

Power-Law Size Distributions

Last updated: 2025/08/25, 14:34:28 EDT

Principles of Complex Systems, Vols. 1, 2, 3D, 4 forever, V for Vendetta
CSYS/MATH 6701, 6713, & a pretend number, 2025–2026

Prof. Peter Sheridan Dodds

Computational Story Lab | Vermont Complex Systems Center
Santa Fe Institute | University of Vermont

 Licensed under the [Creative Commons Attribution 4.0 International](#)

The PoCSverse
Power-Law Size
Distributions
1 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking ⇔
CCDF
References

Outline

Our Intuition

Definition

Examples

Wild vs. Mild

CCDFs

Size rankings and Zipf's law

Size ranking ⇔ CCDF

References

The PoCSverse
Power-Law Size
Distributions
2 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking ⇔
CCDF
References

Two of the many things we struggle with cognitively:

1. Probability.
 Ex. [The Monty Hall Problem.](#)
 Ex. [Daughter/Son born on Tuesday.](#)
(see next two slides; [Wikipedia entry here](#).)
2. Logarithmic scales.

On counting and logarithms:



-  Listen to Radiolab's 2009 piece: ["Numbers."](#)
-  Later: [Benford's Law](#)

Also to be enjoyed: [The Dunning-Kruger effect](#)¹

¹2000 Ig Nobel winners

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.

The PoCSverse
Power-Law Size
Distributions
3 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking ⇔
CCDF
References

Homo probabilisticus?

The set up:

-  A parent has two children.

Simple probability question:

-  What is the probability that both children are girls?

The next set up:

-  A parent has two children.
-  We know one of them is a girl.

The next probabilistic poser:

-  What is the probability that both children are girls?

The PoCSverse
Power-Law Size
Distributions
4 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking ⇔
CCDF
References

Try this one:

-  A parent has two children.
-  We know one of them is a girl born on a Tuesday.

Simple question #3:

-  What is the probability that both children are girls?

Last:

-  A parent has two children.
-  We know one of them is a girl born on December 31.

And ...

-  What is the probability that both children are girls?

The PoCSverse
Power-Law Size
Distributions
5 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking ⇔
CCDF
References

The PoCSverse
Power-Law Size
Distributions
6 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking ⇔
CCDF
References

Wealth distribution in the United States:^[13]

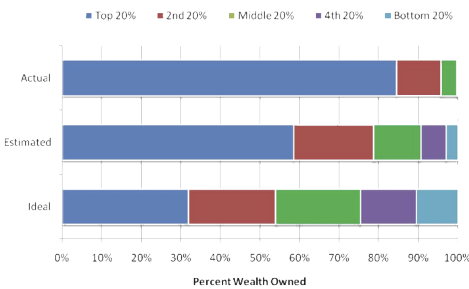


Fig. 2. The actual United States wealth distribution plotted against the estimated and ideal distributions across all respondents. Because of their small percentage share of total wealth, both the "4th 20%" value (0.2%) and the "Bottom 20%" value (0.1%) are not visible in the "Actual" distribution.

"Building a better America—One wealth quintile at a time"

Norton and Ariely, 2011.^[13]

But: [Fraud.](#)

The PoCSverse
Power-Law Size
Distributions
7 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking ⇔
CCDF
References

Wealth distribution in the United States:^[13]

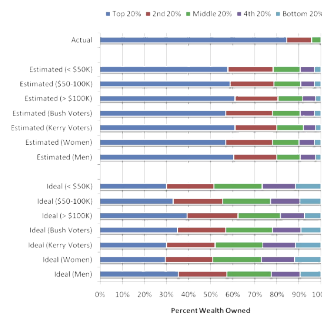


Fig. 3. The actual United States wealth distribution plotted against the estimated and ideal distributions of respondents of different income levels, political affiliations, and genders. Because of their small percentage share of total wealth, both the "4th 20%" value (0.2%) and the "Bottom 20%" value (0.1%) are not visible in the "Actual" distribution.

The PoCSverse
Power-Law Size
Distributions
9 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking ⇔
CCDF
References

The PoCSverse
Power-Law Size
Distributions
10 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking ⇔
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 $\gamma > 1$.

-  $x_{\min}$ = lower cutoff, $x_{\max}$ = upper cutoff
-  Negative linear relationship in log-log space:

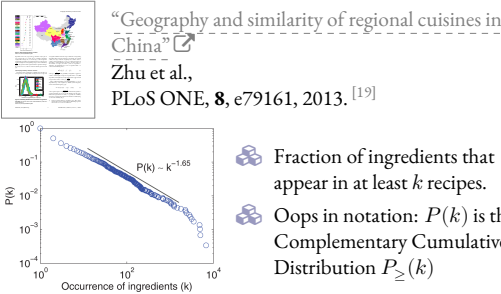
$$\log_{10} P(x) = \log_{10} c - \gamma \log_{10} x$$

-  We use base 10 because we are [good people](#).

My, what big words you have ...



