

Power-Law Size Distributions

Last updated: 2019/08/26, 12:11:53

Principles of Complex Systems | @pocsvox
CSYS/MATH 300, Fall, 2019

Dr. Nick Allgaier | @nallgaier
Prof. Peter Sheridan Dodds | @peterdodds

Dept. of Mathematics & Statistics | Vermont Complex Systems Center
Vermont Advanced Computing Core | University of Vermont



Licensed under the [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License](#).

PoCS
@pocsvox
Power-Law Size
Distributions

Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf $\Leftrightarrow$ CCDF
References



1 of 63

PoCS
@pocsvox
Power-Law Size
Distributions

Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf $\Leftrightarrow$ CCDF
References



2 of 63

PoCS
@pocsvox
Power-Law Size
Distributions

Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf $\Leftrightarrow$ CCDF
References



3 of 63

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?

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?

Let's test our collective intuition:



Money
 $\equiv$
Belief

Two questions about wealth distribution in the United States:

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

Wealth distribution in the United States: ^[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]

Wealth distribution in the United States: ^[13]

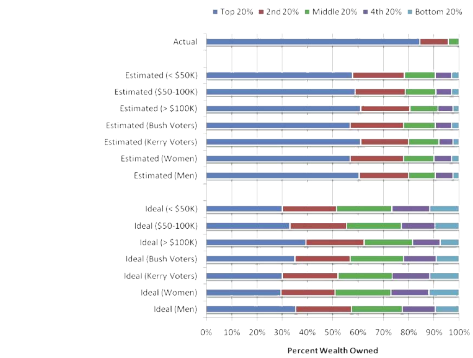


Fig. 3. The actual United States wealth distribution plotted against the estimated and ideal distributions of respondents of different income levels, political affiliations, and genders. Because of their small percentage share of total wealth, both the "4th 20%" value (0.2%) and the "Bottom 20%" value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

A highly watched video based on this research is

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

$$P(\text{size} = x) \sim cx^{-\gamma}$$

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.

PoCS
@pocsvox
Power-Law Size
Distributions

Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf $\Leftrightarrow$ CCDF
References



7 of 63

PoCS
@pocsvox
Power-Law Size
Distributions

Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf $\Leftrightarrow$ CCDF
References



9 of 63

PoCS
@pocsvox
Power-Law Size
Distributions

Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf $\Leftrightarrow$ CCDF
References



12 of 63

Outline

Our Intuition

Definition

Examples

Wild vs. Mild

CCDFs

Zipf's law

Zipf $\Leftrightarrow$ 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 magnificence of the
Dunning-Kruger effect

Size distributions:

Usually, only the tail of the distribution obeys a power law:

$P(x) \sim c x^{-\gamma}$ for x large.

- Still use term ‘power-law size distribution.’
- Other terms:
 - Fat-tailed distributions.
 - Heavy-tailed distributions.

Beware:

- Inverse power laws aren’t the only ones: [lognormals](#), [Weibull distributions](#), ...

Size distributions:

Many systems have discrete sizes k :

- Word frequency
- Node degree in networks: # friends, # hyperlinks, etc.
- # citations for articles, court decisions, etc.

$P(k) \sim c k^{-\gamma}$
where $k_{\min} \leq k \leq k_{\max}$

- Obvious fail for $k = 0$.
- Again, typically a description of distribution’s tail.

Word frequency:

[Brown Corpus](#) (~ 10⁶ words):

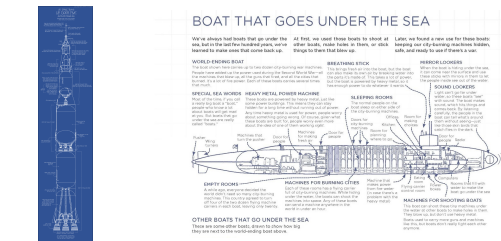
rank	word	% q	rank	word	% q
1.	the	6.8872	1945.	apply	0.0055
2.	of	3.5839	1946.	vital	0.0055
3.	and	2.8401	1947.	September	0.0055
4.	to	2.5744	1948.	review	0.0055
5.	a	2.2996	1949.	wage	0.0055
6.	in	2.1010	1950.	motor	0.0055
7.	that	1.0428	1951.	fifteen	0.0055
8.	is	0.9943	1952.	regarded	0.0055
9.	was	0.9661	1953.	draw	0.0055
10.	he	0.9392	1954.	wheel	0.0055
11.	for	0.9340	1955.	organized	0.0055
12.	it	0.8623	1956.	vision	0.0055
13.	with	0.7176	1957.	wild	0.0055
14.	as	0.7137	1958.	Palmer	0.0055
15.	his	0.6886	1959.	intensity	0.0055

