Analysis of size distributions of fictional characters: from Pokémon to Godzilla

Ken Yamamoto Department of Physics and Earth Sciences, Faculty of Science, University of Ryukyus, Nishihara, Okinawa 903–0213, Japan

In this letter, the size (height and weight) of fictional characters in animations, superhero series, movies, and other media is studied. We find that the distributions of character height and weight approximately follow lognormal distributions in common to five selected works. We propose a mechanism governing this lognormal behavior based on the principle of maximum entropy and the Weber-Fechner law. Moreover, we provide a comparison to the size distributions of real animals. Although the size distributions of fictional characters and real animals are both lognormal, the distributions are essentially different, particularly in the scaling between height and weight.

Size distribution is an important tool in analyzing complex systems. We can compute size distributions simply by collecting and organizing data. In general, obtaining precise microscopic dynamics in complex systems is difficult, and an effective approach for it is a phenomenological description [1, 2]. The properties of size distributions is useful in constructing phenomenological models. In many cases, size distributions follow heavy-tailed probability distributions whose decay in the tail is slower than the exponential functions; thus, the mean, variance, or higher moments diverge [3]. The emergence of heavy-tailed distributions indicates the existence of elements much larger than the ones of typical size. Further, numerous theoretical mechanisms have been proposed to obtain heavy-tailed distributions, such as power-law and lognormal distributions [4, 5].

Generally, the size of the biological organisms typically follows lognormal distribution [6–9]. This is a heavy-tailed distribution where the logarithm of the size is distributed normally [10]. Theoretically, the lognormal behavior of organisms is simply explained by the multiplicative stochastic process

$$X_{t+1} = R_t X_t, \tag{1}$$

where R_t is a random growth rate. In the simplest case, when R_1, R_2, \ldots are independently and identically distributed, $\ln X_t$ approximately follows a normal distribution for sufficiently large t (because of the central limit theorem), and X_t follows a lognormal distribution. This model for the lognormal distribution dates back to Kolmogorov [11].

In this study, we analyzed the sizes (height and weight) of fictional creatures, instead of real organisms. They are characters in creations such as animations, movies, and comics whose sizes are arbitrarily determined by their designers. Nevertheless, we find that size distributions for different works exhibit a common lognormal behavior. (A preliminary result was already reported [12].) We propose a simple mechanism that produces this lognormal distribution based on the principle of maximum entropy and the Weber-Fechner law regarding the perception of numbers. Moreover, we investigate the similarities and differences between fictional characters and real animals from the perspective of size distribution.

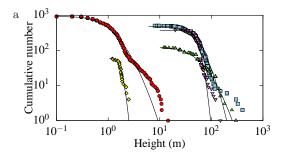
To analyze the size distribution, it is desirable to select works in which the sizes of a large number of characters are officially determined. We analyze the following five groups of characters:

- 1. Pokémon (creatures) from the *Pokémon* series (video games, animations, and other media)
- 2. Zords (robots and vehicles) from the *Power Rangers* series (American superhero television series)
- 3. Giant monsters from the Godzilla series (Japanese movie series),
- 4. Villains (giant monsters and alien invaders) from the *Ultraman* series (Japanese superhero television series)
- 5. Characters from the *Dragon Ball* series (Japanese comics and animations of martial arts experts and alien warriors).

We obtained the sizes of Pokémon creatures from the online official database Pokédex [13] (up to Generation VII), the sizes of Zords from the online database RangerWiki [14] (up to the 25th season, entitled $Super\ Ninja\ Steel$), and the sizes of characters in $Dragon\ Ball$ from the online database $Dragon\ Ball\ Wiki$ [15] in March 2019. We referred to Japanese books [16, 17] to obtain the character sizes in Godzilla (including 29 movie titles from 1954 to 2016) and $Ultraman\ (Showa\ Ultraman\ series\ aired\ in\ Japan\ from\ 1966\ to\ 1981)$. We only analyzed characters whose height and weight were both provided.

Figure 1 shows the height (a) and weight (b) distributions in cumulative form. Each curve is a lognormal distribution:

$$F(x) = \frac{N}{2} \left(1 - \operatorname{erf} \left(\frac{\ln x - \mu}{\sqrt{2}\sigma} \right) \right),$$



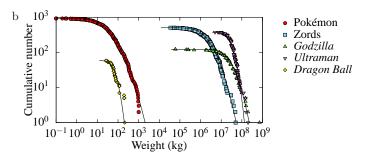


FIG. 1. (Color online) Cumulative height (a) and weight (b) distributions of Pokémon (circles), Zords in *Power Rangers* (squares), *Godzilla* monsters (upper triangles), *Ultraman* villains (lower triangles), and *Dragon Ball* characters (diamonds). The curves present a lognormal fit for each group of characters.

