Explanation of Log-Normal Distributions and Power-Law Distributions in Biology and Social Science

Kai Sun*
Department of Physics, University of Illinois at Urbana-Champaign,
1110 W. Green St. , Urbana, IL 61801-3080, USA

(Dated: May 6, 2004)

In many biological and social systems, where the ideal of heritance and aberrance exist at the same time, normal distribution failed in describing the distributions of variables. In stead, many real data as well as simulations indicate that log-normal distribution and power-law distribution can characterize many properties of this kind of systems. In this paper, the failure of normal distribution is studied and a general phenomenological model including both heritance and aberrance is set up. Using Fokker-Plank equation, it turns out that, log-normal distribution will be favored if the system has an optimum value for the variable we are studying, but power-law distribution will be favored for discrete variables.

---

*Electronic address: `kaisun@uiuc.edu`

I. INTRODUCTION

Although started by physicists, many researches indicate emergent states can also be seen in many biology and social systems. One feature, which exists widely in many systems of this kind, is that the distribution of variables is away from normal distribution.

With the support of central limit theorem, normal distribution is believed to be a good approximation for any independent random variables. In ordinary systems, where interaction is weak, independent-variables is usually a good approximation. As a result, normal distribution is the most widely accepted descriptions of measured data. However, in system where emergent states of matter show up, the independence of variables may be destroyed by strong correlation and collective mode. In this case, normal distribution may break down.

In fact, people have noticed for very long time, many measured data gotten from biology and social science systems cannot be described by normal distribution. In stead, log-normal distribution seems to be favored by the nature[1]. During the last five years, many researches in complex network structure and models based on a network structure find that in network systems, power-law distributions can often be observed[2–5].

In section II, basic phenomena will be introduced. Then in the section III, I will discuss basic ideal about random process and Fokker-Plank equation. Using this knowledge, we will build a phenomenological model in section IV. And we will use this model to explain log-normal distribution and power-law distribution observed in many biological and social systems.

II. PHENOMENA

A log-normal distribution is defined as:

$$\rho(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \ln x_0)^2}{2\sigma^2}} \tag{2.1}$$

For biologist, this distribution is anything but a new discovery. Actually, one of the first observations is reported in 1914 [6] in studying of inheritance of fruit size. During the last 90 years, log-normal distributions have been observed in many systems, including, ecology, human medicine, environment, linguistics, *et al.* Here I list out some example:[1]

- Concentration of elements and their radiation in the Earth's crust[7–9].
- Latent periods of many infectious diseases [10];
- Survival time after cancer diagnosis [11–13];
- Abundance of species in a plant or large communities [14–16];
- Number of words per sentence for writers. [17];
- Age of marriage for women in Denmark [18];
- Family size in England and Wales [19];
- Size of crystals in ice cream [20];

- Size of oil drops in mayonnaise [20];
- Size of holes in cocoa press cake [20].

If we said that log-normal distribution is an old distribution, since it has been observed for such a long time, then power-law distribution should be called a new focus of recent researches in complex system. And this distribution is often, if not all, related with network structure. The definition of power-law distribution is:

$$\rho(x) = Ax^{-\gamma} \tag{2.2}$$

where A and γ are constant. γ is usually called exponent and in most cases, it is a positive number.

It is very obvious that this distribution will diverse at $x = 0$ if $\gamma > 1$, so usually a *cutoff* for small x is needed. But for discrete variables ($x = 1, 2, 3 \ldots$), things become much easier, since we don't need worried about the diverse and cutoff any more in this situation. This is just the case of network model. Here, network help us again, just as it did in QCD, where diverse can be regulated by studying lattice QCD. I believe this is part of the reason why power-law distribution is often linked with network models.

Barabasi and his co-workers have found power-law distributions in many networks[2]. The most famous example may be the distribution of the number of links on web page. The measured exponent for outgoing links is 2.45 and exponent for incoming link is 2.1. Similar studies have also been done to the power-grid of western US, and the exponent here is 4. Same stories also work for the movie actor network of Hollywood with an exponent 2.3. The family name networks in Japan also show a power-law distribution with an exponent 0.65[21]. People have tried many theories to explain this phenomenon, and the most famous example is the model of scale free network, with the rule of "rich getting richer". I will compare this model with the model I raised in section IV at latter time.

