# Power Law Size Distributions

Principles of Complex Systems
CSYS/MATH 300, Fall, 2011

Prof. Peter Dodds

Department of Mathematics & Statistics
Center for Complex Systems
Vermont Advanced Computing Center
University of Vermont









## Outline







## Size distributions

The sizes of many systems' elements appear to obey an inverse power-law size distribution:

$$P(\text{size} = x) \sim c\, x^{-\gamma}$$

$$\text{where} \quad 0 < x_{\min} < x < x_{\max}$$

$$\text{and} \quad \gamma > 1$$

- Exciting class exercise: sketch this function.
- $x_{\min}$ = lower cutoff
- $x_{\max}$ = upper cutoff
- Negative linear relationship in log-log space:

$$\log_{10} P(x) = \log_{10} c - \gamma \log_{10} x$$

- We use base 10 because we are good people.





## Size distributions

Usually, only the tail of the distribution obeys a power law:

$$P(x) \sim c\, x^{-\gamma} \text{ for } x \text{ large.}$$

- Still use term 'power law distribution.'
- Other terms:
  - Fat-tailed distributions.
  - Heavy-tailed distributions.

Beware:

- Inverse power laws aren't the only ones: lognormals (⊞), Weibull distributions (⊞), . . .





## Size distributions

Many systems have discrete sizes $k$:

- Word frequency
- Node degree in networks: # friends, # hyperlinks, etc.
- # citations for articles, court decisions, etc.

$$P(k) \sim c\, k^{-\gamma}$$

$$\text{where} \quad k_{\min} \le k \le k_{\max}$$

- Obvious fail for $k = 0$.
- Again, typically a description of distribution's tail.





## The statistics of surprise—words:

Brown Corpus (⊞) ($\sim 10^6$ words):

| rank | word | % $q$ |
|---|---|---|
| 1. | the | 6.8872 |
| 2. | of | 3.5839 |
| 3. | and | 2.8401 |
| 4. | to | 2.5744 |
| 5. | a | 2.2996 |
| 6. | in | 2.1010 |
| 7. | that | 1.0428 |
| 8. | is | 0.9943 |
| 9. | was | 0.9661 |
| 10. | he | 0.9392 |
| 11. | for | 0.9340 |
| 12. | it | 0.8623 |
| 13. | with | 0.7176 |
| 14. | as | 0.7137 |
| 15. | his | 0.6886 |

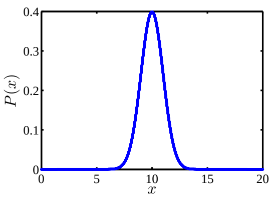
| rank | word | % $q$ |
|---|---|---|
| 1945. | apply | 0.0055 |
| 1946. | vital | 0.0055 |
| 1947. | September | 0.0055 |
| 1948. | review | 0.0055 |
| 1949. | wage | 0.0055 |
| 1950. | motor | 0.0055 |
| 1951. | fifteen | 0.0055 |
| 1952. | regarded | 0.0055 |
| 1953. | draw | 0.0055 |
| 1954. | wheel | 0.0055 |
| 1955. | organized | 0.0055 |
| 1956. | vision | 0.0055 |
| 1957. | wild | 0.0055 |
| 1958. | Palmer | 0.0055 |
| 1959. | intensity | 0.0055 |

## The statistics of surprise—words:

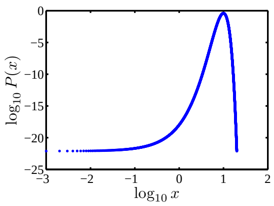
### First—a Gaussian example:

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

linear:




log-log




mean $\mu = 10$, variance $\sigma^2 = 1$.

