Power-Law Size Distributions

Principles of Complex Systems
CSYS/MATH 300, Spring, 2013

Prof. Peter Dodds
@peterdodds

Department of Mathematics & Statistics | Center for Complex Systems | Vermont Advanced Computing Center | University of Vermont

Outline

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

Wealth distribution in the United States: [8]

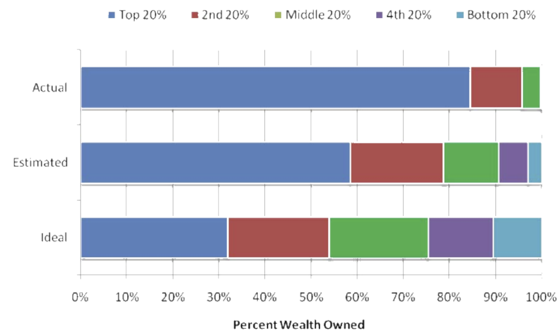


Fig. 2. The actual United States wealth distribution plotted against the estimated and ideal distributions across all respondents. Because of their small percentage share of total wealth, both the "4th 20%" value (0.2%) and the "Bottom 20%" value (0.1%) are not visible in the "Actual" distribution.

"Building a better America—One wealth quintile at a time" Norton and Ariely, 2011. [8]

Wealth distribution in the United States: [8]

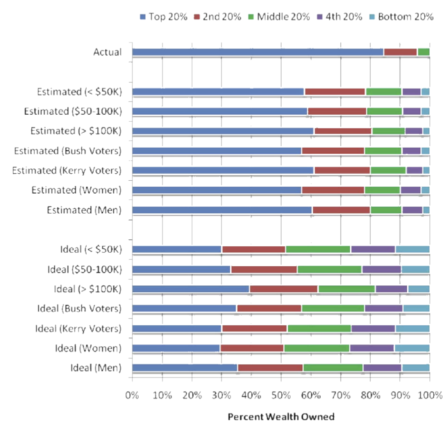
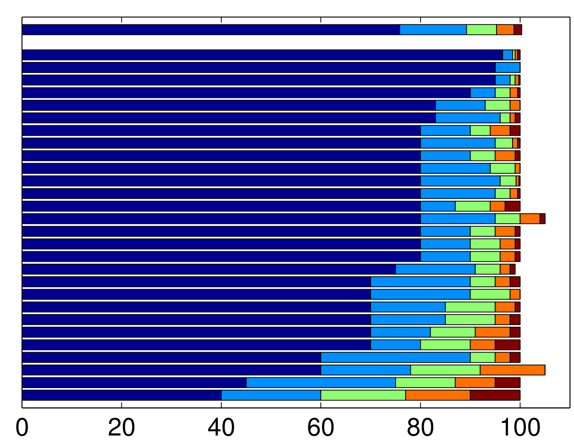
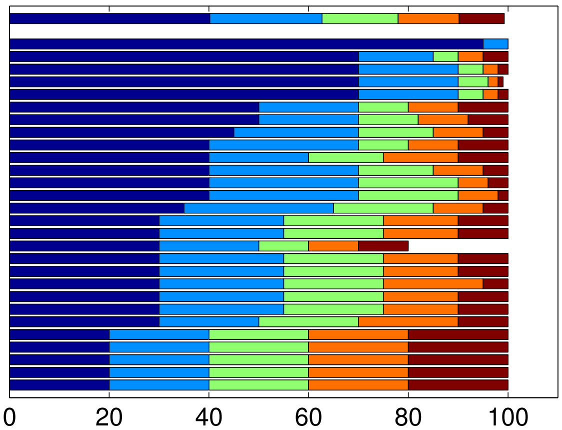


Fig. 3. The actual United States wealth distribution plotted against the estimated and ideal distributions of respondents of different income levels, political affiliations, and genders. Because of their small percentage share of total wealth, both the "4th 20%" value (0.2%) and the "Bottom 20%" value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

Your turn—estimates:



Your turn—ideal:



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

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

$$\text{where} \quad 0 < x_{\min} < x < x_{\max} \quad \text{and} \quad \gamma > 1.$$

- Exciting class exercise: sketch this function.
- $x_{\min}$ = 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.
- power-law decays in probability: The Statistics of Surprise.

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 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.

