Journal of
Economics and Political Economy
econsciences.com
Volume 12 June 2025 Issue 1

The distributions in nature and entropy principle

By Oded KAFRI †

Abstract. The derivation of the maximum entropy distribution of particles in boxes yields two kinds of distributions: a "bell-like" distribution and a long-tail distribution. The first one is obtained when the ratio between particles and boxes is low, and the second one - when the ratio is high. The obtained long tail distribution yields correctly the empirical Zipf law, Pareto's 20:80 rule and Benford's law. Therefore, it is concluded that the long tail and the "bell-like" distributions are outcomes of the tendency of statistical systems to maximize entropy.
Keywords. Maximum Entropy, Long tail distribution, 20:80 Pareto's rule, Zipf Law, Benford's law, Bell-like distribution.
JEL. C62.

1. Introduction

According there are two common distributions in life: The first one is the "bell-like" distribution, which is found in the distribution of IQ, human heights, human age at death etc. This "almost universal" distribution was introduced for the first time by Moivre in the 18th century and explored by Laplace and Gauss around 1800.

As opposed to the bell curve distribution, many quantities are distributed unevenly (Bak, 1996). For example, the probability to live in a big city is higher than the probability to live in a small village. Similarly, the probability to be poor is higher than the probability to be rich. Although intuitively it is logical for cities' population and wealth to have a bell curve distribution, it is not so. Their distributions are uneven and are characterized by a long tail to the right, in which few have a lot and many have quite a little. These distributions were observed by Pareto, Zipf, Newcomb and Benford about a century later and received their name accordingly: Zipf law (Zipf, 1949; Miller, & Newman, 1958), Pareto's rule (Pareto 1897; Jurgan, 1951), and Benford's law (Newcombs, 1881; Benford, 1938).

The first to discover it was Pareto. In 1896 he observed that the ownership of lands in Italy is distributed among the population in the ratio of around 20:80, namely, about 20% of the population own about 80% of the land. From his observations of other countries as well, he concluded that this ratio is general. Mussolini embraced the Italian Marquis Pareto because he believed that the Pareto's rule proves nature's preference of the fittest. Zipf - a Harvard professor of linguistics - found out that the ratio between the first most frequent word and the second one, in any text in many languages, is two. Similarly, the ratio between the second most frequent word and the fourth one

† Kafri Nihul Ltd., Tel Aviv 68165 Israel.

is also two, etc. He claimed that the shortest and most "efficient" words appear more frequently (Zipf, 1949).

Zipf believed in the evolutionary philosophy, i.e. the most "useful" and "efficient" words are the winners, in the spirit of "the survival of the fittest". On the other hand, many people and political movements believe that Pareto's rule is unfair and the wealth should be shared more equally, namely, as in the bell curve distribution. The discovery of Newcomb about the uneven frequency of digits in logarithmic table in 1881 (Newcomb, 1881), (the higher the value of a digit, the lower its frequency) raises some doubts as for the real reason for the uneven distributions. Later, in 1938, Benford confirmed Newcomb's uneven distribution of digits in a wide range of numerical data (Benford, 1938). He attempted, unsuccessfully, to present a formal proof to Newcomb's equation, see Eq. (12). Since then, this distribution was found also in prime numbers (Cohen, 1984), physical constants, Fibonacci numbers and many more (Kossovsky, 2012).

In this paper it is argued that the "bell-like" distribution and the long tail distribution are the boundaries of the same probability distribution. This probability function is obtained by a fair and unbiased random distribution of particles in boxes.

We consider a set of N boxes scoring P particles; it is assumed that all the boxes have an equal probability to score a particle, namely, the probability of a box to score a particle is $= 1/N$. Therefore, the probability to score n particles is $q_n = (\frac{1}{N})^n$. It is clear that $q_n < q$. This is the basic reason why the rich are fewer than the poor. In the case of $P \ll N$, where a multiple score is negligible, the "bell-like" distribution is obtained; and in the case of $P \gg N$, a long tail distribution is obtained.