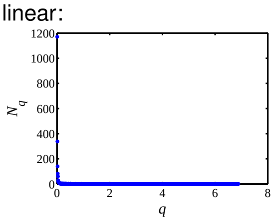




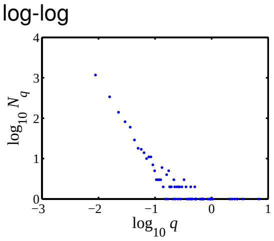
## The statistics of surprise—words:

### Raw 'probability' (binned):

linear:




log-log



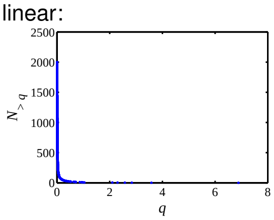




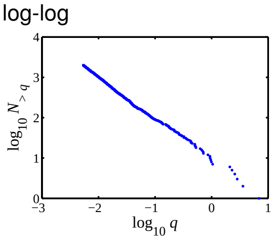

## The statistics of surprise—words:

### 'Exceedance probability':

linear:



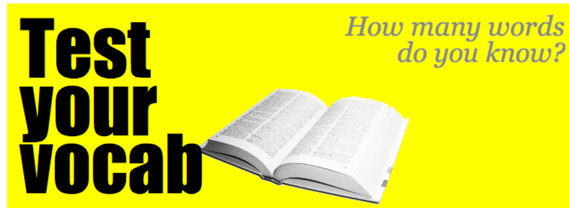

log-log








## My, what big words you have...



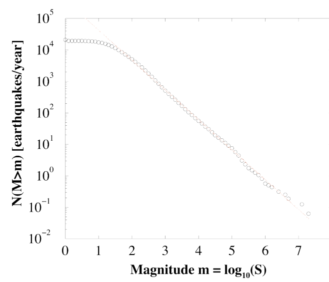

- Test capitalizes on word frequency following a heavily skewed frequency distribution with a decaying power law tail.
- Let's do it collectively... (⊞)





## The statistics of surprise:

Gutenberg-Richter law (⊞)




- Log-log plot
- Base 10
- Slope = -1

$N(M > m) \propto m^{-1}$

- From both the very awkwardly similar Christensen et al. and Bak et al.: "Unified scaling law for earthquakes" [3, 1]





## The statistics of surprise:

From: "Quake Moves Japan Closer to U.S. and Alters Earth's Spin" (⊞) by Kenneth Chang, March 13, 2011, NYT:

What is perhaps most surprising about the Japan earthquake is how misleading history can be. In the past 300 years, no earthquake nearly that large—nothing larger than magnitude eight—had struck in the Japan subduction zone. That, in turn, led to assumptions about how large a tsunami might strike the coast.

"It did them a giant disservice," said Dr. Stein of the geological survey. That is not the first time that the earthquake potential of a fault has been underestimated. Most geophysicists did not think the Sumatra fault could generate a magnitude 9.1 earthquake, . . .

# Great:

## Two things we have poor cognitive understanding of:

1. Probability
   - Ex. The Monty Hall Problem (⊞)
   - Ex. Son born on Tuesday (⊞).
2. Logarithmic scales.

## On counting and logarithms:

- Listen to Radiolab's "Numbers." (⊞).
- Later: Benford's Law (⊞).

$x^{-\gamma}$





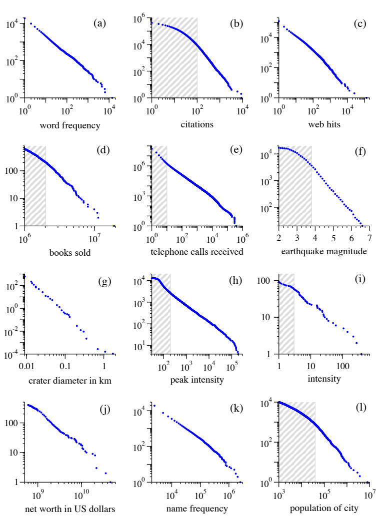



FIG. 4 Cumulative distributions or "rank/frequency plots" of twelve quantities reputed to follow power laws. The distributions were computed as described in Appendix A. Data in the shaded regions were excluded from the calculations of the exponents in Table I. Source references for the data are given in the text. (a) Numbers of occurrences of words in the novel *Moby Dick* by Hermann Melville. (b) Numbers of citations to scientific papers published in 1981, from time of publication until June 1997. (c) Numbers of hits on web sites by 60 000 users of the America Online Internet service for the day of 1 December 1997. (d) Numbers of copies of bestselling books sold in the US between 1895 and 1965. (e) Number of calls received by AT&T telephone customers in the US for a single day. (f) Magnitude of earthquakes in California between January 1910 and May 1992. Magnitude is proportional to the logarithm of the maximum amplitude of the earthquake, and hence the distribution obeys a power law even though the horizontal axis is linear. (g) Diameter of craters on the moon. Vertical axis is measured per square kilometre. (h) Peak gamma-ray intensity of solar flares in counts per second, measured from Earth orbit between February 1980 and November 1989. (i) Intensity of wars from 1816 to 1980, measured as battle deaths per 10 000 of the population of the participating countries. (j) Aggregate net worth in dollars of the richest individuals in the US in October 2003. (k) Frequency of occurrence of family names in the US in the year 1990. (l) Populations of US cities in the year 2000.