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

$$\text{where} \quad k_{\min} \le k \le k_{\max}$$

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

The statistics of surprise—words:

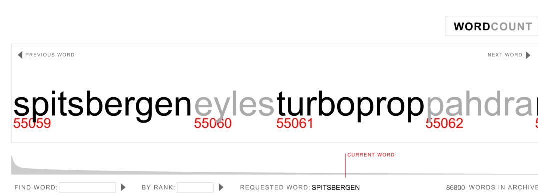
Brown Corpus (⊞) ($\sim 10^6$ words):

rank	word	% q
1.	the	6.8872
2.	of	3.5839
3.	and	2.8401
4.	to	2.5744
5.	a	2.2996
6.	in	2.1010
7.	that	1.0428
8.	is	0.9943
9.	was	0.9661
10.	he	0.9392
11.	for	0.9340
12.	it	0.8623
13.	with	0.7176
14.	as	0.7137
15.	his	0.6886

rank	word	% q
1945.	apply	0.0055
1946.	vital	0.0055
1947.	September	0.0055
1948.	review	0.0055
1949.	wage	0.0055
1950.	motor	0.0055
1951.	fifteen	0.0055
1952.	regarded	0.0055
1953.	draw	0.0055
1954.	wheel	0.0055
1955.	organized	0.0055
1956.	vision	0.0055
1957.	wild	0.0055
1958.	Palmer	0.0055
1959.	intensity	0.0055

Jonathan Harris's Wordcount: (⊞)

A word frequency distribution explorer:

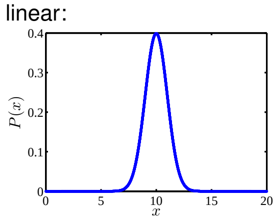


The statistics of surprise—words:

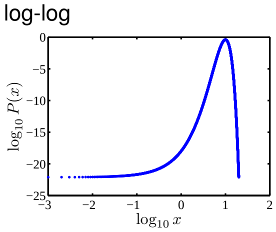
First—a Gaussian example:

$$P(x)\mathrm{d}x = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma}\mathrm{d}x$$

linear:



log-log

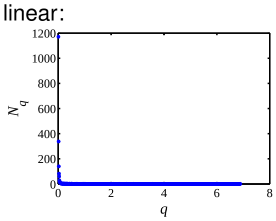


mean $\mu = 10$, variance $\sigma^2 = 1$.

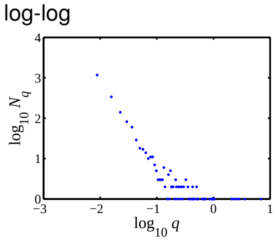
The statistics of surprise—words:

Raw 'probability' (binned) for Brown Corpus:

linear:



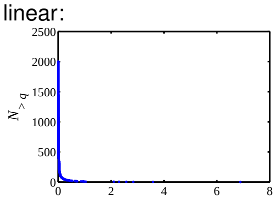
log-log



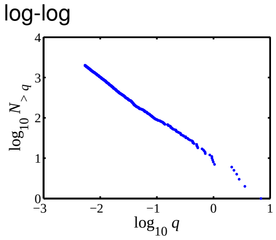
The statistics of surprise—words:

'Exceedance probability' $N_{>q}$:

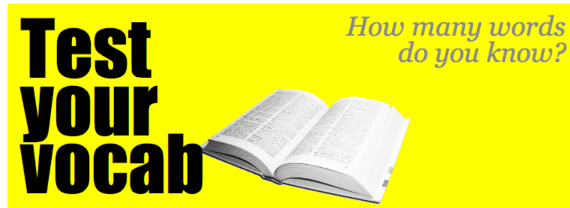
linear:



log-log



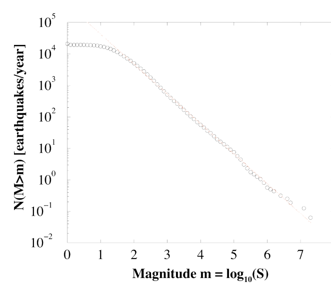
My, what big words you have...



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

The statistics of surprise:

Gutenberg-Richter law (⊞)



- 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" [4, 2]

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, . . . '

Well, that's just great:

Two things we have poor cognitive understanding of:

1. Probability
   - Ex. The Monty Hall Problem (⊞)
   - Ex. Daughter/Son born on Tuesday (⊞) (see asides; Wikipedia entry Boy or Girl Paradox (⊞)here).
2. Logarithmic scales.