TABLE I. The number N of characters and the lognormal parameters μ and σ for the height and weight distributions in each work.

		Height		Weight	
Title	N	$\mu_{ m H}$	$\sigma_{ m H}$	$\mu_{ m W}$	$\sigma_{ m W}$
Pokémon	927	-0.05	0.74	3.34	1.40
Power Rangers	502	3.80	0.52	14.48	1.15
Godzilla	122	3.89	0.69	16.97	0.95
Ultraman	369	3.97	0.22	17.20	0.54
Dragon Ball	58	0.55	0.17	4.21	0.52

where N is the number of characters, erf is the Gauss error function, and μ and σ are parameters. The values of N, μ and σ are listed in Table I. The height and weight distributions are both consistent with a lognormal distribution. As shown in Fig. 1 a, approximately ten of the tallest characters in $Pok\acute{e}mon$ (circles), $Power\ Rangers$ (squares), and Ultraman (lower triangles) deviate from the lognormal distribution. In Fig. 1 b, only the heaviest character in Godzilla (the rightmost upper triangle), SpaceGodzilla Flying Form with a weight of 720 000 tons [16], deviates greatly from the lognormal curve.

To investigate the lognormal distribution of character size in more detail, we applied quantile-quantile (Q-Q) plots [18]. The q-quantile Q_q (for $0 \le q \le 1$) of a random variable X is defined by $P(X \le Q_q) \ge q$ and $P(X \ge Q_q) \ge 1 - q$. The Q-Q plot is a scatter plot of the data and the corresponding theoretical quantiles of a certain reference probability distribution (usually a normal distribution). If the points in the Q-Q plot are linearly aligned, the data are determined to follow the reference distribution well.

Figure 2 shows Q-Q plots for the height and weight of the characters, where log-transformed sizes are compared with the quantiles of a normal distribution. (Note that a normal distribution for the log-transformed size indicates that the size is distributed lognormally.) Among all of the characters, the weight of Zords (b2) and *Ultraman* villains (b4) and the height of *Godzilla* monsters (a3) agree very well with a lognormal distribution. Some of the other Q-Q plots show that a few characters deviate from the lognormal distribution in large- or small-sized regions. These deviations have a common tendency; large (tall or heavy) characters tend to have sizes greater than those expected by the lognormal distribution consistently (points in Fig. 2 are above the solid diagonal lines), and small characters tend to be less than the lognormal distribution (points are below the diagonal line). We surmise that the size of large characters can be overstated and those of small characters can be understated; for example, heavy-looking characters will be given excessively heavy weights to highlight their heaviness. However, as a whole, the character size follows a lognormal distribution, particularly in the intermediate region.

Here, we propose a theoretical mechanism governing the common lognormal behavior of character sizes. We introduce a random variable X for the character size. Because the characters are fictional, their sizes are free from physical, biological, or mechanical constraints. Instead, the character sizes most likely reflect how humans interpret and perceive numbers.

The Weber-Fechner law states that the subjective intensity of a sensation (e.g., vision, hearing, or touch) is proportional not to the stimulus intensity but to its logarithm [19]. In other words, this law implies that our mental number line has a logarithmic scale. Notably, the Weber-Fechner law also governs our recognition of numbers [20]. Furthermore, the Weber-Fechner law has been observed in neural [21] and mathematical [22] studies.

A well-known example of the Weber-Fechner law is the perception of musical pitches. We perceive that the pitch interval between the musical tones by two adjacent keys in piano is constant. However, an octave rise in pitch

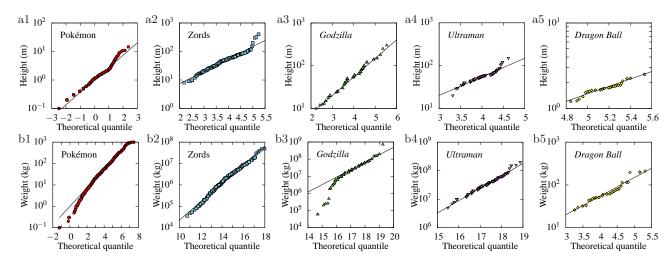


FIG. 2. (Color online) Q-Q plots of the height (a) and weight (b) of Pokémon (1), Zords (2), Godzilla monsters (3), Ultraman villains (4), and Dragon Ball characters (5). The diagonal line in each panel indicates the expected results for the case in which the sizes are lognormally distributed.

