How skew distributions emerge in evolving systems

To cite this article: M. Y. Choi *et al* 2009 *EPL* **85** 30006

A LETTERS JOURNAL EXPLORING THE FRONTIERS OF PHYSICS

February 2009

EPL, **85** (2009) 30006
www.epljournal.org
doi: 10.1209/0295-5075/85/30006

How skew distributions emerge in evolving systems

M. Y. CHOI[1(a)], H. CHOI[2], J.-Y. FORTIN[3] and J. CHOI[4]

[1] *Department of Physics and Astronomy and Center for Theoretical Physics, Seoul National University Seoul 151-747, Korea*
[2] *Department of Mathematics, University of California - Berkeley, CA 94720, USA*
[3] *Laboratoire de Physique des Matériaux, Université Henri Poincaré Nancy 1 BP 239, F-54506 Vandœuvre les Nancy, Cedex, France, EU*
[4] *Department of Physics, Keimyung University - Daegu 704-701, Korea*

received 6 October 2008; accepted in final form 15 January 2009
published online 10 February 2009

PACS 05.40.-a – Fluctuation phenomena, random processes, noise, and Brownian motion
PACS 89.75.Fb – Structures and organization in complex systems
PACS 05.65.+b – Self-organized systems

Abstract – Despite the ubiquitous emergence of skew distributions such as power law, log-normal, and Weibull distributions, there still lacks proper understanding of the mechanism as well as relations between them. It is studied how such distributions emerge in general evolving systems and what makes the difference between them. Beginning with a master equation for general evolving systems, we obtain the time evolution equation for the size distribution function. Obtained in the case of size changes proportional to the current size are the power law stationary distribution with an arbitrary exponent and the evolving distribution, which is of either log-normal or Weibull type asymptotically, depending on production and growth in the system. This master equation approach thus gives a unified description of those three types of skew distribution observed in a variety of systems, providing physical derivation of them and disclosing how they are related.

Introduction. – Many evolving systems, natural and/or man-made, exhibit skew distributions of characteristic spectra, which have long tails to one side. Among them power law, log-normal, and Weibull distributions appear most ubiquitously in a variety of systems. Here the power law distribution manifests scale invariance of the system [1]. The concept of self-organized criticality [2], which is a well-known framework to explain such scale invariance or criticality, has been applied to a number of specific models, mostly on the case-by-case basis [2–4]. As a general theoretical understanding in this direction, the importance of information transfer between the system and the environment has been recognized [5]. On the other hand, the Yule process [6], dubbed the rich-get-richer process, Gibrat's law, or preferential attachment [7,8] in the literature, has long been known to produce successfully power law distributions in a variety of systems, *e.g.*, Pareto's law [9] or Zipf's law [10], usually with exponents greater than two [1]. However, there still remains ambiguity as some systems exhibit power law behaviors but with exponents smaller than two [11]. It is further well known that in some cases the power law regime is rather limited, exhibiting some curvature in the log-log plot and hence other distributions may also well give a reasonable description for particular sets of data [1].

Related to this, log-normal and Weibull distributions, which are observed widely, *e.g.*, by biological system sizes, financial variables, survival or reliability functions, fragmentation sizes of particles, etc., exhibit similar behavior and have long been used as alternatives to the power law distribution [12–15]. The origin of the log-normal distribution is usually attributed to a large number of small multiplicative factors which are independent. However, in an evolving system, such assumption of independent multiplicative factors corresponds to growth rates varying randomly at every instant of time (thus forming a Gaussian distribution), which is neither substantiated nor realistic. On the other hand, derivation of the Weibull distribution relies upon extreme value statistics of correlated random variables [16] or empirical assumption such as fractal cracking [17]. Accordingly, it is not likely to give an accurate description of the evolving system at the fundamental level. In particular, it should be noted that in many cases those distributions may not be discerned,