On counting and logarithms:

- Listen to Radiolab's "Numbers." (⊞).
- Later: Benford's Law (⊞).

Size distributions:

Examples:

- Number of citations to papers:[9, 10] $P(k) \propto k^{-3}$.
- 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:[1] $P(F) \propto F^{-5/2}$. (see the Holtsmark distribution (⊞) and stable distributions (⊞))
- Diameter of moon craters:[7] $P(d) \propto d^{-3}$.
- Word frequency:[12] e.g., $P(k) \propto k^{-2.2}$ (variable)

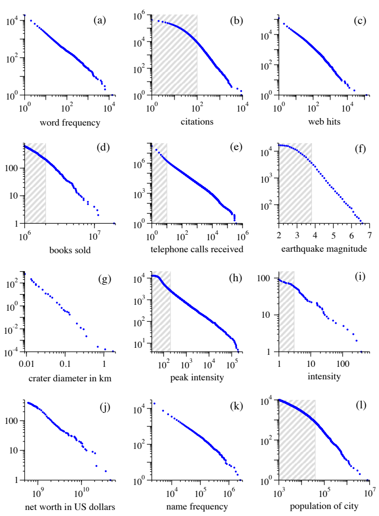
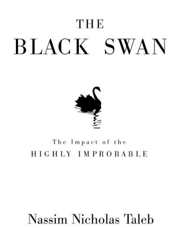


FIG. 4 Cumulative distributions or "rank/frequency plots" of twelve quantities reputed to follow power laws. The distributions were computed as described in Appendix A. Data in the shaded regions were excluded from the calculations of the exponents in Table I. Source references for the data are given in the text. (a) Numbers of occurrences of words in the novel *Moby Dick* by Hermann Melville. (b) Numbers of citations to scientific papers published in 1981, from time of publication until June 1997. (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 copies of bestselling books sold in the US between 1895 and 1965. (e) Number of calls received by AT&T telephone customers in the US for a single day. (f) Magnitude of earthquakes in California between January 1910 and May 1992. Magnitude is proportional to the logarithm of the maximum amplitude of the earthquake, and hence the distribution obeys a power law even though the horizontal axis is linear. (g) Diameter of craters on the moon. Vertical axis is measured per square kilometre. (h) Peak gamma-ray intensity of solar flares in counts per second, measured from Earth orbit between February 1980 and November 1989. (i) Intensity of wars from 1816 to 1980, measured as battle deaths per 10 000 of the population of the participating countries. (j) Aggregate net worth in dollars of the richest individuals in the US in October 2003. (k) Frequency of occurrence of family names in the US in the year 1990. (l) Populations of US cities in the year 2000.

power-law distributions

Gaussians versus power-law distributions:

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



- See "The Black Swan" by Nassim Taleb.[13]

Size distributions:

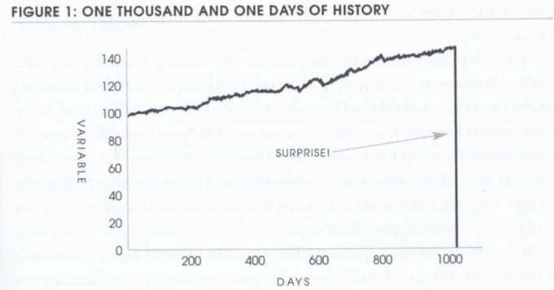
Examples:

- Earthquake magnitude (Gutenberg-Richter law (⊞)):[6, 2] $P(M) \propto M^{-2}$
- Number of war deaths:[11] $P(d) \propto d^{-1.8}$
- Sizes of forest fires[5]
- Sizes of cities:[12] $P(n) \propto n^{-2.1}$
- Number of links to and from websites[3]

- See in part Simon[12] and M.E.J. Newman[7] "Power laws, Pareto distributions and Zipf's law" for more.
- Note: Exponents range in error

Turkeys...

FIGURE 1: ONE THOUSAND AND ONE DAYS OF HISTORY



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"[13]

Taleb's table [13]

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

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 (⊞).