Jonathan Harris’s [Wordcount](#):

A word frequency distribution explorer:



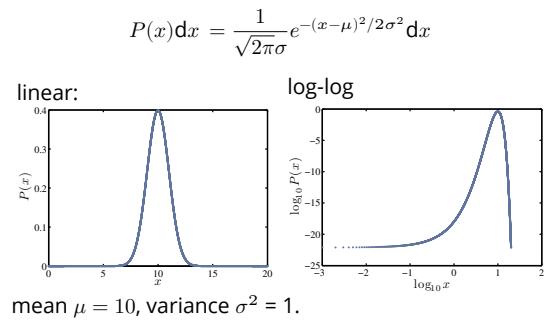
“[Thing Explainer: Complicated Stuff in Simple Words](#)”
by Randall Munroe (2015).^[11]



[Up goer five](#)

The statistics of surprise—words:

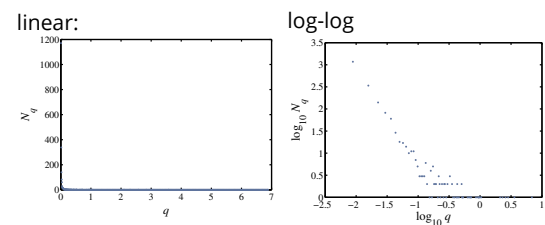
First—a Gaussian example:



Activity: Sketch $P(x) \sim x^{-1}$ for $x = 1$ to $x = 10^7$.

The statistics of surprise—words:

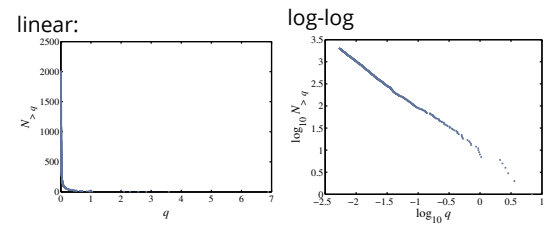
Raw ‘probability’ (binned) for Brown Corpus:



- q_w = normalized frequency of occurrence of word w (%).
- N_q = number of distinct words that have a normalized frequency of occurrence q .
- e.g. $q_{\text{the}} \simeq 6.9\%$, $N_{q_{\text{the}}} = 1$.

The statistics of surprise—words:

Complementary Cumulative Probability Distribution $N_{>q}$:



Also known as the ‘Exceedance Probability.’

My, what big words you have ...

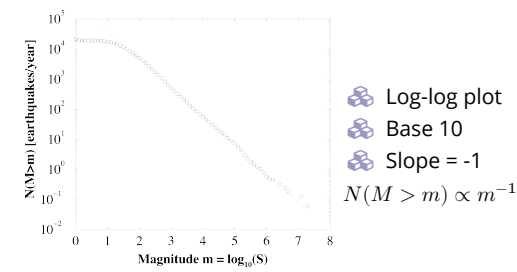


Test capitalizes on word frequency following a heavily skewed frequency distribution with a decaying power-law tail.

- This [Man Can Pronounce Every Word in the Dictionary](#) (story [here](#))
- Best of Dr. Bailly

The statistics of surprise:

Gutenberg-Richter law



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

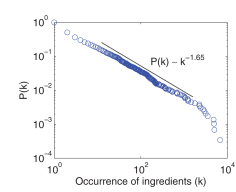
The statistics of surprise:

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, ...’