(a)E-mail: mychoi@snu.ac.kr

$$\frac{\mathrm{d}}{\mathrm{d}t}P(x_1,x_2,\ldots,x_N;t)=\sum_j\sum_{x'_j}[w_j(x'_j\to x_j)P(x_1,\ldots,x'_j,\ldots,x_N;t)-w_j(x_j\to x'_j)P(x_1,\ldots,x_j,\ldots,x_N;t)] \quad (1)$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}[Nf(x,t)] &= \sum_{i,j}\sum_{x_1,\ldots,x_N,x'_j}\delta(x_i-x)[w_j(x'_j\to x_j)P(x_1,\ldots,x'_j,\ldots;t)-w_j(x_j\to x'_j)P(x_1,\ldots,x_j,\ldots;t)]\\ &= \sum_{i}\sum_{x_1,\ldots,x_N,x'_i}\delta(x_i-x)[w_i(x'_i\to x_i)P(x_1,\ldots,x'_i,\ldots;t)-w_i(x_i\to x'_i)P(x_1,\ldots,x_i,\ldots;t)], \end{aligned} \quad (3)$$

$$\frac{\partial}{\partial t}f(x,t)=-rf(x,t)+\frac{1}{N}\sum_{i}\sum_{x_1,\ldots,x_N,x'_i}\delta(x_i-x)[w_i(x'_i\to x_i)P(x_1,\ldots,x'_i,\ldots;t)-w_i(x_i\to x'_i)P(x_1,\ldots,x_i,\ldots;t)]. \quad (5)$$

indicating possible connections between them; this has never been addressed. Above all, those distributions as well as the power law distribution have been considered only at stationarity, and it would be of interest to explore a non-stationary distribution and its time evolution.

Motivated by these, we in this work attempt to provide a general framework to describe the skew distributions, based on the master equation approach to evolving systems. We consider a master equation, describing the time evolution of the probability for the system configuration, and probe the resulting distributions in the common case that size changes are proportional to the current size. Observed is the evolving (non-stationary) distribution, asymptotically of the log-normal or the Weibull type, as well as the stationary distribution of the power law type, where the exponent can take any value. In such a way this approach gives a unified description of the skew distributions observed in a variety of systems: It provides physical derivation of them and clarifies how they are related, thus resolving the long-standing puzzle.

Master equation. – Consider an evolving system consisting of N elements, the i-th of which is characterized by, say, "size" x_i. We begin with the master equation for the probability $P(x_1, x_2, \ldots, x_N;t)$ for the system in configuration $\{x_1, x_2, \ldots, x_N\}$ at time t:

see eq. (1) *above*

with the appropriate transition rate $w_j(x_j \to x'_j)$. We are interested in the time evolution equation for the normalized distribution function $f(x,t)$ for size x at time t, which is defined by

$$f(x,t)\equiv\frac{1}{N}\sum_{x_1,\ldots,x_N}\sum_i\delta(x_i-x)P(x_1,\ldots,x_N;t). \quad (2)$$

To probe the time evolution of the size distribution $f(x,t)$, we multiply eq. (1) by $\delta(x_i-x)$ and take the summation over i and over all configurations. This leads to

see eq. (3) *above*

where it has been noted on the first line that for $i \neq j$, the first term in the summation of the right-hand side cancels out the second term in the summation, thus leaving only terms of $i=j$.

We write

$$\frac{\mathrm{d}}{\mathrm{d}t}[Nf(x,t)]=\frac{\mathrm{d}N}{\mathrm{d}t}f(x,t)+N\frac{\partial f(x,t)}{\partial t} \quad (4)$$

and suppose that the total number of elements changes according to $\mathrm{d}N/\mathrm{d}t = rN$, which should apply to a wide variety of real situations and has been employed in deriving the Yule-type distribution [7,8]. A typical example includes the system in which each element tends to produce a new element of unit size with rate r [15]. In principle, the production rate r may depend on other parameters of the system, which may be taken into consideration; for simplicity, here it will be assumed to be constant. Equation (3) then reads

see eq. (5) *above*

We consider the case that each element changes its size ("grows") by the amount proportional to the current size, which corresponds to the Yule process, Gibrat's law, or preferential attachment [6–8]. In this case the transition rate reads

$$w_j(x_j\to x'_j)=\lambda\delta(x'_j-x_j-bx_j), \quad (6)$$

where λ is the mean growth rate and b the growth factor. This leads eq. (5), upon summed over x'_i, to take the form

$$\frac{\partial}{\partial t}f(x,t)=-(r+\lambda)f(x,t)+\lambda f(x-\delta,t) \quad (7)$$

with $\delta\equiv bx/(1+b)$, which governs the time evolution of the distribution function $f(x,t)$.

