Fractals and Power Laws

Renjie Yang

COMPHI LAB for Data Science

April 2020

Outline

Coastline Paradox

An example of the coastline paradox. If the coastline of Great Britain is measured using units 100 km (62 mi) long, then the length of the coastline is approximately 2,800 km (1,700 mi). With 50 km (31 mi) units, the total length is approximately 3,400 km (2,100 mi), approximately 600 km (370 mi) longer.

Fractals

- The coastline paradox is the counterintuitive observation that the coastline of a landmass does not have a well-defined length.
- This results from the fractal-like properties of coastlines, i.e., the fact that a coastline typically has a fractal dimension
- Self-similar: small parts of the object are similar to the whole.
- This self-similarity extends over many scales.

Fir

Broccoli

https://commons.wikimedia.org/wiki/File:Fractal_Broccoli.jpg

Tree

https://commons.wikimedia.org/wiki/File:Tree_Fractal.jpeg

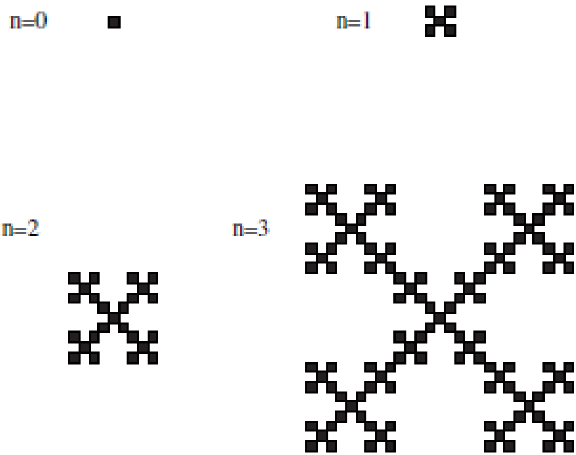
Mountain Ridge

https://commons.wikimedia.org/wiki/File:
Blue_ridge_Mountains_layered.jpg

Mountain Ridge

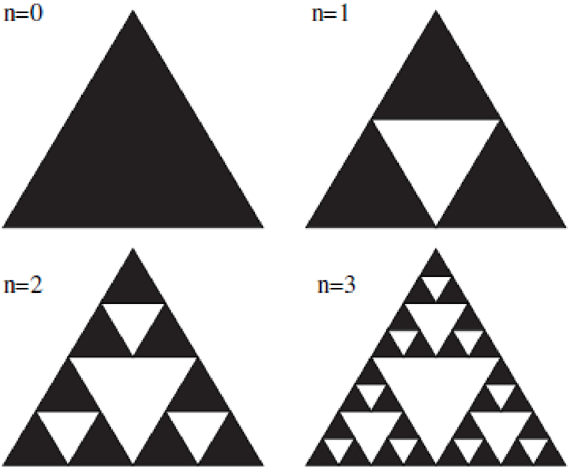
https://commons.wikimedia.org/wiki/File:
Blue_ridge_Mountains_layered.jpg