All these examples, I just mentioned about power-law distribution, are the distributions of quantities reflecting the structure of a network. But this is only part of the whole story. Power-law dynamics are also observed in many dynamical models based on network structure. For example, the social influence model[4]. This model is studied by Castellano and his coworkers with computer simulations in 2000. It has two basic assumptions:

1. One society has a better chance to communicate with the others who share more identical attributes with it.
2. The more communications exist between two societies, the more common cultures they are going to share.

They studied this model in a 2-D square lattice. Although this model only has these two simple disciplines, the simulation shows very rich dynamical behaviors. A phase transition from uniform phase (where almost all societies have one common culture) to random phase (where each society intents to have its own unique culture) will be observed, if one increases the initial culture diversity of the system. They also observed a power-law distribution of the size of culture clusters (number of societies sharing the same culture) at critical point.

But my coworkers and I studied this model in detail[5], and two new features are founded.

1. The phase transition can also be seen in a small-work network and random network. And a phase transition from random phase to uniform phase can be seen if we increase the randomness of the network (or say, transfer the network from a regular one to a small-world one.)

2. We observed the power-law distribution under almost any control parameters, instead of only seen at critical point and they all have similar exponent. This means that the power-law is not a critical phenomenon.

Another example of power-law distributions observed in dynamical models based on a network structure is the size distribution of business companies[3], if we considered the business relations between companies as a complex network.

III. BASIC IDEAL ABOUT RANDOM PROCESS AND FOKKER-PLANK EQUATION

Usually, a random process can be expressed as:

$$\frac{\mathrm{d}\vec{x}(t)}{\mathrm{d}t} = \vec{f}(\vec{x},t) + \vec{g}(\vec{x},t)\Gamma(t) \tag{3.1}$$

where t is time; $\Gamma(t)$ is a random function (or called noise); $\vec{f}(\vec{x},t)$ and $\vec{g}(\vec{x},t)$ are known functions of x and t. Let's suppose x has only one component for simplicity:

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = f(x,t) + g(x,t)\Gamma(t) \tag{3.2}$$

In order to solve this kind of problems, we must have some information about the noise. Most commonly used noise is called white noise, which has properties:

$$\langle\Gamma(t)\rangle = 0 \tag{3.3}$$

$$\langle\Gamma(t)\Gamma(t')\rangle = 2D\delta(t-t') \tag{3.4}$$

where D is a constant. However these two equations don't provide us enough information about the noise, although in most cases, these two properties are enough for the calculation of most quantities we want. The full information about a noise should include the correlation functions of all orders. For example, Gaussian white noise contains enough information about the noise:

$$\langle\Gamma(t_1)\Gamma(t_2)\ldots\Gamma(t_{2n-1})\rangle = 0 \tag{3.5}$$

$$\langle\Gamma(t_1)\Gamma(t_2)\ldots\Gamma(t_{2n})\rangle = (2D)^n\sum\delta(t_{i1}-t_{i2})\delta(t_{i3}-t_{i4})\ldots\delta(t_{i2n-1}-t_{i2n}) \tag{3.6}$$

Suppose $f(x,t)$ is a linear function of x, $g(x,t)$ is 1 and $\Gamma(t)$ is Gaussian white noise, then this equation is called Langevin Equation. We can take an average on both side and got:

$$\frac{\mathrm{d}\langle x(t)\rangle}{\mathrm{d}t} = \alpha\langle x(t)\rangle \tag{3.7}$$

This is a very great feature of linear system, because in this case, we can simply ignore the noise and the solution we got will just be the solution for the mean value of x. We can get correlations of $x(t)$ in similar way. We can let this equation time itself, and then, take a average:

$$\frac{\partial^2\langle x(t)x(t')\rangle}{\partial t\partial t'} + \alpha\frac{\partial\langle x(t)x(t')\rangle}{\partial t} + \alpha\frac{\partial\langle x(t)x(t')\rangle}{\partial t'} + \alpha^2\langle x(t)x(t')\rangle = \langle\Gamma(t)\Gamma(t')\rangle \tag{3.8}$$

In fact, we can calculate to any order, as long as we have enough information about the noise. Usually, people only concern about first 2 orders, that's why we usually just use white noise, instead of Gaussian white noise.