corresponds to a doubling of frequency [23]. The frequency f of the pitch n octave above the frequency f_0 is written as $f = 2^n f_0$. Its logarithm, $\ln f = n \ln 2 + \ln f_0$, indicates that $\ln f$ increases at a constant rate $\ln 2$ with the increase in n. Therefore, the logarithm of frequency matches the perception of pitches, which is a manifestation of the Weber-Fechner law.

A typical character size appears to exist for each work; for instance, the sizes of Pokémon are similar to those of human children. Hence, in the simplest view, the size of a character is given by this typical size with fluctuations due to the individual differences. Combined with the Weber-Fechner law, in which character sizes are mentally measured as $\ln X$, this condition is expressed as

$$E[\ln X] = \mu, \quad V[\ln X] = \sigma^2, \tag{2}$$

where E and V represent the average and variance, respectively, and μ and σ are associated with the typical character size and fluctuation, respectively.

According to the principle of maximum entropy [24], when we have no knowledge about X except for Eq. (2), the best probability density function f(x) of X is given by maximizing the entropy $-\int_0^\infty f(x) \ln f(x) dx$ under constraint (2). We employ the method of Lagrange multipliers and set

$$H[f] = -\int_0^\infty f(x) \ln f(x) dx - \lambda_0 \left(\int_0^\infty f(x) dx - 1 \right)$$
$$-\lambda_1 \left(\int_0^\infty f(x) \ln x dx - \mu \right)$$
$$-\lambda_2 \left(\int_0^\infty f(x) (\ln x - \mu)^2 dx - \sigma^2 \right).$$

Multipliers λ_1 and λ_2 correspond to constraints shown in Eq. (2), and λ_0 is needed for the normalization $\int_0^\infty f(x)dx = 1$. To find f such that the functional H[f] takes a stationary value, we calculate

$$\frac{\delta H}{\delta f} = -\ln f - 1 - \lambda_0 - \lambda_1 \ln x - \lambda_2 (\ln x - \mu)^2 = 0,$$

which directly leads to

$$f(x) = e^{-1-\lambda_0} \frac{1}{x} \exp(-\lambda_2(\ln x - \mu)^2 + (1 - \lambda_1) \ln x).$$

The exponent of the exponential function is a quadratic of $\ln x$; thus, f is a lognormal-type function. By determining the multipliers λ_0, λ_1 , and λ_2 , the function f is found to be the probability density of the lognormal distribution with parameters μ and σ :

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

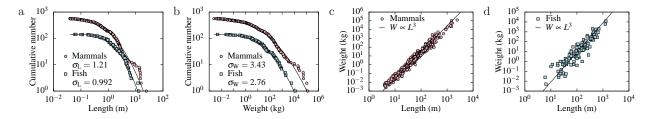


FIG. 3. (Color online) (a) Cumulative length distributions of 564 mammals (circles) and 124 fish (squares). The solid curves are lognormal distributions, with $\mu_{\rm L}=-0.565$ and $\sigma_{\rm L}=1.21$ for mammals and $\mu_{\rm L}=0.051$ and $\sigma_{\rm L}=0.992$ for fish. (b) Cumulative weight distributions of mammals and fish. The lognormal parameters are $\mu_{\rm W}=1.61$ and $\sigma_{\rm W}=3.43$ for mammals and $\mu_{\rm W}=2.36$ and $\sigma_{\rm W}=2.76$ for fish. (c) Correlation between the length and weight of mammals. The solid line represents the scaling $W \propto L^3$ based on the principle of similitude. (d) The length and weight of fish.

Therefore, the principle of maximum entropy combined with the Weber-Fechner law can yield a lognormal distribution. Other mechanisms for lognormal distributions based on the Weber-Fechner law have been proposed for the length of user posts on the Internet [25], the length of English words and the number of strokes in Chinese characters [26], and the size of human tumors [27].