Fractal Snowflake



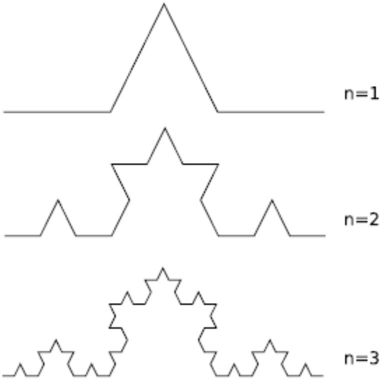
Chaos and Fractals, Figure 15.3

Sierpiński Triangle



Chaos and Fractals, Figure 16.4

Koch Curve

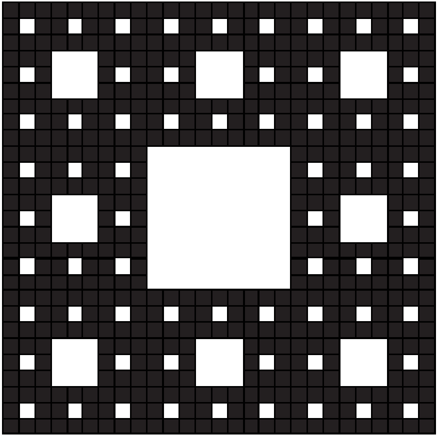


Chaos and Fractals, Figure 16.5

Cantor Set

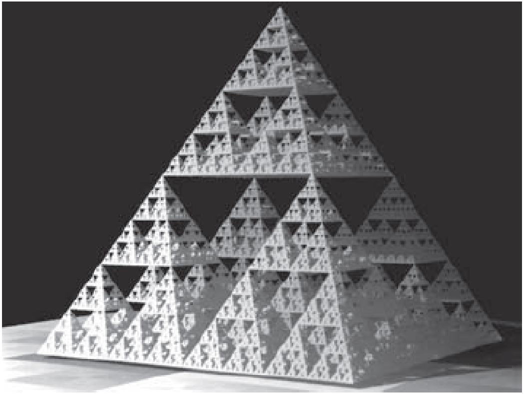
Chaos and Fractals, Figure 16.3

Sierpiński Carpet



Chaos and Fractals, Figure 16.6

Sierpiński Pyramid



Chaos and Fractals, Figure 16.8

Generations of Fractals

- Mandelbrot set
- Fir
- Irregular Koch curve
- Sierpiński Triangle
- Random Koch curve
- The chaos game

Fractals and Scaling

- If an object is self-similar, it is scale free.
- In a fractal, if you were shrunk, you could not tell, because there are no objects that set a size scale.
- Ex: There is no typical size of the bumps in a Koch curve that sets a scale. In contrast, there is a typical size to a tomato.
- Real fractals are not self-similar forever, the way mathematical fractals are.

Fractals “Defined”

- Fractals are self-similar across many scales
- They are not well-described by classical geometry: circles, cubes, etc.
- There is not an airtight, standard definition of fractal.
- They have a **self-similarity dimension** that is strictly larger than the topological dimension.

Outline

1. Fractals
2. **Dimensions**
3. Power Laws

Exponent and Log Functions

- Exponentiation means successive multiplication: $x^2 = x \times x$
- $log_b(x) = a$ means $b^a = x$.
- Therefore $b^{log_b(x)} = x$
- $x^a \times x^b = x^{a+b}$
- $1/x^a = x^{-a}$
- $log(ab) = log(a) + log(b)$
- $log(a/b) = log(a) - log(b)$
- $log(a^D) = D \times log(a)$

Self-Similarity Dimension

- Example: line, square, cube
- No. of small copies = (magnification factor)^D
- Sierpiński Triangle: No. of small copies = 3, mag factor = 2, $D \approx 1.585$
- Snowflake: $D = log5/log3 \approx 1.465$

Self-Similarity Dimension

- The length of the Koch curve: $(4/3)^n$
- Dimension of Koch curve: $D = log4/log3 \approx 1.262$
- The area of Sierpiński Triangle: $(3/4)^n A$

"In-between" Dimension

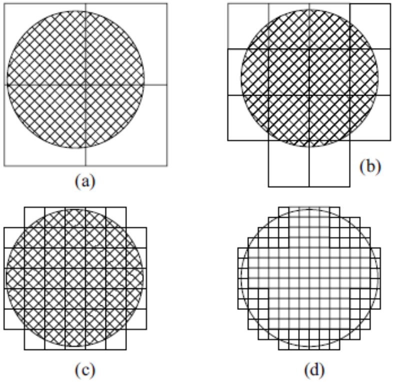
- The Koch curve has both 1-dimensional and 2-dimensional qualities.
- Infinite length in a finite area.
- The Sierpiński triangle has zero area but infinite perimeter.

Dimension and Scaling

- Increase in "size" $=$ (scale factor)^D
- Ex: If sphere (3D) is stretched by a factor of two, it is now 8 times larger, since $2^3 = 8$.
- The dimension tells you how the size of an object changes as it is scaled up.

