A Crash Course in Fractals and Scaling

David Feldman
Complex Systems Summer School
June 13, 2019

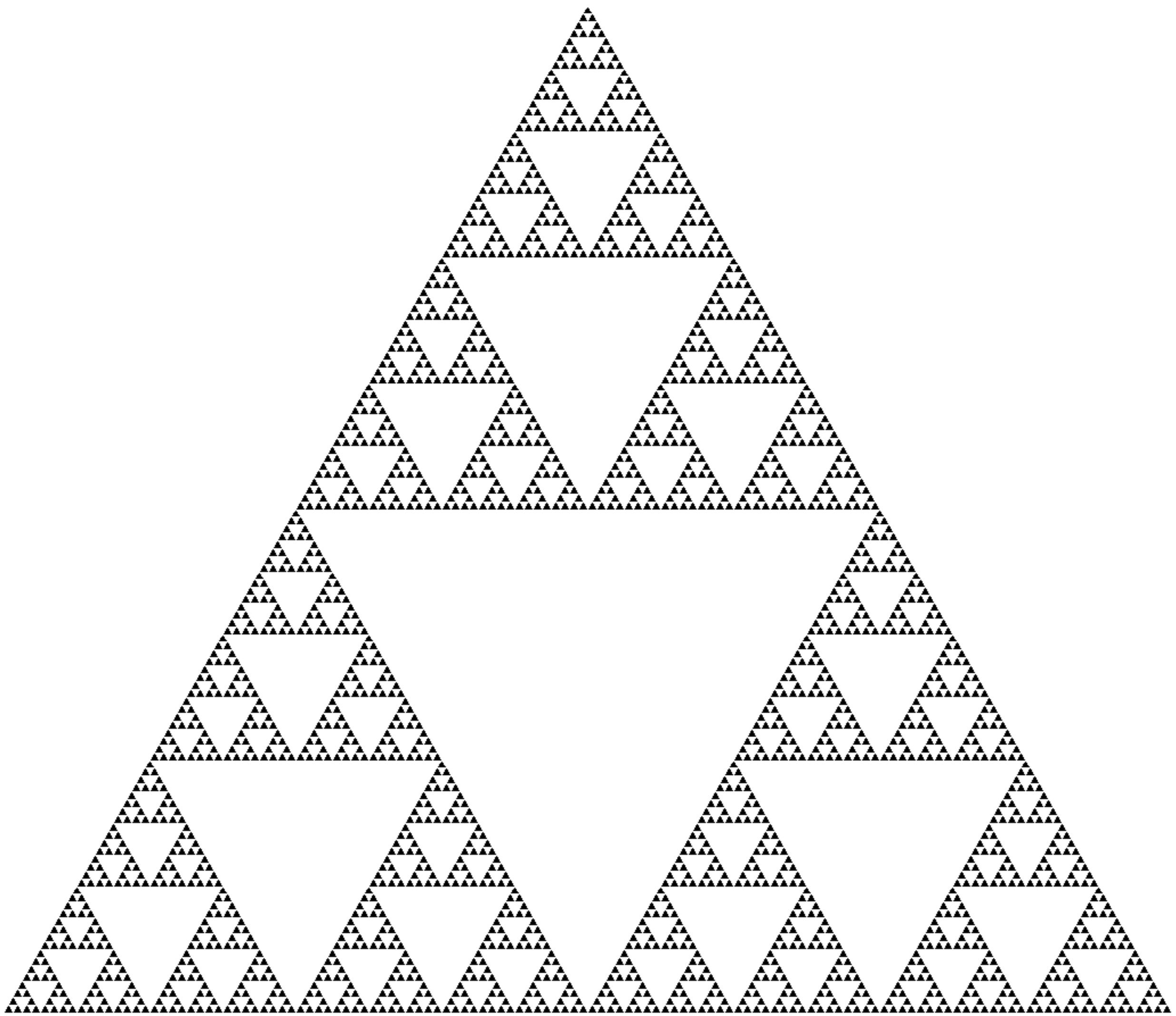
Outline & Goals

- Give overview of scaling:
  - What scaling is,
  - How to describe and detect it
  - What it means
- Aimed at those who have little experience with the idea of scaling.
- Based on the ComplexityExplorer MOOC:

  http://fractals.complexityexplorer.org

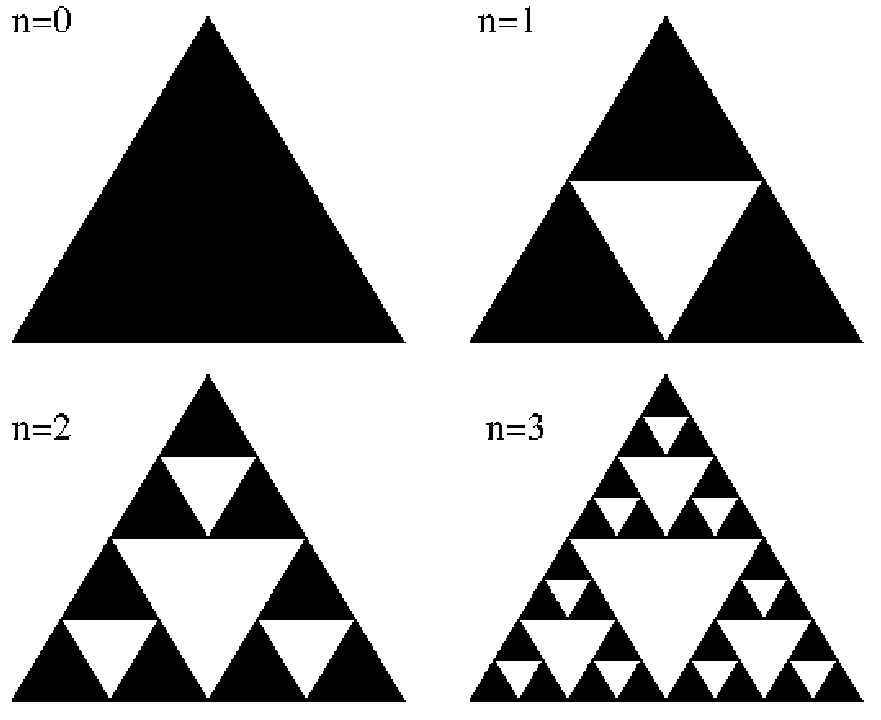
Fractals

- Self-similar geometric objects
- “Scale Free”



Self-Similarity Dimension

- **No. of small copies = (magnification factor)^D**
- Ex: No. of small copies = 3, mag factor = 2
- D = log(3)/log(2). (approx = 1.585.)



