Various Properties of the Log-Normal Distribution

Cal Abel

Abstract

The Log-Normal distribution is found in many places biological systems, income distributions, and even bitcoin transactions. This paper explores some basic properties of the Log-Normal distribution and provide some results of conducting analysis within this dynamical framework.

Introduction

Log-Normal distributions are found in many different fields of study: economics, metrology, biology, neuroscience, and engineering. While we are able to gain considerable understanding from fitting our observations to these distributions, we tend to miss what happens when parameters within these distributions change. This paper explores the consequences of examining these dynamical changes within these systems.

This paper is composed of three parts. The first part is a generic analysis–derivation of the dynamical relationships. The second part examines the relationship between the Log-Normal distribution and the Lorenz curve. The final part is an application of the Log-Normal distribution in policy analysis.

While the conclusions are profound, they are only derived properties from the Log-Normal density function. The findings are a consequence of the distribution, nothing more. The approach taken here was inspired by the work of J. Willard Gibbs and of Edwin T. Jaynes.[1, 2] Without their elementary insights, any understanding here would not be possible.

Part 1

We start with the univariate Log-Normal distribution,

$$f(x)=\frac{1}{x\sqrt{2\pi\sigma^2}}e^{-\frac{\left(\ln[x]-\xi\right)^2}{2\sigma^2}} \tag{1}$$

with an information entropy,

$$s=\frac{1}{2}\ln\left[2\pi e\sigma^2\right]+\xi \tag{2}$$

where,

$$\xi = \ln\left[\langle x \rangle\right] - \frac{1}{2}\sigma^2 \tag{3}$$

$$\langle x \rangle = E[x]$$

and,

$$\sigma^2 = \ln\left[1 + \frac{\mathrm{Var}[x]}{\langle x \rangle^2}\right] \tag{4}$$

Recall,

$$\frac{\mathrm{Var}[x]}{\langle x \rangle^2} \propto \frac{1}{N}$$

where N is the number of degrees of freedom of the system.[3] Let c_k be a positive Real constant of proportionality such that,

$$\frac{\mathrm{Var}[x]}{\langle x \rangle^2} = \frac{1}{c_k N} \tag{5}$$

Now, define a new variable,

$$k = 1 + \frac{1}{c_k N} \tag{6}$$

that is roughly inversely proportional to N over the domain $k \in (0, \infty)$. Equation (6) leaves us with the relationship,

$$\sigma^2 = \ln[k] \tag{7}$$

Substituting (3) and (7) into (2) results in,

$$s = \frac{1}{2}\ln[2\pi e] + \ln\left[\left(\frac{\ln[k]}{k}\right)^{\frac{1}{2}} \langle x \rangle\right] \tag{8}$$

The information entropy of a Log-Normal distribution is a separable function of the size of the system, k, and its expectation, $\langle x \rangle$. Because of this logical independence the size and expectation are additive quantities of entropy. Figure 1 plots k's contribution to entropy (a) and $\partial s / \partial k$ (b). We see that the system's entropy is maximized when $k = e$.

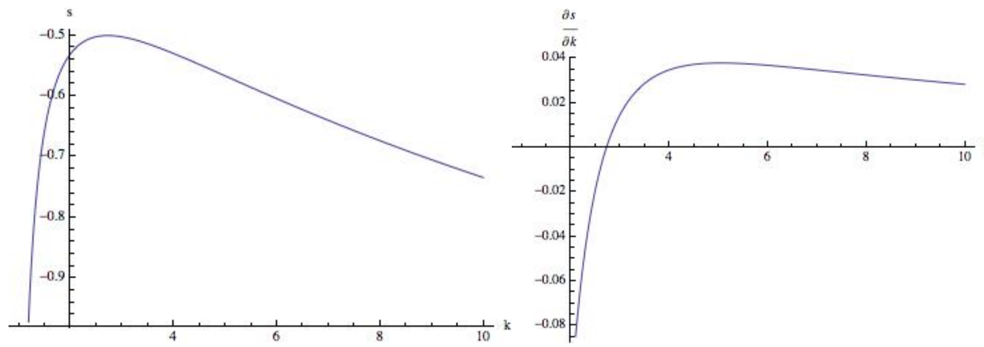


Figure 1 k's contribution to entropy (a) and $\partial s/\partial k$ (b).