The Box-Counting Dimension

- The idea is to cover the figure using small boxes.
- Let N(s) = number of boxes of side s needed to cover an object.



The Box-Counting Dimension

- The idea is to cover the figure using small boxes.
- Let $N(s)$ = number of boxes of side s needed to cover an object.

s	$N(s)$
$\frac{1}{2}$	4
$\frac{1}{4}$	14
$\frac{1}{8}$	45
$\frac{1}{16}$	162

The Box-Counting Dimension

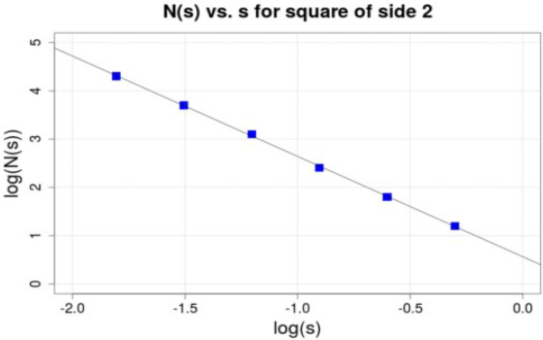
- Box counting for lines, squares
- Circle
- Sierpiński triangle
- Cantor Set

The Box-Counting Dimension

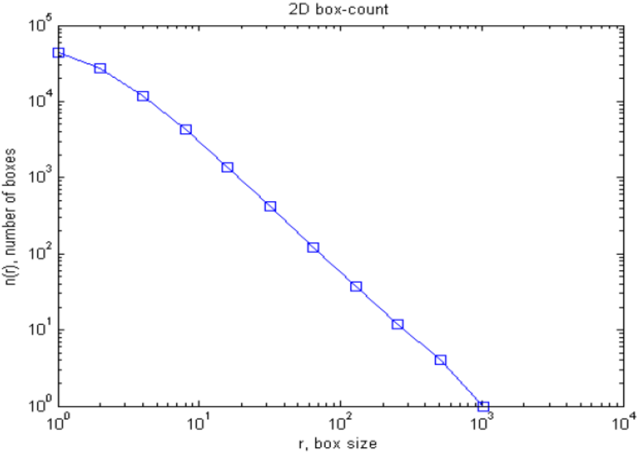
- The idea is to cover the figure using small boxes.
- Let $N(s) =$ number of boxes of side s needed to cover an object.
- $N(s) = c\left(\frac{1}{s}\right)^D$ as $s \to 0$
- D is the box-counting dimension

Log-Log Plot

- $N(s) = c\left(\frac{1}{s}\right)^D$ as $s \to 0$
- $log(N(s)) = -Dlog(s) + log(c)$
- This is a line!



Log-Log Plot in Practice

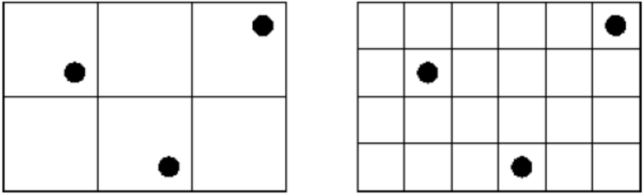


Log-Log Plot in Practice

- The linear equation does not hold for large box size s.
- For small s, the shape may no longer be selfsimilar, and/or we may run out of data.
- The equation $log(N(s)) = -Dlog(s) + log(c)$ is only true if the object is self-similar across scales. So a nonself- simialr shape will not have a linear log-log box-counting plot.
- So in practice we often see a log-log plot that is linear over a somewhat ambiguous middle region.

Log-Log Plot in Practice

- Problems with Small s: Once one point is in each box, making boxes smaller will not increase N(s).



Different Dimensions

- There are a handful of different dimensions (box-counting, self-similarity, Hausdorf, and others).
- All involve looking at how an object behaves when the scale of analysis is changed.
- These dimensions give the same number for most cases of interest.