- ☞ Test ☞ capitalizes on word frequency following a heavily skewed frequency distribution with a decaying power-law tail.
- ☞ This Man Can Pronounce Every Word in the Dictionary (story here) ☞
- ☞ Best of Dr. Bailly ☞



- ☞ Fraction of ingredients that appear in at least k recipes.
- ☞ Oops in notation: $P(k)$ is the Complementary Cumulative Distribution $P_{\geq}(k)$

Size distributions:

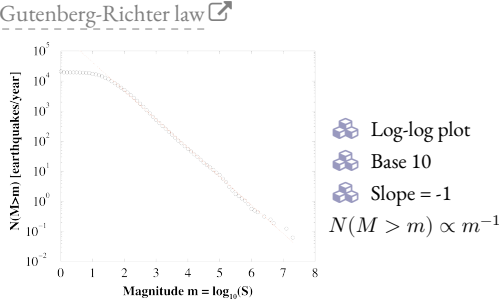
- Some examples:
- ☞ Earthquake magnitude (Gutenberg-Richter law ☞): $P(M) \propto M^{-2}$
- ☞ # war deaths: $P(d) \propto d^{-1.8}$
- ☞ Sizes of forest fires $P(f) \propto f^{-2.1}$
- ☞ Sizes of cities: $P(n) \propto n^{-2.1}$
- ☞ # links to and from websites $P(l) \propto l^{-2.5}$
- ☞ Note: Exponents range in error

The PoCServe
Power-Law Size
Distributions
20 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and Zipf's law
Size ranking ☞ CCDF
References

The PoCServe
Power-Law Size
Distributions
23 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and Zipf's law
Size ranking ☞ CCDF
References

The PoCServe
Power-Law Size
Distributions
26 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and Zipf's law
Size ranking ☞ CCDF
References

The statistics of surprise:



- ☞ From both the very awkwardly similar Christensen et al. and Bak et al.:
“Unified scaling law for earthquakes” $[4, 1]$

“On a class of skew distribution functions”
Herbert A. Simon,
Biometrika, **42**, 425–440, 1955. $[16]$

“Power laws, Pareto distributions and Zipf's law”
M. E. J. Newman,
Contemporary Physics, **46**, 323–351, 2005. $[12]$

“Power-law distributions in empirical data”
Clauset, Shalizi, and Newman,
SIAM Review, **51**, 661–703, 2009. $[5]$

Size distributions:

- More examples:
- ☞ # 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: $P(F) \propto F^{-5/2}$. (See the Holtzmark distribution ☞ and stable distributions ☞.)
- ☞ Diameter of moon craters: $P(d) \propto d^{-3}$.
- ☞ Word frequency: $P(k) \propto k^{-2.2}$ (variable).
- ☞ # religious adherents in cults: $P(k) \propto k^{-1.8 \pm 0.1}$.
- ☞ # sightings of birds per species (North American Breeding Bird Survey for 2003): $P(k) \propto k^{-2.1 \pm 0.1}$.
- ☞ # species per genus: $P(k) \propto k^{-2.4 \pm 0.2}$.

The PoCServe
Power-Law Size
Distributions
21 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and Zipf's law
Size ranking ☞ CCDF
References

The PoCServe
Power-Law Size
Distributions
24 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and Zipf's law
Size ranking ☞ CCDF
References

The PoCServe
Power-Law Size
Distributions
27 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and Zipf's law
Size ranking ☞ CCDF
References

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, ...’

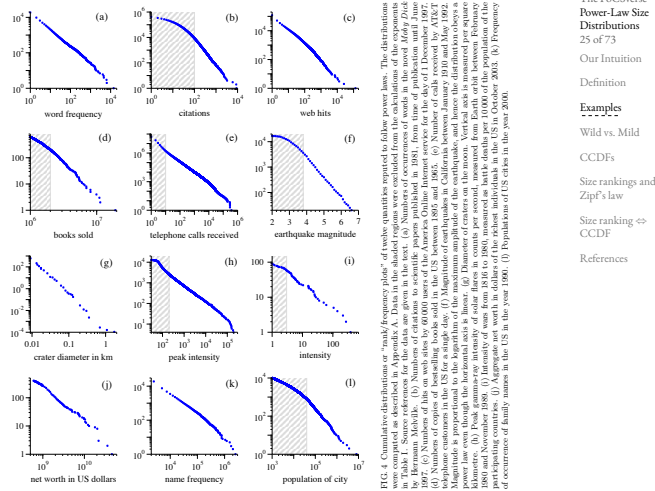


Table 3 from Clauset, Shalizi, and Newman $[5]$:

Basic parameters of the data sets described in section 6, along with their power-law fits and the corresponding p-values (statistically significant values are denoted in bold).									
Quantity	n	$\langle x \rangle$	σ	x_{max}	x_{min}	α	n_{tail}	p	
count of word use	18,855	11.14	1.48 (3)	14,086	7 ± 2	1.95(2)	2958 ± 987	0.49	
protein interaction degree	1846	2.54	3.05	56	5 ± 2	3.1(3)	204 ± 263	0.31	
metabolic degree	1641	5.68	17.81	468	4 ± 1	2.8(1)	748 ± 136	0.00	
Internet degree	22,688	5.63	37.83	2583	21 ± 9	2.12(9)	770 ± 1124	0.29	
telephone calls received	51,360,423	3.88	179.09	375,746	120 ± 49	2.09(1)	102,592 ± 210,147	0.63	
intensity of wars	115	15.70	49.97	382	2.1 ± 3.5	1.7(2)	70 ± 14	0.20	
terrorist attack severity	9101	4.35	31.58	2749	12 ± 4	2.4(2)	547 ± 1663	0.68	
HTTP size (kilobytes)	226,386	7.96	57.94	10,971	36.25 ± 22.74	2.48(5)	6794 ± 2232	0.00	
species per genus	569	5.59	6.94	56	4 ± 2	2.4(2)	233 ± 136	0.10	
bird species sightings	591	3384.36	10952.34	138,705	6679 ± 2463	2.1(2)	66 ± 41	0.55	
blackouts (×10 ³)	211	253.87	610.31	7500	230 ± 90	2.3(3)	59 ± 35	0.62	
sales of books (×10 ³)	633	1966.67	1396.60	10,077	2400 ± 430	3.7(3)	139 ± 115	0.66	
population of cities (×10 ³)	19,447	9.00	77.83	8,009	52.46 ± 11.88	2.37(8)	580 ± 177	0.76	
email address books size	4581	12.45	21.49	333	57 ± 21	3.5(6)	196 ± 449	0.16	
forest fire size (acres)	203,785	0.90	20.99	4121	6324 ± 3487	2.2(3)	521 ± 6801	0.05	
solar flare intensity	12,773	689.41	6520.59	231,300	323 ± 89	1.79(2)	1711 ± 384	1.00	
quake intensity (×10 ³)	19,302	24.54	563.83	63,096	0.794 ± 0.198	1.6(4)	11,697 ± 2159	0.00	
religious followers (×10 ⁶)	103	27.36	136.64	1050	3.85 ± 1.60	1.8(1)	39 ± 26	0.42	
freq. of surnames (×10 ³)	2753	50.59	113.99	2562	111.02 ± 40.67	2.3(2)	239 ± 215	0.20	
net worth (mil. USD)	400	2388.69	4167.35	46,000	900 ± 364	2.3(1)	302 ± 77	0.00	
citations to papers	415,229	16.17	44.02	8904	160 ± 35	3.16(6)	3455 ± 1859	0.20	
papers authored	401,445	7.21	16.52	1416	133 ± 13	4.3(1)	988 ± 377	0.80	
hits to web sites	119,724	9.83	392.52	129,641	2 ± 13	1.81(8)	50,981 ± 16,898	0.00	
links to web sites	241,428,853	9.15	106,871.65	1,199,460	3684 ± 151	2.336(9)	28,986 ± 1560	0.00	

- ☞ We'll explore various exponent measurement techniques in assignments.

power-law size distributions

Gaussians versus power-law size distributions:

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

THE BLACK SWAN



Nassim Nicholas Taleb

- See “The Black Swan” by Nassim Taleb.
- Terrible if successful framing: Black swans are not that surprising ...

The PoCServe Power-Law Size Distributions 29 of 73

Our Intuition

Definition

Examples

Wild vs. Mild

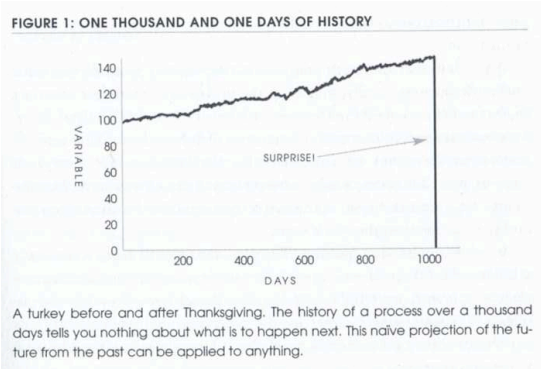
CCDFs

Size rankings and Zipf's law

Size ranking ↔ CCDF

References

Turkeys ...



From “The Black Swan”

The PoCServe Power-Law Size Distributions 30 of 73

Our Intuition

Definition

Examples

Wild vs. Mild

CCDFs

Size rankings and Zipf's law

Size ranking ↔ CCDF

References

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

The PoCServe Power-Law Size Distributions 31 of 73

Our Intuition

Definition

Examples

Wild vs. Mild

CCDFs

Size rankings and Zipf's law

Size ranking ↔ CCDF

References

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, see later).
- Term used especially by practitioners of the Dismal Science.

The PoCServe Power-Law Size Distributions 32 of 73

Our Intuition

Definition

Examples

Wild vs. Mild

CCDFs

Size rankings and Zipf's law

Size ranking ↔ CCDF

References

Devilish power-law size 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 assignment question

The PoCServe Power-Law Size Distributions 33 of 73

Our Intuition

Definition

Examples

Wild vs. Mild

CCDFs

Size rankings and Zipf's law

Size ranking ↔ CCDF

References

Moments:

- If $n \neq \gamma - 1$:

$$\langle x^n \rangle = \int_{x_{\min}}^{x_{\max}} x^n P(x) dx = \frac{c}{n - \gamma + 1} \left(x_{\max}^{n-\gamma+1} - x_{\min}^{n-\gamma+1} \right),$$

where $c = \frac{\gamma - 1}{a^{-(\gamma-1)} - b^{-(\gamma-1)}}.$

- Because both $n - \gamma + 1$ and $(x_{\max}^{n-\gamma+1} - x_{\min}^{n-\gamma+1})$ are either negative or positive, we can write:

$$\langle x^n \rangle = \frac{c}{|n - \gamma + 1|} \left| x_{\max}^{n-\gamma+1} - x_{\min}^{n-\gamma+1} \right|.$$

- If $n = \gamma - 1$:

$$\langle x^n \rangle = c \frac{x_{\max}}{x_{\min}}.$$

The PoCServe Power-Law Size Distributions 34 of 73

Our Intuition

Definition

Examples

Wild vs. Mild

CCDFs

Size rankings and Zipf's law

Size ranking ↔ CCDF

References

“The horror, the horror ...”

