Power Law Size Distributions
Principles of Complex Systems
CSYS/MATH 300, Fall, 2010

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

Overview
- Introduction
- Examples
- Wild vs. Mild
- CCDFs
- Zipf’s law
- Zipf ⇔ CCDF

References
Add section on stable distributions
Add an assignment question or two
convolve distributions
Cauchy
Inverse Gaussian

Outline

Overview

Introduction
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf ⇔ CCDF

References
Size distributions—Assignment 1 recap:

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

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

where \( x_{\min} < x < x_{\max} \)

and \( \gamma > 1 \)

- \( x_{\min} \) = lower cutoff
- \( x_{\max} \) = upper cutoff

Negative linear relationship in log-log space:

\[ \log P(x) = \log c - \gamma \log x \]
Size distributions—Assignment 1 recap:

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

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

where \( x_{\text{min}} < x < x_{\text{max}} \)

and \( \gamma > 1 \)

- \( x_{\text{min}} = \) lower cutoff
- \( x_{\text{max}} = \) upper cutoff

Negative linear relationship in log-log space:

\[ \log P(x) = \log c - \gamma \log x \]
Size distributions—Assignment 1 recap:

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

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

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

- \( x_{\text{min}} = \) lower cutoff
- \( x_{\text{max}} = \) upper cutoff

Negative linear relationship in log-log space:

\[ \log P(x) = \log c - \gamma \log x \]
Size distributions—Assignment 1 recap:

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

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

where \( x_{\text{min}} < x < x_{\text{max}} \)

and \( \gamma > 1 \)

- \( x_{\text{min}} \) = lower cutoff
- \( x_{\text{max}} \) = upper cutoff

- Negative linear relationship in log-log space:

\[ \log P(x) = \log c - \gamma \log x \]
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’
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’
Size distributions

Many systems have discrete sizes $k$:

- Word frequency
- Node degree (as we have seen): # hyperlinks, etc.
- number of citations for articles, court decisions, etc.

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

where $k_{\text{min}} \leq k \leq k_{\text{max}}$
Power law size distributions are sometimes called Pareto distributions (مادة) after Italian scholar Vilfredo Pareto.

- Pareto noted wealth in Italy was distributed unevenly (80–20 rule).
- Term used especially by economists.
Devilish power law distribution details:

From assignment 1, we know many nasty things.

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

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

- 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.
Devilish power law distribution details:

From assignment 1, we know many nasty things.

Exhibit A:

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

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

- 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.
Devilish power law distribution details:

From assignment 1, we know many nasty things.

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

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

- 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.
Devilish power law distribution details:

From assignment 1, we know many nasty things.

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

- 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.
Devilish power law distribution details:

From assignment 1, we know many nasty things.

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

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

- 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.
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.
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 ‘infinitesimal’ (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.
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.
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.
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)
- $\sigma^2 = \text{variance is ‘infinite’ (depends on upper cutoff)}$
- Width of distribution is ‘infinite’
- If $\gamma > 3$, distribution is less terrifying and may be easily confused with other kinds of distributions.
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.
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.
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.
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 unpleasant analytically...
- We still speak of infinite ‘width’ if \(\gamma < 3\).
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 unpleasant analytically...
- We still speak of infinite ‘width’ if \( \gamma < 3 \).
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 unpleasant analytically...
- We still speak of infinite ‘width’ if \(\gamma < 3\).
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 unpleasant analytically...
- We still speak of infinite ‘width’ if \(\gamma < 3\).
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 unpleasant analytically...

  We still speak of infinite ‘width’ if \( \gamma < 3 \).
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 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

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

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

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

  $$x_1 \gtrsim \frac{1}{\lambda} \ln n.$$
How sample sizes grow...

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

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

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

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

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

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

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

Overview
  Introduction
  Examples
  Wild vs. Mild
  CCDFs
  Zipf’s law
  Zipf $\Leftrightarrow$ CCDF

References
Size distributions

Examples:

- Earthquake magnitude (Gutenberg Richter law):
  \[ P(M) \propto M^{-3} \]
- 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
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)

Note: Exponents range in error; see M.E.J. Newman
arxiv.org/cond-mat/0412004v3
Size distributions

Power-law distributions are..

▷ often called ‘heavy-tailed’
▷ or said to have ‘fat tails’

Important!:

▷ Inverse power laws aren’t the only ones:
  ▷ lognormals, stretched exponentials, ...
Size distributions

Power-law distributions are..

