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

References

1 of 36
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

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

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

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

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 (⊩), . . .
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

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

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 (\(\square\)), Weibull distributions (\(\square\)), …
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.
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 (福田) (~ 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>
The statistics of surprise—words:

First—a Gaussian example:

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

linear:

\[ x \quad P(x) \]

\[ 0 \quad 0.1 \quad 0.2 \quad 0.3 \quad 0.4 \]

\[ 0 \quad 5 \quad 10 \quad 15 \quad 20 \]

log-log:

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

\[ -3 \quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \]

\[ -25 \quad -20 \quad -15 \quad -10 \quad -5 \quad 0 \]

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

Raw ‘probability’ (binned):

linear:

\[
\begin{align*}
N_q & \quad 0 \quad 200 \quad 400 \quad 600 \quad 800 \\
q & \quad 0 \quad 2 \quad 4 \quad 6 \quad 8
\end{align*}
\]
The statistics of surprise—words:

Raw ‘probability’ (binned):

**linear:**

**log-log**

- **Zipf’s law**
  - Zipf $\Leftrightarrow$ CCDF
The statistics of surprise—words:

‘Exceedance probability’:

linear:

\[
N > q
\]

log-log:

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

Zipf’s law

\[
\text{Zipf} \Leftrightarrow \text{CCDF}
\]
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... (⊞)

My, what big words you have...
Test your vocab

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

How many words do you know?
The statistics of surprise:

Gutenberg-Richter law (𝐺)  

- 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, . . .
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.
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 copies of bestselling books sold in the US between January 1910 and May 1965. (b) Numbers of citations to scientific papers published in 1981, from time of publication until June 1997. (c) Numbers of web hits on websites by 60,000 users of the America Online Internet service for the day of 1 December 1997. (d) Numbers of calls received by AT&T telephone customers in the US for a single day. (e) Magnitude of earthquakes in California between January 1910 and May 1992. (f) Number of days of call received by AT&T. (g) Peak gamma-ray intensity of solar flares in counts per second, measured from Earth orbit between February 1896 and November 1980. (h) Maximum amplitude of solar flares in counts per square kilometre. (i) Magnitude of earthquakes on the moon. (j) Intensity of wars from 1816 to 1980, measured as battle deaths per 100,000 of the population of the participating countries. (k) Aggregate net worth in dollars per 100,000 of the richest individuals in the US in October 2003. (l) Populations of US cities in the year 2000.

Power Law Size Distributions

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:

- Earthquake magnitude (Gutenberg-Richter law): \[ P(M) \propto M^{-2} \]
- 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

References:

1. [1]
2. [2]
3. [3]
4. [4]
5. [5]
6. [6]
7. [7]
8. [8]
9. [9]
10. [10]
Size distributions

Examples:
- Number of citations to papers: $P(k) \propto k^{-3}$.
- 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: $P(F) \propto F^{-5/2}$.
- Diameter of moon craters: $P(d) \propto d^{-3}$.
- Word frequency: e.g., $P(k) \propto k^{-2.2}$ (variable).
Power law distributions

Gaussians versus power-law distributions:

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

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]
# 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
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mediocristan/Extremistan</strong></td>
<td></td>
</tr>
<tr>
<td>▶ Most typical member is <em>mediocre</em>/Most typical is either giant or tiny</td>
<td></td>
</tr>
<tr>
<td>▶ Winners get a small segment/Winner take almost all effects</td>
<td></td>
</tr>
<tr>
<td>▶ When you observe for a while, you know what’s going on/It takes a <em>very long time</em> to figure out what’s going on</td>
<td></td>
</tr>
<tr>
<td>▶ Prediction is <em>easy</em>/Prediction is <em>hard</em></td>
<td></td>
</tr>
<tr>
<td>▶ History crawls/History makes jumps</td>
<td></td>
</tr>
<tr>
<td>▶ Tyranny of the collective/Tyranny of the rare and accidental</td>
<td></td>
</tr>
</tbody>
</table>
Taleb’s table

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
### Taleb’s table \[^{[11]}\]

<table>
<thead>
<tr>
<th>Mediocristan/Extremistan</th>
</tr>
</thead>
<tbody>
<tr>
<td>▶ Most typical member is mediocre/Most typical is either giant or tiny</td>
</tr>
<tr>
<td>▶ Winners get a small segment/Winner take almost all effects</td>
</tr>
<tr>
<td>▶ 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</td>
</tr>
<tr>
<td>▶ Prediction is easy/Prediction is hard</td>
</tr>
<tr>
<td>▶ History crawls/History makes jumps</td>
</tr>
<tr>
<td>▶ Tyranny of the collective/Tyranny of the rare and accidental</td>
</tr>
</tbody>
</table>
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
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
### Taleb’s table

