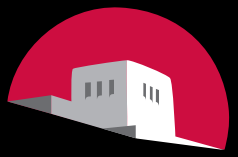
CS 365
Introduction to Scientific Modeling
Lecture 4: Power Laws and Scaling

Stephanie Forrest
Dept. of Computer Science
Univ. of New Mexico
Fall Semester, 2014



THE UNIVERSITY *of*
NEW MEXICO

Readings

- Mitchell, Ch. 15 – 17
- http://www.complexityexplorer.org/online-courses/11
  - Units 9, 10
- Newman Ch. 8
- *Metabolic Ecology Ch. 24* "Beyond biology"

Topics

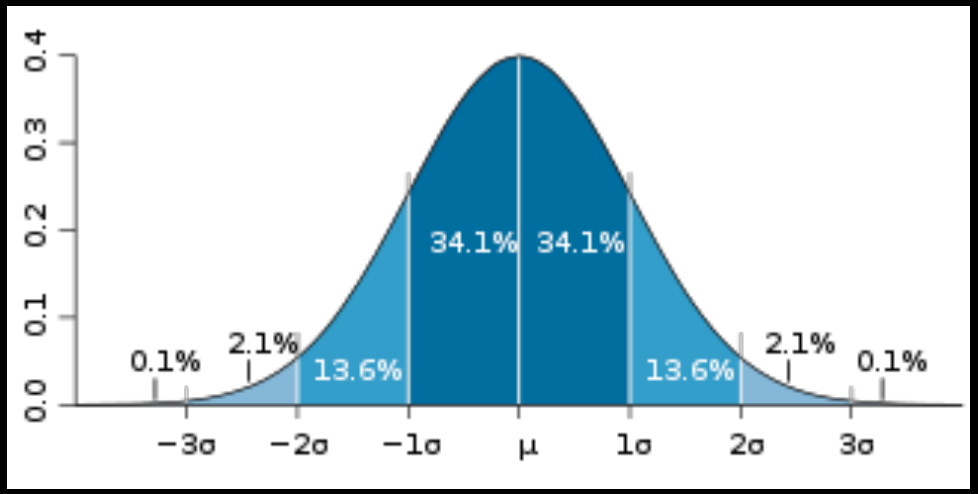
- Statistical distributions
  - Normal
  - Exponential
  - Power law
- Power laws in nature and computing
  - Complex networks
- Detecting power laws in data
- How power laws are generated
- Special topics
  - Metabolic scaling theory

Reflection

“I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error". The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement, amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshaled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.”

Sir Francis Galton (Natural Inheritance, 1889)

The Normal Distribution



So: Wikipedia

- Many psychological and physical phenomena (e.g., noise) are well approximated by the normal distribution:
  - Height of a man (woman)
  - Velocity in any direction of a molecule in a gas
  - Error made in measuring a physical quantity
  - Many small independent effects contribute **additively** to each observation
  - Abraham de Moivre (1733) used the normal distribution to approximate the distribution of the number of heads resulting from many tosses of a fair coin.

The Central Limit Theorem

- Let $X_1, X_2, X_3, ..., X_n$ be a sequence of *n* independent and identically distributed (i.i.d.) random variables each having finite values of expectation μ and variance $\sigma^2 > 0$.

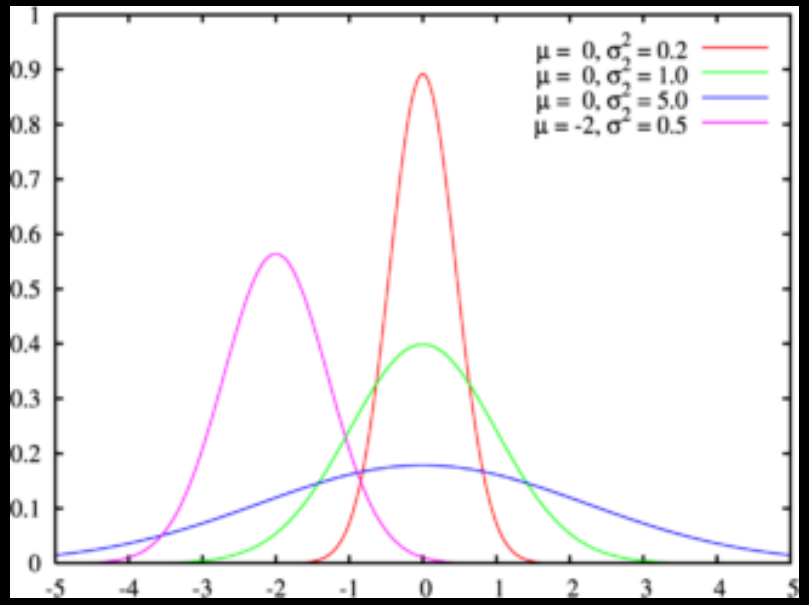
- Th: As the sample size *n* increases, the distribution of the sample **average** of these random variables approaches the normal distribution with a mean μ and variance σ^2/n regardless of the shape of the original distribution.

- $Var(X) = E[(X - \mu)^2 = E[X^2] - \mu^2 = \sigma^2$

$$Var(X) = \sum_{i=1}^{n} p_i (x_i - \mu)^2$$

The Normal (Gaussian) Distribution

- The "Bell curve"
  - A histogram of samples
  - Peaked around a typical value
- Standard Normal Distribution:
  - Mean: 0.0
  - Standard deviation: 1.0
- Variance:
  - 68.3% of values are within +/- 1 σ of mean
  - 95.5% of values are withing +/- 2 σ of mean
  - 99.7% of values are within +/- 3 σ of mean

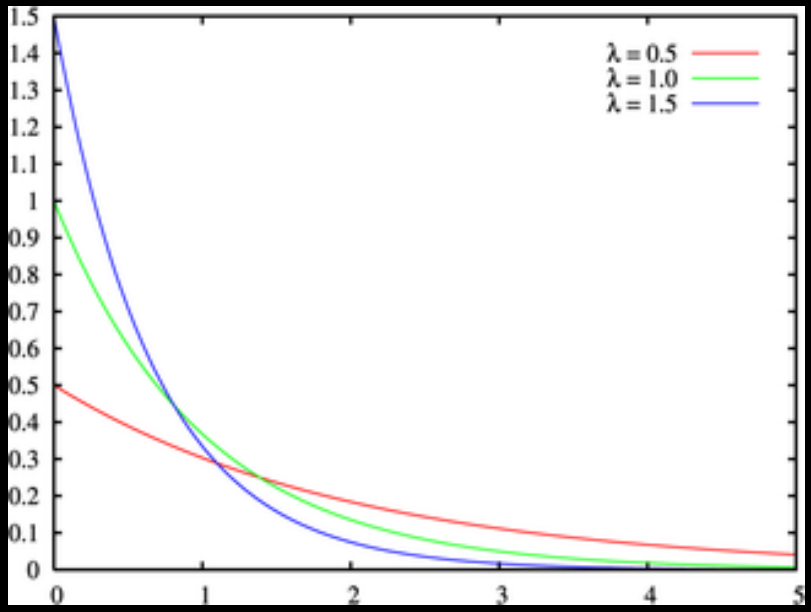


$$P(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2}$$

Exponential Distribution

- In practice, the distribution of the length of time until some specific event occurs:
  - Until an earthquake occurs
  - A new war breaks out
  - A telephone call I receive turns out to be the wrong number

$$p(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$



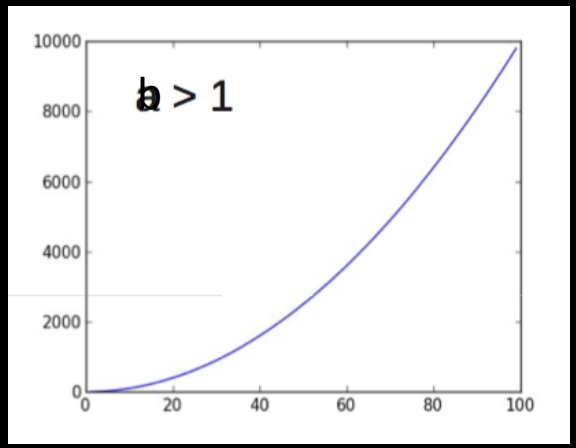
$$\lambda = \frac{1}{\text{avg. value of random var.}}$$

