Chaos, Complexity, and Inference (36-462)

Lecture 13: Heavy-Tailed Distributions, Especially Power Laws

Cosma Shalizi

26 February 2008

Heavy Tailed Distributions, Especially Power Laws

Heavy tails The difference between light and heavy tails; some examples

Pure power laws: Pareto and Zipf distributions

Impure power laws

Further reading: Newman (2005) — but this was assigned, so you already read it, right?; Schroeder (1991) (fun); Arnold (1983) (reference); Resnick (2006) for the really ambitious

R files for these lectures:
`http://www.santafe.edu/~aaronc/powerlaws/`

Highly Skewed Distributions and Heavy Tails

Recall that the **skew** of a random variable X is

$$s = \frac{\mathbf{E}\left[(X - \mathbf{E}[X])^3\right]}{(\mathrm{Var}[X])^{3/2}}$$

Distributions with $s = 0$ are symmetric
think about positive s
"long thin tail to the right"
much more probability mass at extreme values than one would expect from a Gaussian or exponential

Survival Function/Upper CDF

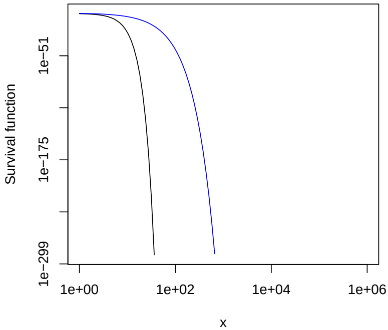
Usual or lower CDF

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

Upper CDF or **survival function**

$$F^{\uparrow}(x) = F^{+}(x) = \Pr(X \geq x) = \int_{x}^{\infty} f(y)dy$$

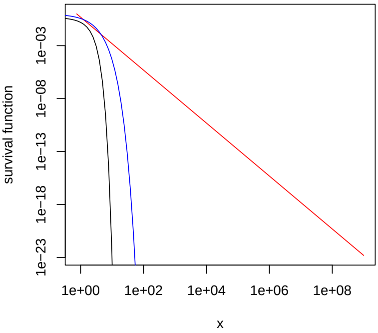
What You Are Used To



standard Gaussian (black) and standard exponential (blue); log-log scale

With extremely high probability, all observations fall within some bounded typical range

Suppose the tails decay slower than exponential



Red = Pareto distribution with $x_{\min} = 0.75$, $\alpha = 3.5$

Red distribution has mean = variance, just like exponential (= 1.25)

much higher probability of being very far from mean

Heavy Tails

In a loose sense, **heavy tailed** means slower-than-exponential decay of the survival function

In a stricter sense, it means that for some $a > 1$,

$$F^{\uparrow}(x) = O(x^{-a+1})$$

or

$$f(x) = O(x^{-a})$$

Heavy tails $\Rightarrow$ high probability of very large values

Heavy tails $\not\Rightarrow$ high mean, high variance, etc.

Light-tailed Well-described by mean (or median) and variance, typical observations within a few standard deviations of the mean

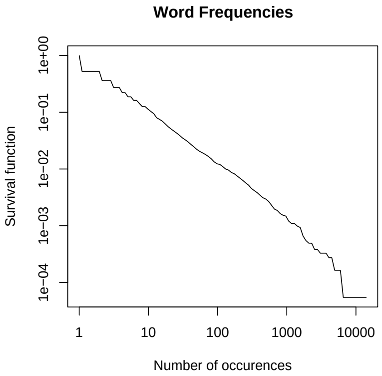
Heavy-tailed If mean and variance exist, not necessarily representative, lots of probability mass far from the mean

Chebyshev inequality:

$$\Pr\left(|X - \mathbf{E}[X]| \geq \epsilon\right) \leq \frac{\mathrm{Var}[X]}{\epsilon^2}$$

this is *very slow*
Heavy-tailed in the strict sense: $\mathbf{E}[X^m]$ exists only if $m < a - 1$
assumes X can get arbitrarily large, but so does the Gaussian!
Let's look at some examples — for data sources see Clauset *et al.* (2009).