Real Fractals

- Cannot be self-similar forever
- How wide a range of self-similarity needed to be a fractal?

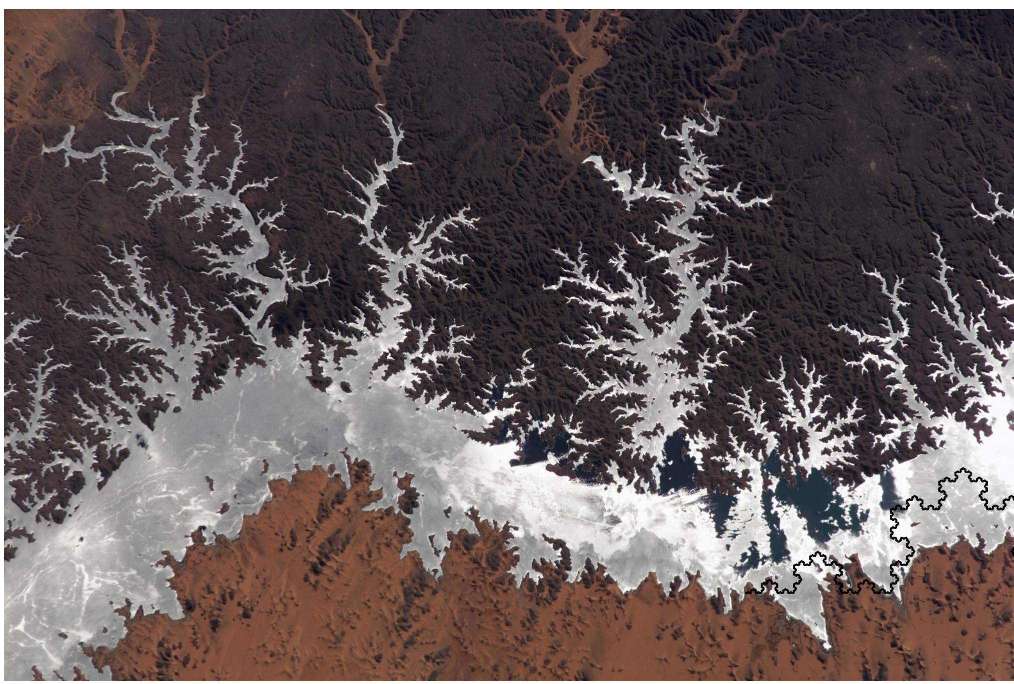
Photo by Hagerty Ryan, U.S. Fish and Wildlife Service.

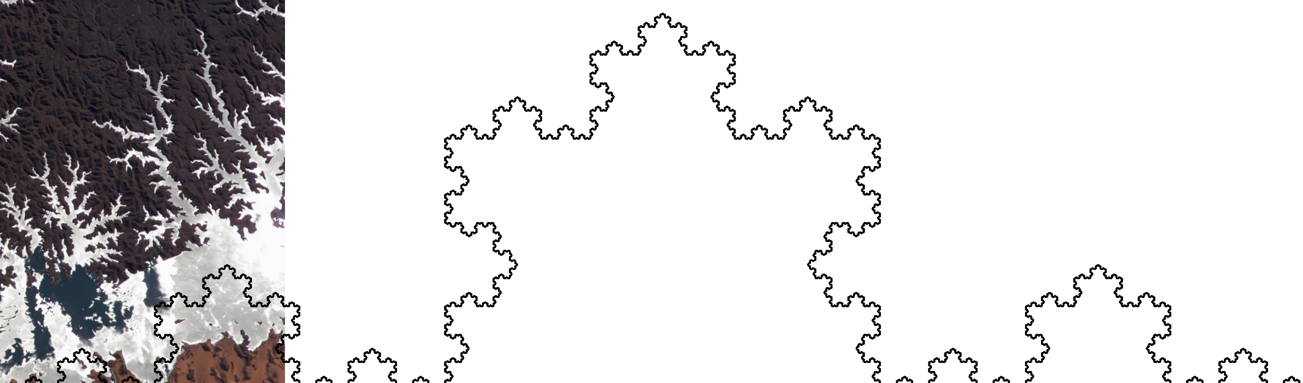
https://commons.wikimedia.org/wiki/File:Looking_up_into_the_silhouetted_branches_and_gree_foliage_of_sycamore_tree_platanus_occidentalis.jpg

Fractals “Defined”

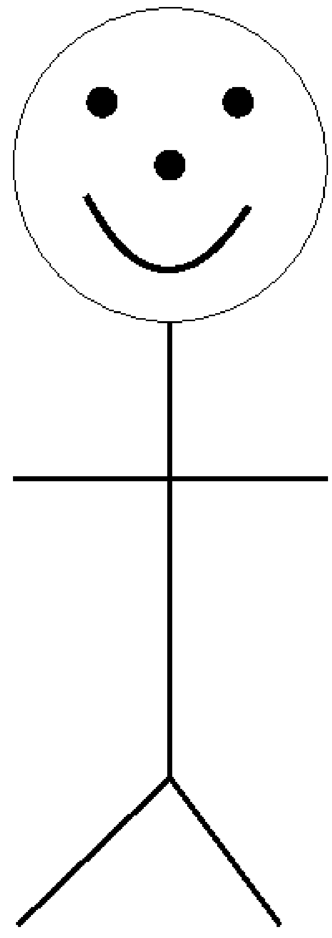
- Fractals are self-similar across many scales
- For physical objects, fractal-ness is a notion, not a strict category.
- Some objects are more or less fractal-like.

https://earthobservatory.nasa.gov/IOTD/view.php?id=5988



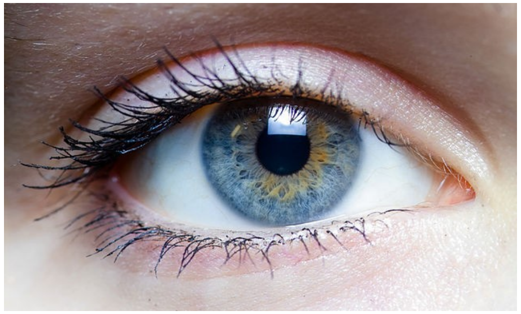


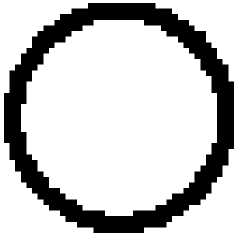
What is this a picture of?

