Chaos, Complexity, and Inference (36-462)

Lecture 14: How the Distributions Got Their Tails

Cosma Shalizi

26 February 2009

Where Do Heavy Tails Come From?

- Change of Variables — Some very boring explanations
- Growing by Multiplying — A somewhat boring explanation
- Critical Fluctuations — An exciting and mysterious explanation

More reading: Newman (2005); Mitzenmacher (2004); Schroeder (1991) for some fun examples I can't fit in here

Change of Variables: Take Logarithms

Suppose $X \sim \mathrm{Pareto}(\alpha, x_{\min})$
Define $Y \equiv \ln X/x_{\min}$

$$
\begin{aligned}
F^{\uparrow}(y) &= \Pr(Y \geq y) = \Pr(X \geq x_{\min} e^{y}) \\
&= \left(\frac{x_{\min} e^{y}}{x_{\min}}\right)^{-(\alpha-1)} = e^{-(\alpha-1)y} \\
Y &\sim \mathrm{Exp}(\alpha - 1)
\end{aligned}
$$

Conclusion: things only look heavy-tailed because you're measuring the exponential of what you should be measuring
Makes sense sometimes... but hard to get behind the idea of "log population" or "log money"

Change of Variables: Small Denominators

After Sornette (2002)
Let $X \sim \mathrm{Whatever}$
$Y \equiv X^{-1/a} \Rightarrow$

$$f_Y(y) = \alpha \frac{f_X(y^{-\alpha})}{y^{1+\alpha}}$$

If $f_X(x) \rightarrow c$ as $x \rightarrow 0$ then for large y

$$f_Y(y) = O(y^{1+\alpha})$$

Story: we measure the reciprocal of something which should be sensibly-distributed; flat distribution near zero gets turned into a heavy tail towards infinity

Mixtures of Exponentials

(Maguire *et al.*, 1952; Beck, 2005)
Exponential variables, with Γ-distributed rates:

$$\begin{aligned} X|\Lambda &\sim \mathrm{Exp}(\Lambda/s) \\ \Lambda &\sim \Gamma(\alpha, 1) \end{aligned}$$

What is the distribution of X?

$$\begin{aligned} \Pr\left(X \geq x\right) &= \int_0^\infty d\lambda\, \lambda^{\alpha-1} \frac{e^{-\lambda}}{\Gamma(\alpha)} \int_x^\infty dy\, \frac{\lambda}{s} e^{-\lambda y/s} \\ &= \int_0^\infty d\lambda\, \lambda^{\alpha-1} \frac{e^{-\lambda}}{\Gamma(\alpha)} e^{-\lambda x/s} \\ &= \int_0^\infty d\lambda\, \lambda^{\alpha-1} \frac{e^{-\lambda(1+x/s)}}{\Gamma(\alpha)} \\ &= \int_0^\infty d\mu (1+x/s)^{-1}\, \mu^{\alpha-1} (1+x/s)^{-(\alpha-1)} \frac{e^{-\mu}}{\Gamma(\alpha)} \\ &= (1+x/s)^{-\alpha} \end{aligned}$$

which is the “Pareto II” distribution

Multiplicative Growth: Lognormal

Recall central limit theorem: X_i all IID, $\mathbf{E}[X_i] = \mu$, $\mathrm{Var}[X_i] = \sigma^2$, then

$$\sum_{i=1}^{n} X_i \rightsquigarrow \mathcal{N}(n\mu, n\sigma^2)$$

Now let $Y_i = e^{X_i}$:

$$\prod_{i=1}^{n} Y_i \rightsquigarrow e^{\mathcal{N}(n\mu, n\sigma^2)}$$

The exp function is continuous

Issue: CLT is really

$$\frac{1}{n}\sum_{i=1}^{n} X_i \rightsquigarrow \mathcal{N}(\mu, \sigma^2/n)$$

so

$$\prod_{i=1}^{n} Y_i \rightsquigarrow e^{n\mathcal{N}(\mu,\sigma^2/n)+o(n)}$$

and $e^{o(n)}$ is not necessarily small!
Put a little differently, the center of the distribution will become log-normal much faster than the tails

Multiplicative Growth: Exponential Growth with Random Origins

Reed and Hughes (2002)
Imagine many piles
Each pile grows exponentially

$$X_i(t) = x_0 e^{\lambda(t - T_i)}$$

piles start growing at random times
= with constant probability per unit time

$$t - T_i \sim \mathrm{Exp}(\mu)$$

What is the distribution of pile sizes?