We find that the lognormal distribution is a standard distribution for fictional character sizes. We also note that the above conclusion of lognormal distribution is derived from the simplest assumption. Because character sizes are set arbitrarily, a designer can intentionally determine the character sizes that do not follow a lognormal distribution. We believe that the lognormal distribution is natural, in the sense that the character-size distribution tends to be lognormal when the size is set without substantial thought.

We compare the size distributions of fictional characters and real animals. Figures 3 a and b show the length and weight distributions of 564 species of mammals (circles) and 124 species of fish (squares), obtained from a visual guide of animals [28]; we included only species for which both the length and weight were given. Although the circle points in Fig. 3 a and b deviate slightly from the solid lognormal fitting curves in the large-sized regions (the author does not know why the deviation occurs), the overall distributions can be approximated by a lognormal distribution.

Equation (1) indicates that the lognormality of real animals of a single species arises from the randomness in their growth. However, this equation cannot be directly applied to the sizes of different species, as shown in Fig. 3 a and b. Similarly, the lognormal life-span distribution of invertebrates has been reported [29]. The possibility of the existence of a multiplicative stochastic effect in evolutionary processes is indicated as the cause of this lognormality; however, this hypothesis has not been completely verified. Although the principle behind the lognormal distribution in Fig. 3 a and b has not been accurately determined, it is evident that growth and evolution are not associated with the lognormality of fictional characters. In fact, characters usually do not grow or age; child characters occasionally grow, but they do not grow continuously as described by Eq. (1). Incidentally, the "evolution" of some Pokémon does not cause a multiplicative effect, because Pokémon can evolve, at most, only twice, which is insufficient to yield a lognormal distribution. Therefore, although both fictional characters and real animals exhibit lognormal size distribution, their principles of generation are completely different.

One might assume that the lognormal behavior of fictional characters arises from the imitation of real creatures, whose sizes are distributed lognormally; for instance, Pikachu in $Pok\acute{e}mon$ is modeled from a mouse [13]. To determine whether this hypothesis is valid, we focus on the value of the lognormal fitting parameter σ . Whereas $(\sigma_L, \sigma_W) = (1.21, 3.43)$ for mammals and (0.992, 2.76) for fish, the values of (σ_H, σ_W) for fictional characters are much smaller (see Table I). (Note that σ_H for character height is the counterpart of σ_L for animal length because height and length have linear dimensions.) By comparing the parameter σ , the lognormality of characters is found to be independent of the size of real creatures.

The difference between the size of fictional characters and that of real animals is further clarified by the correlation between their weight and height (or length). If animals had a homothetic body shape and a constant density, their weight W and length L would satisfy $W \propto L^3$, known as the principle of similitude [30]. As shown in Fig. 3 c and d, the scatter plots of the length and weight for mammals and fish are both consistent with the principle of similitude (solid lines). Taking the logarithm, we have $\ln W = 3 \ln L + \text{constant}$. Because the parameters σ_L and σ_W of the lognormal distribution represent the standard deviations of $\ln L$ and $\ln W$, respectively, we have $\sigma_W = 3\sigma_L$. In other words, the ratio σ_W/σ_L is a good estimate for the dimensionality of animals. Indeed, for the data in Fig. 3, $\sigma_W/\sigma_L = 2.86$ for mammals and $\sigma_W/\sigma_L = 2.78$ for fish, which are both close to three.

Scatter plots for the height H and weight W of characters are shown in Fig. 4. Only the *Dragon Ball* characters (e) follow the scaling $W \propto H^3$ ($\sigma_{\rm W}/\sigma_{\rm H}=3.06$). We believe that *Dragon Ball* characters can easily be assigned realistic height and weight, because many of the characters are humans or human-shaped aliens. The other plots,

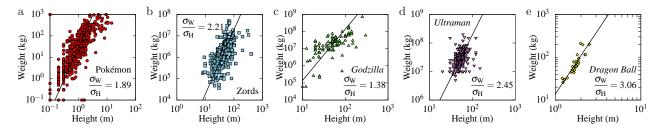


FIG. 4. (Color online) Scatter plots of the height and weight of Pokémon (a), Zords (b), Godzilla monsters (c), Ultraman villains (d), and Dragon Ball characters (e). The solid line in each panel is based on the principle of similitude, $W \propto H^3$.

a–d, exhibit a roughly positive correlation, but the fluctuations are large compared with those of mammals and fish, as shown in Fig. 3 c and d. Thus, small but heavy characters and large but light characters are included. The most extreme character is Cosmoem in $Pok\acute{e}mon$ [13], which is 0.1 m tall but 999.9 kg heavy. Such unrealistic combinations of height and weight may help emphasize the peculiarity of a character. The estimated $\sigma_{\rm W}/\sigma_{\rm H}$ of dimensionality is considerably smaller than three (close to two), in contrast to the finding for real animals.