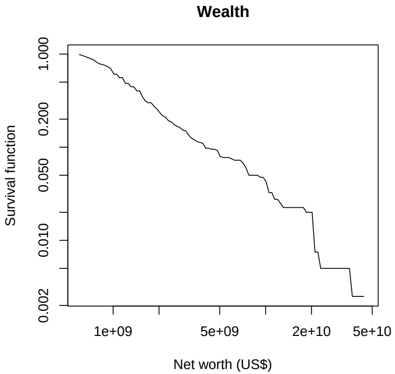
Word Frequencies — Zipf's Law



$a \approx 2$ (*Moby Dick* — but this is typical)

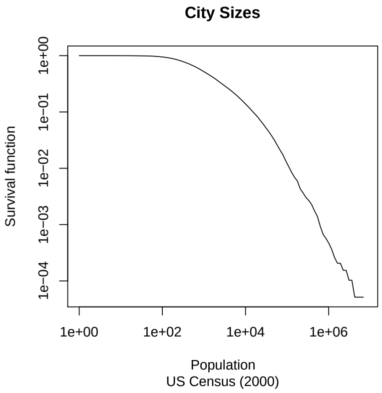
the (14086), of (6414), and (6260), a (4573), to (4484), in (4040), that (2917), his (2483), it (2374), i (1942), ...

Net Worth — Pareto's Law



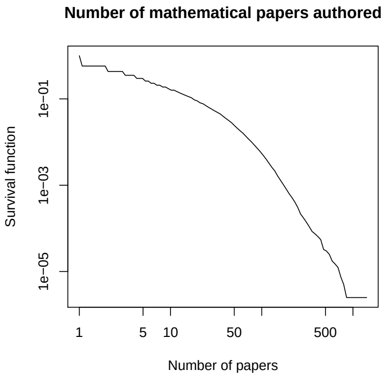
US, 2003, richest 400 — other countries, times, income, . . . similar

Sizes of Cities — Zipf's Law



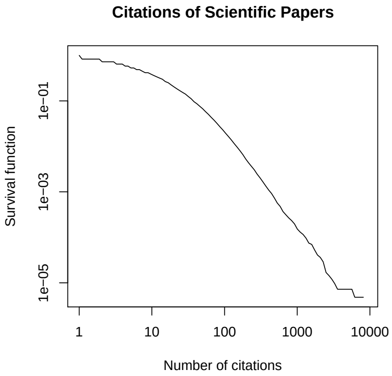
$a \approx 2$ Similarly for other countries and times

Papers per Author — Lotka's Law



Papers authored or co-authored, listed in American Mathematical Society's MathSciNet database

Citations of Scientific Papers — Price's Law

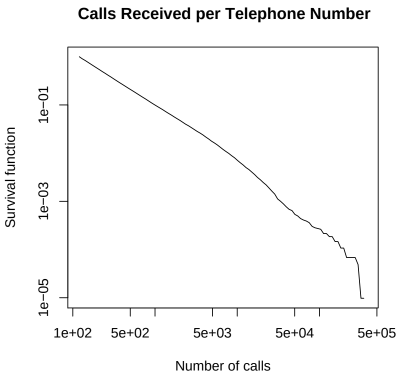


1981–1997, Science Citation Index

omitting 368110 papers, out of 6716198, never cited

Sales of U.S. Bestselling Books

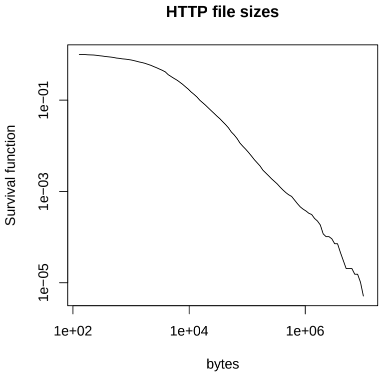
Calls Received by Telephone Numbers



Number of calls received in one day by AT&T customers

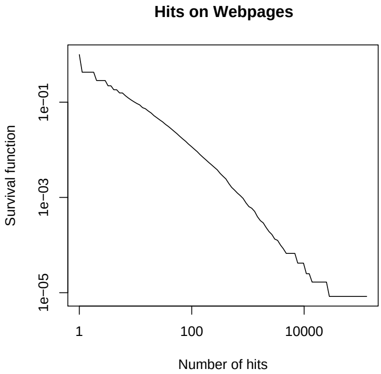
Only showing part ≥ 120 so it doesn't take forever to plot