We define a new variable,

$$z \equiv \sigma^2 e^{-\sigma^2} \tag{9}$$

and rearrange (8) in terms of $\langle x \rangle$,

$$\langle x \rangle = \frac{1}{\sqrt{2\pi e z}} e^s \tag{10}$$

$$\frac{\partial \langle x \rangle}{\partial s} = \langle x \rangle = T \tag{11}$$

$$\frac{\partial \langle x \rangle}{\partial z} = -\frac{T}{2z} \tag{12}$$

This results in a total differential of,

$$\mathrm{d}\langle x \rangle = \langle x \rangle \mathrm{d}s - \frac{\langle x \rangle}{2z} \mathrm{d}z \tag{13}$$

The plotting the scale parameter z against $c_k N$, Figure 2 shows $\min[z] = 1/e$ when $\sigma^2 = 1$. This corresponds to the point of maximum entropy of the distribution for any given location $\langle x \rangle$. We see that z has two regimes. In the first regime, $\sigma^2 > 1$, z increases as $c_k N$ becomes larger, σ^2 shrinking. In the second regime, $\sigma^2 < 1$, $z \propto 1/N$.

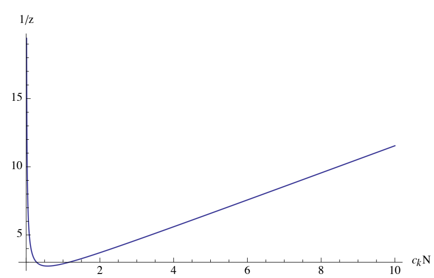


Figure 2 Relationship between $1/z$ and $c_k N$.

We now look at a special scenario with an additional random variable. This special bivariate case is where there is no covariance of the two random variables. The random variables are logically independent. The second constraint is that impact of the scale parameter of the system differs only by a constant of proportionality $\sigma_X^2 = a^2\sigma^2$ and $\sigma_M^2 = b^2\sigma^2$, where a and $b \in [0,1]$. Equations (8) and (10) become,

$$s = \frac{1}{2}\ln\left[2\pi e\right] + \ln\left[z^{\frac{1}{2}}\langle x\rangle^{a}\langle m\rangle^{-b}\right] \tag{14}$$

$$\langle x\rangle = \left(2\pi e z\right)^{-\frac{1}{2a}}\langle m\rangle^{-\frac{b}{a}} e^{\frac{s}{a}} \tag{15}$$

Expanding upon (11) and (12), we have,

$$T = \frac{\partial\langle x\rangle}{\partial s} = \frac{\langle x\rangle}{a} \tag{16}$$

$$\mu = \frac{\partial\langle x\rangle}{\partial z} = -\frac{T}{2z} \tag{17}$$

and,

$$\lambda = \frac{\partial\langle x\rangle}{\partial\langle m\rangle} = b\frac{T}{\langle m\rangle} \tag{18}$$

Equation (18) becomes the familiar,

$$\lambda\langle M\rangle = bNT \tag{19}$$

We see that for an isentropic process of constant N, that we derive the familiar polytropic process directly from (15),

$$\langle x\rangle\langle m\rangle^{\frac{b}{a}} = Const \tag{20}$$

We identify the adiabatic exponent as,

$$\gamma = \frac{b}{a} \tag{21}$$

The total differential of the system is given by

$$\mathrm{d}\langle x\rangle = T\,\mathrm{d}s - \lambda\,\mathrm{d}\langle m\rangle + \mu\,\mathrm{d}z \tag{22}$$

Because the system is at statistical equilibrium, we use the Gibbs-Duhem relationship to determine the relationship between a and b.

