Lognormals and friends

Principles of Complex Systems
CSYS/MATH 300, Fall, 2011

Prof. Peter Dodds

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

Outline

Alternative distributions

There are other 'heavy-tailed' distributions:

1. The Log-normal distribution (⊞)

$$P(x) = \frac{1}{x\sqrt{2\pi}\sigma} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$$

2. Weibull distributions (⊞)

$$P(x)\mathrm{d}x = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{\mu-1} e^{-(x/\lambda)^{\mu}}\mathrm{d}x$$

CCDF = stretched exponential (⊞).

3. Gamma distributions (⊞), and more.

lognormals

The lognormal distribution:

$$P(x) = \frac{1}{x\sqrt{2\pi}\sigma} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$$

- $\ln x$ is distributed according to a normal distribution with mean μ and variance σ.
- Appears in economics and biology where growth increments are distributed normally.

lognormals

- Standard form reveals the mean μ and variance σ^2 of the underlying normal distribution:

$$P(x) = \frac{1}{x\sqrt{2\pi}\sigma} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$$

- For lognormals:

$$\mu_{\text{lognormal}} = e^{\mu+\frac{1}{2}\sigma^2}, \qquad \text{median}_{\text{lognormal}} = e^{\mu},$$

$$\sigma_{\text{lognormal}} = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}, \qquad \text{mode}_{\text{lognormal}} = e^{\mu-\sigma^2}.$$

- All moments of lognormals are finite.

Derivation from a normal distribution

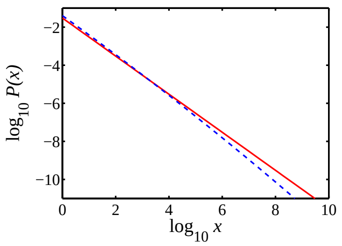
Take Y as distributed normally:

- $$P(y)\mathrm{d}y = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{d}y \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right)$$

Set $Y = \ln X$:

- Transform according to $P(x)\mathrm{d}x = P(y)\mathrm{d}y$:
- $$\frac{\mathrm{d}y}{\mathrm{d}x} = 1/x \Rightarrow \mathrm{d}y = \mathrm{d}x/x$$
- $$\Rightarrow P(x)\mathrm{d}x = \frac{1}{x\sqrt{2\pi}\sigma} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)\mathrm{d}x$$

Confusion between lognormals and pure power laws



Near agreement over four orders of magnitude!

- For lognormal (blue), $\mu = 0$ and $\sigma = 10$.
- For power law (red), $\gamma = 1$ and $c = 0.03$.

Confusion

What's happening:

- $$\ln P(x) = \ln \left\{ \frac{1}{x\sqrt{2\pi}\sigma} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2} \right) \right\}$$
- $$= -\ln x - \ln\sqrt{2\pi} - \frac{(\ln x - \mu)^2}{2\sigma^2}$$
- $$= -\frac{1}{2\sigma^2}(\ln x)^2 + \left(\frac{\mu}{\sigma^2} - 1\right) \ln x - \ln\sqrt{2\pi} - \frac{\mu^2}{2\sigma^2}.$$
- $\Rightarrow$ If $\sigma^2 \gg 1$ and μ,
- $$\boxed{\ln P(x) \sim -\ln x + \text{const.}}$$

Confusion

- Expect -1 scaling to hold until $(\ln x)^2$ term becomes significant compared to $(\ln x)$.
- This happens when (roughly)
- $$-\frac{1}{2\sigma^2}(\ln x)^2 \simeq 0.05 \left(\frac{\mu}{\sigma^2} - 1\right) \ln x$$
- $$\Rightarrow \log_{10} x \lesssim 0.05 \times 2(\sigma^2 - \mu) \log_{10} e$$
- $$\simeq 0.05(\sigma^2 - \mu)$$
- $\Rightarrow$ If you find a -1 exponent,
  you may have a lognormal distribution...

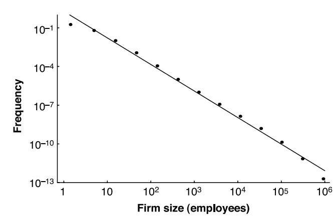
Generating lognormals:

Random multiplicative growth:

- $$x_{n+1} = r x_n$$
  where $r > 0$ is a random growth variable
- (Shrinkage is allowed)
- In log space, growth is by addition:
  $$\ln x_{n+1} = \ln r + \ln x_n$$
- $\Rightarrow \ln x_n$ is normally distributed
- $\Rightarrow x_n$ is lognormally distributed

Lognormals or power laws?

- Gibrat [2] (1931) uses preceding argument to explain lognormal distribution of firm sizes ($\gamma \simeq 1$).
- But Robert Axtell [1] (2001) shows a power law fits the data very well with $\gamma = 2$, not $\gamma = 1$ (!)
- Problem of data censusing (missing small firms).



Freq $\propto$ (size)$^{-\gamma}$
$\gamma \simeq 2$

- One mechanistic piece in Gibrat's model seems okay empirically: Growth rate r appears to be independent of firm size. [1].