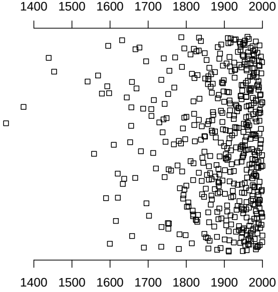
$$\begin{aligned}
\Pr(X_i \geq x) &= \Pr\left(e^{\lambda(t-T_i)} \geq x/x_0\right) = \Pr(\lambda(t-T_i) \geq \ln x/x_0) \\
&= \Pr\left(t - T_i \geq \frac{\ln x/x_0}{\lambda}\right) \\
&= e^{-\mu\frac{\ln x/x_0}{\lambda}} = \left(\frac{x}{x_0}\right)^{-\mu/\lambda} \\
X &\sim \mathrm{Pareto}(\mu/\lambda + 1, x_0)
\end{aligned}$$

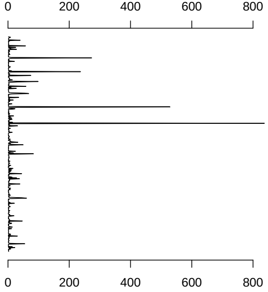
Still works if it's only *average* size that grows exponentially

Imagine doing this for cities; $\lambda = \mu = 1/100$

```
> t.start = 2000-rexp(500,rate=1/100)
> summary(t.start)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1327    1852    1927    1892    1969    2000
> plot.new()
> plot.window(xlim=c(min(t.start),2000),ylim=c(1,500),xlab="starting t
> points(t.start,1:500,pch=22)
> axis(3)
> axis(1)
> sizes.now = exp((1/100)*(2000-t.start))
> plot.new()
> plot.window(xlim=c(0,max(sizes.now)),ylim=c(1,500))
> lines(sizes.now,1:500)
> axis(1)
> axis(3)
> plot.survival.loglog(sizes.now,xlab="present size",ylab="survival fu
> curve(ppareto(x,1,2,lower.tail=FALSE),col="blue",add=TRUE)
```

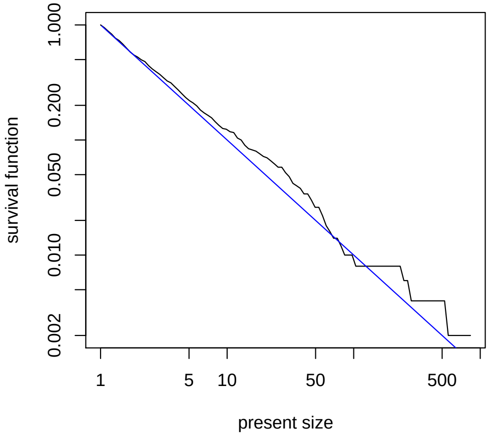
For US, min $T_i = 1327$ is not crazy, oldest city is Acoma, N.M., from 12th century — but otherwise?





starting times

sizes at $t = 2000$


survival function
1.000
0.200
0.050
0.010
0.002
1
5
10
50
500
present size

Problems: Acoma is not the largest city in the US; largest city is 8×10^6 larger than smallest, not 8×10^2 larger
Another model: make $X(t)$ log-normal

$$\ln X(t)/x_0 \sim \mathcal{N}((\mu - \frac{\sigma^2}{2})(t - t_i), \sigma^2(t - t_i))$$

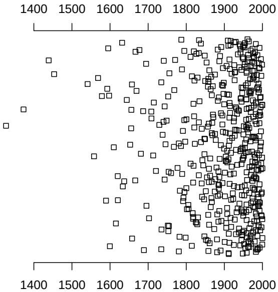
Then $\mathbf{E}[X(t)] = x_0 e^{\mu(t-t_i)}$
This comes from a simple multiplicative growth model, **geometric Brownian motion**

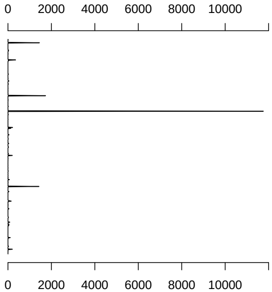
$$\frac{dX}{dt} = \mu X + \sigma X \xi$$

with $\xi =$ white noise

Unfortunately a real explanation needs stochastic calculus
set $\mu = \sigma^2 = 0.01$

```
> sizes.gbm = rlnorm(500,(0.01-0.005)*(2000-t.start),0.01*(2000-t.star
```

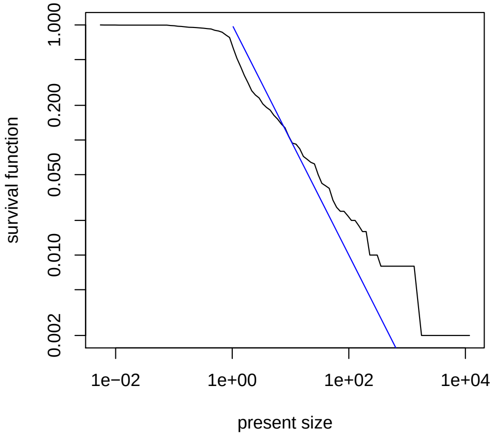




starting times

sizes at $t = 2000$

$\max X / \min X$ now 2×10^6, a bit small but in the right ballpark


survival function
1.000
0.200
0.050
0.010
0.002
1e−02
1e+00
1e+02
1e+04
present size