$x^{-\gamma}$





# Size distributions

## Examples:

- Earthquake magnitude (Gutenberg-Richter law (⊞)):[1] $P(M) \propto M^{-2}$
- Number of war deaths:[9] $P(d) \propto d^{-1.8}$
- Sizes of forest fires[4]
- Sizes of cities:[10] $P(n) \propto n^{-2.1}$
- Number of links to and from websites[2]

- See in part Simon[10] and M.E.J. Newman[6] "Power laws, Pareto distributions and Zipf's law" for more.
- Note: Exponents range in error

$x^{-\gamma}$





# Size distributions

## Examples:

- Number of citations to papers:[7, 8] $P(k) \propto k^{-3}$.
- Individual wealth (maybe): $P(W) \propto W^{-2}$.
- Distributions of tree trunk diameters: $P(d) \propto d^{-2}$.
- The gravitational force at a random point in the universe:[5] $P(F) \propto F^{-5/2}$.
- Diameter of moon craters:[6] $P(d) \propto d^{-3}$.
- Word frequency:[10] e.g., $P(k) \propto k^{-2.2}$ (variable)

$x^{-\gamma}$

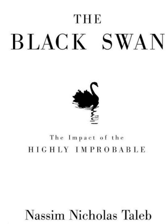




# Power law distributions

## Gaussians versus power-law distributions:

- Mediocristan versus Extremistan
- Mild versus Wild (Mandelbrot)
- Example: Height versus wealth.




- See "The Black Swan" by Nassim Taleb.[11]

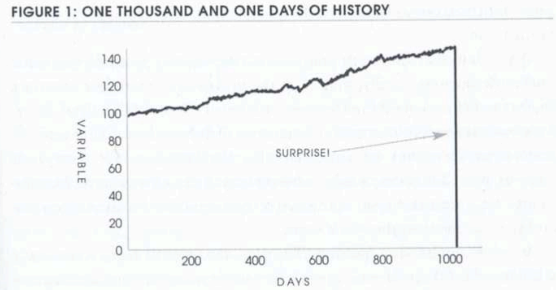
$x^{-\gamma}$





# Turkeys...

FIGURE 1: ONE THOUSAND AND ONE DAYS OF HISTORY




A turkey before and after Thanksgiving. The history of a process over a thousand days tells you nothing about what is to happen next. This naïve projection of the future from the past can be applied to anything.

From "The Black Swan"[11]

$x^{-\gamma}$

## Taleb's table [11]



### Mediocristan/Extremistan

- Most typical member is mediocre/Most typical is either giant or tiny
- Winners get a small segment/Winner take almost all effects
- When you observe for a while, you know what's going on/It takes a very long time to figure out what's going on
- Prediction is easy/Prediction is hard
- History crawls/History makes jumps
- Tyranny of the collective/Tyranny of the rare and accidental



## Size distributions



Power law size distributions are sometimes called Pareto distributions (⊞) after Italian scholar Vilfredo Pareto. (⊞)

- Pareto noted wealth in Italy was distributed unevenly (80–20 rule; misleading).
- Term used especially by practitioners of the Dismal Science (⊞).



## Devilish power law distribution details:



### Exhibit A:

- Given $P(x) = cx^{-\gamma}$ with $0 < x_{\min} < x < x_{\max}$, the mean is ($\gamma \neq 2$):

$$\langle x \rangle = \frac{c}{2-\gamma}\left(x_{\max}^{2-\gamma} - x_{\min}^{2-\gamma}\right).$$

- Mean 'blows up' with upper cutoff if $\gamma < 2$.
- Mean depends on lower cutoff if $\gamma > 2$.
- $\gamma < 2$: Typical sample is large.
- $\gamma > 2$: Typical sample is small.