HTTP File Sizes



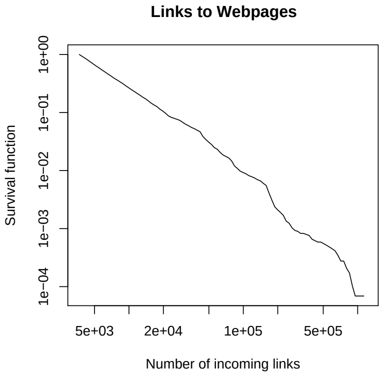
Bytes received, one day in 1996, one lab

Web Downloads



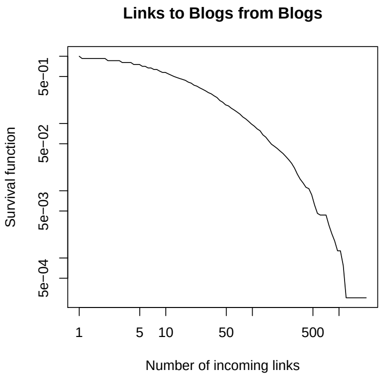
Number of downloads of given URLs by AOL users, one day in 1999

Web Links



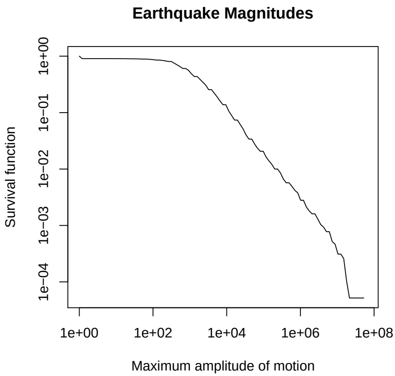
Incoming links to 2×10^8 web-pages in 1997 (only those ≥ 3680)

Blog Links



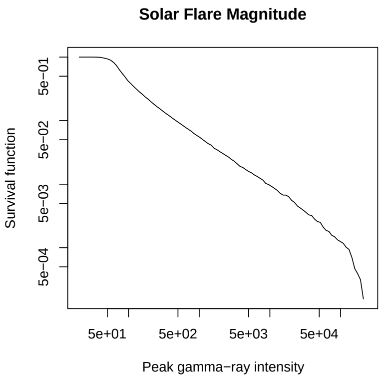
Incoming links to weblogs, late 2003 (Farrell and Drezner, 2008)

Earthquakes — Gutenberg-Richter Law



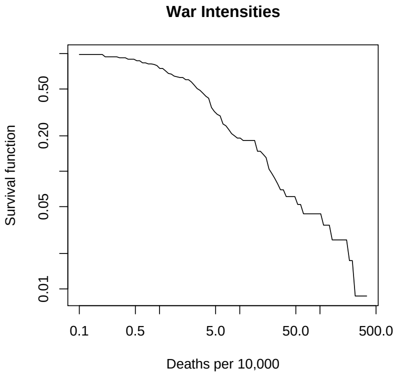
Maximum amplitude, California, 1910–1992

Solar Flares



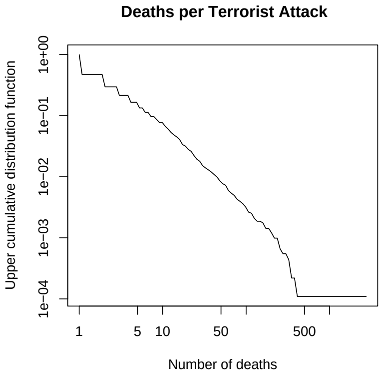
Peak γ-ray intensity, 1980–1989

Wars — Richardson's Law



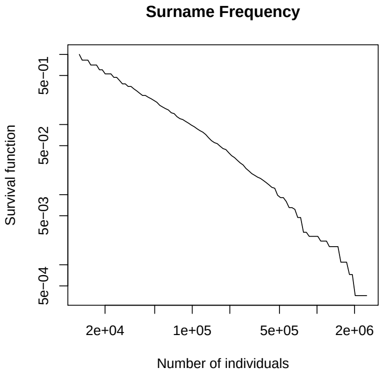
Deaths per 10^4 population, 1816–1980

Terrorism



Total deaths per incident, 1968–2006

Surnames



Per 1990 US Census

Pure Power Laws — Pareto and Zipf

Pareto Distribution Continuous x, two parameters, $x_{\min}$ and α, range $[x_{\min}, \infty)$

$$
\begin{aligned}
f(x) &= (\alpha-1)x_{\min}^{\alpha-1}x^{-\alpha} = \frac{\alpha-1}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\alpha} \\
F^{\uparrow}(x) &= \left(\frac{x}{x_{\min}}\right)^{-(\alpha-1)}
\end{aligned}
$$