“Geography and similarity of regional cuisines in China”
Zhu et al.,
PLoS ONE, 8, e79161, 2013. [18]



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

PoCS
@pocsvox
Power-Law Size Distributions

Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf’s law
Zipf $\Leftrightarrow$ CCDF
References

22 of 63

PoCS
@pocsvox
Power-Law Size Distributions

23 of 63

PoCS
@pocsvox
Power-Law Size Distributions

Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf’s law
Zipf $\Leftrightarrow$ CCDF
References

24 of 63

“On a class of skew distribution functions”
Herbert A. Simon,
Biometrika, 42, 425–440, 1955. [15]

“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]

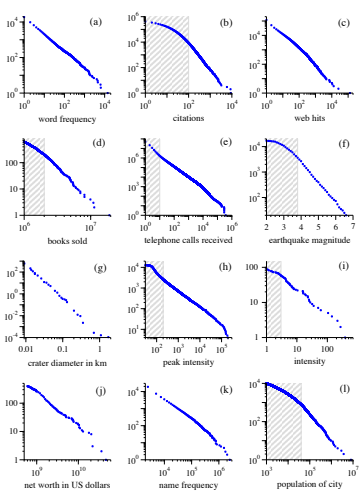


FIG. 4. Cumulative distributions or “rank-frequency plots” of twelve quantities are shown in the figure. The distributions were computed as described in Appendix A. Data in the shaded regions were excluded from the calculations of the exponents in Table 1. Source references for the data are given in the text. (a) Numbers of occurrences of words in the novel Moby Dick (1851). (b) Numbers of occurrences of words in the novel War and Peace (1869). (c) Numbers of hits on web sites by 60,000 users of the America Online Internet service for the day of 1 December 1997. (d) Numbers of books sold in the United States in the year 1998. (e) Numbers of telephone calls received in the United States in the year 1998. (f) Magnitude of earthquakes in California between January 1910 and May 1992. (g) Crater diameter in km. (h) Peak intensity of earthquakes in the United States between January 1910 and May 1992. (i) Intensity of earthquakes in the United States between January 1910 and May 1992. (j) Net worth in US dollars. (k) Name frequency. (l) Population of city in the year 1998. (m) Frequency of occurrence of family names in the US in the year 1998. (n) Population of US cities in the year 2000.

Size distributions:

Some examples:

- Earthquake magnitude (Gutenberg-Richter law): [9, 1] $P(M) \propto M^{-2}$
- # war deaths: [14] $P(d) \propto d^{-1.8}$
- Sizes of forest fires [8]
- Sizes of cities: [15] $P(n) \propto n^{-2.1}$
- # links to and from websites [2]
- Note: Exponents range in error

Size distributions:

More examples:

- # citations to papers: [6, 7] $P(k) \propto k^{-3}$.
- 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: [10] $P(F) \propto F^{-5/2}$. (See the Holtzmark distribution and stable distributions.)
- Diameter of moon craters: [12] $P(d) \propto d^{-3}$.
- Word frequency: [15] e.g., $P(k) \propto k^{-2.2}$ (variable).
- # religious adherents in cults: [5] $P(k) \propto k^{-1.8 \pm 0.1}$.
- # sightings of birds per species (North American Breeding Bird Survey for 2003): [5] $P(k) \propto k^{-2.1 \pm 0.1}$.
- # species per genus: [17, 15, 5] $P(k) \propto k^{-2.4 \pm 0.2}$.