Scaling and Dimension

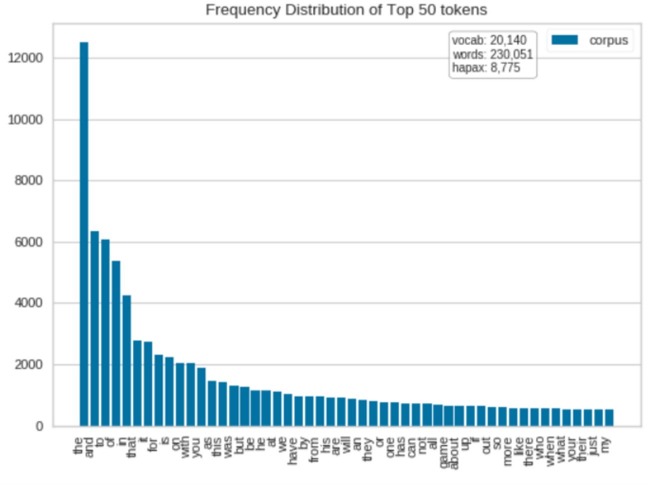
- $N(s) = c\left(\frac{1}{s}\right)^{D}$
- $log(N(s)) = -Dlog(s) + log(c)$
- If either of them holds for a range of s, then we say that the object exhibits scaling.
- If it exists, the dimension is telling us that something stays the same as the scale is changed.

Outline

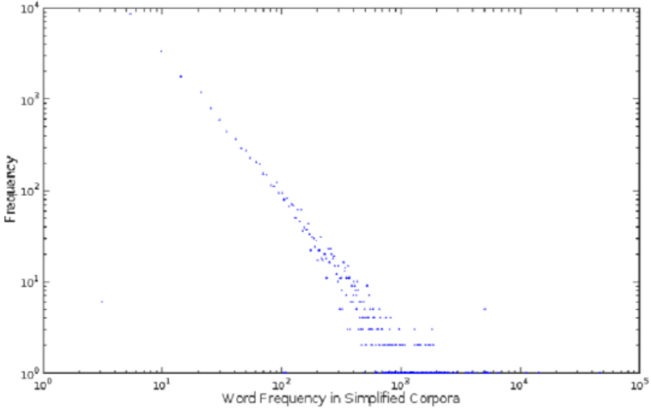
Power Law

- Box-Counting dimension: $N(s) = c\left(\frac{1}{s}\right)^{D}$
- Power law: $p(x) = Ax^{(-\alpha)}$
- If equation is true, there is self-similarity, and we see a line on a log-log plot.
- Reverse logic: If we see linear behavior on log-log plot, there must be selfsimilarity.

Word Frequency

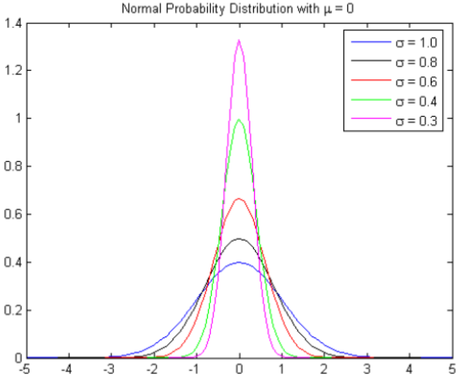


Word Frequency



Normal Probability Disribution

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

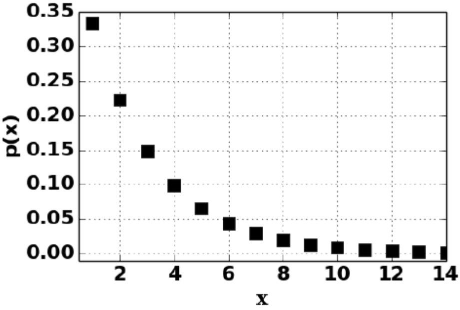


Central Limit Theorem

- Let X_i be a set of random variables
- Then the sum of N such random variables is normally distributed as N gets large
- with one condition: the distribution of the X_i's has finite variance.
- A variable that is the result of a number of additive influences should be normal.