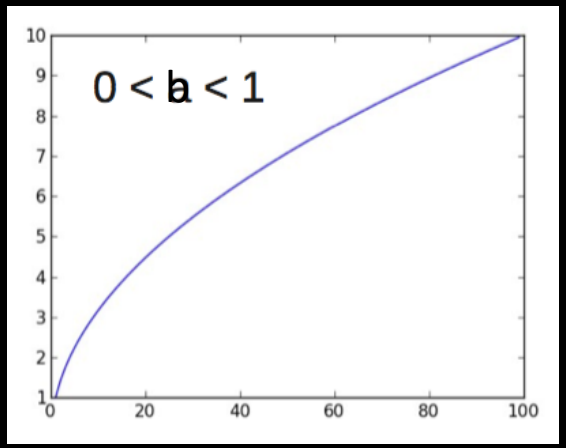
- Memoryless:

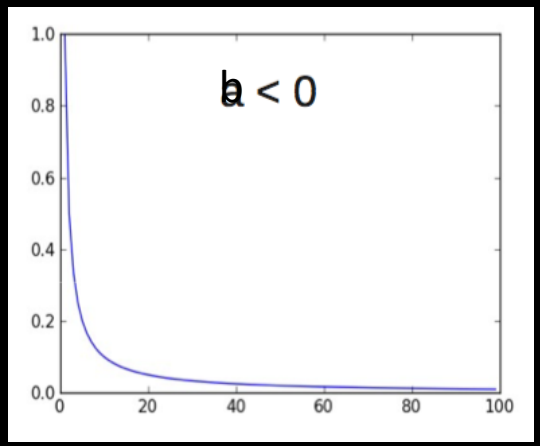
$$p\{X > s + t \mid X > t\} = p\{X > s\} \text{ for } s,t \geq 0$$

Power Law Distribution

- Polynomial: $p(x) = ax^b$
- Scale invariant: $p(cx) = a(cx)^b = c^b p(x) \propto p(x)$
- Distribution can range over many orders of magnitude
  - Ratio of largest to smallest sample





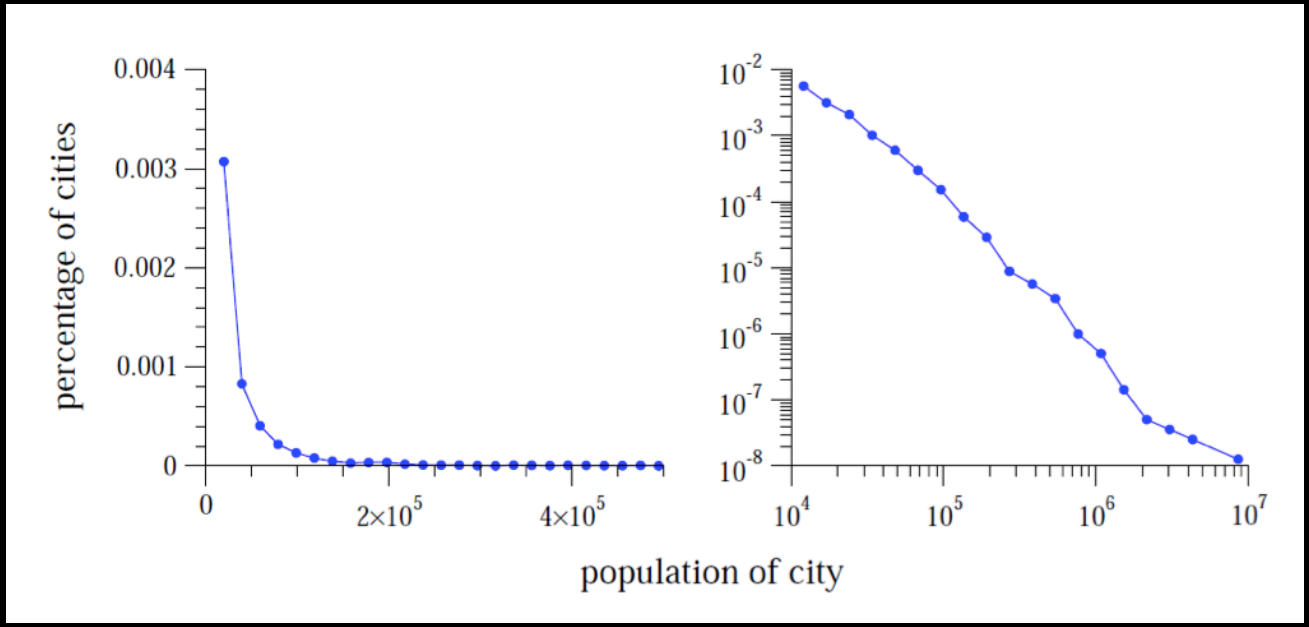


What happens if we plot on log-log scale?

What happens if we plot on log-log scale?

$$\log(p(x)) = \log(ax^b) = b\log x + \log a$$

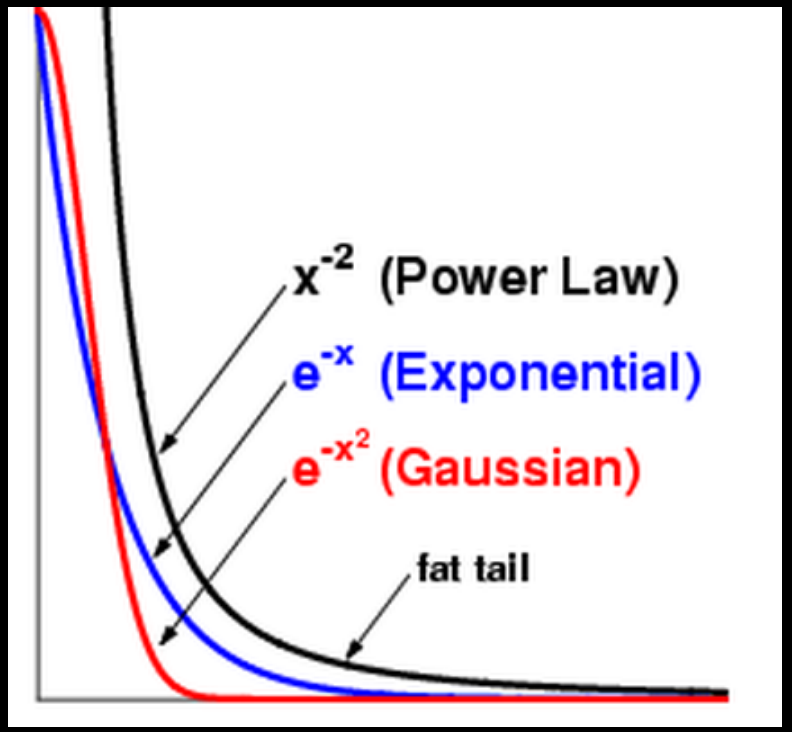
Histogram of the populations of all US cities US with population 10000 or more.