Zipf Distribution Discrete x, again $x_{\min}$ and α

$$
\begin{aligned}
p(x) &= \frac{x^{-\alpha}}{\zeta(\alpha, x_{\min})} \\
\zeta(\alpha, x_{\min}) &\equiv \sum_{k=x_{\min}}^{\infty} k^{-\alpha}
\end{aligned}
$$

this is **Hurwicz zeta function**

Variants — Pareto II and Yule

Pareto II

$$f(x) \propto \left(1 + \frac{x-m}{s}\right)^{-\alpha}$$

or $1 + (X-m)/s \sim \mathrm{Pareto}(\alpha, 1)$

Zipf II As Pareto II but for pmf

Yule/Yule-Simon discrete

$$p(x) = (\alpha - 1)\frac{\Gamma(x)\Gamma(\alpha)}{\Gamma(x+\alpha)}$$

mean $(\alpha-1)/(\alpha-2)$, mean square $(\alpha-1)^2/[(\alpha-2)(\alpha-3)]$, etc.

In all these cases, the density/pmf is $\propto x^{-\alpha}$ for very large x

Some Properties of Power Law Distributions

(stick with Pareto for simplicity)
Log-log plots are linear:

$$\begin{aligned} f(x) &= Cx^{-\alpha} \\ \log f(x) &= \log C - \alpha \log x \end{aligned}$$

Moments:

$$\mathbf{E}\left[X^m\right] = \frac{\alpha - 1}{\alpha - 1 - m} x_{\min}^m$$

EXERCISE: Calculate the variance. What does your answer mean when $\alpha < 3$?

Sample maximum:

$$
\begin{aligned}
X_{(n)} &\equiv \max_{i=1,\ldots n} X_i \\
f_{(n)}(x) &= n\frac{\alpha-1}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\alpha}\left(1-\left(\frac{x}{x_{\min}}\right)^{-(\alpha-1)}\right)^{n-1} \\
\mathbf{E}\left[X_{(n)}\right] &= n x_{\min} B(n, \frac{\alpha-2}{\alpha-1}) \\
&= n x_{\min}\frac{\Gamma(n)\Gamma(\frac{\alpha-2}{\alpha-1})}{\Gamma(n+\frac{\alpha-2}{\alpha-1})} \\
&\approx x_{\min} n^{1/(a-1)}
\end{aligned}
$$

note: rises *rapidly* forever

Scale-Free

Fix $x_0 > x_{\min}$, pick any $x \geq x_0$

$$\begin{aligned}
\Pr\left(X \geq x | X \geq x_0\right) &= \frac{\Pr\left(X \geq x,\ X \geq x_0\right)}{\Pr\left(X \geq x_0\right)} \\
&= \frac{\Pr\left(X \geq x\right)}{\Pr\left(X \geq x_0\right)} \\
&= \frac{\left(\frac{x}{x_{\min}}\right)^{-(\alpha-1)}}{\left(\frac{x_0}{x_{\min}}\right)^{-(\alpha-1)}} = \left(\frac{x}{x_0}\right)^{-(\alpha-1)}
\end{aligned}$$

The conditional distribution looks just like the marginal distribution, only with x_0 in place of $x_{\min}$
In a sense there is no natural "scale" to the distribution
(In another sense: most samples are close to the minimum)

The 80/20 Rule

"80% of the things have 20% of the stuff"
Larger half of population must have at least half the stuff — how much more?
Median = $\tilde{x} = 2^{1/(\alpha-1)} x_{\min}$
EXERCISE: Prove this

$$\frac{\int_{\tilde{x}}^{\infty} x f(x) dx}{\int_{x_{\min}}^{\infty} x f(x) dx} = 2^{-(\alpha-2)/(\alpha-1)}$$