2. How *P* particles are distributed in *N* boxes?

The answer to it is not new: the particles are distributed in a way that maximizes the entropy (Planck, 1901).

According to Boltzmann, entropy is proportional to the maximum possible number of the different configurations (microstates) of a set. Namely,

$$S = \ln \Omega \tag{1}$$

(we take here the Boltzmann constant $k_B \equiv 1$). A microstate is one possible distinguishable configuration of a set of boxes and particles. Boltzmann entropy is obtained from the Gibbs-Shannon entropy by assuming that all the microstates have an equal probability. The Gibbs-Shannon entropy is given by:

$$S = -\sum_{j=1}^{\Omega} p_j \ln p_j \tag{2}$$

where p_j is the probability of the microstate j and Ω is the number of microstates to be maximized. If all the microstates have an equal probability, namely, $p_j = 1/\Omega$, Boltzmann entropy $\ln \Omega$ is obtained.

Therefore, the distribution of particles that maximizes Boltzmann entropy means an equal probability to any configuration as well as an equal probability to any particle to be in any box.

The number of microstates (different configurations) of P particles in N states is given by the Planck expression (Planck, 1901) namely,

$$\Omega(P,N) = \frac{(N+P-1)!}{p!(N-1)!} \tag{3}$$

To visualize the problem we start with a numerical example; namely, calculating the distribution of 3 particles in 3 boxes that maximizes entropy. According to Eq. (3) the number of microstates $\Omega(3,3) = 10$ as follows:

|300| 030| 003| 210| 201| 120| 021| 102| 012| and |111|.

We see that although each box has an equal chance to score 1, 2, or 3 particles, the boxes with 1 particle appear 9 times, those with 2 particles appear 6 times, and those with 3 particles appear 3 times. The relative frequency of the boxes with one particle in a set of three boxes is therefore $f(1)$=0.5; with two particles $f(2)$=0.333and with three particles $f(3)$=0.166.

To calculate the relative frequencies $f(n)$, we designate $n = P/N$, where$n$ is the number of particles in a box, and apply the Stirling's formula $\ln N! \cong N \ln N - N$. We obtain (Planck, 1901) from Eqs.(1) and (3) that,

$$S \cong N\{(1+n)\ln(1+n) - n\ln n\} \cong \sum_{n=1}^{N}\{(1+n)\ln(1+n) - n\ln n\} \tag{4}$$

Now we write the Lagrange equation,

$$F(n) \cong \sum_{n=1}^{N}\{(1+n)\ln(1+n) - n\ln n\} - \beta\{P - \sum_{n=1}^{N} n\phi(n)\} \tag{5}$$

The first term on the RHS is the entropy and the second term is theconstraint of the number of particles. Namely, $P = \sum_{n=1}^{N} n\phi(n)$ is the number of particles, $\phi(n)$ is the number of boxes that scored n particles and β is a

Lagrange multiplier.$\phi(n)$ can be interpreted as the probability of a box to have n particles. The normalized (n) , $f(n)$ is the relative frequency of the boxes that scored n particles. From $\frac{\partial(F(n)}{\partial n} = 0$ one obtains,

$$\phi(n) = \beta^{-1}\ln(1+\frac{1}{n}) \tag{6}$$

Eq. (6) is the analogue of Planck equation (Kafri, 2007, 2009, 2016), namely,

$$n = 1/[e^{\beta\phi(n)} - 1] \tag{7}$$

Hereafter, we examine three cases:

In the first case we assume that n>> 1. Here one can expect to find a large number of particles (limited by P) in any of the boxes. For example, if we conduct a popularity poll between the N words among P authors, and there are many more authors than words, then the maximum entropy distribution of the votes between the words is shown to be the Zipf law.

In the second case we consider the intermediate zone where n is in the range of the number of the boxes. This case fits well to the distribution of ranks, namely, Pareto's rule and Benford's law.