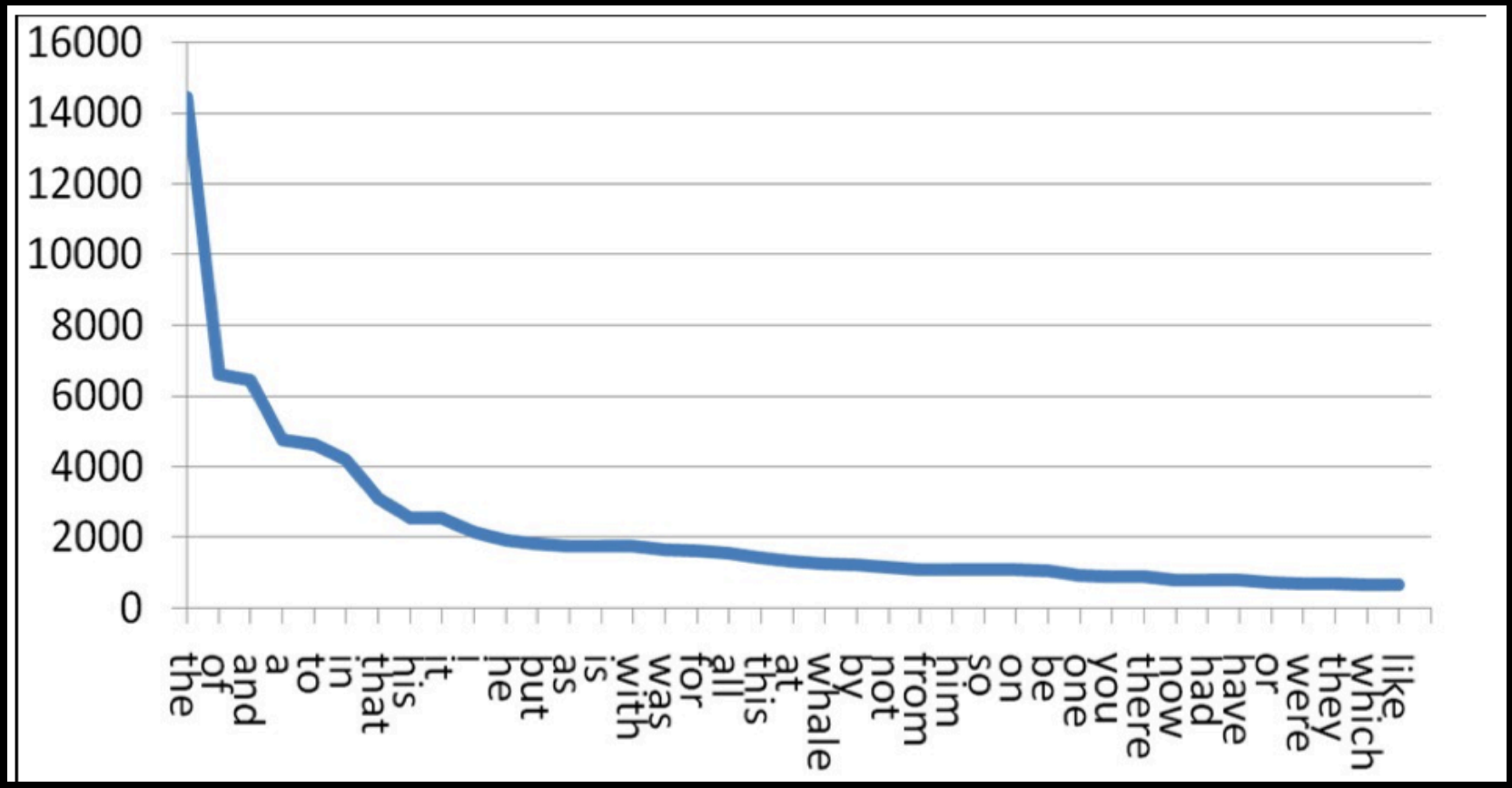
- Slope of line gives scaling exponent b
- y-intercept gives the constant a

Comparison

- Exponential Distributions:
  - aka single-scale
  - Have form $P(x) = e^{-ax}$
  - Use Gaussian to approximate exponential because differentiable at 0.
  - Plot on log-linear scale to see straight line.
- Power-law Distributions:
  - aka scale-free or polynomial
  - Have form $P(x) = x^{-a}$
  - Fat (heavy) tail is associated with power law because it decays more slowly.
  - Plot on log-log scale to see straight line.

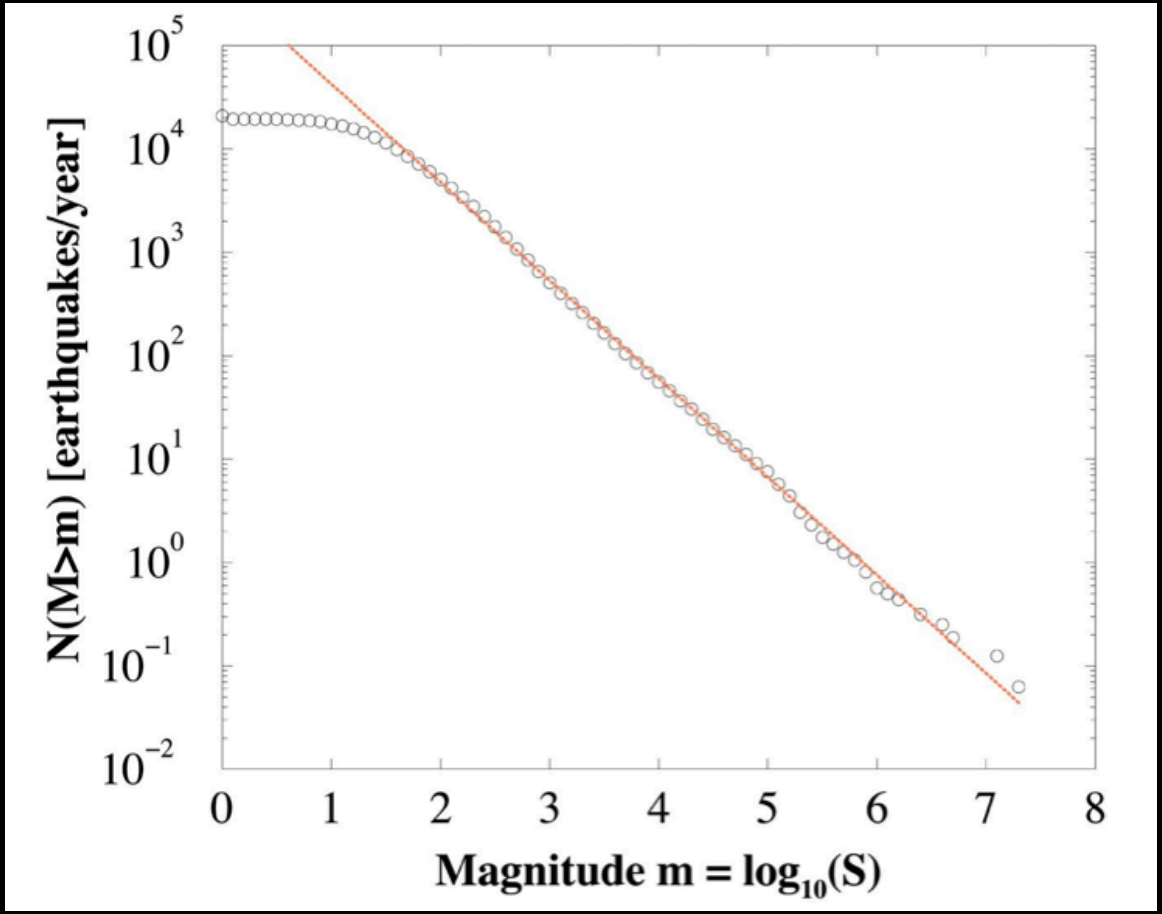


Zipf's Law



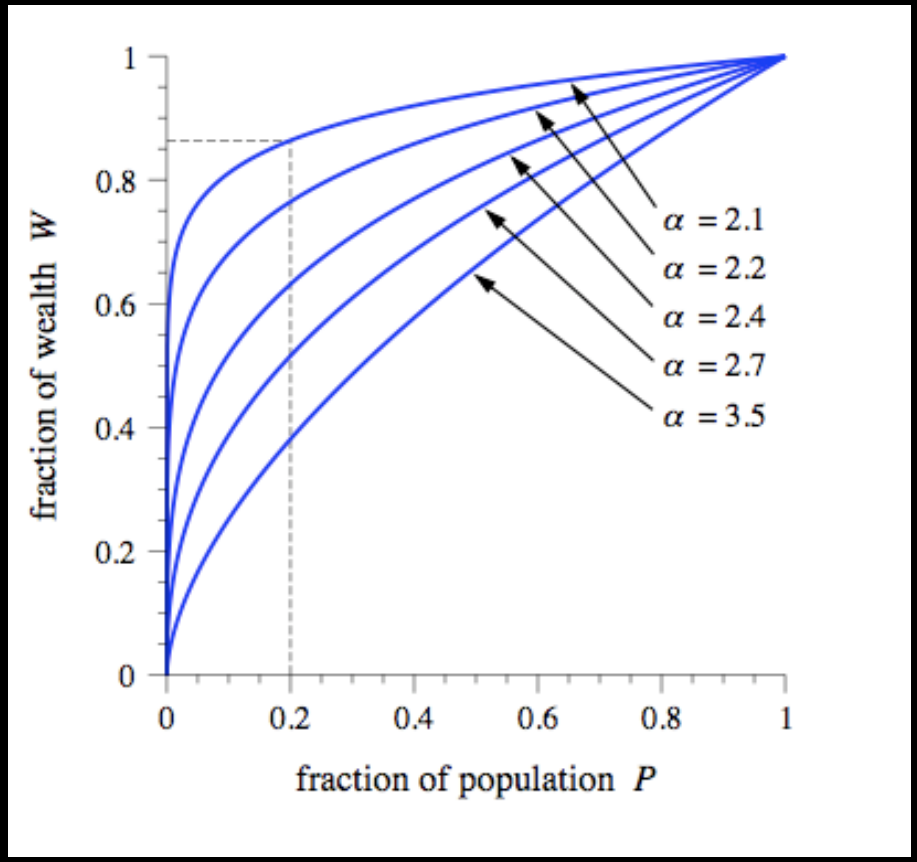
Number of occurrences of words in the book Moby Dick