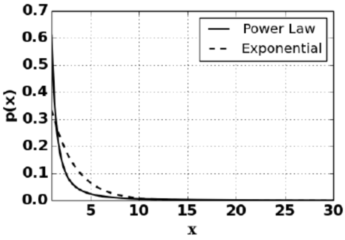
Exponential Distribution

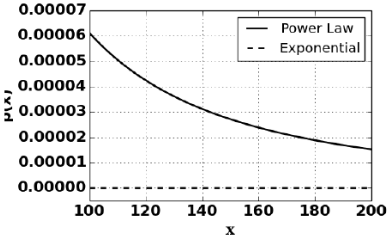
- Also called geometric distribution
- Large range of outcomes, but probability decreases very quickly
- Waiting times between events that happen with constant probability.



Power Laws have Long Tails

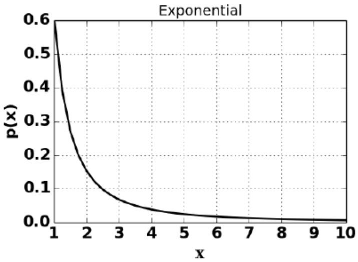
- Very large x values, while rare, are still observed
- Exponential: p(50) = 0.00000000078
- Power Law: p(50) = 0.000244

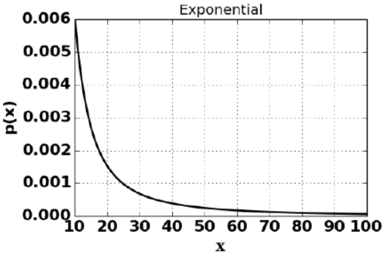




Power Laws are Scale Free

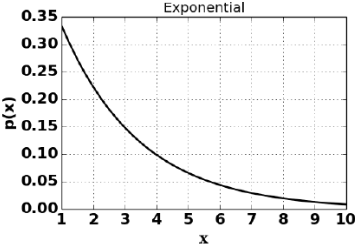
- Power laws look like the same at all scales.

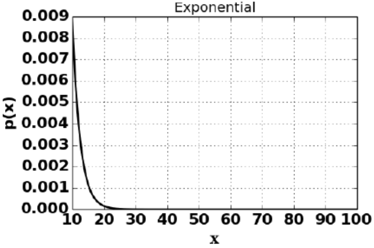




Exponentials are not Scale Free

- Exponential functions do not look the same at all scales



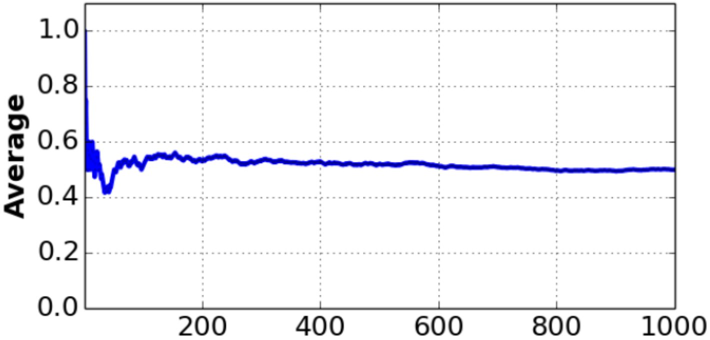


Power Laws are Scale Free

Power laws are the only distribution that is scale free.