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)
- σ^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 assignment question

The PoCServe Power-Law Size Distributions 35 of 73

Our Intuition

Definition

Examples

Wild vs. Mild

CCDFs

Size rankings and Zipf's law

Size ranking ↔ CCDF

References

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 $\gamma < 3$.

The PoCServe Power-Law Size Distributions 36 of 73

Our Intuition

Definition

Examples

Wild vs. Mild

CCDFs

Size rankings and Zipf's law

Size ranking ↔ CCDF

References

How sample sizes grow ...

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

- We can show that after n samples, we expect the largest sample to be²

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

- 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 assignment question

The PoCServe Power-Law Size Distributions 37 of 73

Our Intuition

Definition

Examples

Wild vs. Mild

CCDFs

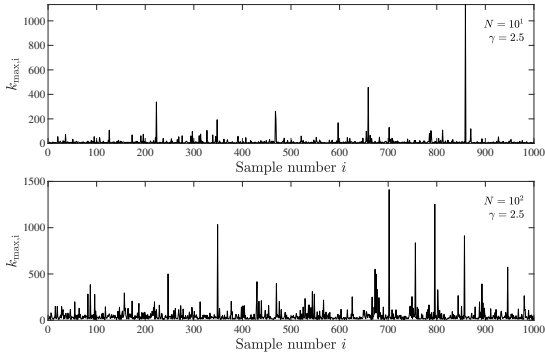
Size rankings and Zipf's law

Size ranking ↔ CCDF

References

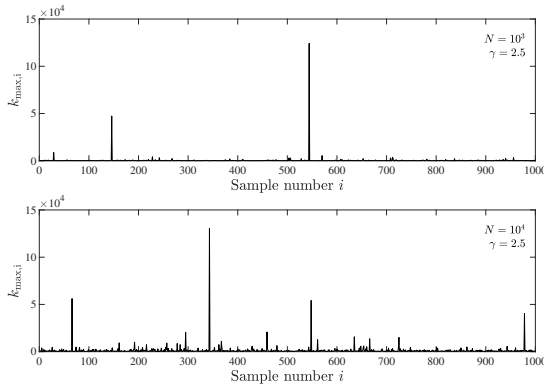
²Later, we see that the largest sample grows as n^α where α is the size-ranking exponent

$\gamma = 5/2$, maxima of N samples, $n = 1000$ sets of samples:



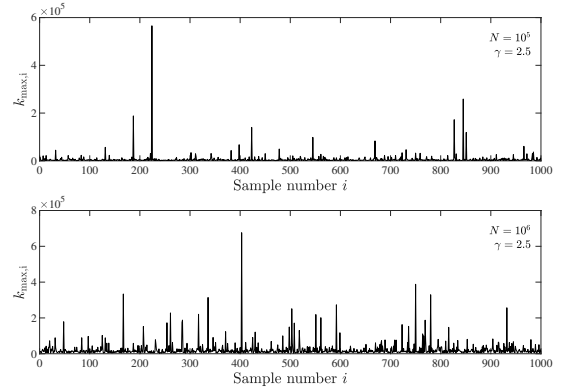
The PoCSverse
Power-Law Size
Distributions
38 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking $\leftrightarrow$
CCDF
References

$\gamma = 5/2$, maxima of N samples, $n = 1000$ sets of samples:



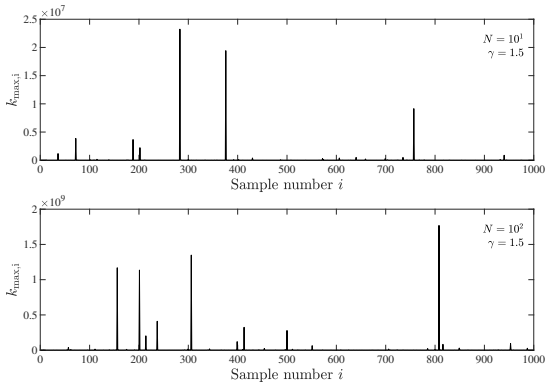
The PoCSverse
Power-Law Size
Distributions
39 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking $\leftrightarrow$
CCDF
References

$\gamma = 5/2$, maxima of N samples, $n = 1000$ sets of samples:



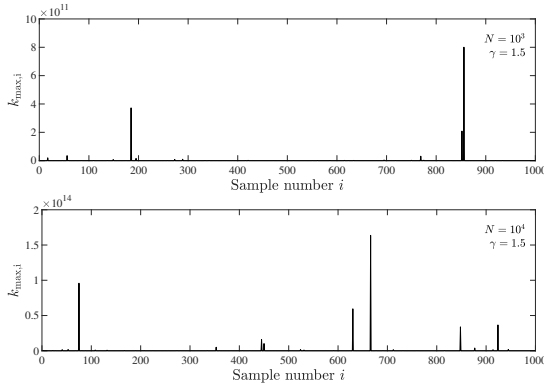
The PoCSverse
Power-Law Size
Distributions
40 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking $\leftrightarrow$
CCDF
References

$\gamma = 3/2$, maxima of N samples, $n = 1000$ sets of samples:



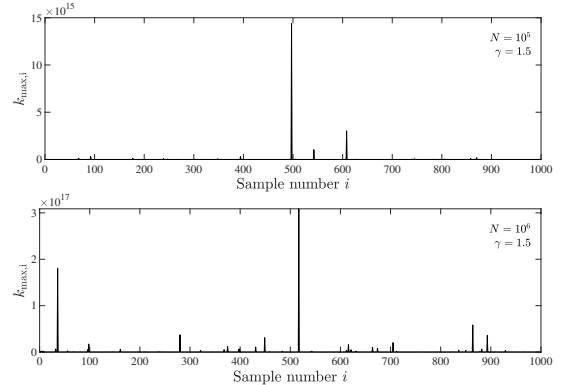
The PoCSverse
Power-Law Size
Distributions
41 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking $\leftrightarrow$
CCDF
References

$\gamma = 3/2$, maxima of N samples, $n = 1000$ sets of samples:



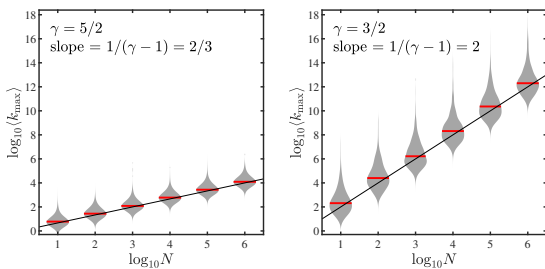
The PoCSverse
Power-Law Size
Distributions
42 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking $\leftrightarrow$
CCDF
References

$\gamma = 3/2$, maxima of N samples, $n = 1000$ sets of samples:



The PoCSverse
Power-Law Size
Distributions
43 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking $\leftrightarrow$
CCDF
References

Scaling of expected largest value as a function of sample size N :



The PoCSverse
Power-Law Size
Distributions
44 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking $\leftrightarrow$
CCDF
References

Back to understanding the 80/20 rule:

- Imagine a population of n people with variable x for individual wealth.
- Define $N(x) = cx^{-\gamma}$ as the distribution of wealth x .
- Must have $\int_{x_{\min}}^{\infty} dx N(x) = n$.
- Total wealth W in the system:
 $W = \int_{x_{\min}}^{\infty} dx xN(x)$.
- Imagine that the bottom $100\theta_{\text{pop}}$ percent of the population holds $100\theta_{\text{wealth}}$ percent of the wealth.
- Find γ depends on θ_{pop} and θ_{wealth} as

$$\gamma = 1 + \frac{\ln \frac{1}{(1-\theta_{\text{pop}})}}{\ln \frac{1}{(1-\theta_{\text{pop}})} - \ln \frac{1}{(1-\theta_{\text{wealth}})}}. \quad (1)$$

Pleasant detail: $x_{\min}$ does not matter.

Insert assignment question

The PoCSverse
Power-Law Size
Distributions
45 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking $\leftrightarrow$
CCDF
References

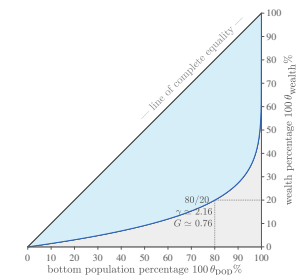
80/20, γ , and the Gini coefficient G :

Gini coefficient G :

Ratio of blue shape's area to triangle's area.

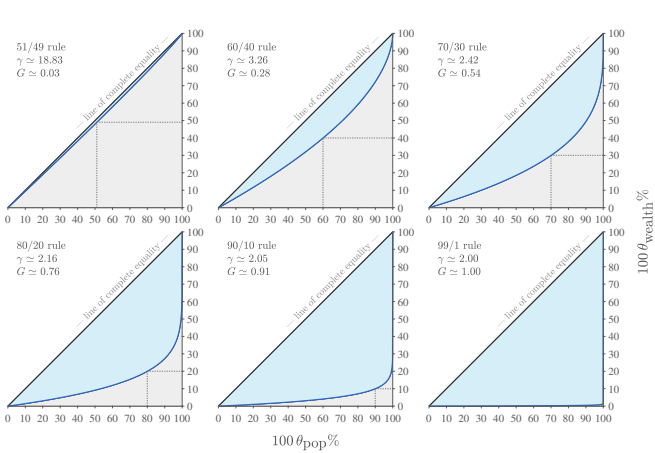
$0 \leq G \leq 1$

Blue line is the "Lorenz curve."