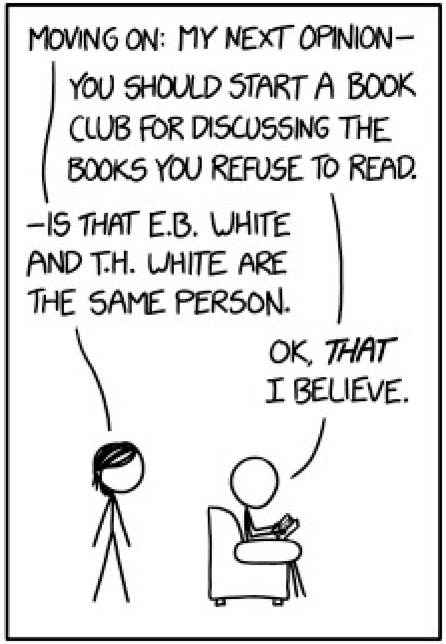


Circles

$$x^2 + y^2 = r^2$$







Circles are abstractions.
There are no true or perfect circles in nature.

Where do fractals come from?

- Deterministic and stochastic iterated processes can make fractals
- Fractals look “complex”, but can be “simple” to build
- Fractals are “generic” shapes, in that there are many fairly simple ways to make them.
- We see similar structures in very different systems.

There are many simple ways to make fractals.

Power Law Example: Word Frequencies

The scientist does not study nature because it is useful; he studies it because he delights in it, and he delights in it because it is beautiful. If nature were not beautiful it would not be worth knowing, and if nature were not worth knowing, life would not be worth living. (Henri Poincaré)

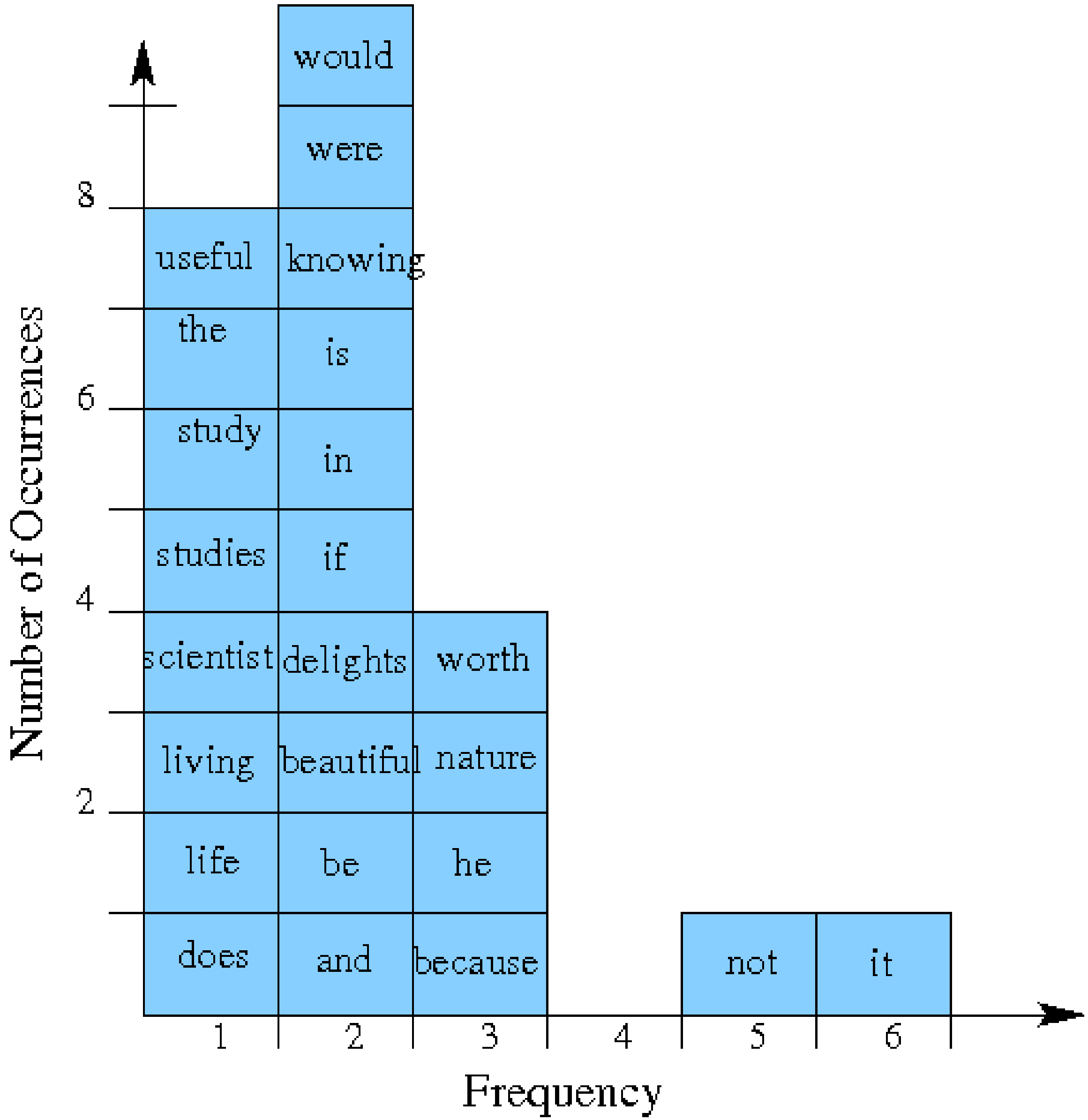
- it (6)
- not (5)
- because, he, nature, worth (3)
- and, be, beautiful, delights, if, in, is, knowing, were, would (2)
- does, life, living, scientist, study, studies, the, useful (1)