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	P_{max}	x_{min}	$\hat{\alpha}$	n_{tail}	P
count of word use	18,855	11.14	1.48(3)	11,086	7 ± 2	1.85(2)	2958 $\pm$ 987	0.49
protein interaction degree	18,46	2.34	3.05	56	5 ± 2	3.1(3)	204 $\pm$ 263	0.31
metabolic degree	16,41	5.68	17.81	468	4 ± 1	2.8(1)	748 $\pm$ 136	0.00
Internet degree	22,088	3.63	37.83	2583	21 ± 9	2.12(9)	770 $\pm$ 1124	0.29
telephone calls received	51,360,423	3.88	179.09	375,746	120 ± 40	2.09(1)	102,292 $\pm$ 210,147	0.63
intensity of wars	115	15.70	49.97	382	2.1 ± 3.5	1.7(2)	70 $\pm$ 14	0.20
terrorist attack severity	9101	4.35	31.58	2749	12 ± 4	2.4(2)	547 $\pm$ 1663	0.68
HTTP size (kilobytes)	226,386	7.36	57.94	10,971	36.25 ± 22.74	2.48(5)	6734 $\pm$ 2232	0.00
species per genus	509	5.59	6.94	56	4 ± 2	2.4(2)	233 $\pm$ 138	0.10
bird species sightings	591	3384.36	10,952.34	138,705	6679 ± 2463	2.1(2)	66 $\pm$ 41	0.55
blackouts ($\times 10^3$)	211	253.87	610.31	7500	230 ± 90	2.3(3)	59 $\pm$ 35	0.62
sales of books ($\times 10^3$)	633	1986.67	1396.60	19,077	2400 ± 430	3.7(3)	139 $\pm$ 115	0.66
population of cities ($\times 10^3$)	19,447	9.00	77.83	8009	52.46 ± 11.88	2.37(8)	580 $\pm$ 177	0.76
email address books size	4581	12.45	21.49	333	57 ± 21	3.5(6)	196 $\pm$ 449	0.16
forest fire size (acres)	203,785	0.90	20.99	4121	6324 ± 3487	2.2(3)	521 $\pm$ 6801	0.05
solar flare intensity	12,773	689.41	6520.59	231,300	323 ± 89	1.79(2)	1711 $\pm$ 384	1.00
quake intensity ($\times 10^3$)	19,302	24.54	563.83	63,096	0.794 ± 80.198	1.64(4)	11,697 $\pm$ 2159	0.00
religious followers ($\times 10^3$)	103	27.36	136.64	1050	3.85 ± 1.09	1.8(1)	39 $\pm$ 26	0.42
freq. of surnames ($\times 10^3$)	2753	50.59	113.99	2502	111.02 ± 40.67	2.5(2)	239 $\pm$ 215	0.20
net worth (mil. USD)	400	2388.69	4167.35	46,000	960 ± 364	2.3(1)	302 $\pm$ 77	0.00
citations to papers	412,229	16.17	44.02	8904	160 ± 35	3.16(6)	3455 $\pm$ 1859	0.20
papers authored	401,445	7.21	16.52	1416	133 ± 13	4.3(1)	988 $\pm$ 377	0.90
hits to web sites	119,724	9.83	392.52	129,641	2 ± 13	1.81(8)	50,981 $\pm$ 16,898	0.00
links to web sites	241,428,853	9.15	106,871.65	1,199,466	3684 ± 151	2.336(9)	28,986 $\pm$ 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.



Nassim Nicholas Taleb

See “The Black Swan” by Nassim Taleb. [16]
Terrible if successful framing: Black swans are not that surprising ...

PoCS
@pocsvox
Power-Law Size Distributions

Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf’s law
Zipf $\Leftrightarrow$ CCDF
References

28 of 63

PoCS
@pocsvox
Power-Law Size Distributions

26 of 63

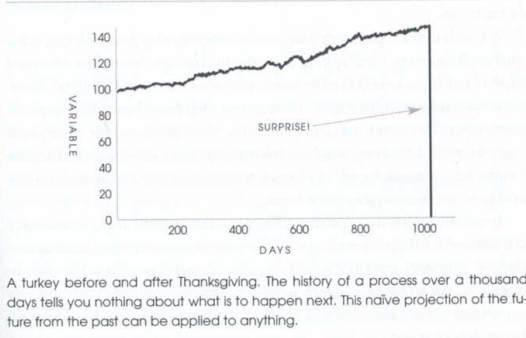
PoCS
@pocsvox
Power-Law Size Distributions

Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf’s law
Zipf $\Leftrightarrow$ CCDF
References

30 of 63

Turkeys ...

FIGURE 1: ONE THOUSAND AND ONE DAYS OF HISTORY



From “The Black Swan”^[16]

Taleb’s table^[16]

- 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

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

Devilish power-law size distribution details:

Exhibit A:

- Given $P(x) = cx^{-\gamma}$ with $0 < x_{\min} < x < x_{\max}$, 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 question from assignment 2

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)
- 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 question from assignment 3