The PoCSverse
Power-Law Size
Distributions
46 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking $\leftrightarrow$
CCDF
References

The top 1% owns 52.3%, the top 0.1% 38.4%, the top 0.01% 27.9%, the top 10^{-7} % 5.6%, ...



99/1 rule:

$\gamma \approx 2.002$.

$100 \theta_{\text{pop}}$	$100 \theta_{\text{wealth}}$	$100(1 - \theta_{\text{pop}})$	$100(1 - \theta_{\text{wealth}})$
20	0.05	80	99.95
50	0.15	50	99.85
80	0.35	20	99.65
$100 - 10^{-1}$	0.50	10^1	99.50
99	1	1	99
$100 - 10^{-1}$	1.50	10^{-1}	98.50
$100 - 10^{-2}$	1.99	10^{-2}	98.01
$100 - 10^{-3}$	2.48	10^{-3}	97.52
$100 - 10^{-4}$	2.97	10^{-4}	97.03
$100 - 10^{-5}$	3.46	10^{-5}	96.54
$100 - 10^{-6}$	3.94	10^{-6}	96.06
$100 - 10^{-7}$	4.42	10^{-7}	95.58
$100 - 10^{-8}$	4.90	10^{-8}	95.10
$100 - 10^{-9}$	5.38	10^{-9}	94.62
$100 - 10^{-10}$	5.85	10^{-10}	94.15
$100 - 10^{-11}$	6.32	10^{-11}	93.68
$100 - 10^{-12}$	6.79	10^{-12}	93.21
$100 - 10^{-13}$	7.26	10^{-13}	92.74

Complementary Cumulative Distribution Function:

CCDF:



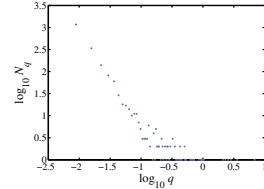
$$P_{\geq}(x) \propto x^{-(\gamma-1)}$$

Use when tail of P follows a power law.

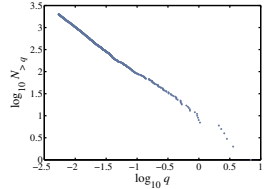
Increases exponent by one.

Useful in cleaning up data.

PDF:



CCDF:



The PoCSverse
Power-Law Size
Distributions
50 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking $\leftrightarrow$
CCDF
References

Gini coefficient:

$$G = \begin{cases} \frac{1}{1+2(\gamma-2)} & \text{if } 1 < \gamma \leq 2, \\ \frac{1}{1+2(\gamma-2)} & \text{if } \gamma > 2. \end{cases} \quad (2)$$

Insert assignment question

Complementary Cumulative Distribution Function:

Same story for a discrete variable: $P(k) \sim ck^{-\gamma}$.



$$P_{\geq}(k) = P(k' \geq k) = \sum_{k'=k}^{\infty} P(k') \propto k^{-(\gamma-1)}$$

Use integrals to approximate sums.

The PoCSverse
Power-Law Size
Distributions
48 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking $\leftrightarrow$
CCDF
References

80/20 rule:

$\gamma \approx 2.16$.

$100 \theta_{\text{pop}}$	$100 \theta_{\text{wealth}}$	$100(1 - \theta_{\text{pop}})$	$100(1 - \theta_{\text{wealth}})$
20	3.05	80	96.95
50	9.16	50	90.84
80	20	20	80
90	27.33	10	72.67
99	47.19	1	52.81
$100 - 10^{-1}$	61.62	10^{-1}	38.38
$100 - 10^{-2}$	72.11	10^{-2}	27.89
$100 - 10^{-3}$	79.73	10^{-3}	20.27
$100 - 10^{-4}$	85.27	10^{-4}	14.73
$100 - 10^{-5}$	89.30	10^{-5}	10.70
$100 - 10^{-6}$	92.22	10^{-6}	7.78
$100 - 10^{-7}$	94.35	10^{-7}	5.65
$100 - 10^{-8}$	95.89	10^{-8}	4.11
$100 - 10^{-9}$	97.02	10^{-9}	2.98
$100 - 10^{-10}$	97.83	10^{-10}	2.17
$100 - 10^{-11}$	98.42	10^{-11}	1.58
$100 - 10^{-12}$	98.85	10^{-12}	1.15
$100 - 10^{-13}$	99.17	10^{-13}	0.83

Complementary Cumulative Distribution Function:

CCDF:



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



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



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



$$= \frac{1}{-\gamma + 1} (x')^{-\gamma+1} \Big|_{x'=x}^{\infty}$$



$$\propto x^{-(\gamma-1)}$$

“Zipfian” size-rank 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



“Human Behaviour and the Principle of Least-Effort” by G. K. Zipf (1949).^[20]



We'll study Zipf's law in depth ...