Word Frequencies

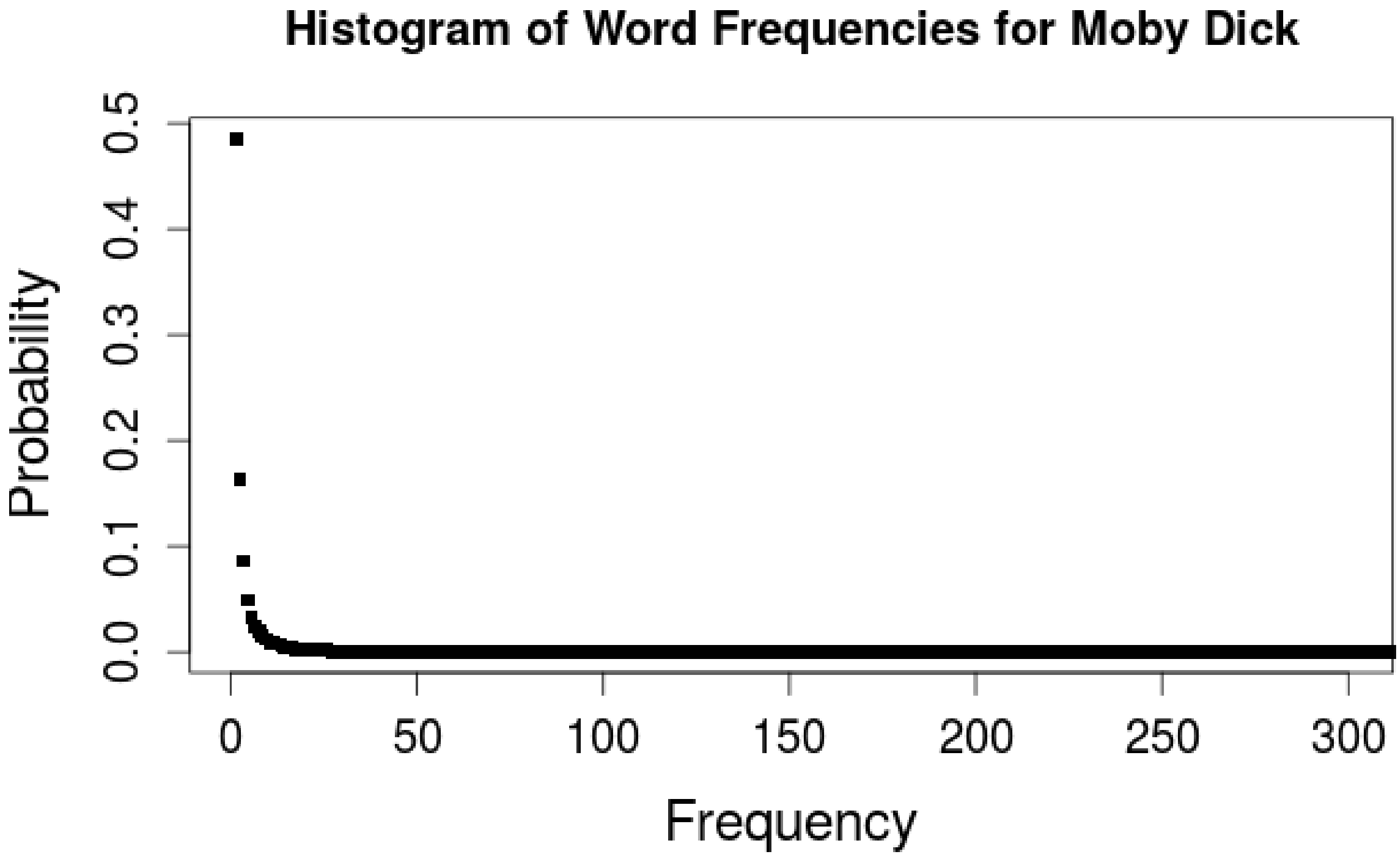
The scientist does not study nature because it is useful; he studies it because he delights in it, and he delights in it because it is beautiful. If nature were not beautiful it would not be worth knowing, and if nature were not worth knowing, life would not be worth living. (Henri Poincaré)

6, 5, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1

Word Frequencies: Histogram



Word Frequencies in *Moby Dick*



- **There are 18,855 different words. There is one word that appears 14,086 times. There are 9161 words that appear only once.**
- Data from: http://tuvalu.santafe.edu/~aaronc/powerlaws/data.htm

A Power Law

- Probability of word appearing x times:

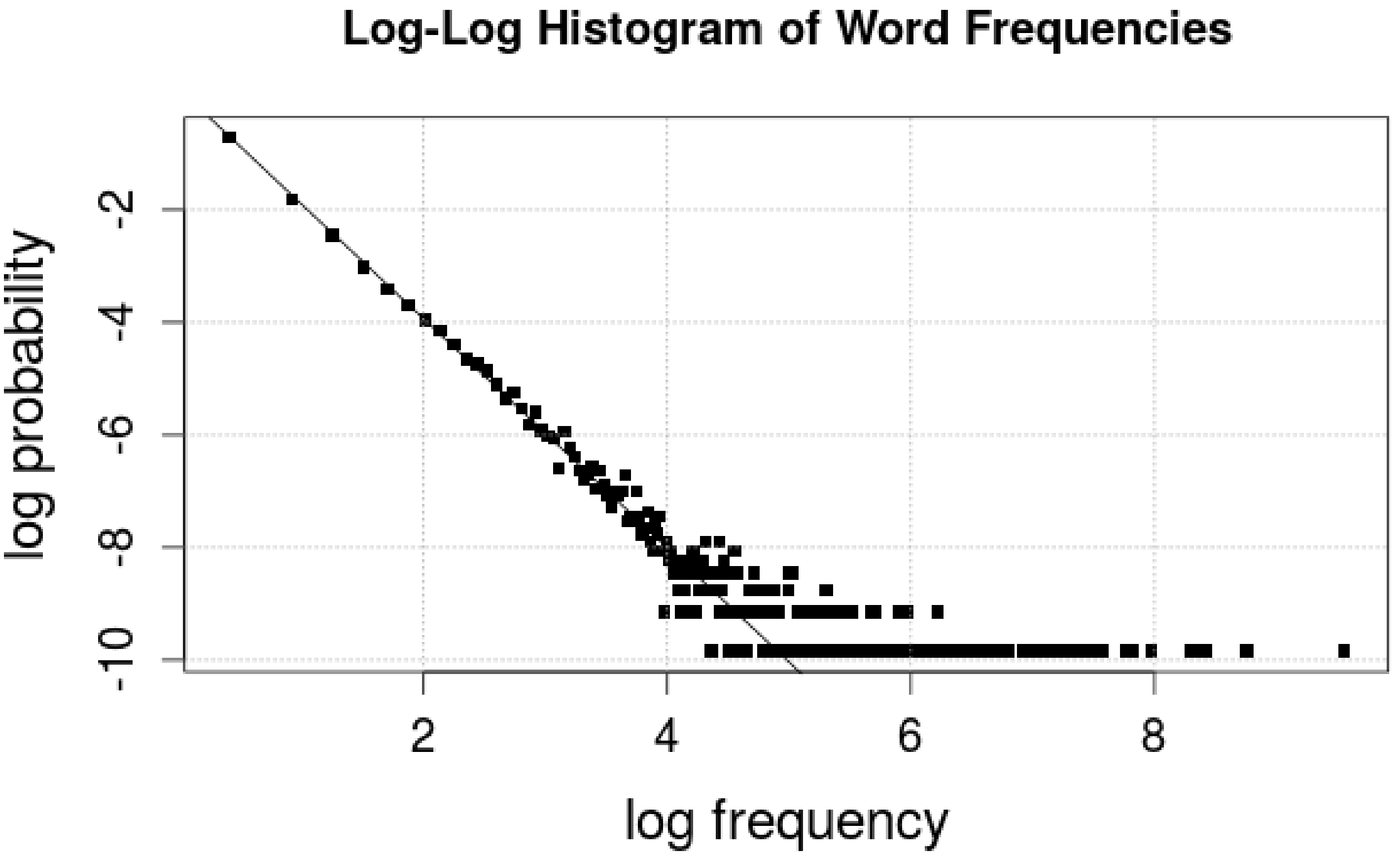
$$p(x) = Ax^{-\alpha}$$

- Take the log of both sides:

$$\log p(x) = \log A - \alpha \log x$$

- So we expect a log-log plot of p(x) vs x to be linear.

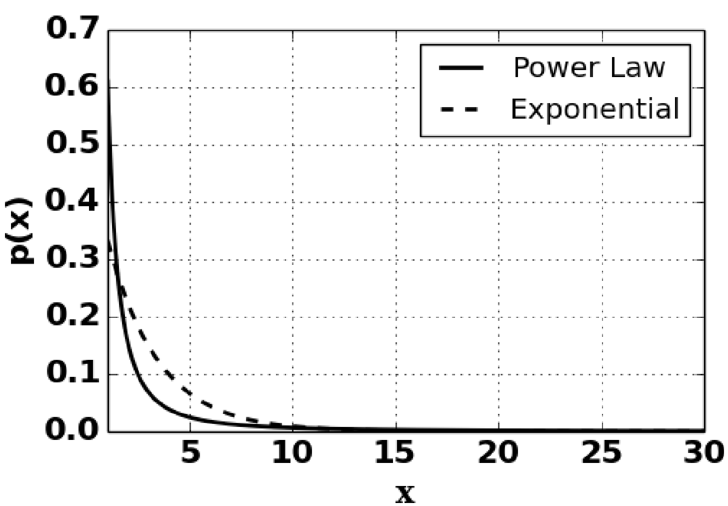
Word Frequencies in *Moby Dick*

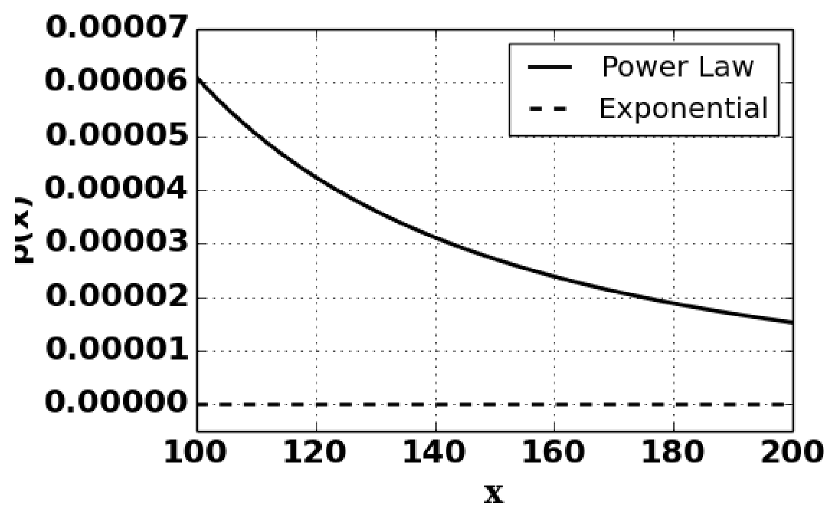


- **There are 18,855 different words. There is one word that appears 14,086 times. There are 9161 words that appear only once.**
- Data from: http://tuvalu.santafe.edu/~aaronc/powerlaws/data.htm

Power Laws have Long Tails

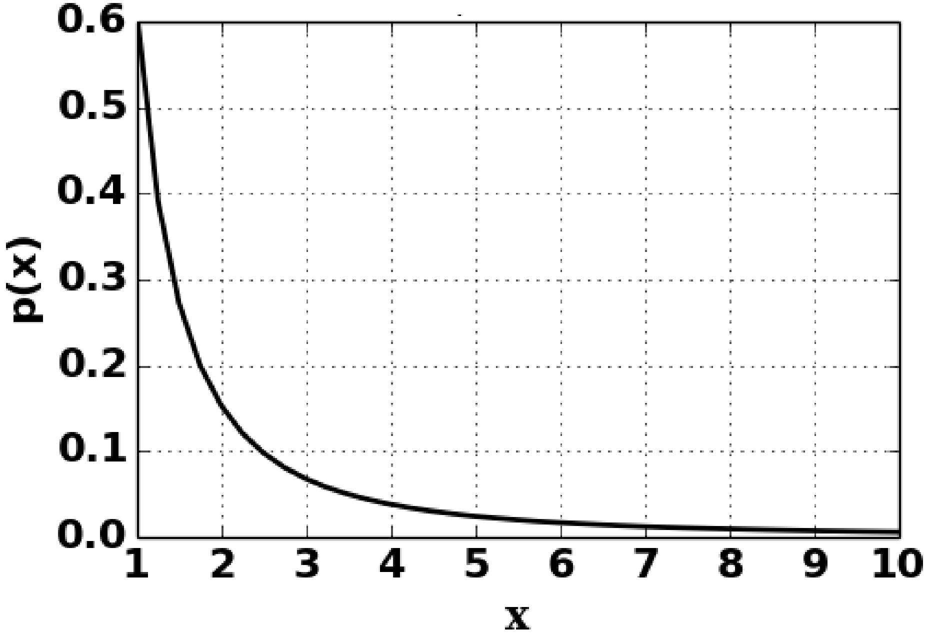
- Power laws decay much more slowly than exponentials.
- Very large x values, while rare, are still observed.
- Exponential: p(50) = 0.00000000078
- Power Law: p(50) = 0.000244