But all these features are not true for non-linear systems. This is simply because, for non-linear $f(x,t)$, $\langle f(x,t)\rangle \neq f(\langle x\rangle, t)$. Therefore, new methods must be found to handle non-linear problems. Fokker-Plank equation is just the right way.

By the way, here I would like to emphasize one key conception. If we simply ignore the noise, for linear system we got the equation for mean value of x, but for non-linear system, this equation has no real physical meaning, and it doesn't stands for mean value any more. This is the key difference between linear system and non-linear system. For non-linear system, ignore noise is a very dangerous thing, but unfortunately, it is in many cases what people did.

Fokker-Plank equation doesn't give us the value of x, but it can give us the distribution of x. Let's define the distribution of x at time t as $\rho(x,t)$ and suppose the noise here is white noise. Then, Fokker-Plank equation for system described by Eq.(3.2) will be:

$$\frac{\partial \rho(x,t)}{\partial t} = -\frac{\partial}{\partial x}[f(x,t) + Dg'(x,t)g(x,t)]\rho(x,t) + D\frac{\partial^2}{\partial x^2}[g(x,t)^2\rho(x,t)] \tag{3.9}$$

Due to different regulations of delta-function, there is also another version of Fokker-Plank equation:

$$\frac{\partial \rho(x,t)}{\partial t} = -\frac{\partial}{\partial x}[f(x,t)]\rho(x,t) + D\frac{\partial^2}{\partial x^2}[g(x,t)^2\rho(x,t)] \tag{3.10}$$

The difference between these two versions lies in whether we keep the term $Dg'(x,t)g(x,t)$ or not. If $g(x,t)$ is a constant, then there would be no difference between these two versions.

There is no way to tell which version is better. Usually people use both these two equations and compare the result with real data to decide which one is chosen to use under certain condition. Here I will use the first one with no particular reason.

IV. MODEL

The model I set up is an evolutional model with both inheritance and aberrance. Suppose we have a system which contains large number of individuals and each individual have a lifespan. During its living time, one individual will procreate and then die. Let's concentrate on the evolution of one variable (call it x). Obviously, the x of one individual should be related with its parent, but also has certain aberrance. Let's describe this process by:

$$x_{i+1} = x_i + f(x_i) + g(x_i)\Gamma \tag{4.1}$$

Here i and $i+1$ means the i^{th} and the $(i+1)^{th}$ generation. Γ is a white noise. I will take a continue limit to the number of generation, since I just interested in the long-term behavior of the system:

$$x'(t) = f(x(t),t) + g(x(t),t)\Gamma(t) \tag{4.2}$$

Now I will make two assumptions:

1. $f(x,t)$ is 0, which means the aberrance has no particular preference.
2. $g(x,t)$ is a linear function of x.

The first assumption is very usual one, but I'd like to make some comment on the second one. Usually, the assumption for $g(x,t)$ is: it is a constant, which means the absolute aberrance is a constant no matter what x it has. But my assumption is different. I assume the relative aberrance, $\delta x/x$, is a constant. The traditional assumption makes the calculation much easier, which is just a linear system, but it is obviously wrong. We can imagine that if a human adult have a height $1cm$ away from his or her parents, it is totally normal, but if a single cell creature have a $1cm$ difference, in size, from its parents, it would be total impossible.

With these assumptions, we can start to write the Fokker-Plank equation. But here is one more problem. Individuals in my model don't necessarily just have one child. In fact the number of child is not fixed. As a result, one more term is needed to describe the effect of population changing. This situation is very similar to the situation of chemical reactions. The equations describing chemical reactions are reaction-diffusion equations. Reaction-diffusion equations are just diffusion equations plus proper reaction terms. Here, I did the same thing. The Fokker-Plank equation is just our diffusion equation, and we need to add "reaction term" which describes population changing. The revised equation becomes:

$$\frac{\partial \rho(x,t)}{\partial t} = (h(x,t) - \langle h(x,t) \rangle)\rho(x,t) - \frac{\partial}{\partial x}(Dx\rho(x,t)) + D\frac{\partial^2}{\partial x^2}(x^2\rho(x,t)]) \tag{4.3}$$

Here the first term one the right side is just the "reaction term" I added to Fokker-Plank equation, where $h(x,t)$ is the growth rate at each x. Here, I assume the non-linear effect of ρ to the "reaction term" is small. I also let $h(x,t)$ subtract its own mean value. This is simply a renormalization to $\rho(x,t)$, since the integration of $\rho(x,t)$ over all possible x should give us 1 by definition.