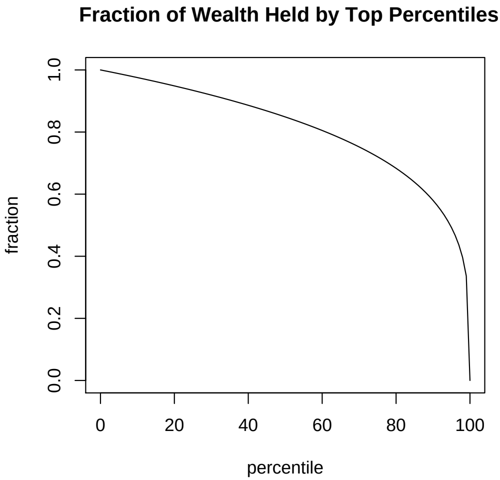
EXERCISE: Prove this, too
Nothing special about $1/2$; the top P fraction holds a fraction W of the stuff,

$$W = P^{(\alpha-2)/(\alpha-1)}$$

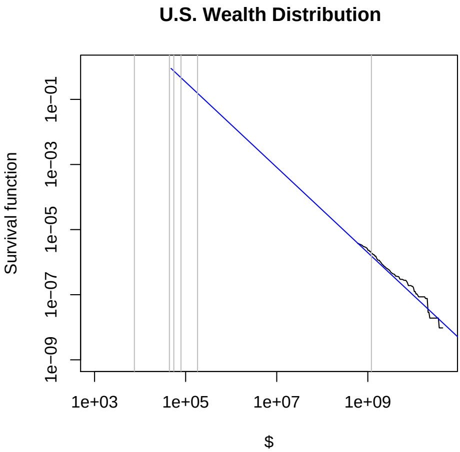
so the literal 80/20 rule means $\alpha = 2.16$

Application of 80/20 Rule and Scale-Freedom: Inequality

α for US wealth ≈ 2.3 (maximum likelihood estimate)
Median household net worth in 2000, \$ 5.5×10^4
`http://www.census.gov/prod/2003pubs/p70-88.pdf`
7.9×10^4 for white households, 7.5×10^3 for black
$\approx 10^8$ households in US
Total household net worth $\approx 1.9 \times 10^{13}$
400 richest Americans (2003), smallest net worth 6.0×10^8, total net worth 9.5×10^{11}
Or: top 4.0×10^{-6} holds 5.0×10^{-2} of the wealth
So $\alpha = 2.31 = \mathrm{MLE}$
Also: $x_{\mathrm{min}} = 4.4 \times 10^4$, $\tilde{x} = 7.4 \times 10^4$ — pretty good
inaccuracies: household vs. individual, 2000 vs. 2003, error in total *because* there are a small number of really rich people



Fraction of wealth held by top percentiles, assuming $\alpha = 2.31$



Black: data on 400 richest Americans; blue: extrapolated Pareto distribution

Grey (L to R): median black household worth; Pareto $x_{\min}$; median; white median; mean; Richard Mellon Scaife

Scale-free: from any point, looking right, the curve looks the same!
Population-based surveys miss the tail completely
Middle class vs. upper-middle class vs. upper class vs. rich vs. really rich vs. really, really rich vs. . . .

> *If you're not inside, you are outside, OK? I'm not talking about some $400,000-a-year Wall Street stiff, flying first class and being comfortable. I'm talking about liquid. Rich enough to have your own jet. Rich enough not to waste time. Fifty, a hundred million dollars, Buddy.*
>
> *"Gordon Gekko" in* Wall Street

All market economies are highly unequal...
but some are more unequal than others
there have never been any rich, equal, non-market economies

Non-Power Laws

The pure power law is implausible

some moments are infinite

positive probability of an American richer than the US

So: distributions which are heavy-tailed in the loose sense but not so badly behaved

Truncated power law

Stretched exponential/Weibull

Log-normal

Truncated Power Law/Power Law with Exponential Cut-Off

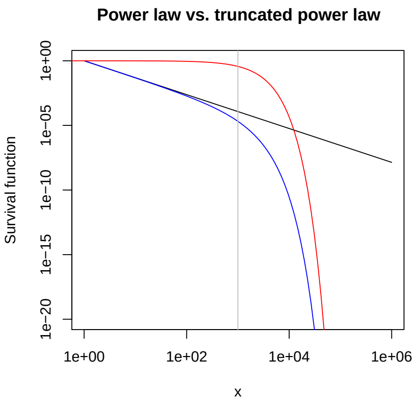
$$\begin{aligned} f(x) &\propto x^{-\alpha} e^{-\lambda x} \\ &= \frac{\lambda}{\Gamma(1-\alpha, \lambda x_{\min})} (\lambda x)^{-\alpha} e^{-\lambda x} \end{aligned}$$