Stationary and non-stationary solutions. – We first consider the stationary solution and set $\partial f/\partial t = 0$ in eq. (7), which leads to

$$(r+\lambda)f(x) = \lambda f(x-\delta). \tag{8}$$

It is straightforward to obtain the solution in the form

$$f(x) \propto x^{-\alpha}, \tag{9}$$

which is the power law distribution with the exponent $\alpha \equiv \ln(1+r/\lambda)/\ln(1+b)$. Note that the exponent α in this power law distribution, determined by the production rate, growth rate, and growth factor, can take any value including values *less than two*. This contrasts with most approaches predicting the exponent α greater than two and may thus provide an explanation of the cases of α smaller than two [1,11].

To probe the non-stationary solution of eq. (7), we, for convenience, expand the right-hand side and write eq. (7) in the form

$$\frac{\partial}{\partial t} f(x,t) = -rf(x,t) + \lambda \sum_{n=1}^{\infty} (-1)^n \frac{\delta^n}{n!} \frac{\partial^n}{\partial x^n} f(x,t). \tag{10}$$

It shall be shown that the solution takes the form of the log-normal or Weibull distribution. In the former case, with the ansatz

$$f(x,t) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2} \tag{11}$$

which depends on time through the mean μ and the variance σ^2, we have the derivatives

$$\frac{1}{f}\frac{\partial f}{\partial t} = -\frac{\dot\sigma}{\sigma} + \frac{\dot\mu}{\sigma^2}(\ln x - \mu) + \frac{\dot\sigma}{\sigma^3}(\ln x - \mu)^2 \tag{12}$$

and

$$\frac{x^n}{f}\frac{\partial^n f}{\partial x^n} = \sum_{k=0}^{n} a_{nk} \left(\frac{\ln x - \mu}{\sigma^2}\right)^k \tag{13}$$

for nonnegative integer n, which can be proved by mathematical induction. The coefficients a_{nk} are of the order unity and given in table 1.

Plugging these into eq. (10) with $\delta = bx/(1+b)$, we obtain the equations for μ and σ:

$$\begin{aligned} \dot\mu &= A + O(\sigma^{-2}), \\ \dot\sigma &= B\sigma + O(\sigma^{-1}), \end{aligned} \tag{14}$$

where A and B are constants depending on r, λ, and b: $A = \lambda \sum_{n=1}^{\infty} (-1)^n (n!)^{-1} a_{n1} b^n (1+b)^{-n} = \lambda b + (\lambda/2)b^2 + O(b^3)$ and $B = r - \lambda \sum_{n=1}^{\infty} (-1)^n (n!)^{-1} a_{n0} b^n (1+b)^{-n} = r - \lambda b + O(b^3)$. Equation (14) is valid for σ growing sufficiently faster than μ. To the leading order, the solution reads

$$\begin{aligned} \mu &= At, \\ \sigma &= \sigma_0 e^{Bt}, \end{aligned} \tag{15}$$

Table 1: Coefficients in the derivatives of the distribution functions.

(n,k)	a_{nk} (log-normal)	$\tilde a_{nk}$ (Weibull)
(0, 0)	1	1
(1, 0)	-1	$\gamma-1$
(1, 1)	-1	$-\gamma$
(2, 0)	$2-\sigma^{-2}$	$(\gamma-1)(\gamma-2)$
(2, 1)	3	$-3\gamma(\gamma-1)$
(2, 2)	1	γ^2
(3, 0)	$-6+6\sigma^{-2}$	$(\gamma-1)(\gamma-2)(\gamma-3)$
(3, 1)	$-11+3\sigma^{-2}$	$\gamma(\gamma-1)(7\gamma-11)$
(3, 2)	-6	$6\gamma^2(\gamma-1)$
(3, 3)	-1	$-\gamma^3$

which, for $B>0$, indeed manifests the validity in the limit $t\to\infty$. It is thus concluded that for $B>0$ or $r > \lambda b + O(b^3)$, the log-normal distribution in eq. (11) is asymptotically exact, with the mean and the variance growing linearly and exponentially, respectively, in time.