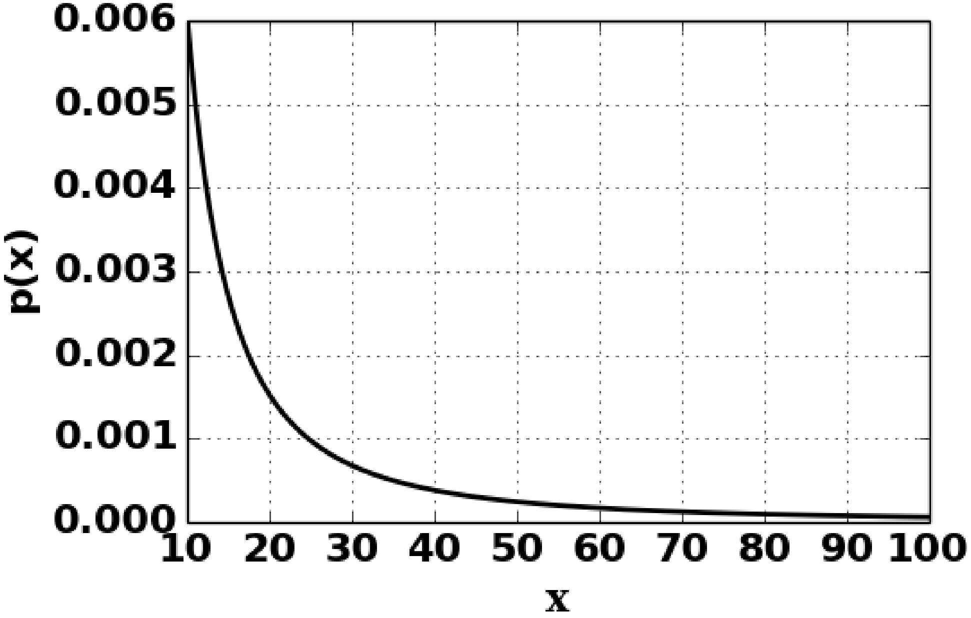




Power Laws are Scale Free

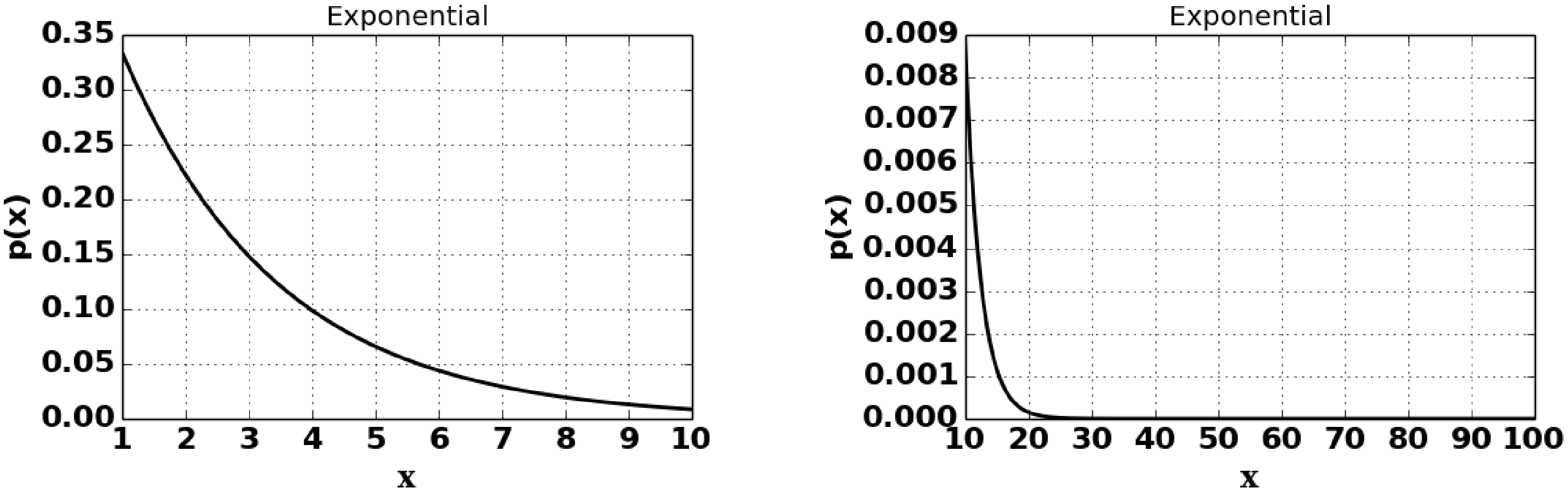
- Power laws look the same at all scales.





Exponentials are not Scale Free

- Exponential functions do not look the same at all scales



Power Laws are Scale Free

- Power Law

$$p(x) = 0.61x^{-2}$$

$$\frac{p(x)}{p(2x)} = \frac{0.61x^{-2}}{0.61(2x)^{-2}} = 4$$

- Same ratio no matter what x is.

- Exponential

$$p(x) = (1/3)(2/3)^{x-1}$$

$$\frac{p(x)}{p(2x)} = \frac{(1/3)(2/3)^{x-1}}{(1/3)(2/3)^{2x-1}}$$

$$\frac{p(x)}{p(2x)} = (2/3)^{-x}$$

Ratio depends on x

Some Mathematical Notes

- Power laws are the **only** distribution that is scale free
- Discrete and Continuous probability distributions are different mathematical entities and need to be handled differently.
- However, from “20,000 feet” the difference is usually not crucial.

Summary of Power Laws so far...