over-all dimension of λ, right for a density

Looks like a power law if $x_{\min} \leq x \ll \lambda^{-1}$

Looks like an exponential if $x \gg \lambda^{-1}$

λ^{-1} acts like upper limiting scale



black: Pareto, $x_{\min} = 1, \alpha = 2.31$

blue: truncated Pareto, $x_{\min} = 1, \alpha = 2.31, \lambda = 10^{-3}$

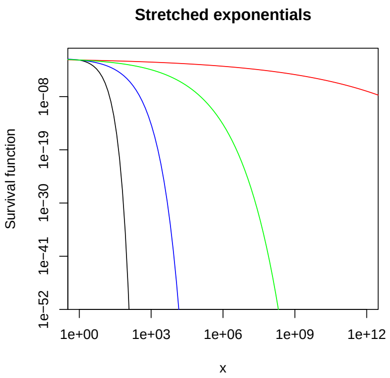
red: exponential, $\lambda = 10^{-3}$; grey: $1/\lambda$

Stretched Exponential or Weibull Distribution

$x^\beta \sim \mathrm{Exp}(\lambda)$, "stretched" if $\beta < 1$

$$f(x) = \beta\lambda x^{\beta-1} e^{-\lambda x^\beta}$$

$\lambda^{1/\beta}$ must have same units as x, so over-all units are x^{-1}
Typical scale of X is $\lambda^{-1/\beta}$



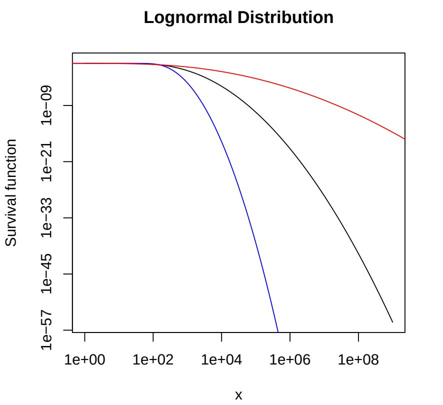
Stretched exponentials with $\lambda = 1$
black: $\beta = 1$ (ordinary exponential); blue: $\beta = 0.5$; green: $\beta = 0.25$; red: $\beta = 0.1$
note: more and more of a slope to the tail

Lognormal Distribution

$\ln X \sim \mathcal{N}(\mu, \sigma^2)$

$$f(x) = \frac{1}{x}\sqrt{\frac{1}{2\pi\sigma^2}}e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

$$\begin{aligned}
\ln f(x) &= -\ln x + C - \frac{(\ln x - \mu)^2}{2\sigma^2} \\
&= C - \frac{\mu^2}{2\sigma^2} + \left(\frac{\mu}{\sigma^2} - 1\right)\ln x - \frac{(\ln x)^2}{2\sigma^2} \\
&\approx \ln f(x_0) + \left[\frac{\mu}{\sigma^2} - 1 + \frac{\ln x_0}{\sigma^2}\right](\ln x - \ln x_0) + \text{h.o.t.}
\end{aligned}$$



all with $\mu = 5$; black: $\sigma^2 = 1$; blue: $\sigma^2 = 0.5$; red: $\sigma^2 = 2$

Coming attractions:

Generating mechanisms a.k.a. origin myths for heavy-tailed distributions

Estimation

Testing Comparing different heavy-tailed distributions, with more general morals

Arnold, Barry C. (1983). *Pareto Distributions*. Fairland, Maryland: International Cooperative Publishing House.

Clauset, Aaron, Cosma Rohilla Shalizi and M. E. J. Newman (2009). “Power-law distributions in empirical data.” *SIAM Review*, **forthcoming**. URL http://arxiv.org/abs/0706.1062.

Farrell, Henry and Daniel Drezner (2008). “The Power and Politics of Blogs.” *Public Choice*, **134**: 15–30. URL http://www.utsc.utoronto.ca/~farrell/blogpaperfinal.pdf.

Newman, M. E. J. (2005). “Power laws, Pareto distributions and Zipf’s law.” *Contemporary Physics*, **46**: 323–351. URL http://arxiv.org/abs/cond-mat/0412004.

Resnick, Sidney I. (2006). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer-Verlag.

Schroeder, Manfred (1991). *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. San Francisco: W. H. Freeman.