Then what about the case $r<\lambda b$? As will be shown below, the solution is given by the Weibull distribution

$$f(x,t) = \frac{\gamma}{x}\left(\frac{x}{\eta}\right)^{\gamma} e^{-(x/\eta)^\gamma} \tag{16}$$

with the shape parameter γ and the scale parameter η. Taking this ansatz with time-dependent η and constant γ leads to the derivatives

$$\frac{1}{f}\frac{\partial f}{\partial t} = -\gamma\left[1 - \left(\frac{x}{\eta}\right)^{\gamma}\right]\frac{\dot\eta}{\eta} \tag{17}$$

and

$$\frac{x^n}{f}\frac{\partial^n f}{\partial x^n} = \sum_{k=0}^{n} \tilde a_{nk}\left(\frac{x}{\eta}\right)^{k\gamma}, \tag{18}$$

which is again proved by mathematical induction. The coefficients $\tilde a_{nk}$, taking the form of a degree-n polynomial in γ, are given in table 1. These are plugged into eq. (10) to yield, for large η,

$$\left[1 - \left(\frac{x}{\eta}\right)^{\gamma}\right]\frac{\dot\eta}{\eta} = C - D\left(\frac{x}{\eta}\right)^{\gamma} + O\left(\frac{x}{\eta}\right)^{2\gamma} \tag{19}$$

with constants $C = r/\gamma - (\lambda/\gamma)\sum_{n=1}^{\infty}(-1)^n(n!)^{-1}\tilde a_{n0} b^n \times (1+b)^{-n} = r/\gamma + \lambda(\gamma-1)[b/\gamma - b^2/2 + O(b^3)]$ and $D = (\lambda/\gamma)\sum_{n=1}^{\infty}(-1)^n (n!)^{-1}\tilde a_{n1} b^n (1+b)^{-n} = \lambda[b + (1/2 - 3\gamma/2)b^2 + O(b^3)]$. Equation (19) allows a solution for η if $C=D$. This gives γ as a function of r, λ, and b:

$$\gamma = \sqrt{-\frac{r}{\lambda b^2} + \frac{1}{b} + O(b)}, \tag{20}$$

which is valid for $r < \lambda b + O(b^3)$. If this holds, it is easy to see that $D>0$ and the scale parameter

$$\eta = \eta_0 e^{Dt} \tag{21}$$

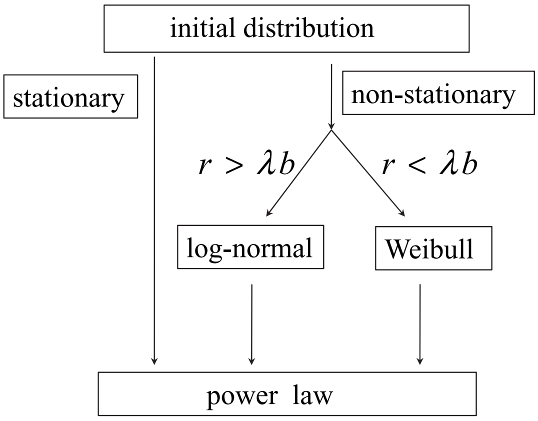

Fig. 1: Evolution of distributions in evolving systems.

indeed grows exponentially in time, justifying eq. (19) and this solution for long time t. Accordingly, for $r < \lambda b + O(b^3)$, the Weibull distribution in eq. (16) is asymptotically exact, where the shape parameter γ is constant and the scale parameter η grows exponentially with time constant $D^{-1} = (\lambda b)^{-1}[1-(1/2-3\gamma/2)b + O(b^2)] \approx 1/\lambda b$.