In the third case we consider $n<<1$, where the number of particles is negligible as compared to the number of boxes. This case fits well to the probability of guessing correctly the IQ of a person in a single guess based only on the knowledge of the average. This case yields the "bell-like" distribution.

3. Zipf law

Consider the case where $P \gg N$ where $\gg 1$. In this case $\beta\phi \ll 1$, therefore from Eq. (7) $\phi(n)$ can be approximated to,

$$n\phi(n) = \frac{1}{\beta} \tag{8}$$

Eq.(8) is the Zipf law. Namely, the ratio in the frequencies between n=1 (the most frequent word) and n=2 (the second most frequently word) is 2 which is identical to the ratio between $n = 2$ and $n = 4$ etc. This ratio is not a function of β as $\frac{\phi(1)}{\phi(2)} = \frac{\phi(2)}{\phi(4)} = \cdots \frac{\phi(n)}{\phi(2n)} = \cong 2$.

4. Pareto's rule

to calculate the relative frequency of Eq.(6), namely, $f(n)$ we have to divide $\phi(n)$ by the sum over all the M occupied boxes $M \leq N$, namely,

$$\sum_{i=1}^{M} \phi(n) = \frac{1}{\beta}\left(\ln\frac{2}{1} + \ln\frac{3}{2} + \cdots .. + \ln\frac{M+1}{M}\right) = \frac{1}{\beta}\ln\frac{M+1}{M} \tag{9}$$

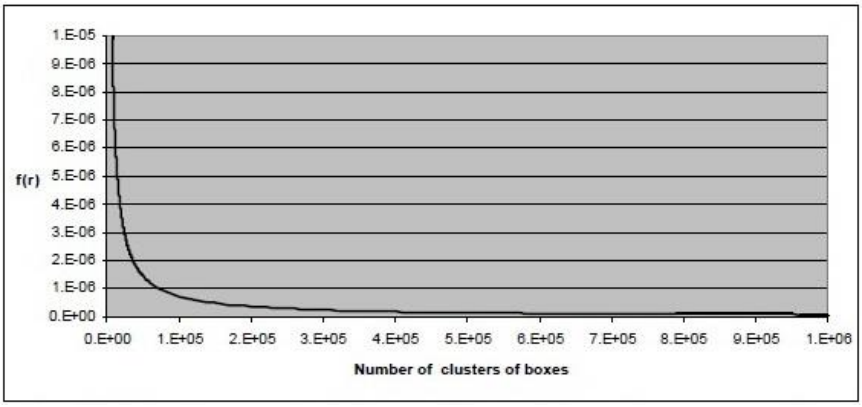
Therefore,

$$f(n) = \frac{\ln(1+\frac{1}{n})}{\ln(M+1)} \tag{10}$$

Like in the Zipf law, for integer n's, the relative frequency is not a function of β .We define a rank $r \equiv nN/P$where $r = 1, 2, 3, \ldots . R$. By defining the ranks we combined the boxes into clusters of boxes such that each cluster will contain $r = 1, 2, 3, \ldots . R$ groups of P/N particles. Therefore $r = 10$ means 10 times more particles than r= 1. We can repeat the calculation of the frequency again but instead of using n, we will use r, and obtain;

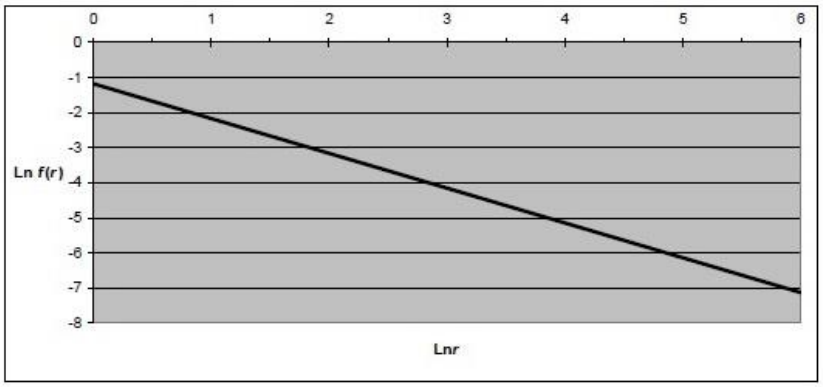
$$f(r) = \frac{\ln(1+\frac{1}{r})}{\ln(R+1)} \tag{11}$$

