Power-Law Size Distributions
Principles of Complex Systems
CSYS/MATH 300, Spring, 2013

Prof. Peter Dodds
@peterdodds

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

Licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License.
Let’s test our collective intuition:

Money ≡ Belief

Two questions about wealth distribution in the United States:

1. Please estimate the percentage of all wealth owned by individuals when grouped into quintiles.
2. Please estimate what you believe each quintile should own, ideally.
3. Extremes: 100, 0, 0, 0, 0 and 20, 20, 20, 20, 20
Wealth distribution in the United States: [8]

Fig. 2. The actual United States wealth distribution plotted against the estimated and ideal distributions across all respondents. Because of their small percentage share of total wealth, both the “4th 20%” value (0.2%) and the “Bottom 20%” value (0.1%) are not visible in the “Actual” distribution.

“Building a better America—One wealth quintile at a time” Norton and Ariely, 2011. [8]
Wealth distribution in the United States: [8]

Fig. 3. The actual United States wealth distribution plotted against the estimated and ideal distributions of respondents of different income levels, political affiliations, and genders. Because of their small percentage share of total wealth, both the “4th 20%” value (0.2%) and the “Bottom 20%” value (0.1%) are not visible in the “Actual” distribution.
Your turn—estimates:
Power-Law Size Distributions

Your turn—ideal:

0 20 40 60 80 100

Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf’s law
Zipf ⇔ CCDF
References
The sizes of many systems’ elements appear to obey an inverse power-law size distribution:

\[ P(\text{size} = x) \sim c x^{-\gamma} \]

where \( 0 < x_{\min} < x < x_{\max} \) and \( \gamma > 1 \).

- 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.

- power-law decays in probability: The Statistics of Surprise.
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 size 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.

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

where $k_{\text{min}} \leq k \leq k_{\text{max}}$

- Obvious fail for $k = 0$.
- Again, typically a description of distribution’s tail.
The statistics of surprise—words:

Brown Corpus ( دقيقة ) (∼ 10^6 words):

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

<table>
<thead>
<tr>
<th>rank</th>
<th>word</th>
<th>% q</th>
</tr>
</thead>
<tbody>
<tr>
<td>1945.</td>
<td>apply</td>
<td>0.0055</td>
</tr>
<tr>
<td>1946.</td>
<td>vital</td>
<td>0.0055</td>
</tr>
<tr>
<td>1947.</td>
<td>September</td>
<td>0.0055</td>
</tr>
<tr>
<td>1948.</td>
<td>review</td>
<td>0.0055</td>
</tr>
<tr>
<td>1949.</td>
<td>wage</td>
<td>0.0055</td>
</tr>
<tr>
<td>1950.</td>
<td>motor</td>
<td>0.0055</td>
</tr>
<tr>
<td>1951.</td>
<td>fifteen</td>
<td>0.0055</td>
</tr>
<tr>
<td>1952.</td>
<td>regarded</td>
<td>0.0055</td>
</tr>
<tr>
<td>1953.</td>
<td>draw</td>
<td>0.0055</td>
</tr>
<tr>
<td>1954.</td>
<td>wheel</td>
<td>0.0055</td>
</tr>
<tr>
<td>1955.</td>
<td>organized</td>
<td>0.0055</td>
</tr>
<tr>
<td>1956.</td>
<td>vision</td>
<td>0.0055</td>
</tr>
<tr>
<td>1957.</td>
<td>wild</td>
<td>0.0055</td>
</tr>
<tr>
<td>1958.</td>
<td>Palmer</td>
<td>0.0055</td>
</tr>
<tr>
<td>1959.</td>
<td>intensity</td>
<td>0.0055</td>
</tr>
</tbody>
</table>

\[ P(x) \sim x^{-\gamma} \]
Power-Law Size Distributions

Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf ⇔ CCDF
References

A word frequency distribution explorer:

Jonathan Harris's Wordcount: (Submit)

\[ P(x) \sim x^{-\gamma} \]
The statistics of surprise—words:

First—a Gaussian example:

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

linear:

\[ P(x) \]

log-log:

\[ \log_{10} P(x) \]

mean \( \mu = 10 \), variance \( \sigma^2 = 1 \).
The statistics of surprise—words:

Raw ‘probability’ (binned) for Brown Corpus:

linear:

\[ N_q \]

log-log:

\[ \log_{10} N_q \]

\[ \log_{10} q \]

\[ P(x) \sim x^{-\delta} \]
The statistics of surprise—words:

‘Exceedance probability’ $N_{>q}$:

linear:

log-log

$P(x) \sim x^{-\gamma}$
Test your vocab

How many words do you know?

- Test capitalizes on word frequency following a heavily skewed frequency distribution with a decaying power-law tail.
- Let’s do it collectively... (押金)
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” [4, 2]
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, . . . ’
Well, that’s just great:

Two things we have poor cognitive understanding of:

1. Probability
   - Ex. The Monty Hall Problem (_enqueue)
   - Ex. Daughter/Son born on Tuesday (_enqueue) (see asides; Wikipedia entry Boy or Girl Paradox (_enqueue)here).