In summary, the height and weight of fictional characters in animations, movies, and other media approximately follow a lognormal distribution, even though character sizes are set arbitrarily for each work and are not associated with the size of real animals. In this letter, we propose that this lognormal behavior can be explained by entropy maximization under the constraints where mean and variance of the size, determined by applying the Weber-Fechner law, are fixed. This study implies that statistical regularity and commonality arise from the process of perception and generation of numbers by humans. We expect this work to motivate further studies on the complex thought process and behavior of humans.

The present study was supported by the Hayao Nakayama Foundation for Science & Technology and Culture (H29-B-41) and by a Grant-in-Aid for Scientific Research (C) (19K03656) from the Japan Society for the Promotion of Science.

- [1] C. Castellano, Rev. Mod. Phys. 81, 591 (2009).
- [2] M. Buchanan, Ubiquity (Weidenfeld and Nicolson, London, 2000).
- [3] S. Foss, D. Korshunov, and S. Zachary, An Introduction to Heavy-Tailed and Subexponential Distributions (Springer, New York, 2011).
- [4] M. Mitzenmacher, Internet Math. 1, 226 (2004).
- [5] M. E. J. Newman, Contemp. Phys. 46, 323 (2005).
- [6] A. L. Koch, J. Theor. Biol. 12, 276 (1966).
- [7] E. Limpert, W. A. Stahel, and M. Abbt, BioScience 51, 341 (2001).
- [8] K. Yamamoto and J. Wakita, J. Phys, Soc. Jpn. 85, 074004 (2016).
- [9] K. Koyama, K. Yamamoto, and M. Ushio, Proc. R. Soc. B 284, 20162395 (2017).
- [10] N. Kobayashi, H. Kuninaka, J. Wakita, and M. Matsushita, J. Phys. Soc. Jpn. 80, 072001 (2011).
- [11] A. N. Kolmogorov, Dokl. Akad. Nauk. 31, 99 (1941).
- [12] K. Yamamoto, JPS Conf. Proc. 16, 011015 (2017).
- [13] Pokédex https://www.pokemon.com/us/pokedex/
- [14] RangerWiki https://powerrangers.fandom.com/wiki/RangerWiki
- [15] Dragon Ball Wiki https://dragonball.fandom.com/wiki/Main_Page
- [16] Kodansha (ed.), Ketteiban Godzilla DaiKaijuu Perfect Chou Hyakka (Kodansha, Tokyo, 2016). [in Japanese]
- [17] Kodansha (ed.), Ketteiban Zen Ultra Kaijuu Kanzen Chou Hyakka (Kodansha, Tokyo, 2005). [in Japanese]
- 18] H. C. Thode, Jr., Testing for Normality (Marcel Dekker, New York, 2002).
- [19] E. Maor, e: The Story of a Number (Princeton University Press, Princeton, New Jersey, 2011).
- [20] S. Dehaene, The Number Sense (Oxford University Press, Oxford, 2011).
- [21] S. Dehaene, Trends in Cognitive Sciences 4, 145 (2003).
- [22] R. D. Portugal and B. F. Svaiter, Minds and Machines 21, 73 (2011).
- [23] E. Agmon, The Languages of Western Tonality (Springer, Berlin, 2013).
- [24] E. T. Jaynes, Probability Theory: The Logic of Science (Cambridge University Press, Cambridge, 2003).
- [25] P. Sobkowicz, M. Thelwall, K. Buckley, G. Paltoglou, and A. Sobkowicz, EPJ Data Sci. 2, 2 (2013).
- [26] J. Huang, PLoS One 8, e74515 (2013).
- [27] J. S. Spratt, Jr., J. Surg. Res. 9, 151-157 (1969).
- [28] D. E. Wilson and D. Burnie, Animal: The Definitive Visual Guide to the World's Wildlife (Dorling Kindersley, London,

2001).

- [29] N. Kobayashi, Y. Sasaki, O. Moriyama, S. Matsushita, and M. Matsushita, Nonlinear Phenomena in Complex Systems 9, 276 (2006).
- [30] D. Thompson, On Growth and Form (Cambridge University Press, Cambridge, 1961).