In Graph.1 The relative frequencies $f(r)$ for a set of R=10^6 clusters and r= 1,2,3,...., Raccording to Eq.(11) is plotted. A long tail distribution is demonstrated.



Graph 1. *A million clusters and their probabilities. The rank increases as its probability decreases.*

Eq.(11) "behaves" as a power law, this is so because a plot of the logarithm of the cluster *r* versus the logarithm of its probability yields a straight line as demonstrated for a million ranks.



Graph 3. *Log-Log plot of frequency versus the rank for R=million is a straight line.*

The Pareto's 20:80 rule of thumb was proved to be correct not only in wealth distribution but in many other phenomena as well. For example, it is believed that 20 percent of customers yields 80 percent of the revenue; 20 percent of the drivers cause 80 percent of the accidents; etc (Jurgan, 1951). In order to find the ratio obtained from Eq. (11) we divide the boxes into 10 ranks. Each rank contains 1, 2, 3,….,9, 10 equal groups of particles. We construct the table below from $f(r) = \frac{\ln(1+\frac{1}{r})}{\ln(11)}$

Table 1. *The relative frequencies of 10 ranks*

r	10	9	8	7	6	5	4	3	2	1
f(r)%	4	4.4	4.9	5.6	6.6	7.6	9.3	12	16.9	28.9

The total number of groups is $\sum_{r=1}^{10} r = 55$. However, the richest five ranks contain $\sum_{r=6}^{10} r = 40$ groups. Their total frequencies are $\sum_{r=6}^{10} f(r) = 25.5\%$, which means that about 73% of the packages are in the hands of about 25% of the boxes. This is a typical behavior of the Pareto's rule but with a small deviation from the empirical rule of thumb of 20:80, namely, a 25:75 rule.

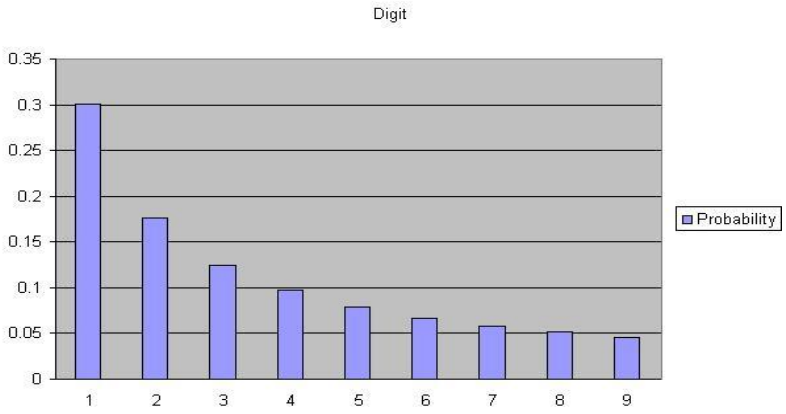
5.Benford's Law

Another application of Eq. (11) is Benford's law. Newcomb suggested Benford's law in 1881 from observations of the physical tear and wear of books containing logarithmic tables (Newcomb, 1881). Benford further explored the phenomenon in 1938, and empirically checked it for a wide range of numerical data. The main application of Benford's distribution is based on its existence in numerous random numerical files like financial data, street addresses, etc. Since one intuitively expects to obtain an even distribution of digits, as would be in the case of an unbiased lottery, some income tax authorities are looking at balance sheets for digit distributions in order to detect fraud detection. If the balance sheets don't fit to Benford's law, a further inspection is done (Nigrini, 1996).