$$z\,\mathrm{d}\mu = -s\,\mathrm{d}T + \langle m\rangle\,\mathrm{d}\lambda \tag{23}$$

Upon reflection of the definition of z, equation (9), we can have a deeper understanding of the significance of this relationship. We let $\sigma^2 = -W(-z)$, where W is the Lambert W function. Substituting this relationship into (9), we recover the Lambert W function definition:

$$z \equiv W(z)e^{W(z)} \tag{24}$$

When we plot z as a function of $c_k N$, figure 3, we see that there are two distinct regions, the first is the W_0 branch (red). The second is the other real branch of the Lambert W function, W_{-1}(blue).

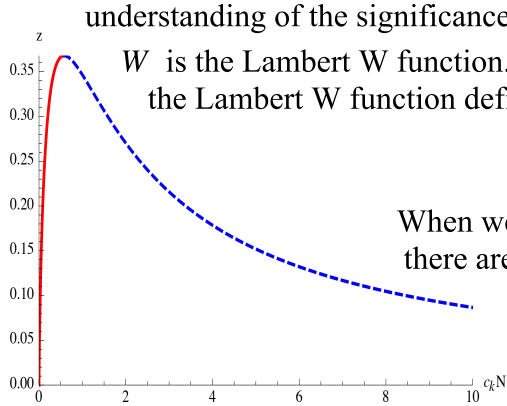


Figure 3 Plot of Lambert W function

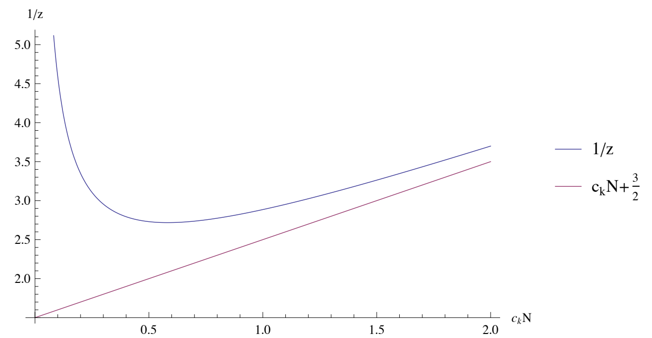


Figure 4 Scale function's approach to linearity

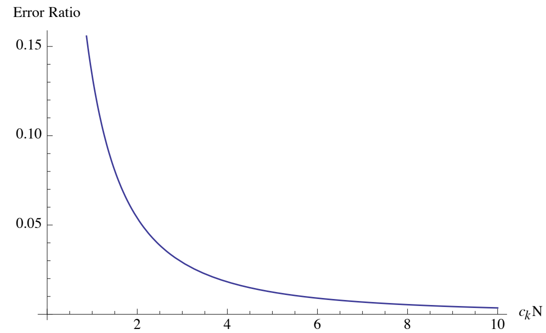


Figure 5 Error in linear model to $\mathbf{1/z}$ function

$\frac{1}{z(c_k N)}$ approaches $\frac{1}{z} = c_k N + \frac{3}{2}$ for large $c_k N$, figure 4, with a rapidly diminishing error, figure 5. The linear model provides a useful approximation for systems of high dimensionality.

Lorenz Curves

The Log-Normal distribution has important characteristics associated with it when evaluated using Lorenz curves. First is the Lorenz asymmetry coefficient is 1 for every Log-Normal distribution. This means that there are not too many small/poor or large/rich members of the study group. The large and small parts are symmetric in their contribution to the overall distribution.

When we look at the contribution of the factors to the Lorenz curve, we find that the shape of the Lorenz curve for a Log-Normal distribution is a function only of σ^2. Correspondingly, we find that by (17) the maximum entropy distribution occurs when $\sigma^2 = 1$. This correlates to a Gini coefficient of 0.521. Any Log-Normal distributed random variate that has a Gini coefficient other than 0.521 will result in lower entropy. In the extremes of σ^2 of 0 and ∞, the entropy goes to zero. When $\sigma^2 \to \infty$ the density of the system becomes so low that it possesses no information. When $\sigma^2 \to 0$ the system possesses no free energy it cools to a point where there is no action or motion in the system because its density is so high. This condition corresponds to a Gini coefficient of 0 and represents perfect "equality".