We stress that the log-normal or the Weibull distribution in eqs. (11) and (16) makes an asymptotically exact solution of eq. (10) for any value of b only if B or D is positive, although analytic expressions of A, B, or $C\,(=D)$ can explicitly be given as series in powers of b. Roughly speaking, the size distribution of the system is determined according to whether production is smaller or larger than growth, and evolves rather quickly in time to either the log-normal distribution or the Weibull distribution. In this sense the two distributions may be considered to belong to one class, thus providing a natural explanation for their similarity. Recall also that part of the log-normal or Weibull distribution appears more or less similar to the power law distribution. The larger the variance or the scale parameter, the wider the power law part of the distribution. As σ or η grows in time, either one thus approaches eventually the power law distribution with the exponent $\alpha = 1$ or $1-\gamma$: At $r = \lambda b + O(b^2)$, we have $\gamma = 0$ and there arises a crossover between the two types of distribution. Figure 1 summarizes the evolution of distribution functions in evolving systems.

Note that this distribution approached in the long-time limit is different from the stationary distribution in eq. (9). Then when does the system display the evolving (log-normal or Weibull) distribution instead of the stationary power law distribution, and vice versa? To seek for an answer, we consider deviation $g(x,t)$ from the power law distribution $f(x)$ in eq. (9), and examine how it evolves in time. Writing $g(x,t) = g(x)e^{\nu t}$ and carrying out a lengthy but straightforward calculation, we obtain the general form $g(x,t) = \sum_\nu g_\nu(x,t)$ with

$$g_\nu(x,t) = \epsilon_\nu x^{-\alpha_\nu} e^{\nu' t} \cos(\kappa_\nu \ln x - \nu'' t + \phi_\nu), \qquad (22)$$

which exhibits oscillations both in size and in time. Here ν' and ν'' denote the real and imaginary parts of ν, respectively. The exponent and the wave number are given by $\alpha_\nu = \frac{1}{2\ln(1+b)} \ln[(1+\frac{r+\nu'}{\lambda})^2 + (\frac{\nu''}{\lambda})^2]$ and $\kappa_\nu = \frac{1}{\ln(1+b)} \tan^{-1} \frac{\nu''}{r+\lambda+\nu'}$, while the phase ϕ_ν as well as the amplitude ϵ_ν of each mode is determined by the initial condition. For stability of $f(x)$, the deviation $g(x,t)$ should decay in time. This is the case for the deviation consisting of modes with exponents less than $\alpha_{\nu=0} = \ln(1+r/\lambda)/\ln(1+b) \equiv \alpha$; otherwise there exist modes with $\nu' > 0$. Accordingly, when the initial distribution is given by power law components with exponents not larger than α, the distribution evolves to the power law distribution in eq. (9). Conversely, if there are components with exponents larger than α in the initial distribution, the power law distribution becomes unstable. In this case the system is expected to display the log-normal or Weibull distribution in eq. (11) or (16), depending on production and growth. (Note that deviation of the above power law form is not allowed at all in the case of the log-normal or Weibull distribution because of the non-negativity condition for all x and t.) Namely, rather a broad initial distribution with a fat tail evolves to the power law distribution. On the other hand, starting from a steep distribution including, *e.g.*, a sharp peak, the system may not evolve directly to the power law distribution; instead it is expected to exhibit the log-normal distribution or the Weibull one.

Discussion. – We have proposed a systematic and unified description of the most ubiquitous skew distributions in evolving systems. Beginning with the master equation for the system configuration probability, we have derived the time evolution equation for the system size distribution. Obtained are the stationary distribution, characterized by power law behavior, and the evolving distribution of the log-normal or Weibull type, depending on the parameters such as the production rate, growth rate, and growth factor. In particular, the exponent of the power law distribution can take values less than two. This master equation approach gives a general description of the characteristic skew distributions observed in evolving systems, revealing the connections between them and clarifying relevant distributions. For example, one may consider the evolving distribution of pancreatic islet sizes [15], where comparison of the fitted data with eq. (14) and with eqs. (20) and (21) excludes the log-normal distribution, thus confirming the Weibull distribution.

Finally, we also point out that in this approach r and b are not necessarily greater than zero. Negative values of r and b, which should describe destruction (rather than production) and degeneration (rather than growth), respectively, of existing elements, are not *a priori* excluded. Therefore results of this work are not limited to growing systems in the narrow sense but applicable to general evolving systems. It is of interest to consider real systems exhibiting skew distributions and examine explicitly the relations between them, which are left for further study.

