N(r) \geq n(-p \log p - q \log q)$

$$\dim \approx \frac{\log N(r)}{-\log r} \geq \frac{-p \log p - q \log q}{p \log 2 + q \log 4} = \frac{\text{entropy}}{\text{average expansion}}$$

Get actual dimension by maximising over (p, q).

Information compression

Entropy measures information content

- Related: how much can data be compressed?

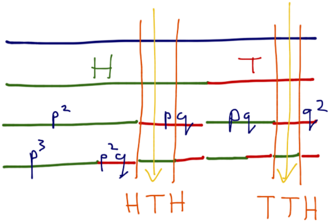
Information compression

Entropy measures information content

- Related: how much can data be compressed?

Shannon's source coding theorem:

If we run n iterates of a process (IID) with entropy h, the results can be stored in nh bits of information, but no fewer.



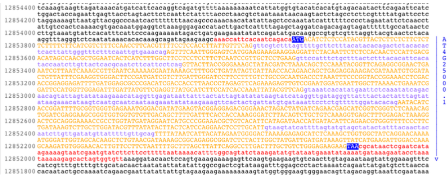
Idea: First n results determine a subinterval of $[0, 1]$

- "Typical" interval has width $p^{pn}q^{qn} = 2^{-nh}$
- Takes n bits to encode that much precision

Information content

Entropy can be used to analyse genetic data.

- Genome: string of symbols A, C, G, T
- Some regions more important than others



Topological entropy and topological pressure

- Quantities related to entropy discussed above
- Can be adapted to study genetic data
- High entropy/pressure $\Rightarrow$ high information content $\Rightarrow$ more likely to be a coding region of the genome