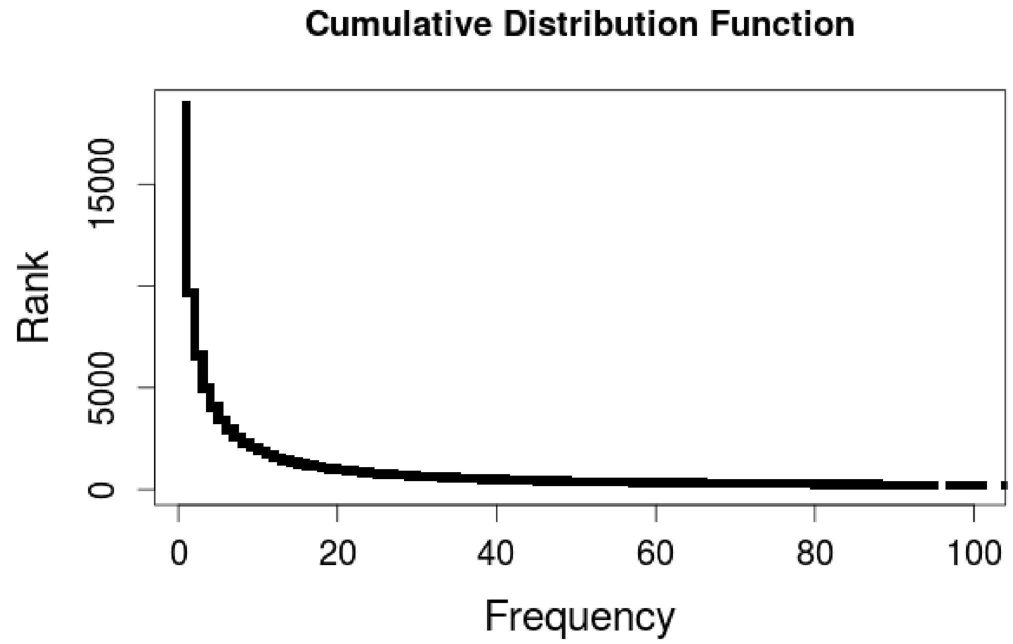
- Long tails.
- Self-similar.
- Sometimes averages or standard deviation does not exist.
- Very different from most distributions we're used to.

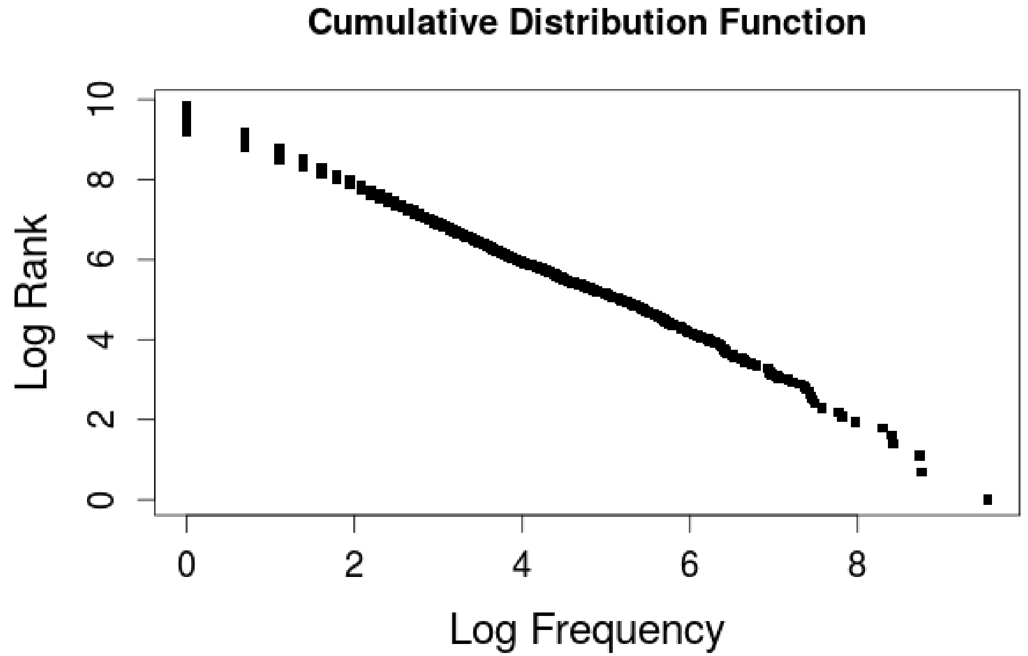
Next: detecting power laws in data….

Cumulative Distribution Function

- CDF: P(x) = fraction of data that has a value of x or greater.
- Sometimes known as the complementary cumulative distribution function.
- Almost always easier to work with than the distribution function p(x).
- Same interpretation for continuous and discrete distributions.

Example: Word Frequencies in *Moby Dick*





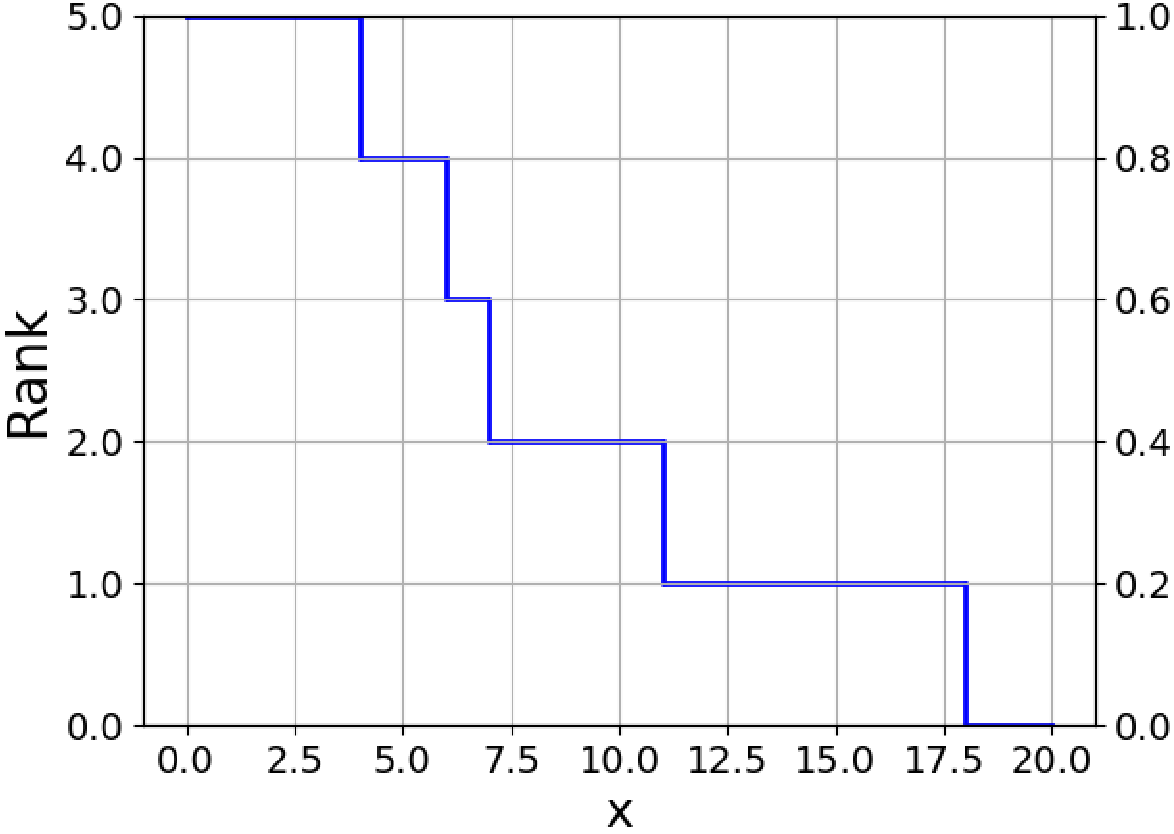
Cumulative Distribution Function

- The CDF is a non-increasing function.
- If x is power-law distributed: $p(x) = Ax^{-\alpha}$
- Then the CDF is also a power law:

$$P(x) = Cx^{-(\alpha-1)}$$

Rank-Frequency Plot

- Sort data and plot it against its rank.
- Data: 4, 6, 7, 11, 18. Ranks: 5, 4, 3, 2, 1



Rank-Frequency Plot

- Equivalent to the CDF
- Is a fast and efficient way to form a CDF
- Originally introduced for word frequency, but is used to plot things that aren't frequencies.
- Sometimes ranks are plotted on the horizontal instead of vertical axis.

Estimating Alpha

- Given a set of data, x_1, x_2 , … x_n, what is the best way to estimate alpha?
- Using least squares to estimate the slope of the line on a log-log plot of either the distribution or the CDF is unreliable.
- Instead, use the maximum likelihood estimator.

Maximum Likelihood Estimator

- Given a set of data, x_1, x_2 , … x_n, and given a distribution with a parameter alpha, what is the best way to estimate alpha?
- Likelihood function:

$$\mathcal{L} = p(x|\alpha) = p(x_1|\alpha)p(x_2|\alpha)\cdots p(x_3|\alpha)$$

L is the probability of data given the alpha.

- Choose alpha that maximizes L.

MLE for alpha

$$\hat{\alpha} = 1 + n \left(\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}} \right)^{-1},$$

where x_{min} is the lower bound for the power-law region.

- This formula is for a continuous x.
- The formula for discrete x is different.
- See, Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. "Power-law distributions in empirical data." *SIAM review.* 51.4 (2009): 661-703.

The MLE is the way to go

Method	Notes	est. α (Discrete)	est. α (Continuous)
LS + PDF	const. width	1.5(1)	1.39(5)
LS + CDF	const. width	2.37(2)	2.480(4)
LS + PDF	log. width	1.5(1)	1.19(2)
LS + CDF	rank-freq.	2.570(6)	2.4869(3)
cont. MLE	–	4.46(3)	**2.50(2)**
disc. MLE	–	**2.49(2)**	2.19(1)

- Results for synthetic data with alpha = 2.5
- The MLE clearly is the best estimator.
- Note that it is important to distinguish between discrete and continuous distributions.
- Table 2 from: Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. "Power-law distributions in empirical data." *SIAM review.* 51.4 (2009): 661-703.

Estimating x_{min}

- Try a series of possible x_{min} values.
- For each, estimate alpha. Calculate distance between data and model $p(x) = Ax^{-\hat{\alpha}}$ with $x_{\min}$
- Choose the x_{min} that leads to a model that minimizes the distance between data and model.
- For distance, use the Kolmogorov-Smirnoff distances between the two CDFs.

Goodness of Fit

- We now know how to estimate the best alpha and x_{min}. But how good is the resulting fit?
- Fit will never be exact.
- How good a fit would we expect?

p-value for Goodness of Fit

1. Determine distance d between data and model with $p(x) = Ax^{-\hat{\alpha}}$, with $\widehat{x_{\min}}$
2. Generate synthetic data by sampling from p(x)
3. Estimate alpha, x_{min} for synthetic data to determine a model $p_i(x)$ for synthetic data .
4. Calculate d_i, the distance between synthetic data and model $p_i(x)$.
5. Repeat 2-4 many times.

The p-value = fraction of $d_i \geq d$.

p-value for Goodness of Fit

- A larger p-value means it was less likely that the original fit distance d was the result of chance.
- Larger p-value implies a better fit and thus stronger evidence in support of the hypothesis that x is distributed according to a power law.
- In general, it is important to have a measure of goodness of fit to accompany any fitting process.

Log-normal Distributions

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

- Arises if positive random variables are multiplied together.
- If x is log-normally distributed, y=ln(x) is normally distributed.
- If y is normally distributed, $x=e^y$ is log-normally distributed.

Comparing among alternatives

- It is important to check other distributions that might be better fits than a power law
- Find optimal parameters for power law and alternative distributions.
- Estimate p-values for alternatives, see if they can be ruled out.
- If not, calculate likelihood ratio.
- See, Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. "Power-law distributions in empirical data." *SIAM review.* 51.4 (2009): 661-703.

Do not use OLS on a log-log plot to estimate alpha.

- Instead, use the Maximum-Likelihood estimator and bootstrap methods described in: Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. "Power-law distributions in empirical data." SIAM review. 51.4 (2009): 661-703.

What processes generate power-laws?

- Rich get richer, preferential attachment
- Exponential growth exponentially sampled
- Multiplicative process with lower threshold
- Optimization
- Phase transitions

In my opinion, phase transitions explain a tiny fraction of observed power laws.

There are many different ways of generating power laws.

So what’s the deal? Are power laws interesting? Do you need to establish that your data IS a power law?

1. **Long Tails**:

- It might be enough to establish that your data has a long tail.
- This would have strong implications for how one thinks about risk, outliers, extreme events, etc.

So what's the deal? ….

2. **Simple Mechanism(?)**:

- Observing a power law may suggest a simple mechanism…
- But says **nothing** about the nature of that mechanism.

So what’s the deal? ….

3. Suggests **Scale may not Matter**:

- Power laws are a way to look for a particular type of pattern…
- One that tells us that there are similarities across scales.

Do you need to establish that your data IS a power law?

- Do you want to establish that there is a scaling pattern? Or...
- Do you want to say that your data **is** a power law?
- Is there any theory that predicts power-law behavior?
- It is not clear to me how strong the evidence needs to be to make a power law claim.

Main Points

- Fractal is not an either-or category. Objects are more or less fractal-like.
- Statistical estimation for power laws is tricky and often done wrong.
- There are many ways to generate fractals and power laws.
- Knowing something is a power law is important and interesting, but does not necessarily reveal its essence.