Insert question from assignment 1 (⊞)



## And in general...



### Moments:

- All moments depend only on cutoffs.
- No internal scale that dominates/matters.
- Compare to a Gaussian, exponential, etc.

### For many real size distributions: $2 < \gamma < 3$

- mean is finite (depends on lower cutoff)
- $\sigma^2$ = variance is 'infinite' (depends on upper cutoff)
- Width of distribution is 'infinite'
- If $\gamma > 3$, distribution is less terrifying and may be easily confused with other kinds of distributions.

Insert question from assignment 1 (⊞)



## Moments



### Standard deviation is a mathematical convenience:

- Variance is nice analytically...
- Another measure of distribution width:

$$\text{Mean average deviation (MAD)} = \langle |x - \langle x \rangle| \rangle$$

- For a pure power law with $2 < \gamma < 3$:

$$\langle |x - \langle x \rangle| \rangle \text{ is finite.}$$

- But MAD is mildly unpleasant analytically...
- We still speak of infinite 'width' if $\gamma < 3$.

Insert question from assignment 2 (⊞)



## How sample sizes grow...



### Given $P(x) \sim cx^{-\gamma}$:

- We can show that after $n$ samples, we expect the largest sample to be

$$x_1 \gtrsim c' n^{1/(\gamma-1)}$$

- Sampling from a finite-variance distribution gives a much slower growth with $n$.
- e.g., for $P(x) = \lambda e^{-\lambda x}$, we find

$$x_1 \gtrsim \frac{1}{\lambda} \ln n.$$

Insert question from assignment 2 (⊞)

## Complementary Cumulative Distribution Function:

### CCDF:

- $$P_{\geq}(x) = P(x' \geq x) = 1 - P(x' < x)$$
- $$= \int_{x'=x}^{\infty} P(x')\mathrm{d}x'$$
- $$\propto \int_{x'=x}^{\infty} (x')^{-\gamma}\mathrm{d}x'$$
- $$= \frac{1}{-\gamma+1}(x')^{-\gamma+1}\Big|_{x'=x}^{\infty}$$
- $$\propto x^{-\gamma+1}$$

$x^{-\gamma}$



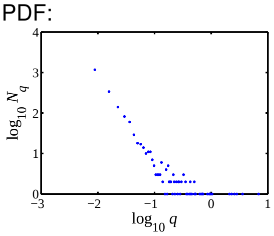


## Complementary Cumulative Distribution Function:

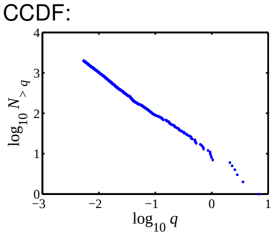
### CCDF:

- $$P_{\geq}(x) \propto x^{-\gamma+1}$$
- Use when tail of $P$ follows a power law.
- Increases exponent by one.
- Useful in cleaning up data.

PDF:




CCDF:




$x^{-\gamma}$





## Complementary Cumulative Distribution Function:

- Discrete variables:

$$P_{\geq}(k) = P(k' \geq k)$$

$$= \sum_{k'=k}^{\infty} P(k)$$

$$\propto k^{-\gamma+1}$$

- Use integrals to approximate sums.

$x^{-\gamma}$





## Zipfian rank-frequency plots

### George Kingsley Zipf:

- Noted various rank distributions followed power laws, often with exponent -1 (word frequency, city sizes...)
- Zipf's 1949 Magnum Opus (⊞): "Human Behaviour and the Principle of Least-Effort" [12]
- We'll study Zipf's law in depth...