The PoCSverse
Power-Law Size
Distributions
49 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking $\leftrightarrow$
CCDF
References

The PoCSverse
Power-Law Size
Distributions
52 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking $\leftrightarrow$
CCDF
References

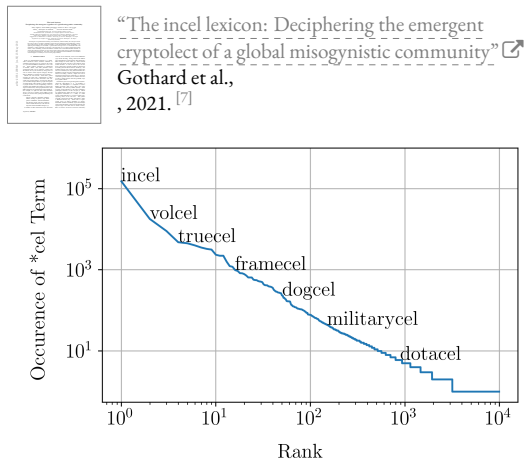
The PoCSverse
Power-Law Size
Distributions
55 of 73
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Size rankings and
Zipf's law
Size ranking $\leftrightarrow$
CCDF
References

“Zipfian” size-rank plots

Zipf’s way:

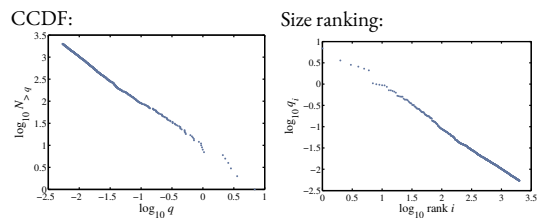
- Given a collection of entities, rank them by size, largest to smallest.
- S_r = the size of the r th ranked entity.
- General term: “Size ranking”
- $r = 1$ corresponds to the largest size.
- Example: S_1 could be the frequency of occurrence of the most common word in a text.
- Zipf’s observation:
$$S_r \propto r^{-\alpha}$$
with α often close to 1.

Incel typology:



Size distributions:

Brown Corpus (1,015,945 words):



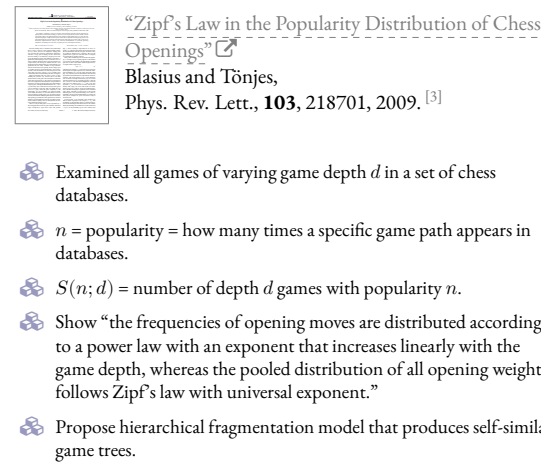
- The, of, and, to, a, ...= ‘objects’
- ‘Size’ = word frequency
- Beep: (Important) CCDF and size-ranking plots are related ...

Misrankings

The “biggest” thing is rank #1, otherwise:

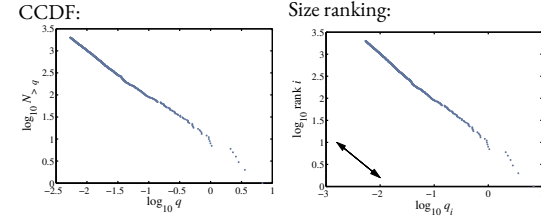
- “USA #195!”⁴
- “USA #195!”
- “USA #195!”
- “USA #195!”
- More:
 - Size distribution connects with ‘#1-is-biggest’ ‘size’ ranking only
 - Main form of ranking by decreasing ‘size’ is robust to low sampling of small ‘size’ entities (the tail ‘fills in’).

⁴As of August 2024. Not simple agreed upon by all.



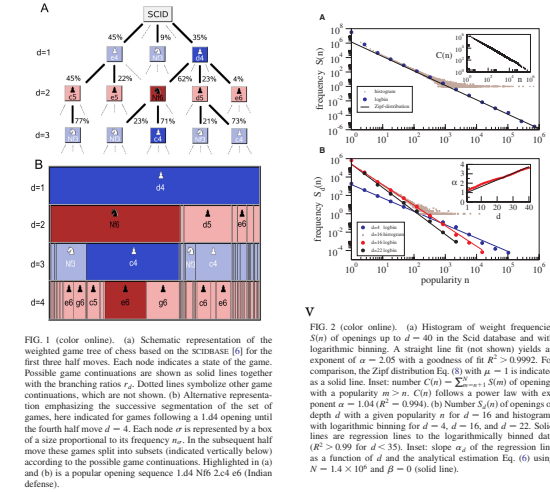
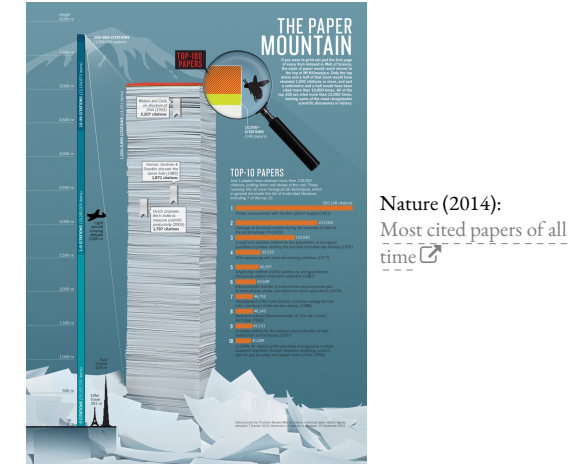
Size distributions:

Brown Corpus (1,015,945 words):



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

Size ranking example:



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 $\gamma = 2$.