Moments

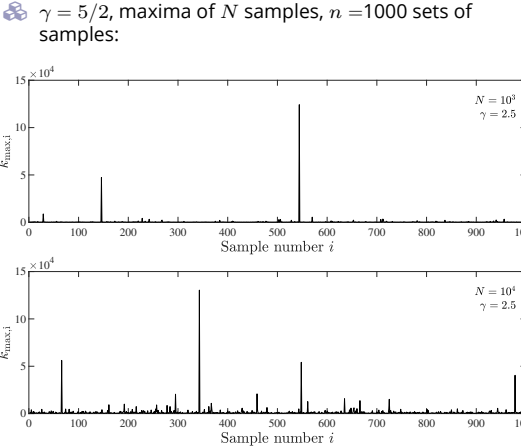
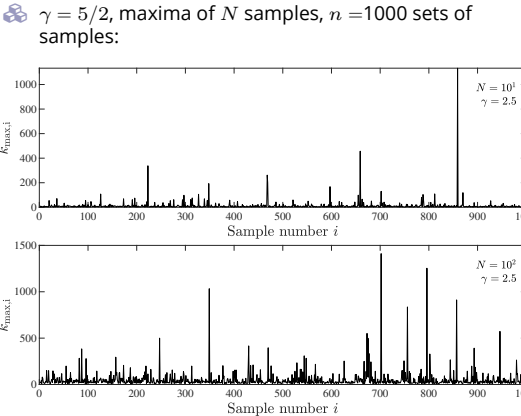
- Standard deviation is a mathematical convenience:
- Variance is nice analytically ...
- Another measure of distribution width:
$$\text{Mean average deviation (MAD)} = \langle |x - \langle x \rangle| \rangle$$
- For a pure power law with $2 < \gamma < 3$:
$$\langle |x - \langle x \rangle| \rangle \text{ is finite.}$$
- But MAD is mildly unpleasant analytically ...
- We still speak of infinite ‘width’ if $\gamma < 3$.

How sample sizes grow ...

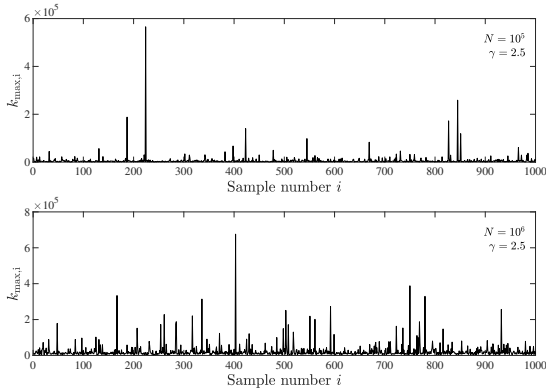
- Given $P(x) \sim cx^{-\gamma}$:
- We can show that after n samples, we expect the largest sample to be¹
$$x_1 \gtrsim c' n^{1/(\gamma-1)}$$
- Sampling from a finite-variance distribution gives a much slower growth with n .
- e.g., for $P(x) = \lambda e^{-\lambda x}$, we find
$$x_1 \gtrsim \frac{1}{\lambda} \ln n.$$

Insert question from assignment 4
Insert question from assignment 6

¹Later, we see that the largest sample grows as n^ρ where ρ is the Zipf exponent

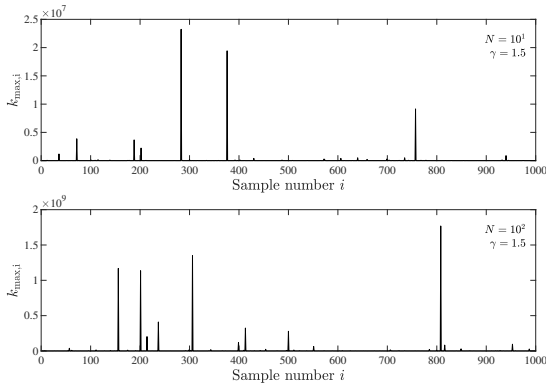


🧠 $\gamma = 5/2$, maxima of N samples, $n = 1000$ sets of samples:



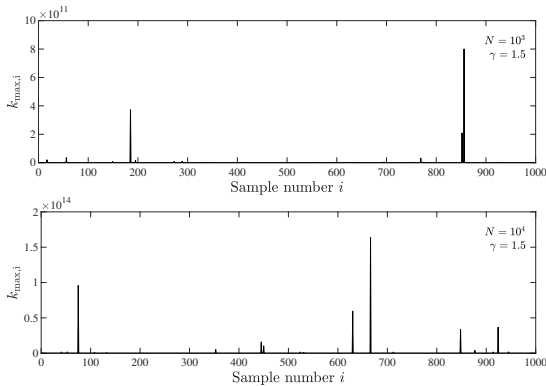
🔗 40 of 63

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



🔗 41 of 63

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

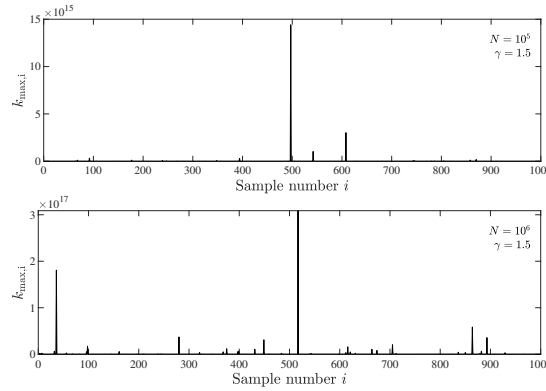


🔗 42 of 63

PoCS
@pocsvox
Power-Law Size
Distributions

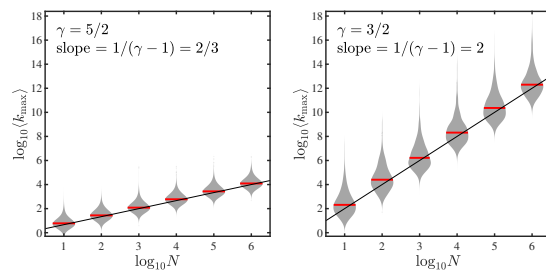
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf $\leftrightarrow$ CCDF
References

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



🔗 43 of 63

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



🧠 Fit for $\gamma = 5/2$: $k_{\max} \sim N^{0.660 \pm 0.066}$ (sublinear)
🧠 Fit for $\gamma = 3/2$: $k_{\max} \sim N^{2.063 \pm 0.215}$ (superlinear)

295% confidence interval

Complementary Cumulative Distribution Function:
CCDF:



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



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



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



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



$$\propto x^{-\gamma + 1}$$

PoCS
@pocsvox
Power-Law Size
Distributions

Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf $\leftrightarrow$ CCDF
References

Complementary Cumulative Distribution Function:
CCDF:



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



Use when tail of P follows a power law.

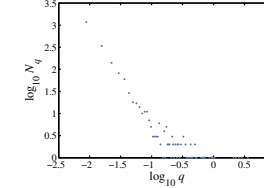


Increases exponent by one.

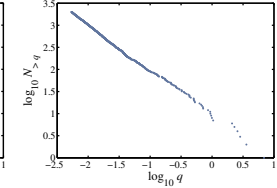


Useful in cleaning up data.

PDF:



CCDF:



🔗 46 of 63

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.



🔗 47 of 63

Zipfian rank-frequency plots

George Kingsley Zipf:



Noted various rank distributions have power-law tails, often with exponent -1 (word frequency, city sizes, ...)



Zipf's 1949 Magnum Opus:



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

PoCS
@pocsvox
Power-Law Size
Distributions

Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf $\leftrightarrow$ CCDF
References



🔗 45 of 63



🔗 49 of 63

Zipfian rank-frequency plots

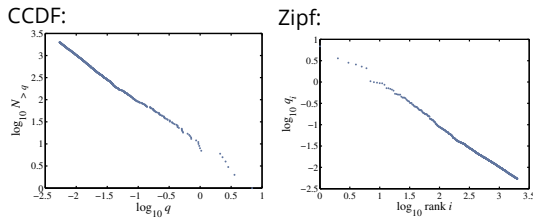
Zipf's way:

- Given a collection of entities, rank them by size, largest to smallest.
- x_r = the size of the r th ranked entity.
- $r = 1$ corresponds to the largest size.
- Example: x_1 could be the frequency of occurrence of the most common word in a text.
- Zipf's observation:

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

Size distributions:

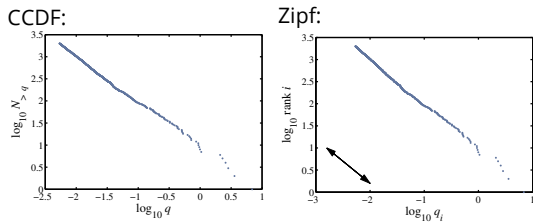
Brown Corpus (1,015,945 words):



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

Size distributions:

Brown Corpus (1,015,945 words):



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

Observe:

- $NP_{\geq}(x)$ = the number of objects with size at least x where N = total number of objects.
 - If an object has size x_r , then $NP_{\geq}(x_r)$ is its rank r .
 - So
- $$x_r \propto r^{-\alpha} = (NP_{\geq}(x_r))^{-\alpha}$$
- $$\propto x_r^{(-\gamma+1)(-\alpha)} \text{ since } P_{\geq}(x) \sim x^{-\gamma+1}.$$

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 $\gamma = 2$.



