

Lognormals and friends

Last updated: 2021/10/06, 23:35:55 EDT

Principles of Complex Systems, Vols. 1 & 2
CSYS/MATH 300 and 303, 2021-2022 | @pocsvox

Prof. Peter Sheridan Dodds | @peterdodds

Computational Story Lab | Vermont Complex Systems Center
Vermont Advanced Computing Core | University of Vermont



Licensed under the [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/).

PoCS
@pocsvox
Lognormals and friends

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

1 of 24

PoCS
@pocsvox
Lognormals and friends

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

2 of 24

PoCS
@pocsvox
Lognormals and friends

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

5 of 24

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.

PoCS
@pocsvox
Lognormals and friends

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

6 of 24

PoCS
@pocsvox
Lognormals and friends

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

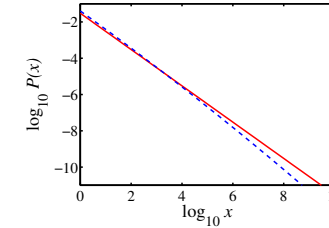
7 of 24

PoCS
@pocsvox
Lognormals and friends

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

8 of 24

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 $c = 0.03$.

PoCS
@pocsvox
Lognormals and friends

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

9 of 24

PoCS
@pocsvox
Lognormals and friends

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

10 of 24

PoCS
@pocsvox
Lognormals and friends

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

11 of 24

Outline

Lognormals

- Empirical Confusability
- Random Multiplicative Growth Model
- Random Growth with Variable Lifespan

References

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

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

CCDF = stretched exponential

3. Also: Gamma distribution, Erlang distribution, and more.

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:

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

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

- All moments of lognormals are finite.

Confusion

What's happening:

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

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

If the first term is relatively small,

$$\ln P(x) \sim -\left(1 - \frac{\mu}{\sigma^2}\right) \ln x + \text{const.} \Rightarrow \gamma = 1 - \frac{\mu}{\sigma^2}$$

Derivation from a normal distribution

Take Y as distributed normally:

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

Set $Y = \ln X$:

Transform according to $P(x)dx = P(y)dy$:

$$\frac{dy}{dx} = 1/x \Rightarrow dy = dx/x$$

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

Confusion

- If $\mu < 0$, $\gamma > 1$ which is totally cool.
- If $\mu > 0$, $\gamma < 1$, not so much.
- If $\sigma^2 \gg 1$ and μ ,

$$\ln P(x) \sim -\ln x + \text{const.}$$

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

Generating lognormals:

Random multiplicative growth:



$$x_{n+1} = rx_n$$

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

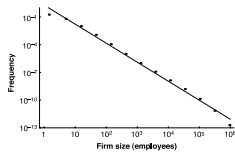
$\Rightarrow x_n$ is lognormally distributed

Lognormals or power laws?

Gibrat^[2] (1931) uses preceding argument to explain lognormal distribution of firm sizes ($\gamma \approx 1$).

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).



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

$$\gamma \approx 2$$

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

An explanation

Axtel cites Malcai et al.'s (1999) argument^[5] for why power laws appear with exponent $\gamma \approx 2$

The set up: N entities with size $x_i(t)$

Generally:

$$x_i(t+1) = rx_i(t)$$

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:

$$x_i(t+1) = \max(rx_i(t), c(x_i))$$

Some math later...

Insert question from assignment 7



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

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.



$$\text{Now, if } c/N \ll 1 \text{ and } \gamma > 2 \quad N = \frac{(\gamma - 2)}{(\gamma - 1)} \left[\frac{-1}{-(c/N)} \right]$$



$$\text{Which gives } \gamma \sim 1 + \frac{1}{1 - c}$$

Groovy... c small $\Rightarrow \gamma \approx 2$

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: $P(t)dt = ae^{-at}dt$ where t = age.

Back to no bottom limit: each x_i follows a lognormal

Sizes are distributed as^[6]

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

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

Now averaging different lognormal distributions.

Averaging lognormals



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

Insert fabulous calculation (team is spared).

Some enjoyable suffering leads to:

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

The second tweak



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

Depends on sign of $\ln \frac{x}{m}$, i.e., whether $\frac{x}{m} > 1$ or $\frac{x}{m} < 1$.



$$P(x) \propto \begin{cases} x^{-1+\sqrt{2\lambda}} & \text{if } \frac{x}{m} < 1 \\ x^{-1-\sqrt{2\lambda}} & \text{if } \frac{x}{m} > 1 \end{cases}$$

'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 and other distributions with long tails. [Proc. Natl. Acad. Sci.](#), 79:3380–3383, 1982. pdf ↗

References III

- [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.