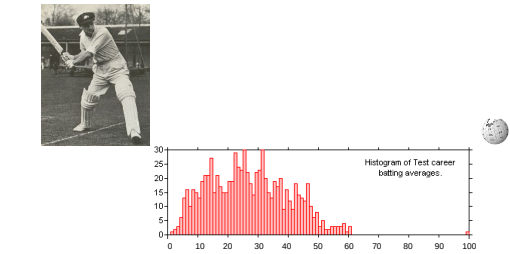
A numerological digression:

- ⌘ We have
$$\alpha = \frac{1}{\gamma - 1} \Leftrightarrow \gamma = 1 + \frac{1}{\alpha}.$$
- ⌘ When does $\alpha = \gamma$?
- ⌘ We set
$$\gamma = \frac{1}{\gamma - 1} \Rightarrow \gamma^2 - \gamma - 1 = 0.$$
- ⌘ Yes, it's the golden ratio φ :⁵ $\Rightarrow \gamma = \varphi = \frac{1+\sqrt{5}}{2} = 1.6180 \dots$
- ⌘ We sing: “ φ is the most irrational number.”
- ⌘ φ is the limiting ratio of consecutive Fibonacci numbers, F_{n+1}/F_n .
- ⌘ Need to convert a Fibonacci number of miles to kilometres? It's as easy as φ because φ is 1.6093 ... $\approx$ 1.6180 ...

⁵ $\gamma = (1 - \sqrt{5})/2 < 1$ is an apysical solution.

The Don.

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 III

[9] B. Gutenberg and C. F. Richter.
Earthquake magnitude, intensity, energy, and acceleration.
Bull. Seism. Soc. Am., 499:105–145, 1942. [pdf](#) 

[10] J. Holtsmark.
Über die verbreiterung von spektrallinien.
Ann. Phys., 58:577–630, 1919. [pdf](#) 

[11] R. Munroe.
Thing Explainer: Complicated Stuff in Simple Words.
Houghton Mifflin Harcourt, 2015.

[12] M. E. J. Newman.
Power laws, Pareto distributions and Zipf's law.
Contemporary Physics, 46:323–351, 2005. [pdf](#) 

Nutshell for power-law size distributions and size-rank orderings:

- ⌘ Heavy-tailed distributions abound.
- ⌘ Some are power-law size distributions.
- ⌘ Continuous: $P(x) \sim x^{-\gamma}$, discrete: $P(k) \sim ck^{-\gamma}$
- ⌘ 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.
- ⌘ Complementary Cumulative Distribution Function (CCDF): $P(x) \propto x^{-(\gamma-1)}$ and $P_{\geq}(k) = k^{-(\gamma-1)}$
- ⌘ Size of largest sample from n samples grows as:

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

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] B. Blasius and R. Tönjes.
Zipf's law in the popularity distribution of chess openings.
Phys. Rev. Lett., 103:218701, 2009. [pdf](#) 

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

References IV

[13] M. I. Norton and D. Ariely.
Building a better America—One wealth quintile at a time.
Perspectives on Psychological Science, 6:9–12, 2011. [pdf](#) 

[14] D. D. S. Price.
A general theory of bibliometric and other cumulative advantage processes.
Journal of the American Society for Information Science, pages 292–306, 1976. [pdf](#) 

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

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

More with the nutshelling:

- ⌘ Size rankings: Order types from “biggest” to “smallest” size S .
- ⌘ Widely observed: S_r is highly skewed.
- ⌘ When scaling is apparent:

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

- ⌘ Claim: α often close to 1. “Zipf's law”:

$$S_r \propto r^{-1}.$$

- ⌘ Scalings of size distribution (γ) and size ranking (α) are connected:

$$\alpha = \frac{1}{\gamma - 1} \text{ and } \gamma = 1 + \frac{1}{\alpha}.$$

- ⌘ Danger Will Robinson point: $\gamma = 2 \Leftrightarrow \alpha = 1$.

References II

[5] A. Clauset, C. R. Shalizi, and M. E. J. Newman.
Power-law distributions in empirical data.
SIAM Review, 51:661–703, 2009. [pdf](#) 

[6] D. J. de Solla Price.
Networks of scientific papers.
Science, 149:510–515, 1965. [pdf](#) 

[7] K. Gothard, D. R. Dewhurst, J. A. Minot, J. L. Adams, C. M. 5-Danforth, and P. S. Dodds.
The incel lexicon: Deciphering the emergent cryptoelect of a global misogynistic community, 2021.
Available online at <https://arxiv.org/abs/2105.12006>. [pdf](#) 

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

References V

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

[18] G. U. Yule.
A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S.
Phil. Trans. B, 213:21–87, 1925. [pdf](#) 

[19] Y.-X. Zhu, J. Huang, Z.-K. Zhang, Q.-M. Zhang, T. Zhou, and Y.-Y. Ahn.
Geography and similarity of regional cuisines in China.
PLoS ONE, 8:e79161, 2013. [pdf](#) 

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

