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

Licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License.

Outline

Definition
Examples
Wild vs. Mild
CCDFs
Zipf’s law
Zipf ⇔ CCDF
References

Size distributions

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

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

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

▶ Exciting class exercise: sketch this function.
▶ \( x_{\text{min}} \) = lower cutoff
▶ \( x_{\text{max}} \) = 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 (⊞), . . .

The statistics of surprise—words:

Brown Corpus (⊞) (∼ 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>

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.

Power Law Size Distributions
The statistics of surprise—words:

First—a Gaussian example:

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

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

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... (III)

The statistics of surprise—words:

Raw ‘probability’ (binned):

Gutenberg-Richter law (III)

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

The statistics of surprise—words:

‘Exceedance probability’:

From: “Quake Moves Japan Closer to U.S. and Alters Earth’s Spin” (III) 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 (iii).
   - Ex. Son born on Tuesday (iii).
2. Logarithmic scales.

On counting and logarithms:

- Listen to Radiolab’s “Numbers,” (iii).
- Later: Benford’s Law (iii).

Size distributions

Examples:
- Earthquake magnitude (Gutenberg-Richter law (iii)): $P(M) \propto M^{-\alpha}$
- Number of war deaths: $P(d) \propto d^{-1.8}$
- Sizes of forest fires: $P(n) \propto n^{-2.1}$
- Sizes of cities: $P(n) \propto n^{-2.1}$
- Number of links to and from websites
- Note: Exponents range in error

Power law distributions

Gaussians versus power-law distributions:
- Mediocrist versus Extremist
- Mild versus Wild (Mandelbrot)
- Example: Height versus wealth.

Turkeys...

Power Law Size Distributions

Definition
Examples
Wild vs. Mild
CCDFs
Zipf’s law
Zipf ⇔ CCDF
References

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.

Size distributions

Power law size distributions are sometimes called Pareto distributions \( \sim cx^{-\gamma} \) after Italian scholar Vilfredo Pareto.\(^{[11]} \)
- Pareto noted wealth in Italy was distributed unevenly (80–20 rule; misleading).
- Term used especially by practitioners of the Dismal Science \( \sim cx^{-\gamma} \).

Devilish power law distribution details:

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

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 $2 < \gamma < 3$:
  \[
  \langle |x - \langle x \rangle| \rangle
  \]
- But MAD is mildly unpleasant analytically...
- We still speak of infinite ‘width’ if $\gamma < 3$.

How sample sizes grow...

Given \( P(x) \sim cx^{-\gamma} \):
- We can show that after $n$ samples, we expect the largest sample to be
  \[
  x_1 \gtrsim c^n n^{1/(\gamma-1)}
  \]
- 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 \( \sim \)
Complementary Cumulative Distribution Function:

CCDF:

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

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

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

\[ = \frac{1}{-\gamma + 1} \int_{x'}^{\infty} (x')^{-\gamma+1} \, dx' \]

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

Zipfian rank-frequency plots

George Kingsley Zipf:

- Noted various rank distributions followed power laws, often with exponent -1 (word frequency, city sizes...)
- We'll study Zipf's law in depth...

Zipf's way:

- Given a collection of entities, rank them by size, largest to smallest.
- \( x_r \) is 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:

\[ \log_{10} q \]

Zipf:

\[ \log_{10} \] rank

The, of, and, to, a, ... = ‘objects’

‘Size’ = word frequency

Beep: CCDF and Zipf plots are related...
Size distributions

Brown Corpus (1,015,945 words):

CCDF: $N_i > q$

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 
  \[
  x_r \propto r^{-\alpha} = (NP_\geq(x_r))^{-\alpha}
  \]
  \[
  \propto x_r^{(-\gamma+1)(-\alpha)} \quad \text{since} \quad P_\geq(x) \propto x^{-\gamma+1}.
  \]
  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 \( \gamma = 2 \).

The Don. (III)

Extreme deviations in test cricket:

- Don Bradman’s batting average (III) = 166% next best.

References I


References II


References III