<table>
<thead>
<tr>
<th><strong>Mediocristan/Extremistan</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>► Most typical member is <em>mediocre</em> / Most typical is either giant or tiny</td>
</tr>
<tr>
<td>► Winners get a small segment / Winner take almost all effects</td>
</tr>
<tr>
<td>► When you observe for a while, you know what’s going on / It takes a <strong>very long time</strong> to figure out what’s going on</td>
</tr>
<tr>
<td>► Prediction is <strong>easy</strong> / Prediction is <strong>hard</strong></td>
</tr>
<tr>
<td>► History crawls / History makes jumps</td>
</tr>
<tr>
<td>► Tyranny of the collective / Tyranny of the rare and accidental</td>
</tr>
</tbody>
</table>
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.
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.
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 \( 0 < x_{\text{min}} < x < x_{\text{max}} \), 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 (导入)
Devilish power law 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 (답)
Devilish power law 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 (⎞)
Devilish power law 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 (迫切)
Devilish power law 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 (□)
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 (□)
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 (⊞)
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 ((topic)
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 (>{$\blacksquare$})
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
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 (田)
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 (●)
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 (△)
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 (□)
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 (■)
Moements

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 (☐)
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 (답)
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 (□)
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 (◻️)
How sample sizes grow...

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

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

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

- Sampling from a finite-variance distribution gives a much slower growth with $n$.

- e.g., for $P(x) = \lambda e^{-\lambda x}$, we find

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

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

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

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

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

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

\[
\propto x^{-\gamma+1}
\]
Complementary Cumulative Distribution Function:

**CCDF:**

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

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

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

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

\[ \propto x^{-\gamma + 1} \]
Complementary Cumulative Distribution Function:

CCDF:

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

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

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

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

\[ \propto x^{-\gamma + 1} \]
Complementary Cumulative Distribution Function:

**CCDF:**

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

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

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

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

\[ \propto x^{-\gamma+1} \]
Complementary Cumulative Distribution Function:

**CCDF:**

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

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

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

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

\[ \propto x^{-\gamma + 1} \]
Complementary Cumulative Distribution Function:

CCDF:

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

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

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

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

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

\[ \log_{10} N_q \]

\[ \log_{10} q \]

**CCDF:**

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

\[ \log_{10} q \]
Complementary Cumulative Distribution Function:

- **Discrete variables:**
  
  \[ P_{\geq}(k) = P(k' \geq k) \]

- Use integrals to approximate sums.
Complementary Cumulative Distribution Function:

- Discrete variables:

\[ P_{\geq}(k) = P(k' \geq k) \]

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

- Use integrals to approximate sums.
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.
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.
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...
George Kingsley Zipf:

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

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

  $$x_r \propto r^{-\alpha}$$
Zipfian rank-frequency plots

Zipf’s way:

- Given a collection of entities, rank them by size, largest to smallest.
- $x_r = \text{the size of the } r\text{th ranked entity.}$
- $r = 1$ 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:

$$x_r \propto r^{-\alpha}$$
Zipfian rank-frequency plots

**Zipf’s way:**

- Given a collection of entities, rank them by size, largest to smallest.
- $x_r = \text{the size of the } r\text{th ranked entity}.$
- $r = 1$ 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} N > q
```

Zipf:

```
log_{10} q_i
```

- The, of, and, to, a, ... = ‘objects’
- ‘Size’ = word frequency
Size distributions

Brown Corpus (1,015,945 words):

**CCDF:**

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

**Zipf:**

\[
\log_{10} q_i \\
\log_{10} \text{rank } i
\]

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

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

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

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

\[ \propto x_r^{(-\gamma+1)(-\alpha)} \] 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 $\gamma = 2$. 
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:

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

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

\[ x \sim x^{-\gamma} \]
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

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

\[ \propto x_r^{(-\gamma+1)(-\alpha)} \] 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 $\gamma = 2$. 

\[ x^{-\gamma} \]
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

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

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

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 \( \gamma = 2 \).
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
  \[
  x_r \propto r^{-\alpha} = (NP_\geq(x_r))^{-\alpha}
  \]

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

  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.
Extreme deviations in test cricket:

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


References II


Power laws, pareto distributions and zipf’s law.  
pdf (↗)

Networks of scientific papers.  
pdf (↗)

A general theory of bibliometric and other cumulative advantage processes.  
References III