When we consider the condition of perfect equality in a dynamical situation we find that there is no free energy in the system. We can rearrange (22) and apply the product rule to derive the relationship,[4]

$$\mathrm{d}A = -S\,\mathrm{d}T - \lambda\,\mathrm{d}\langle M\rangle + \mu N\,\mathrm{d}z \tag{25}$$

Clearly, as $z \to 0$ the free energy of the system disappears. This means that the system has no available wealth to act, through being either too diffuse or too dense. Additionally, we find that the wealth of the system becomes maximal when the entropy is maximized. Returning to equations (6) and (7), we find that the constant c_k determines the **maximum entropic carrying capacity** of the system. In the context of a species distribution a low c_k would represent a biome that could support richer species diversity like a rain forest. Conversely, a relatively high c_k would support a smaller population size, like a desert. Rewriting (6) when $k = e$, we have

$$N\big|_{s_{\max}} = \frac{1}{c_k\left(e-1\right)} \tag{26}$$

For this reason, we will refer to c_k as the systems carrying capacity. The carrying capacity of the system is not necessarily a fixed constant through time. It can depend on other variables. Such treatment significantly complicates the mathematical analysis and is not presented here due to loss of clarity.

We choose to focus on the maximum entropy point of the system for several reasons: it assumes the least amount of information compared to other hypothesis, it is the most probable configuration given the information, distributions of lower entropy are atypical, that they have greater "disorder", or that they are "smoother".[5] I do not prefer to use "disorder" because of the connotations associated with it of a lack of structure. This is quite the contrary; a system at maximum entropy based on its constraints has a very clearly defined structure. The entropy is a measure then of how capable the systems components are "free" to explore the allowed configurations. In this sense, entropy provides a measure of "freedom" or "liberty".

Jaynes provides a justification for adopting the principle of maximum entropy in our analysis through the entropy concentration theorem.[5] For a system to not be at a configuration of maximal entropy, that system has to be perturbed by some "force" from that position of maximal entropy. When we identify such perturbations from what our theory says is maximum entropy, then those instances are where we need to reevaluate our hypothesis to construct a more effective model that explains the observed entropy. This method is how we grow our scientific knowledge. Science seeks to only understand what it can observe. Those ideas that increase our understanding of the world around us explain our observed entropy the best.