- Examined all games of varying game depth d in a set of chess databases.
- n = popularity = how many times a specific game path appears in databases.
- $S(n; d)$ = number of depth d games with popularity n .
- Show "the frequencies of opening moves are distributed according to a power law with an exponent that increases linearly with the game depth, whereas the pooled distribution of all opening weights follows Zipf's law with universal exponent."
- Propose hierarchical fragmentation model that produces self-similar game trees.

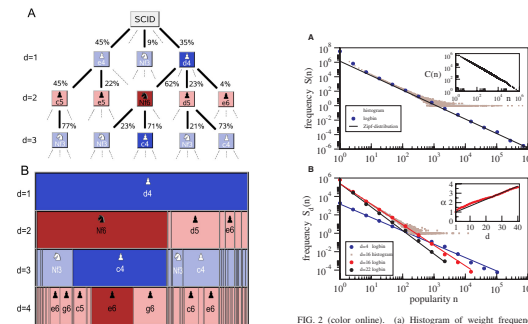


FIG. 1 (color online). (a) Schematic representation of the weighted game tree of chess based on the SCID database [6] for the first three half moves. Each node indicates a state of the game. Possible game continuations are shown as solid lines together with the branching ratios r_d . Dotted lines symbolize other game continuations, which are not shown. (b) Alternative representation emphasizing the successive segmentation of the set of games, here indicated for games following a 1.d4 opening until the fourth half move $d = 4$. Each node σ is represented by a box of size proportional to its frequency n_{σ} . In the subsequent half move these games split into subsets (indicated vertically below) according to the possible game continuations. Highlighted in (a) and (b) is a popular opening sequence 1.d4 Nf6 2.c4 e6 (Indian defence).

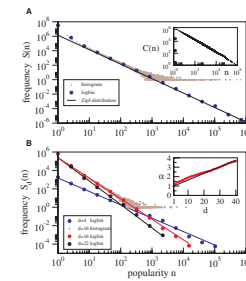
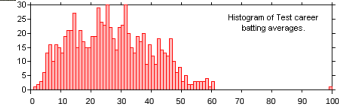


FIG. 2 (color online). (a) Histogram of weight frequencies $S(n)$ of openings up to $d = 40$ in the SCID database and with logarithmic binning. A straight line fit (not shown) yields an exponent of $\alpha = 2.05$ with a goodness of fit $R^2 > 0.9992$. For comparison, the Zipf distribution Eq. (6) with $\mu = 1$ is indicated as a solid line. Inset: number $C(n) = \sum_{m=1}^n S(m)$ of openings with a popularity $m > n$. $C(n)$ follows a power law with exponent $\alpha = 1.04$ ($R^2 = 0.994$). (b) Number $S_d(n)$ of openings of depth d with a given popularity n for $d = 16$ and histograms with logarithmic binning for $d = 4, d = 16$, and $d = 22$. Solid lines are regression lines to the logarithmically binned data ($R^2 > 0.99$ for $d < 35$). Inset: slope α_d of the regression line as a function of d and the analytical estimation Eq. (6) using $N = 1.4 \times 10^6$ and $\beta = 0$ (solid line).

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?

A good eye:

- The great Paul Kelly's tribute to the man who was "Something like the tide"

References I

- P. Bak, K. Christensen, L. Danon, and T. Scanlon. Unified scaling law for earthquakes. *Phys. Rev. Lett.*, 88:178501, 2002. pdf
- A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–511, 1999. pdf
- B. Blasius and R. Tönjes. Zipf's law in the popularity distribution of chess openings. *Phys. Rev. Lett.*, 103:218701, 2009. pdf
- K. Christensen, L. Danon, T. Scanlon, and P. Bak. Unified scaling law for earthquakes. *Proc. Natl. Acad. Sci.*, 99:2509–2513, 2002. pdf

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] D. J. de Solla Price. A general theory of bibliometric and other cumulative advantage processes. [J. Amer. Soc. Inform. Sci.](#), 27:292–306, 1976. pdf

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

PoCS
@pocsvox
Power-Law Size Distributions
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf ⇔ CCDF
References

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–, 1919.

[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

PoCS
@pocsvox
Power-Law Size Distributions
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf ⇔ CCDF
References

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

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

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

PoCS
@pocsvox
Power-Law Size Distributions
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf ⇔ CCDF
References

References V

[17] 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

[18] 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

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

PoCS
@pocsvox
Power-Law Size Distributions
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf ⇔ CCDF
References