$x^{-\gamma}$





## Zipfian rank-frequency plots

### Zipf's way:

- Given a collection of entities, rank them by size, largest to smallest.
- $x_r$ = the size of the $r$th ranked entity.
- $r = 1$ corresponds to the largest size.
- Example: $x_1$ could be the frequency of occurrence of the most common word in a text.
- Zipf's observation:

$$x_r \propto r^{-\alpha}$$

$x^{-\gamma}$

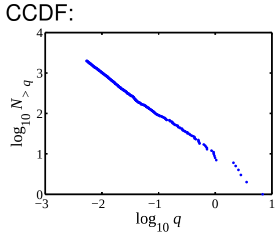




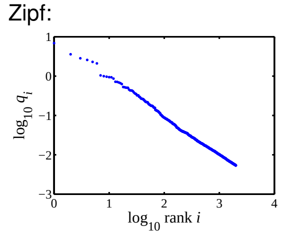
## Size distributions

### Brown Corpus (1,015,945 words):

CCDF:




Zipf:



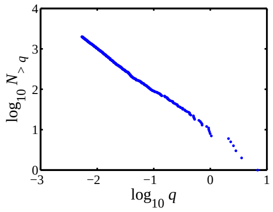

- The, of, and, to, a, ... = 'objects'
- 'Size' = word frequency
- Beep: CCDF and Zipf plots are related...

$x^{-\gamma}$

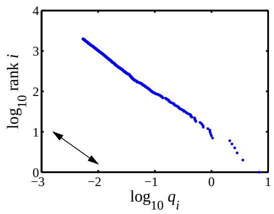
## Size distributions

### Brown Corpus (1,015,945 words):

CCDF:




Zipf (axes flipped):




- The, of, and, to, a, ... = 'objects'
- 'Size' = word frequency
- Beep: CCDF and Zipf plots are related...



### Observe:

- $NP_{\geq}(x)$ = the number of objects with size at least $x$ where $N$ = total number of objects.
- If an object has size $x_r$, then $NP_{\geq}(x_r)$ is its rank $r$.
- So

$$x_r \propto r^{-\alpha} = (NP_{\geq}(x_r))^{-\alpha}$$

$$\propto x_r^{(-\gamma+1)(-\alpha)} \text{ since } P_{\geq}(x) \sim x^{-\gamma+1}.$$

We therefore have $1 = (-\gamma + 1)(-\alpha)$ or:

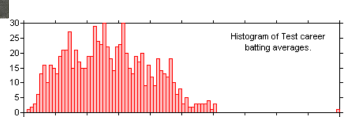
$$\boxed{\alpha = \frac{1}{\gamma - 1}}$$

- A rank distribution exponent of $\alpha = 1$ corresponds to a size distribution exponent $\gamma = 2$.



## The Don. (⊞)

### Extreme deviations in test cricket:




- Don Bradman's batting average (⊞) = 166% next best.



## References I

[1] P. Bak, K. Christensen, L. Danon, and T. Scanlon.
Unified scaling law for earthquakes.
Phys. Rev. Lett., 88:178501, 2002. pdf (⊞)

[2] A.-L. Barabási and R. Albert.
Emergence of scaling in random networks.
Science, 286:509–511, 1999. pdf (⊞)

[3] K. Christensen, L. Danon, T. Scanlon, and P. Bak.
Unified scaling law for earthquakes.
Proc. Natl. Acad. Sci., 99:2509–2513, 2002. pdf (⊞)

[4] P. Grassberger.
Critical behaviour of the Drossel-Schwabl forest fire model.
New Journal of Physics, 4:17.1–17.15, 2002. pdf (⊞)



## References II

[5] J. Holtsmark.
?
Ann. Phys., 58:577–, 1917.

[6] M. E. J. Newman.
Power laws, pareto distributions and zipf's law.
Contemporary Physics, pages 323–351, 2005. pdf (⊞)

[7] D. J. d. S. Price.
Networks of scientific papers.
Science, 149:510–515, 1965. pdf (⊞)

[8] D. J. d. S. Price.
A general theory of bibliometric and other cumulative advantage processes.
J. Amer. Soc. Inform. Sci., 27:292–306, 1976.



## References III

[9] L. F. Richardson.
Variation of the frequency of fatal quarrels with magnitude.
J. Amer. Stat. Assoc., 43:523–546, 1949. pdf (⊞)

[10] H. A. Simon.
On a class of skew distribution functions.
Biometrika, 42:425–440, 1955. pdf (⊞)

[11] N. N. Taleb.
The Black Swan.
Random House, New York, 2007.

[12] G. K. Zipf.
Human Behaviour and the Principle of Least-Effort.
Addison-Wesley, Cambridge, MA, 1949.