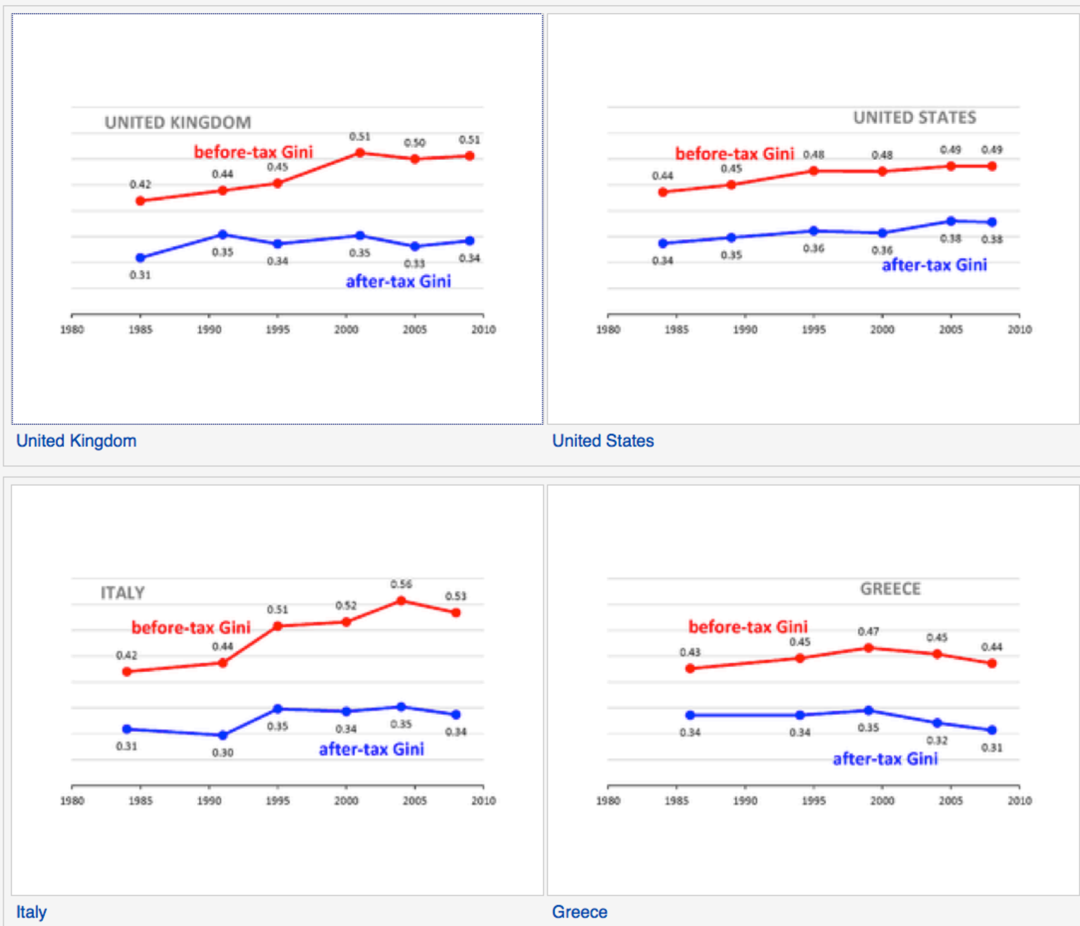
Policy Analysis

By assuming that our observed distribution is at a point of maximum entropy (where we assume the least), we provide a gauge by which to compare our model to observation. Turning our attention to examining income inequality. We recall that the greatest societal wealth occurs when systems entropy is maximized. Forcing the system away from a point of maximal entropy requires a "work" input, or rather information that is not included in the endogenous model. There is some exogenous factor that causes the perturbation from maximum entropy.

A significant perturbation for any society is taxes. Here there are two main types, transfer and services. It is possible to construct service taxes such that the systems entropy is not reduced. In this construct, we call this form of taxation isentropic (entropy remains unchanged). Using a thermodynamic analogy, isentropic taxation is like an isentropic expansion of a gas under a piston (actual relationship is mathematically analogous to a chemical process, but it is difficult to convey extracting free energy in a chemical reaction with a catalyst, here government is the catalyst). These processes extract wealth (Free energy in thermo) and convert it into what is hopefully considered to be useful work.

When we examine transfer/redistributive taxation, the tax code is specifically tailored to affect the distribution of income. In doing so, the tax code affects the system's entropy. If the system was already at a state of maximal entropy for a given average income, then any redistribution lowers the society's entropy and destroys wealth. Progressive taxation, compresses our society destroying entropy (liberty) and free energy (wealth). This inhibits the overall action that we can achieve. It also induces for lack of a better word social stress. Here the stress is in the society being pushed away from its natural state into some entirely arbitrary configuration. It is difficult to envision this as success as more are made worse off than those being made better off.

Gini indexes – before and after taxes between 1980 and 2010[6]



Taxes and social spending in most countries have significant moderating effect on income inequality Gini indices.

Figure 6 Impact of taxation on Gini coeffiicents

Figure 6 taken from Wikipedia's page on the Gini coefficient and based on OECD data[6] shows the impact of progressive tax policy. While the OECD data does not show that income is Log-Normally distributed, examining the Social Security Administration's

Average Wage Index from 1990 to 2011, we find that at least US income is Log-Normally distributed.