Power-Law Size Distributions
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf ⇔ CCDF
References
$P(x) \sim x^{-\gamma}$

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 1 (⊞)

Power-Law Size Distributions
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf ⇔ CCDF
References
$P(x) \sim x^{-\gamma}$

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 1 (⊞)

Power-Law Size Distributions
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf ⇔ CCDF
References
$P(x) \sim x^{-\gamma}$

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 (⊞)

Power-Law Size Distributions
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf ⇔ CCDF
References
$P(x) \sim x^{-\gamma}$

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

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

Insert question from assignment 2 (⊞)

Power-Law Size Distributions
Our Intuition
Definition
Examples
Wild vs. Mild
CCDFs
Zipf's law
Zipf ⇔ CCDF
References
$P(x) \sim x^{-\gamma}$

Complementary Cumulative Distribution Function:

CCDF:

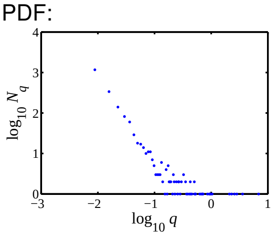
- $$P_{\geq}(x) = P(x' \geq x) = 1 - P(x' < x)$$
- $$= \int_{x'=x}^{\infty} P(x')\mathrm{d}x'$$
- $$\propto \int_{x'=x}^{\infty} (x')^{-\gamma}\mathrm{d}x'$$
- $$= \frac{1}{-\gamma+1}(x')^{-\gamma+1}\Big|_{x'=x}^{\infty}$$
- $$\propto x^{-\gamma+1}$$

Complementary Cumulative Distribution Function:

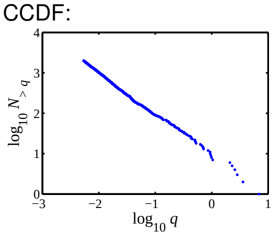
CCDF:

- $$P_{\geq}(x) \propto x^{-\gamma+1}$$
- Use when tail of P follows a power law.
- Increases exponent by one.
- Useful in cleaning up data.

PDF:



CCDF:



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.

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...

Zipfian rank-frequency plots

Zipf's way:

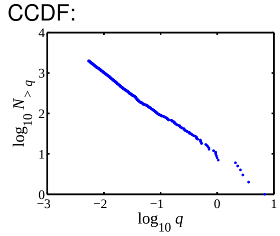
- Given a collection of entities, rank them by size, largest to smallest.
- x_r = the size of the rth ranked entity.
- $r = 1$ corresponds to the largest size.
- Example: x_1 could be the frequency of occurrence of the most common word in a text.
- Zipf's observation:

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

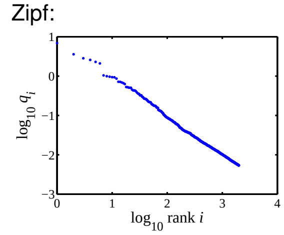
Size distributions:

Brown Corpus (1,015,945 words):

CCDF:



Zipf:

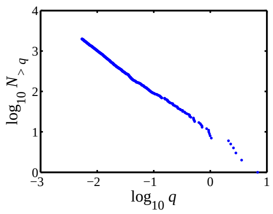


- The, of, and, to, a, ... = 'objects'
- 'Size' = word frequency
- Beep: (Important) CCDF and Zipf plots are related...

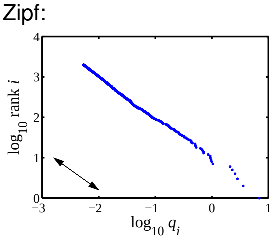
Size distributions:

Brown Corpus (1,015,945 words):

CCDF:



Zipf:



- 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

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

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

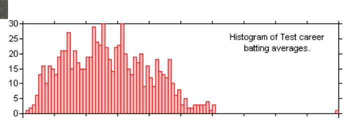
We therefore have $1 = (-\gamma+1)(-\alpha)$ or:

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

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

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 I

[1]

[2] P. Bak, K. Christensen, L. Danon, and T. Scanlon. Unified scaling law for earthquakes. Phys. Rev. Lett., 88:178501, 2002. pdf (⊞)

[3] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. Science, 286:509–511, 1999. 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 (⊞)

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

References II

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

[7] M. E. J. Newman. Power laws, pareto doistributions and zipf's law. Contemporary Physics, pages 323–351, 2005. pdf (⊞)

[8] 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 (⊞)

References III

[9] D. J. d. S. Price. Networks of scientific papers. Science, 149:510–515, 1965. pdf (⊞)

[10] D. J. d. S. Price. A general theory of bibliometric and other cumulative advantage processes. J. Amer. Soc. Inform. Sci., 27:292–306, 1976.

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

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

References IV

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

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

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