Earthquakes



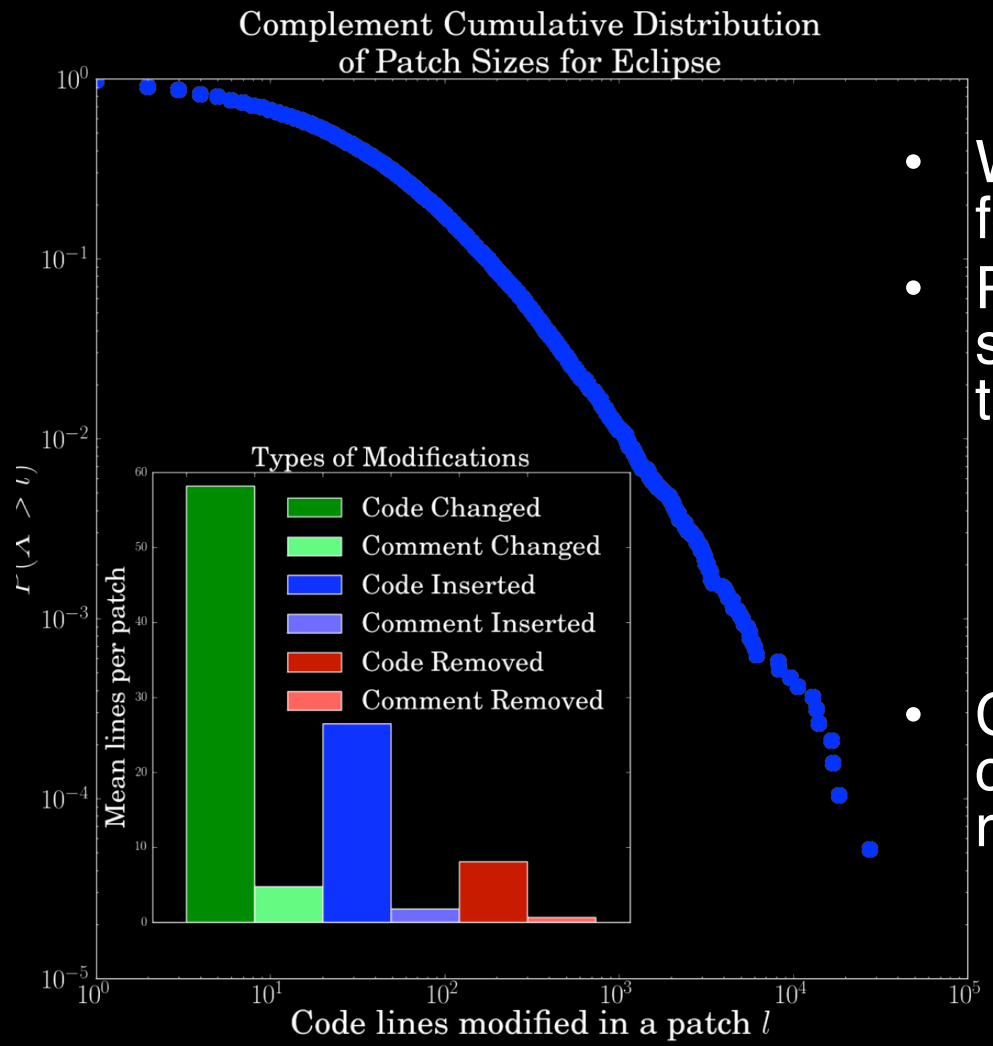
Earthquake magnitude distribution over 6 decades (K Christensen, L. Danon, T. Scanlon, and Per Bak "Unified scaling law for eqrthquakes" *PNAS 2002* 99:2509-2513

Distribution of Wealth



The 80/20 Rule: The fraction W of the total wealth in a country held by the fraction P of the richest people, if wealth is distributed according to a power law with exponent alpha. (Newman, 2006)

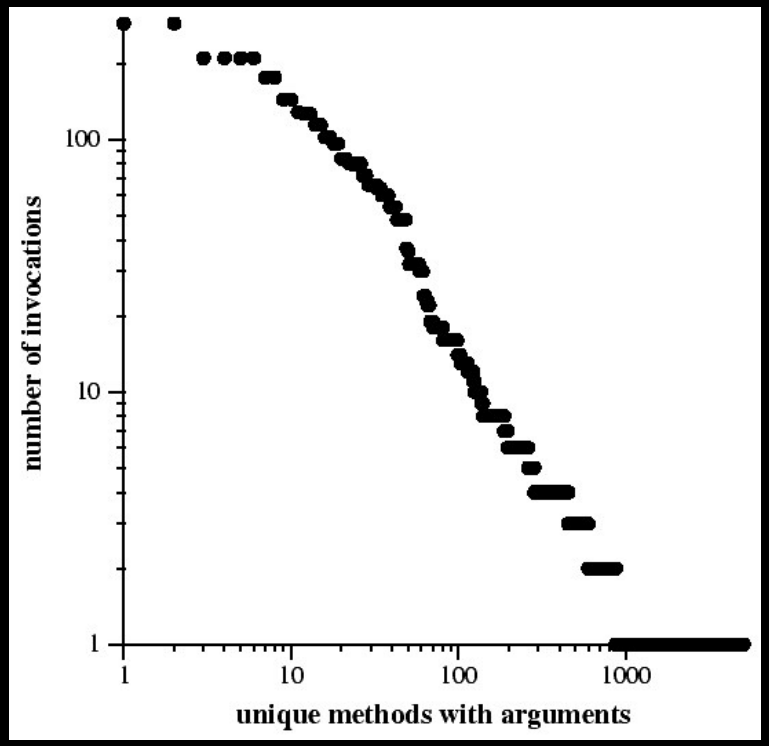
Software Bug Size Distribution



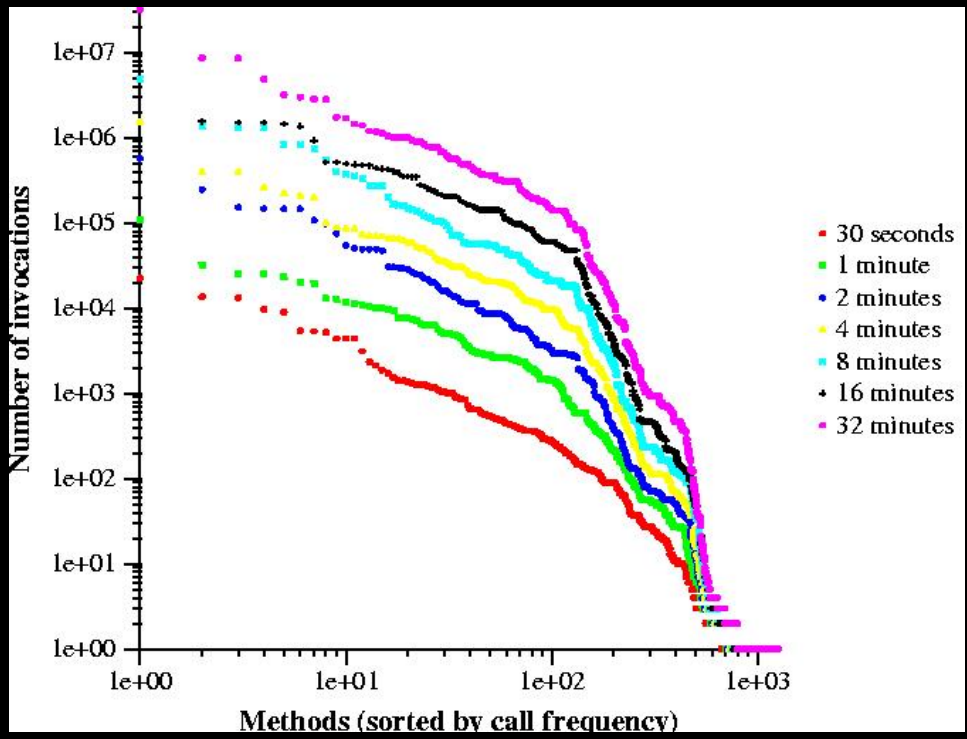
- Weimer studied 20,000 patches for Eclipse (unpublished data)
- Patch: CVS checkin that explicitly states “I am fixing bug #1234 in the log message.
  - 10% of patches are 2 lines or less
  - 20% of patches are 5 lines or less
  - 53% of patches are 25 lines or less
- Changes to lines involving only comments or whitespace were rare