Now, we can make a change of variables to simplify the equation. Define:

$$y = lnx \tag{4.4}$$

$$\tilde{\rho}(y,t) = x\rho(x,t) \tag{4.5}$$

Here the second substitution is necessary, because we need to keep the integration of density over all possible value as 1 ($\int \rho(x,t)\mathrm{d}x = \int \tilde{\rho}(y,t)\mathrm{d}y = 1$). After substitution, we got:

$$\frac{\partial \tilde{\rho}(y,t)}{\partial t} = (h(y,t) - \langle h(y,t) \rangle)\tilde{\rho}(y,t) + D\frac{\partial^2 \tilde{\rho}(y,t)}{\partial y^2} \tag{4.6}$$

V. RESULT

Based on Eq.(4.6), I will try to explain the log-normal distributions and power-law distributions observed in different fields. Now the only problem is figuring out what $h(x,t)$ is.

A. Log-Normal Distribution

We know that $h(x,t)$ is the growth rate for different x. And we can imagine that for most living creatures, there should be one optimum value of x. Although this optimum may depend on many other conditions, we can assume that at mean-field level, we should be able to find one optimum value. Suppose this value is y_0, then we can make a Taylor expansion about this point (Notice since y_0 is optimum value, most individuals will have a y close to this value. Therefore the most contributing part should be around y_0. This is the reason why we can expand it). Obviously, the linear term will be zero, and we keep only leading term, which is quadratic term. We also know that the coefficient of quadratic term should be negative. So we can write:

$$h(y,t) = h(y_0) - \frac{|h''(y_0)|}{2}(y-y_0)^2 + O((y-y_0)^3) \tag{5.1}$$

Now we got an equation which is solvable. I just want to find the steady state, where ρ doesn't depend on time. And the solution is very simple:

$$\tilde{\rho}(y,t) = \frac{1}{\sqrt{2\pi}\delta} e^{-\frac{(y-y_0)^2}{2\delta^2}} \tag{5.2}$$

where

$$\delta^2 = \frac{2D}{|h''(y_0)|} \tag{5.3}$$

If we transfer back from y to x, we will get:

$$\rho(x,t) = \frac{1}{\sqrt{2\pi}\delta x} e^{-\frac{(\ln x - \ln x_0)^2}{2\delta^2}} \tag{5.4}$$

This is just log-normal distributions.

B. Power-Law Distribution

Suppose $h(x,t)$ is a delta function of x. And again we just pay attention to steady state solution where $\frac{\partial \rho}{\partial t} = 0$. The equation will be:

$$0 = A(\delta(y-y_0) - 1)\tilde{\rho}(y,t) + D\frac{\partial^2 \tilde{\rho}(y,t)}{\partial y^2} \tag{5.5}$$

This equation has solution:

$$\tilde{\rho}(y) = \begin{cases} C_1 e^{\sqrt{A/D}y} & (y < y_0) \\ C_2 e^{-\sqrt{A/D}y} & (y > y_0) \end{cases} \tag{5.6}$$

If we transfer back to x, instead of using y, we got:

$$\rho(x) = \begin{cases} C_1 x^{\sqrt{A/D}-1} & (x < x_0) \\ C_2 x^{-(\sqrt{A/D}+1)} & (x > x_0) \end{cases} \tag{5.7}$$

We can see that for $x > x_0$ part, we got a power-law distribution which has a exponent $\sqrt{A/D} + 1$. But here we got 2 problems. First, it is not easy to find a delta function in the real world. Second, for $x \leq x_0$, this distribution will break down and, in fact, these distribution diverse at $x \to x_0$, if we notice that the exponent is larger than 1. We will discuss about these two problems in next part.

VI. CONCLUSION AND DISCUSSION

As we have discussed in previous part, we will get a log-normal distribution as long as these three assumptions are satisfied.