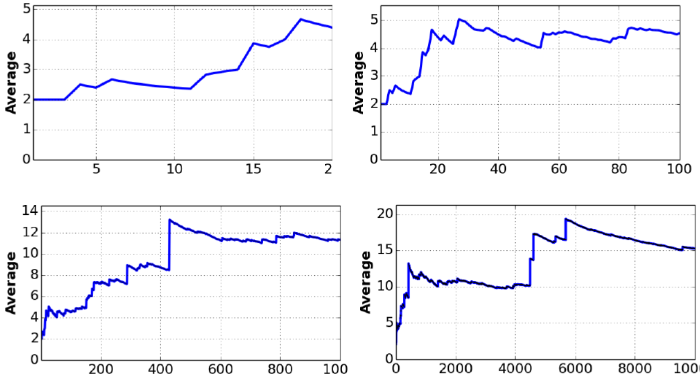
Power Laws and Averages

- Non-power law example: toss coin
- Average winnings quickly approaches



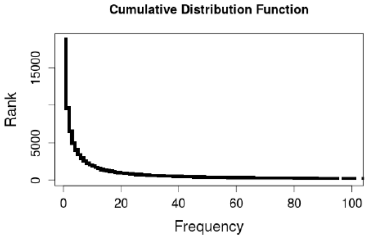
St. Petersberg Game

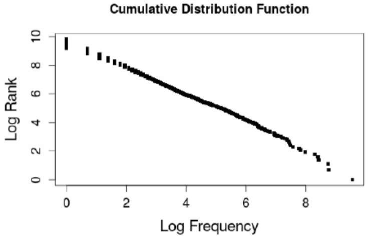
- Keep tossing a coin until you get Heads. You win 2^x, where x is the number of tosses.
- The average winnings does not exist: it is infinite.



Complementary Cumulative Distribution Function.

- CDF: P(x) = fraction of data that has a value of x or greater.
- $p(x) = Ax^{-\alpha}$
- $P(x) = Cx^{-(\alpha-1)}$

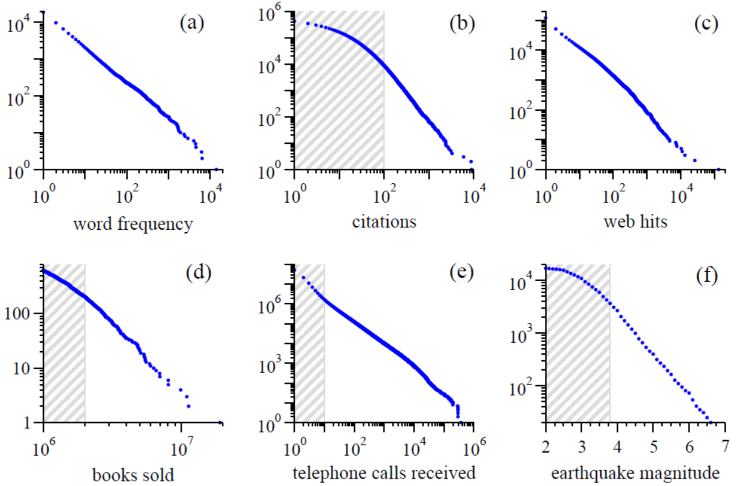




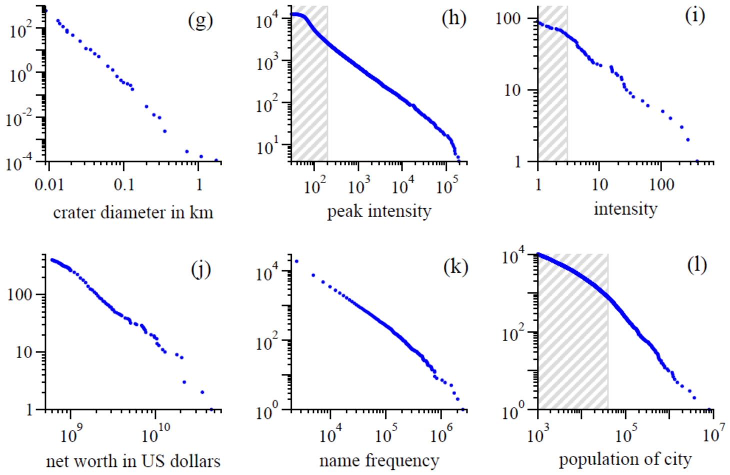
Power Laws

- Long tails
- Self-similar
- Sometimes averages or standard deviation does not exist
- Very different from most distributions we're used to

Power Laws are Everywhere



Power Laws are Everywhere



References

- David P. Feldman. (2012) *Chaos and Fractals: An Elementary Introduction*, Oxford University Press.
- David P. Feldman's slides on fractals and scaling.
- Wikimedia Commons https://commons.wikimedia.org/