Scaling in Software Execution

H. Inoue

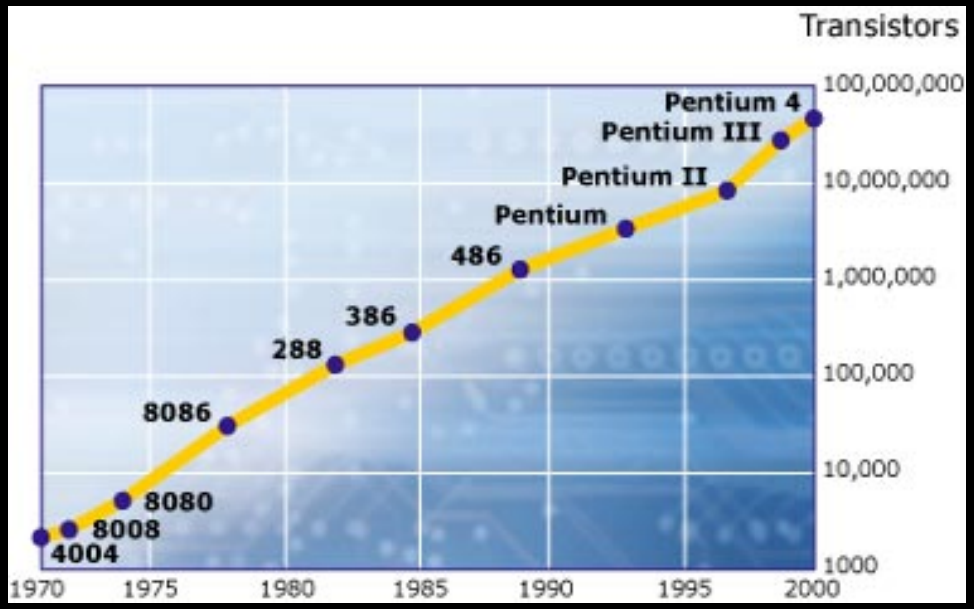


HelloWorld: Unique Function Calls vs. Invocation Freq

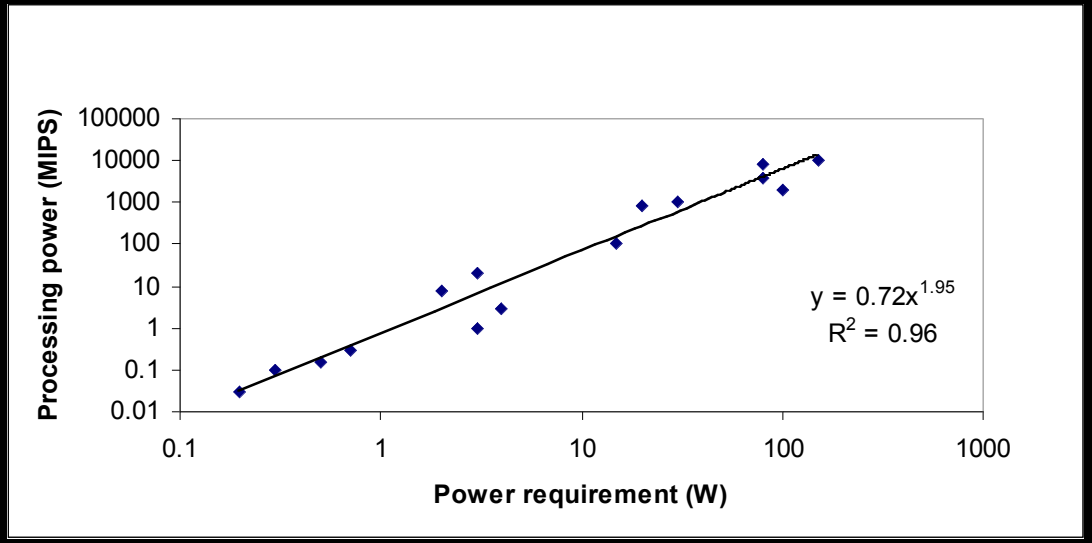


LimeWire Behavior

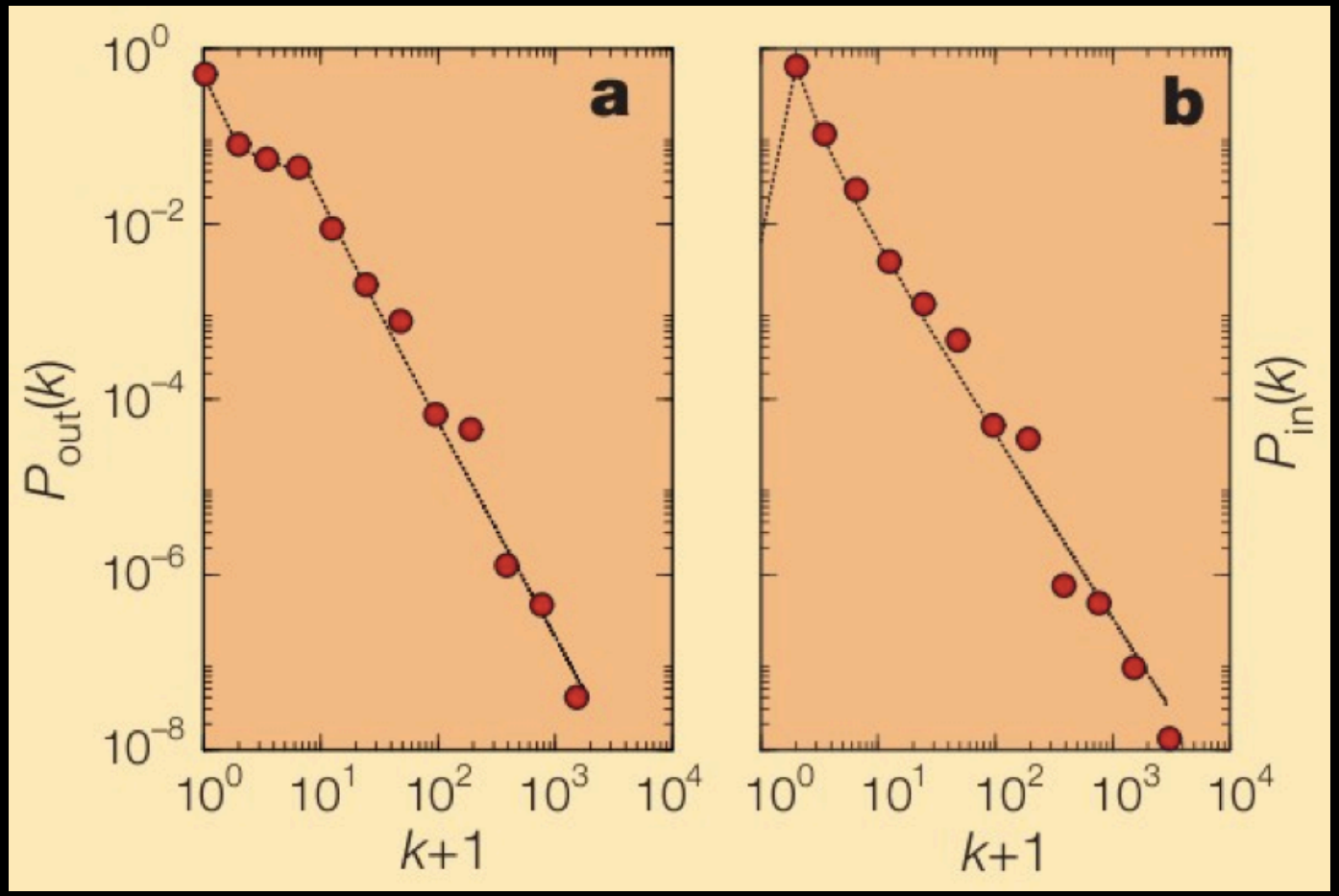
Power Scaling



1970: 100 Watts powers 15 MIPS
2005: 100 Watts powers 6700 MIPS
Wire scaling prevents better returns