2. Logarithmic scales.

On counting and logarithms:

- Listen to Radiolab’s “Numbers.” (enqueue).
- Later: Benford’s Law (enqueue).
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 are given in the text. (a) Numbers of occurrences of words in the novel Moby Dick by Herman Melville. (b) Numbers of citations to scientific papers published in 1981. (c) Numbers of web 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 the earthquake, and hence the distribution obeys a power law even though the horizontal axis is linear. (g) Diameter of craters on the moon. (h) Magnitude of solar flares. (i) Peak gamma-ray intensity of solar flares. (j) Intensity of wars from 1816 to 1980. (k) Numbers of hits on web sites by 60,000 users of the America Online Internet service for the day of 1 December 1997. (l) Populations of US cities in the year 2000. (b) (c) (d) (e) (f) (h) (i) (j) (k) (l). Power Laws as Distribution Functions

Definition

Examples

Wild vs. Mild

CCDFs

Zipf’s law

Zipf ↔ CCDF

References

Power-Law Size Distributions

Our Intuition

Definitions

Examples

Wild vs. Mild

CCDFs

Zipf’s law

Zipf ↔ CCDF

References
Size distributions:

Examples:

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


Note: Exponents range in error
Size distributions:

Examples:

- Number of citations to papers: $[9, 10] P(k) \propto k^{-3}$.
- 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: $[1] P(F) \propto F^{-5/2}$. (see the Holtsmark distribution and stable distributions)
- Diameter of moon craters: $[7] P(d) \propto d^{-3}$.
- Word frequency: $[12]$ e.g., $P(k) \propto k^{-2.2}$ (variable)
Gaussians versus power-law distributions:

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

See “The Black Swan” by Nassim Taleb. [13]
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”\cite{13}
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
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 size distribution details:

Exhibit A:

- Given \( P(x) = cx^{-\gamma} \) with \( 0 < x_{\text{min}} < x < x_{\text{max}} \), the mean is \( (\gamma \neq 2) \):
  \[
  \langle x \rangle = \frac{c}{2 - \gamma} \left( x_{\text{max}}^{2-\gamma} - x_{\text{min}}^{2-\gamma} \right).
  \]

- 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 (موت)
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)
- $\sigma^2 = \text{variance is ‘infinite’} \text{ (depends on upper cutoff)}$
- 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:

  Mean average deviation (MAD) = \( \langle |x - \langle x \rangle| \rangle \)

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

  \( \langle |x - \langle x \rangle| \rangle \) is finite.

- But MAD is mildly unpleasant analytically...
- We still speak of infinite ‘width’ if \( \gamma < 3 \).

Insert question from assignment 2 ( miệng)
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
  \[ x_1 \gtrsim \frac{1}{\lambda} \ln n. \]
Complementary Cumulative Distribution Function:

CCDF:

\[ P_{\geq}(x) = P(x' \geq x) = 1 - P(x' < x) \]

\[ = \int_{x' = x}^{\infty} P(x')dx' \]

\[ \propto \int_{x' = x}^{\infty} (x')^{-\gamma}dx' \]

\[ = \left. \frac{1}{-\gamma + 1} (x')^{-\gamma + 1} \right|_{x' = x}^{\infty} \]

\[ \propto x^{-\gamma + 1} \]
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:**

$$P(x) \sim x^{-\gamma}$$
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.
Zipfian rank-frequency plots

George Kingsley Zipf:

- Noted various rank distributions have power-law tails, often with exponent -1 (word frequency, city sizes...)
- Zipf’s 1949 **Magnum Opus** (นโยบาย): We’ll study Zipf’s law in depth...
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:

\[
x_r \propto r^{-\alpha}
\]
Size distributions:

Brown Corpus (1,015,945 words):

CCDF:

Zipf:

- The, of, and, to, a, ... = ‘objects’
- ‘Size’ = word frequency
- Beep: (Important) CCDF and Zipf plots are related...

\[ P(x) \sim x^{-\gamma} \]
Size distributions:

Brown Corpus (1,015,945 words):

CCDF:

\[
\log_{10} N > q
\]

Zipf:

\[
\log_{10} \text{rank}_i
\]

- The, of, and, to, a, ... = 'objects'
- ‘Size’ = word frequency
- **Beep:** (Important) 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

\[
x_r \propto r^{-\alpha} = \left( NP_\geq(x_r) \right)^{-\alpha}
\]

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

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

\[
\alpha = \frac{1}{\gamma - 1}
\]

- A rank distribution exponent of \( \alpha = 1 \) corresponds to a size distribution exponent \( \gamma = 2 \).
Extreme deviations in test cricket:

- Don Bradman’s batting average = 166% next best.
- That’s pretty solid.
- Later in the course: Understanding success—is the Mona Lisa like Don Bradman?
References I

[1]  


References II


References III


References IV

The Black Swan.  

Human Behaviour and the Principle of Least-Effort.  
Addison-Wesley, Cambridge, MA, 1949.