An explanation

- Axtel (mis?)cites Malcai et al.'s (1999) argument [5] for why power laws appear with exponent $\gamma \simeq 1$
- The set up: N entities with size $x_i(t)$
- Generally:
  $$x_i(t+1) = r x_i(t)$$
  where r is drawn from some happy distribution
- Same as for lognormal but one extra piece.
- Each x_i cannot drop too low with respect to the other sizes:
  $$x_i(t+1) = \max(r x_i(t), c\,\langle x_i \rangle)$$

An explanation

Some math later... Insert question from assignment 6 (⊞)

- $$\text{Find} \quad P(x) \sim x^{-\gamma}$$
- where γ is implicitly given by

$$N = \frac{(\gamma-2)}{(\gamma-1)}\left[\frac{(c/N)^{\gamma-1}-1}{(c/N)^{\gamma-1}-(c/N)}\right]$$

  N = total number of firms.

- $$\text{Now, if } c/N \ll 1, \quad N = \frac{(\gamma-2)}{(\gamma-1)}\left[\frac{-1}{-(c/N)}\right]$$
- $$\text{Which gives} \quad \gamma \sim 1 + \frac{1}{1-c}$$
- Groovy... c small $\Rightarrow \gamma \simeq 2$

The second tweak

Ages of firms/people/... may not be the same

- Allow the number of updates for each size x_i to vary
- Example: $P(t)\mathrm{d}t = ae^{-at}\mathrm{d}t$ where t = age.
- Back to no bottom limit: each x_i follows a lognormal
- Sizes are distributed as [6]

$$P(x) = \int_{t=0}^{\infty} ae^{-at}\frac{1}{x\sqrt{2\pi t}}\exp\left(-\frac{(\ln x-\mu)^2}{2t}\right)\mathrm{d}t$$

  (Assume for this example that $\sigma \sim t$ and $\mu = \ln m$)

- Now averaging different lognormal distributions.

Averaging lognormals

- $$P(x) = \int_{t=0}^{\infty} ae^{-at}\frac{1}{x\sqrt{2\pi t}}\exp\left(-\frac{(\ln x/m)^2}{2t}\right)\mathrm{d}t$$
- Insert question from assignment 6 (⊞)
- Some enjoyable suffering leads to:

$$P(x) \propto x^{-1}e^{-\sqrt{2\lambda(\ln x/m)^2}}$$

The second tweak

- $$P(x) \propto x^{-1}e^{-\sqrt{2\lambda(\ln x/m)^2}}$$
- Depends on sign of $\ln x/m$, i.e., whether $x/m > 1$ or $x/m < 1$.
- $$P(x) \propto \begin{cases} x^{-1+\sqrt{2\lambda}} & \text{if } x/m < 1 \\ x^{-1-\sqrt{2\lambda}} & \text{if } x/m > 1 \end{cases}$$
- 'Break' in scaling (not uncommon)
- Double-Pareto distribution (⊞)
- First noticed by Montroll and Shlesinger [7, 8]
- Later: Huberman and Adamic [3, 4]: Number of pages per website

Summary of these exciting developments:

- Lognormals and power laws can be awfully similar
- Random Multiplicative Growth leads to lognormal distributions
- Enforcing a minimum size leads to a power law tail
- With no minimum size but a distribution of lifetimes, the double Pareto distribution appears
- Take-home message: Be careful out there...

References I

[1] R. Axtell.
Zipf distribution of U.S. firm sizes.
Science, 293(5536):1818–1820, 2001. pdf (⊞)

[2] R. Gibrat.
Les inégalités économiques.
Librairie du Recueil Sirey, Paris, France, 1931.

[3] B. A. Huberman and L. A. Adamic.
Evolutionary dynamics of the World Wide Web.
Technical report, Xerox Palo Alto Research Center, 1999.

[4] B. A. Huberman and L. A. Adamic.
The nature of markets in the World Wide Web.
Quarterly Journal of Economic Commerce, 1:5–12, 2000.

References II

[5] O. Malcai, O. Biham, and S. Solomon.
Power-law distributions and lévy-stable intermittent fluctuations in stochastic systems of many autocatalytic elements.
Phys. Rev. E, 60(2):1299–1303, 1999. pdf (⊞)

[6] M. Mitzenmacher.
A brief history of generative models for power law and lognormal distributions.
Internet Mathematics, 1:226–251, 2003. pdf (⊞)

[7] E. W. Montroll and M. W. Shlesinger.
On $1/f$ noise aned other distributions with long tails.
Proc. Natl. Acad. Sci., 79:3380–3383, 1982. pdf (⊞)

References III

Lognormals and friends

Lognormals
Empirical Confusability
Random Multiplicative Growth Model
Random Growth with Variable Lifespan
References

[8] E. W. Montroll and M. W. Shlesinger.
Maximum entropy formalism, fractals, scaling phenomena, and $1/f$ noise: a tale of tails.
J. Stat. Phys., 32:209–230, 1983.