1. No preference heritance ($f(x,t) = 0$);
2. Linear aberrance ($g(x,t) \propto x$);
3. Existence of optimum value x_0 ($h(y,t) = h(y_0) - \frac{|h''(y_0)|}{2}(y-y_0)^2 + O((y-y_0)^3)$).

These assumptions are, at least, qualitatively reasonable in many biology and social systems. Of course, qualitatively reasonable doesn't guarantee this theory is right, but at least we got a theory which is not obvious wrong.

Power-law distribution can be reached if we keep the first two assumptions and change the least one into $h(x,t) = A\delta(x - x_0)$. But as we mentioned before, there are two major problems in this argument. In order to conquer these two problems, discrete variable seems to be a very good choice.

Suppose we have a model where variable can only be positive integers and also suppose we continuously bring into this system new members which have $x = 1$, then in this model, we have $h(x,t) = A\delta_{x,1}$. In continue limit, it will become $h(x) = A\delta(x-1)$. And since x_0 here is 1, which is just the smallest possible values for integer x, we don't need to worry about $x < x_0$ any more.

This is just the scale-free network model. To set up a scale-free network, two things are necessary:

1. Rich getting richer law;
2. Continuously bring new nodes into the system, and new nodes have very small number of links.

The first condition is just our linear aberrance assumption, and the second one is just the argument I made for discrete models. Therefore, we have pretty good reasons to believe that my model is equivalent with scale-free network in the continuous limit.

Besides, this equation may be more than just equivalent to scale-free network. It may also be a good quantitative description of it. Although computer simulations and real data show that there are different power-law exponent for different systems, the mean-field theory of scale free network can give us only exponent 3. But my model can give any exponent larger than 1.

We also notice that this explanation of power-law doesn't use any feature about critical point, which agrees with our simulation result , which shows that these distributions can be observed in almost any control parameter, instead of only at critical point [5].

Acknowledgments

I am grateful to X. Mao, X. Cheng and Q. Ouyang for helpful discussions.

[1] E. Limpert, W.A. Stahel and M.Abbt, BioScience **51**, 341 (2001).
[2] A.L. Barabasi, and R.Albert, Science **286**, 509 (1999).
[3] H. Takayasu, and K.Okuyama, Fractals **6**, 67 (1998).
[4] C. Castellano, M. Marsili, and A.Vespignani, Phys. Rev. Lett. **85**, 3536 (2000).
[5] K. Sun, X.M. Mao, and Q.Ouyang, Chin. Phys. **11**, 1280 (2002).
[6] B.H.A. Groth, Science **39**, 581 (1914).
[7] N.K. Razumovsky, del' Académie des Sciences de l'URSS **9**, 814 (1940)
[8] L.H. Ahrens, Geochim Cosmochim AC **5**, 49 (1954)
[9] A. Malance, L. Gaidolfi, V. Pessina and G.Dallara, J. Environ. Radioactiv. **30**, 55 (1996).
[10] K. Kondo, Japanese J. Human Gen. **21**, 217 (1997).
[11] J.W. Boag, J. Roy. Stat. Soc. B. Met. **11**, 15 (1949).
[12] M. Feinleib and B.McMahon, Blood **15**, 332 (1960).
[13] R.D. Horner, Am. J. Epidemiol **126**, 409 (1987).
[14] F.W. Preston, Ecology **29**, 254 (1948).
[15] R.M. May, *Theoretical Ecology: Principles and Applications* (Blackwell, Oxford, UK, 1981).
[16] A.E. Magurran, *Ecological Diversity and its Measurement* (Croom Helm, London, UK, 1988).
[17] C.B. Williams, Biometrika **31**, 356 (1940).
[18] F.W. Preston, Ecology **62**, 355 (1981).
[19] P. Allanson, J. Agr. Econ. **43**, 137 (1992).
[20] E. Limpert, M. Abbt, R. Asper W.K. Graber F. Godet W.A. Stahel and E.J. Windhab, *Theoretical Ecology: Principles and Applications* (Blackwell, Oxford, UK, 1981).
[21] S. Miyazima, Y. Lee, T. Nagamine, and H.Miyajima, Physica A **278**, 282 (2000).