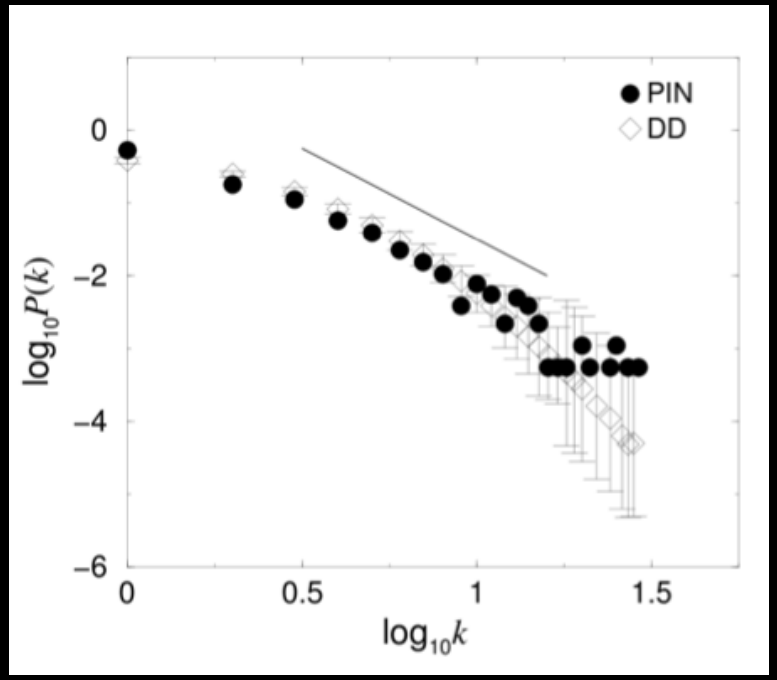


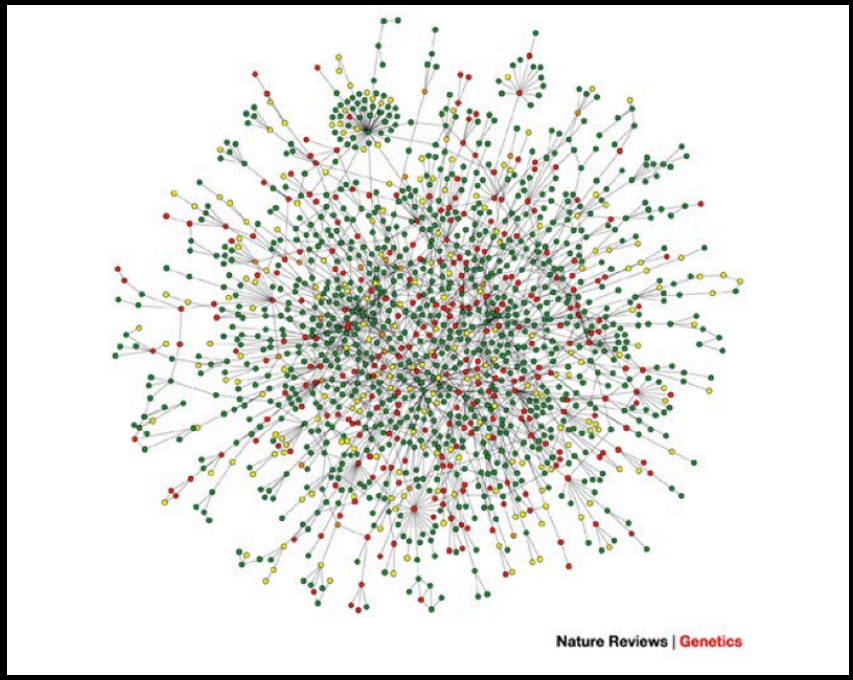
The World Wide Web



Albert, Jeong, Barabasi (1999)

Protein Interaction Networks

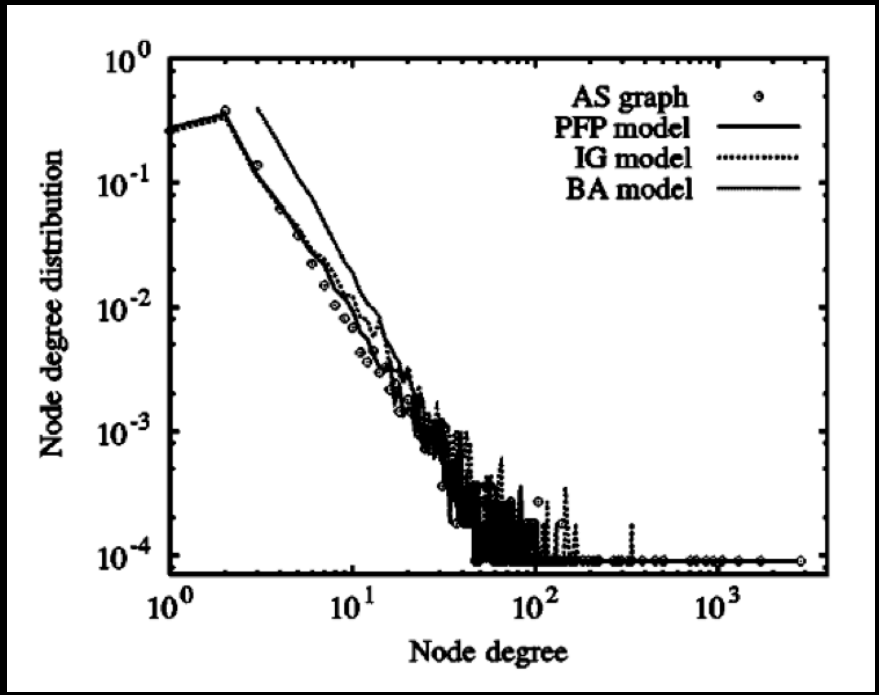


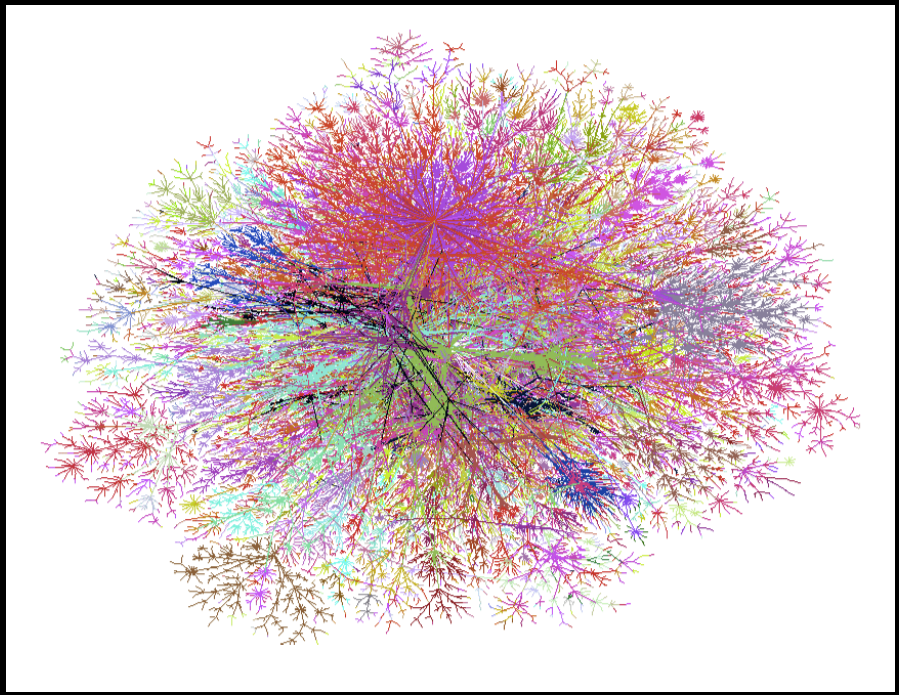


Complex Networks Examples

- Social networks
- Airline connections
- Scientific collaborations
- Email connections
- Metabolic networks
- Actors
- Semantic networks
- Neuronal networks
- Gene regulatory networks
- Terrorist networks
- Software call networks
- Food webs

The Physical Structure of the Internet





Power laws in nature

- Thermal noise in electronic devices.
- Flashing of fireflies.
- Sizes of forest fires.
- Distribution of earthquake sizes (Gutenberg-Richter law).
- Distribution of solar flares and sunspots.
- Size distributions of initial masses of stars (Salpeter law).
- Allometric scaling relationships.
- Zipf's Law (frequency of word use in any language)
- Number of papers that scientists write (and cite)
- Stock market activity.
- Sizes of towns and cities.
- Number of hits on web pages.
- Distributions of function invocations in Java programs.