Yule-Simon Mechanism

Simon (1955); Ijiri and Simon (1977)
a.k.a. “the rich get richer”, “Matthew Effect”, “preferential attachment” ...
Again with the piles, but now discrete
One lump arrives each time step
Starts new pile with probability ρ
Otherwise joins an existing pile, probability of joining some pile of size k is $\propto k$

not necessarily equally likely to join *every* pile of the same size

What is the limiting distribution of pile sizes?
$N_k(t) =$ number of piles of size k, after t time-steps
Assume $N_k(t) \rightarrow p_k t$

Look at how the distribution can change
If $k \geq 2$,

$$\Pr\left(N_k(t+1) = N_k(t)+1\right) = (1-\rho)(k-1)\frac{N_{k-1}(t)}{t}$$
$$\Pr\left(N_k(t+1) = N_k(t)-1\right) = (1-\rho)k\frac{N_k(t)}{t}$$
$$\mathbf{E}\left[N_k(t+1)\right] - N_k(t) = (1-\rho)\frac{(k-1)N_{k-1}(t) - kN_k(t)}{t}$$

As $t \rightarrow \infty$, we want $N_k(t) \rightarrow p_k t$

$$
\begin{aligned}
p_k(t+1) - p_k t &= (1-\rho)\frac{(k-1)p_{k-1}t - kp_k t}{t} \\
p_k &= (1-\rho)\left((k-1)p_{k-1} - kp_k\right) \\
p_k(1+(1-\rho)k) &= (1-\rho)(k-1)p_{k-1} \\
\frac{p_k}{p_{k-1}} &= \frac{(1-\rho)(k-1)}{1+(1-\rho)k}
\end{aligned}
$$

Define $\alpha = 1/(1-\rho)$

$$\begin{aligned}
\frac{p_k}{p_{k-1}} &= \frac{k-1}{\alpha+k} \\
p_k &= \frac{k-1}{\alpha+k} p_{k-1} \\
&= \frac{k-1}{\alpha+k}\frac{k-2}{\alpha+k-1} p_{k-2} \\
&= \frac{(k-1)(k-2)\ldots 2\cdot 1}{(\alpha+k)(\alpha+k-1)\cdot(\alpha+1)} p_1 \\
&= \frac{\Gamma(k)\Gamma(\alpha+1)}{\Gamma(k+\alpha+1)} p_1 = B(k,\alpha+1) p_1
\end{aligned}$$

Using normalization,

$$p_k = \alpha B(k, \alpha+1) = O(k^{-\alpha+1})$$

Why physicists expect Gaussian fluctuations around equilibrium

Probability of macroscopic variables M having value m (Einstein fluctuation formula):

$$\Pr(M = m) \propto e^{S(m)}$$

Equilibrium m^* = state of *maximum* entropy, so $\partial S / \partial m = 0$ at m^*; Taylor expansion in the exponent:

$$\begin{aligned}\Pr(M = m^*) &\propto e^{S(m^*) + \frac{1}{2}\frac{\partial^2 S(m^*)}{\partial m^2}(m - m^*)^2 + \mathrm{h.o.t.}} \\ &\propto e^{\frac{1}{2}\frac{\partial^2 S(m^*)}{\partial m^2}(m - m^*)^2 + \mathrm{h.o.t.}}\end{aligned}$$

drop the $\mathrm{h.o.t.}$

$$M \sim \mathcal{N}(m^*, -\frac{\partial^2 S(m^*)}{\partial m^2})$$

What's really going on

correlations are short range
$\Rightarrow$ rapid approach to independence, exponential mixing
$\Rightarrow$ central limit theorem for averages over space (and time)
$\Rightarrow$ Gaussians

Phase Transitions

See Yeomans (1992) for nice introduction

Basically, bifurcations: behavior changes suddenly as temperature (or pressure or other control variable) crosses some threshold

First order: entropy is discontinuous at critical point

Examples: ice/water at 273K (and standard pressure); water/steam at 373K

order parameter is discontinuous

Second order: *derivative* of entropy is discontinuous

Example: "Curie point", permanent magnetization/not in iron 1043K

order parameter continuous but with sharp kink

like amplitude of limit cycle in period-doubling

Focus on continuous (second-order) case

Critical fluctuations

Entropy story breaks down because derivatives $\to \pm\infty$
Competition between two phases, no preference, one can pop up in the middle of the other
Fluctuations get arbitrarily large $\Rightarrow$ long-range correlations $\Rightarrow$ slow mixing (if any)
Assemblage becomes self-similar: magnify a small part and it looks just like the whole thing ("renormalization")
only strictly true for infinitely big assemblages
averaging must lead to a self-similar distribution
Power laws are self-similar (scale-free)
Conclusion: at critical point, expect to see power law distributions