▶ often called ‘heavy-tailed’
▶ or said to have ‘fat tails’

Important!:

▶ Inverse power laws aren’t the only ones:
  ▶ lognormals, stretched exponentials, ...
Size distributions

Power-law distributions are..

▶ often called ‘heavy-tailed’
▶ or said to have ‘fat tails’

Important!:

▶ Inverse power laws aren’t the only ones:
  ▶ lognormals, stretched exponentials, …
Size distributions

Power-law distributions are..

- often called ‘heavy-tailed’
- or said to have ‘fat tails’

Important!:

- Inverse power laws aren’t the only ones:
  - lognormals, stretched exponentials, ...
Power law distributions

Gaussians versus power-law distributions:

- Example: Height versus wealth.
- Mild versus Wild (Mandelbrot)
- Mediocristan versus Extremistan
  (See “The Black Swan” by Nassim Taleb[1])
Gaussians versus power-law distributions:

- Example: Height versus wealth.
- Mild versus Wild (Mandelbrot)
- Mediocristan versus Extremistan
  (See “The Black Swan” by Nassim Taleb \cite{1})
Power law distributions

Gaussians versus power-law distributions:

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

(See "The Black Swan” by Nassim Taleb[1])
Gaussians versus power-law distributions:

- Example: Height versus wealth.
- Mild versus Wild (Mandelbrot)
- Mediocristan versus Extremistan
  (See “The Black Swan” by Nassim Taleb[1])
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”[1]
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

References

[1] 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 [1]

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[1]

<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>
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
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
Complementary Cumulative Distribution Function:

**CCDF:**

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

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

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

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

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

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

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

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

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

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

**CCDF:**

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

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

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

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

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

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

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

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

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

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

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

\[ P_>(x) \propto x^{-\gamma+1} \]

- Use when tail of \( P \) follows a power law.
- Increases exponent by one.
- Useful in cleaning up data.
Complementary Cumulative Distribution Function:

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.
Outline

Overview
Introduction
Examples
Wild vs. Mild
CCDFs
Zipf’s law
Zipf ⇔ CCDF

References
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...
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...
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...
Zipfian rank-frequency plots

Zipf’s way:

- $s_r =$ the size of the $r$th ranked object.
- $r = 1$ corresponds to the largest size.
- Example: $s_1$ could be the frequency of occurrence of the most common word in a text.

Zipf’s observation:

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

Zipf’s way:

- $s_r$ = the size of the $r$th ranked object.
- $r = 1$ corresponds to the largest size.
- Example: $s_1$ could be the frequency of occurrence of the most common word in a text.
- Zipf’s observation:

$$s_r \propto r^{-\alpha}$$
Outline

Overview
  Introduction
  Examples
  Wild vs. Mild
  CCDFs
  Zipf’s law
  Zipf ⇔ CCDF

References
Size distributions

Brown Corpus (1,015,945 words):

CCDF:

```
log_{10} N > n
```

Zipf:

```
log_{10} n_i
```

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

---

Overview
Introduction
Examples
Wild vs. Mild
CCDFs
Zipf’s law
Zipf ⇔ CCDF
References
Size distributions

Brown Corpus (1,015,945 words):

CCDF:

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

\[ \log_{10} n \]

Zipf:

\[ \log_{10} n_i \]

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

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

Observe:

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

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

\[
\propto x_r^{(-\gamma+1)(-\alpha)}
\]

Since \( P(x) \sim x^{-\gamma+1} \),

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

- A rank distribution exponent of \( \alpha = 1 \) corresponds to a size distribution exponent \( \gamma = 2 \).
Size distributions

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 $P_\geq(x) \sim x^{-\gamma+1}$,

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

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

Size distributions

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

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

Since \( P_\geq(x) \sim x^{-\gamma+1} \),

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

► A rank distribution exponent of \( \alpha = 1 \) corresponds to a size distribution exponent \( \gamma = 2 \).
Size distributions

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 \( P_{\geq}(x) \sim x^{-\gamma+1} \),

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

- A rank distribution exponent of \( \alpha = 1 \) corresponds to a size distribution exponent \( \gamma = 2 \).
Size distributions

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 \( P_{\geq}(x) \sim x^{-\gamma+1} \),

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

- A rank distribution exponent of \( \alpha = 1 \) corresponds to a size distribution exponent \( \gamma = 2 \).
Size distributions

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 \( P \geq (x) \sim x^{-\gamma+1}, \)

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

- 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.
The Don

Extreme deviations in **test cricket**

Don Bradman’s batting average = 166% next best.
References

The Black Swan.

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