Power Laws in Physical Systems

- Power spectral density distribution (1/f noise)
  - f = frequency, P(f) is the power at frequency f

$$P(f) \sim \frac{1}{f^{\alpha}}$$

- Size distributions (allometry)
  - s = size of event, N(s) = frequency of event s

$$N(s) \sim \frac{1}{s^{t}}$$

- Temporal distributions of events (e.g., sandpile models)
  - τ is either the duration of an event or the time between events

$$N(\tau) \sim \frac{1}{\tau^{\gamma}}$$

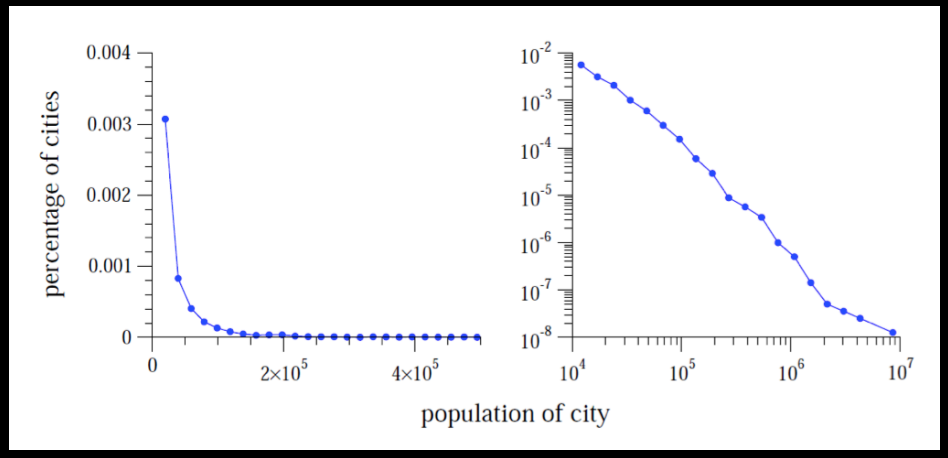

NO G-20
REPEAL POWER LAWS

How do power laws arise?

Review: Power Law Distribution

- Polynomial: $p(x) = ax^b$
- Scale invariant: $p(cx) = a(cx)^b = c^b p(x) \propto p(x)$
- Take the log of both sides of the equation:

$$\log(p(x)) = \log(ax^b) = b\log x + \log a$$

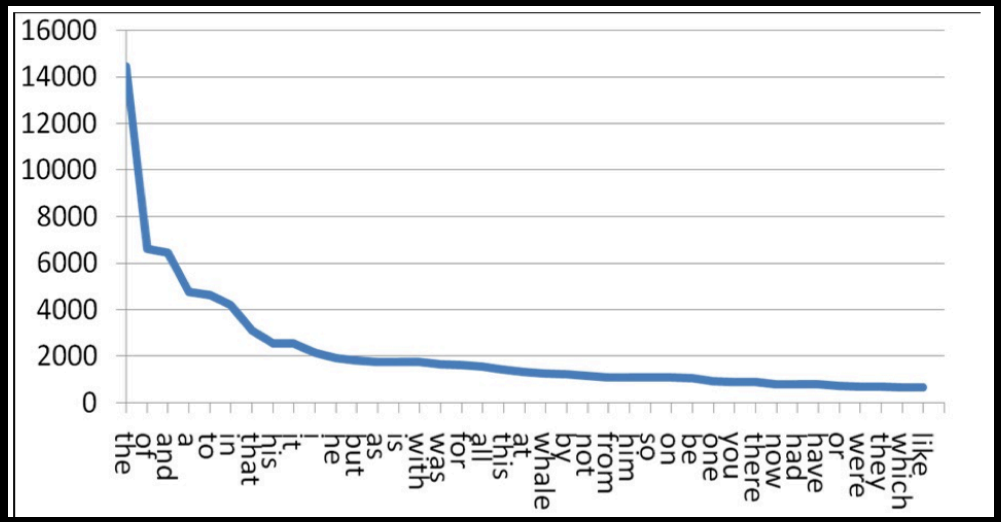


Slope of line gives scaling exponent b
y-intercept gives the constant a

Why do we care?

Measuring Power Laws

- Data-driven modeling
  - I give you some data
  - What do you do to determine if it follows a power law, or some other distribution?

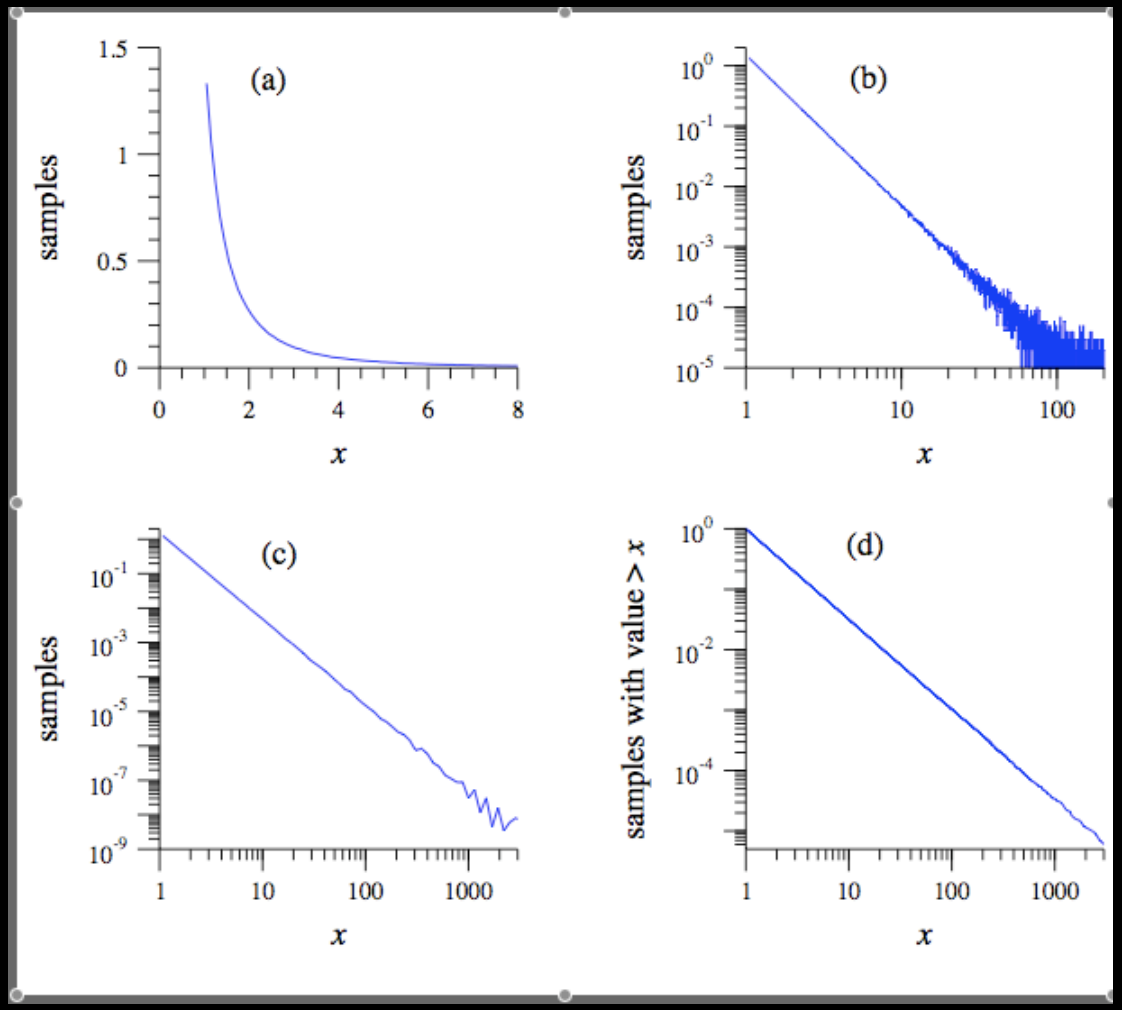


Is this a power law?