The path of making the poor better off and no one worse off is through increasing the real (not nominal) income of individuals. A collapsing average income in a society is a definite sign of a contracting economy (directly analogous to cooling, where wealth is leaving the nation). This is particularly troubling for Japan, which since 2010 has collapsing income while drastically increasing the money supply.[7, 8] The Japanese seemingly are doing everything in their power to reduce the entropy of their society. The collapsing entropy is indicative (unknown direction of causation) of the extremely low birthrates in Japan. This is consistent with others observations attributing low birth rates to "exorbitant living costs, elevated stress and diminished confidence. Even after two decades of deflation, prices in Japan for everything from rent to food to entertainment remain among the highest in the world. Economic stagnation and changes in labor laws have restrained wage growth and enabled companies to swap employees into low-paying part-time jobs with few benefits."[9] If Japan is to recover, they need to implement policies that will increase the societal entropy, not reduce it.

Conclusion

When we examine the Log-Normal distribution in closer detail, we find that we garner valuable information from the results. Perhaps the most important consequence of all of this exposition is that we cannot avoid the second law of thermodynamics. This is an iron law, and one that cannot be violated. Stated another way. Good intentions are not good enough to achieve favorable outcomes.

We can only hope to achieve favorable outcomes if we examine the change in distribution of the things that we measure. The information content of the distribution, entropy, and its trend is as important as the average value and more important than any other metric. As entropy is purely a statistical metric, it exists for anything that we observe. Thermodynamics is formally a study of how the distribution of particles in phase space change. It is perhaps the most powerful mathematical framework we have for any quantitative analysis.

Reflection

"If someone points out to you that your pet theory of the universe is in disagreement with Maxwell's equations—then so much the worse for Maxwell's equations. If it is found to be contradicted by observation—well these experimentalists do bungle things sometimes. But if your theory is found to be against the second law of thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation." -Sir Arthur Eddington [10]

References

1. Gibbs, J.W., *Elementary Principles in Statistical Mechanics Developed with Especial Reference to The Rational Foundation of Thermodynamics*. 1902, New York: Charles Scribner's Sons. 207.
2. Jaynes, E.T., *Probability Theory: The Logic of Science*. 1st ed, ed. G.L. Bretthorst. 2003: Cambridge University Press. 727.
3. Landau, L. and E. Lifshitz, *Statistical Physics, Part 1: Volume 5 (Course of Theoretical Physics, Volume 5)*. 1980, Butterworth-Heinemann, Oxford.
4. Saslow, W.M., *An Economic Analogy to Thermodynamics.* American Journal of Physics, 1999. **67**(12): p. 1239-1247.
5. Jaynes, E.T., *Concentration of Distributions at Entropy Maxima*, in *E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, R.D. Rosenkrantz, Editor. 1989, Kluwer Academic Publishers: Boston. p. 317-336.
6. OECD, *OECD.StatExtracts*. 2013.
7. Mayger, J. and M. Iwamoto, *Japan Salaries Extend Slide as Inflation Begins to Take Root* in *Bloomberg News*. 2013: http://www.bloomberg.com/news/2013-12-03/japan-salaries-extend-tumble-as-inflation-begins-to-take-root.html.
8. Ujkane, K. and A. Sharp, *Japan Salaries Extend Fall as Abe Urges Companies to Raise Wages*. 2013, Bloomberg News: http://www.bloomberg.com/news/2013-10-31/japan-salaries-extend-fall-as-abe-urges-companies-to-raise-wages.html.
9. Pesek, W., *The Lust Beneath Japan's Sex Drought*, in *Bloomberg*. 2013: http://www.bloomberg.com/news/2013-10-25/the-lust-beneath-japan-s-sex-drought-.html.
10. Eddington, A.S., *The Nature of the Physical World.* Gifford Lectures, 1927. **74**.