* * *

MYC thanks the Laboratoire de Physique Théorique (CNRS) for hospitality, where part of this work was accomplished, and acknowledges the support by the KOSEF-CNRS Cooperative Program and by the KOSEF/MOST grant to NCRC for Systems Bio-Dynamics.

REFERENCES

[1] See, *e.g.,* Newman M. E. J., *Contemp. Phys.*, **46** (2005) 323 and references therein.

[2] Bak P., Tang C. and Wiesenfeld K., *Phys. Rev. Lett.*, **59** (1987) 381; *Phys. Rev. A*, **38** (1988) 364.

[3] Jensen H. J., Christensen K. and Fogedby H. C., *Phys. Rev. B*, **40** (1989) R7425; Kertesz J. and Kiss L. B., *J. Phys. A*, **23** (1990) L433; Bonabeau E. and Lederer P., *J. Phys. A*, **27** (1994) L243.

[4] Bak P. and Tang C., *J. Geophys. Res.*, **94** (1989) 15635; Drossel B. and Schwabl F., *Phys. Rev. Lett.*, **69** (1992) 1629; Bak P. and Sneppen K., *Phys. Rev. Lett.*, **71** (1993) 4083; Fernandez J., Plastino A. and Diambra L., *Phys. Rev. E*, **52** (1995) 5700; Jo J., Kang H., Yoon B., Choi J.-G. and Choi M. Y., *J. Phys. A*, **41** (2008) 145101.

[5] Choi M. Y., Kim B. J., Yoon B.-G. and Park H., *Europhys. Lett.*, **69** (2005) 503.

[6] Yule G. U., *Philos. Trans. R. Soc. London, Ser. B*, **213** (1925) 21.

[7] Gibrat R., *Les Inégalités Économiques* (Sirey, Paris) 1932; Simon H. A., *Biometrika*, **42** (1955) 425; Sutton J., *J. Econ. Lit.*, **35** (1997) 40.

[8] See, *e.g.*, Albert R. and Barabási A.-L., *Rev. Mod. Phys.*, **74** (2002) 47 and references therein.

[9] Pareto V., *Cours d'Économie Politique* (Droz, Geneva) 1896.

[10] Zipf G. K., *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge) 1949.

[11] Montoya J. M. and Solé R. V., *J. Theor. Biol.*, **214** (2002) 405; Fujiwara Y., Guilmi C. D., Aoyama H., Gallegati M. and Souma W., *Physica A*, **335** (2004) 197; Lee K., Jung W.-S., Park J. S. and Choi M. Y., *Physica A*, **387** (2008) 6231.

[12] Nelson W., *Applied Life Data Analysis* (Wiley, New York) 1982; Kececioglu D., *Reliability and Life Testing Handbook* (Prentice Hall, Englewood Cliffs) 1993.

[13] Oexle K., *J. Theor. Biol.*, **190** (1998) 369; Laherrère J. and Sornette D., *Eur. Phys. J. B*, **2** (1998) 525; Paredes V. R. and Alvarado V., *Phys. Rev. E*, **58** (1998) 771; Limbert E., Stahel W. A. and Abbt M., *Bioscience*, **51** (2001) 341; Aragao G. M. F., Corradini M. G., Normand M. D. and Peleg M., *Int. J. Food Microbiol.*, **119** (2007) 243.

[14] Luchinsky D. G. *et al.*, *Phys. Rev. Lett.*, **82** (1999) 1806; Kunes K. C., Cox D. L. and Singh R. R. P., *Phys. Rev. E*, **72** (2005) 051915; Chessa A. G. and Murre J. M. J., *Physica A*, **366** (2006) 539.

[15] Jo J., Choi M. Y. and Koh D. S., *Biophys. J.*, **93** (2007) 2655.

[16] Bertin E. and Clusel M., *J. Phys. A*, **39** (2006) 7607.

[17] Brown W. K. and Wohletz K. H., *J. Appl. Phys.*, **78** (1995) 2758.