Landau and Lifshitz (1980); Keizer (1987) are good on details but advanced

Theory of phase transitions / critical phenomena / order parameters / renormalization one of the key developments in physics over the last half century (Yeomans, 1992; Domb, 1996)
$\Rightarrow$ physicists think criticality is Very Cool
Criticality $\Rightarrow$ power law distributions
so physicists tend to think:

$$\text{(i) } \neg \text{ power laws} \Rightarrow \neg \text{ critical} \Rightarrow \text{Bored Now}$$
$$\text{(ii) power laws} \Rightarrow \text{critical} \rightarrow \text{Very Cool}$$

(ii) is called "the fallacy of affirming the consequent"

Self-Organized Criticality

See Miller and Page; papers: Bak *et al.* (1987); Carlson and Swindle (1995); Dickman *et al.* (2000); Bak (1996) if *heavily* salted

No externally set control parameter

Instead, external driving + interactions tend to keep the system towards a critical point

Turns out (Dickman *et al.*, 2000) that this is another version of the same story, only with the driving rate tuned very low

Morals

1. There are many ways to obtain heavy-tailed distributions, with or without power law tails
2. Some of these mechanisms make different predictions about the distributions
3. Even if they do not, they make different predictions about the dynamics
4. Both distributions and dynamics can be used to learn about mechanisms

Bak, Per (1996). *How Nature Works: The Science of Self-Organized Criticality*. New York: Copernicus.

Bak, Per, C. Tang and Kurt Wiesenfeld (1987). "Self-Organized Criticality: An explanation of 1/f noise." *Physical Review Letters*, **59**: 381–384.

Beck, Christian (2005). "Superstatistics: Recent developments and applications." In *Complexity, Metastability and Nonextensiviity* (Christian Beck and G. Benedek and A. Rapisarda and Constantino Tsallis, eds.). Singapore: World Scientific. URL `http://arxiv.org/abs/cond-mat/0502306`.

Carlson, Jean M. and G. H. Swindle (1995). "Self-organized criticality: Sandpiles, singularities, and scaling." *Proceedings of the National Academy of Sciences (USA)*, **92**: 6712–6719. URL `http://www.pnas.org/cgi/content/abstract/92/15/6712`.

Dickman, Ronald, Miguel A. Munoz, Alessandro Vespignani and Stefano Zapperi (2000). "Paths to Self-Organized Criticality." *Brazilian Journal of Physics*, **30**: 27–41. URL http://arxiv.org/abs/cond-mat/9910454. doi:10.1590/S0103-97332000000100004.

Domb, Cyril (1996). *The Critical Point: A Historical Introduction to the Modern Theory of Critical Phenomena*. London: Taylor and Francis.

Ijiri, Yuji and Herbert A. Simon (1977). *Skew Distributions and the Sizes of Business Firms*. Amsterdam: North-Holland. With Charles P. Bonini and Theodore A. van Wormer.

Keizer, Joel (1987). *Statistical Thermodynamics of Nonequilibrium Processes*. New York: Springer-Verlag.

Landau, L. D. and E. M. Lifshitz (1980). *Statistical Physics*. Oxford: Pergamon Press.

Maguire, B. A., E. S. Pearson and A. H. A. Wynn (1952). "The time intervals between industrial accidents." *Biometrika*, **39**: 168–180. URL http://www.jstor.org/pss/2332475.

Mitzenmacher, Michael (2004). "A brief history of generative models for power law and lognormal distributions." *Internet Mathematics*, **1**: 226–251. URL http://www.internetmathematics.org/volumes/1/2/pp226_251.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.

Reed, William J. and Barry D. Hughes (2002). "From Gene Families and Genera to Incomes and Internet File Sizes: Why Power Laws are so Common in Nature." *Physical Review E*, **66**: 067103.

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

Simon, Herbert A. (1955). "On a Class of Skew Distribution Functions." *Biometrika*, **42**: 425–440. URL http://www.jstor.org/pss/2333389.

Sornette, Didier (2002). "Mechanism for Powerlaws without Self-Organization." *International Journal of Modern Physics C*, **13**: 133–136. URL http://arxiv.org/abs/cond-mat/0110426.

Yeomans, Julia M. (1992). *Statistical Mechanics of Phase Transitions*. Oxford: Clarendon Press.