Measuring Power Laws

- Plot histogram of samples on log-log axes (a):
  - Test for linear form of data on plot
  - Measure slope of best-fit line to determine scaling exponent
- Problem: Noise in right-hand side of distribution (b)
  - Each bin on the right-hand side of plot has few samples
  - Correct with logarithmic binning (c)
  - Divide # of samples by width of bin to get
    - Count per unit interval of x

M. Newman Power laws, Pareto Distributions and Zipf's Law (2006)
1 million random numbers, with b = 2.5



A better approach

The Cumulative Density Function (CDF)

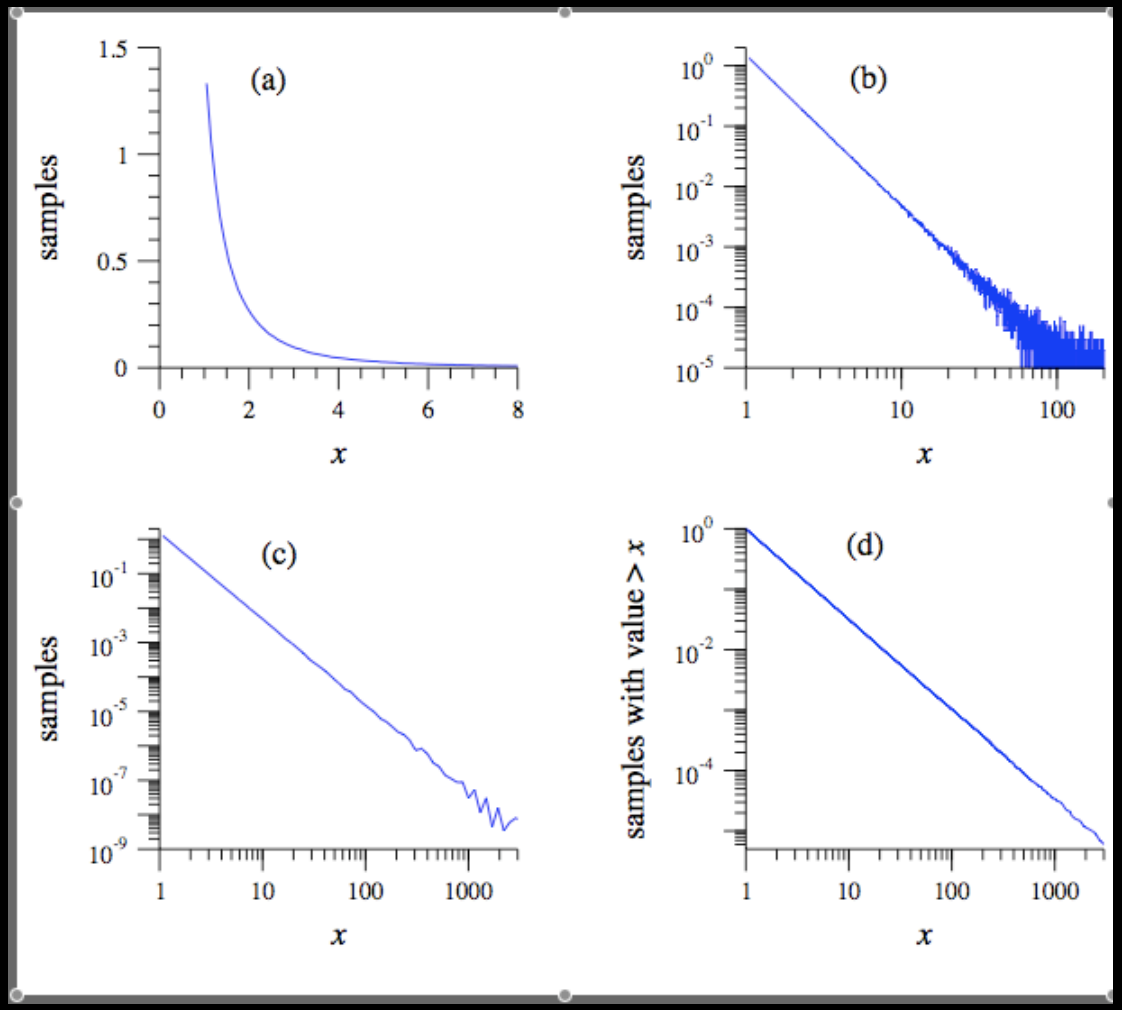
- Complement Cumulative Distribution function (d)

$$P(x) = \int_{x}^{\infty} p(y)dy$$

  - Probability P(x) that x has a value greater than y (1 - CDF)
  - Also follows power law but with the exponent –b + 1
  - No need to use logarithmic binning
  - Sometimes called rank/frequency plots
- For power laws

$$P(x) = \int_x^\infty p(y)dy = a\int_x^\infty y^{-b}dy = \frac{a}{b-1}y^{-b+1}\,|_x^\infty = 0 - a\frac{x^{-b+1}}{(b-1)}$$

M. Newman Power laws, Pareto Distributions and Zipf's Law (2006)
1 million random numbers, with b = 2.5

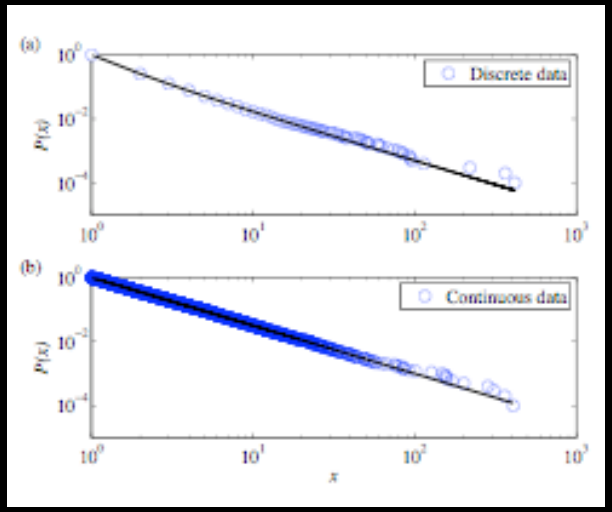


Determining the Scaling Exponent (slope)

- Fit a "best fit" line through the (logged) data and measure its slope.
  - See next two slides for one method of computing the best fit.
- For distributional data: Use the following formula (Newman 2006):

$$b = 1 + n\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}}\right]^{-1}$$

  - This is an estimator for b, based on maximum likelihood estimation
  - Consider a sequence of observations x_i, e.g., 0, 0, 0, 1, 1, 1, 3.2, 3.3, …
  - Throw out all 0s or negative numbers
    - Power law distribution not defined for neg. values
    - OK because of scale-free property
  - We apply this formula instead of creating the histogram $P(x_i)$



Points represent the cumulative density functions P(x) for synthetic datasets distributed according to: (a) a discrete powerlaw and (b) a continuous power law, both with α=2.5 and x_{min}=1. Solid lines represent best fits to the data using the methods described in the text. (Clauset et al. 2007)

Linear Empirical Models

- An *empirical model* is a function that captures the trend of observed data:
  - It *predicts* but does *not explain* the system that produced the data.
- A common technique is to fit a line through the data:

$$y = mx + b$$

- Assume Gaussian distributed errors.
  - Note: For logged data, we assume that the errors are log-normally distributed.

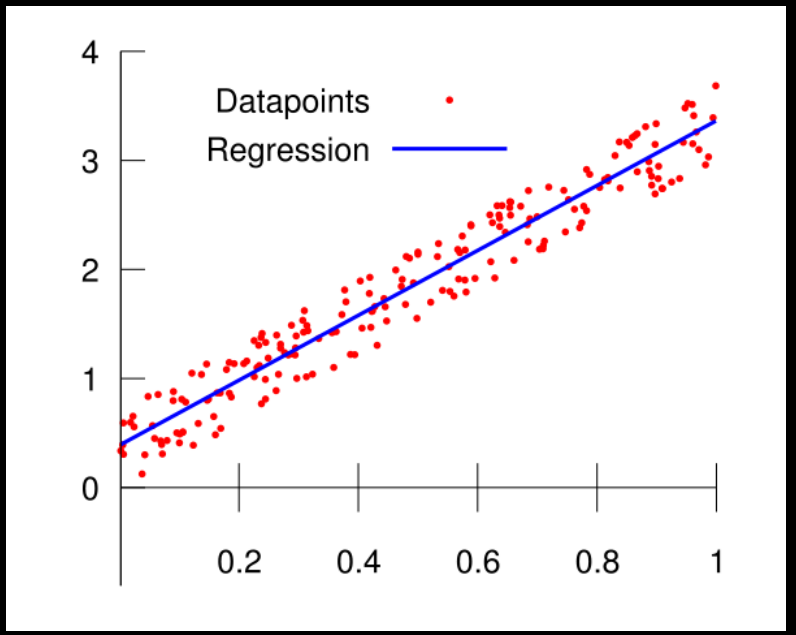


Image downloaded from Wikipedia Sept. 11, 2007

Linear Regression

- Least-squares fit uses linear regression
  - Goal: Find the line y = mx+b that minimizes the sum of squares

$$\min(\sum_{i=1}^{n}(mx_i + b - y_i)^2)$$

  - M and b computed using the following formulae:

$$m = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{\sum x_i^2 \sum y_i - \sum x_i y_i \sum x_i}{n\sum x_i^2 - (\sum x_i)^2}$$