In the derivation of Benford's law we assume that a digit is a box with n particles. This assumption is logical as a digit, unlike a word, has an absolute meaning as compared to other digits, exactly as the meaning of the number of particles in a box. There is a constraint though: the number of particles in a digit cannot exceed 9. The digit zero does not appear in Benford's law distribution of the first order. In Eq. (11) r may have any number. In digits, per definition, $r \leq 9$, therefore, it is legitimate to calculate the equilibrium distribution of the occupied boxes and to add as many empty boxes without affecting the distribution. In this case R is 9 and Eq. (11) yields the relative frequency,

$$f(r) = \frac{\ln\left(1+\frac{1}{r}\right)}{\ln(10)} = \log(1+\frac{1}{r}) \qquad (12)$$

This is the Benford's law.



Graph 3. *Benford's law predicts a decreasing frequency of first digits, from 1through 9.*

6. "Bell-like" distribution

Zipf law, Pareto's rule and Benford's lawoccurs where the number of particles is larger than the number of boxes. Hereafter, the case where $P<<N$ is considered.

In this case $n<<1$, we neglect the boxes that scored several particles, because, practically there are no such boxes. We want to find the probability

distribution of *N* boxes to score one particle. In this limit, $e^{\beta\phi} \gg 1$ and Eq. (7) can be approximated to,

$$n_i = e^{-\beta\phi_i} \tag{13}$$

Here n_i is the fraction of a particle in a box and the frequency ϕ_i=ϕ (n_i) is the probability to find this fraction. The total number of particles *P* is given by the same expression that we used in the Lagrange equation (5) namely,

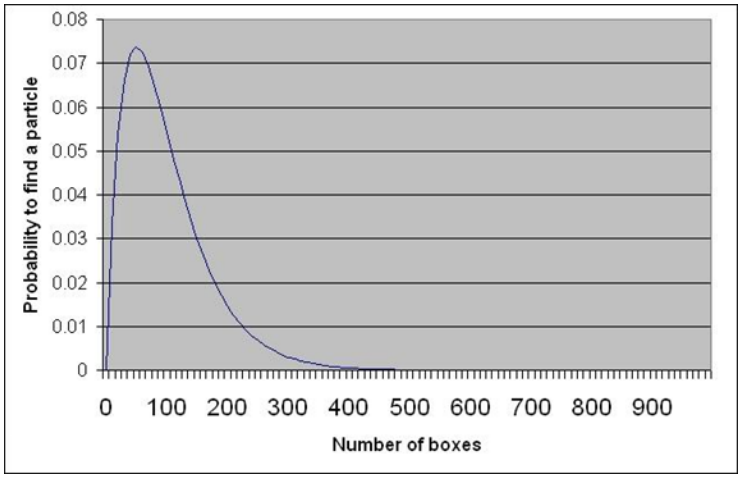
$$P = N\phi_i n_i = N\phi_i e^{-\beta\phi_i} \tag{14}$$

in the limit $\beta\rightarrow$ 0 one obtains that all the frequencies ϕ_i of the boxes are equal, namely ϕ_i=P/N. This is an even distribution. The even distribution isthe intuitive distribution that one expects to find in a distribution of particles in boxes. This distribution causes us to believe that uneven distributions are counterintuitive.

In the case where β is finite

$$P = \sum_{i=1}^{N} \phi_i e^{-\beta\phi_i} = \sum_{i=1}^{N} P(\phi_i, \beta) \tag{15}$$

$\frac{P(\phi_i,\beta)}{P}$is the relative probability to find a particle in a box. From Eq. (15) it is seen that $P(\phi_i, \beta)$ has two components, the first is the frequency ϕ_i of the fraction $n_j$of the particle and the second is the fraction of particles. As opposed to the case where *P*>>*N*, the frequency $\phi(n)$ itself is not the probability to find *n* particles but the probability to find a fraction of a particle. To find the probability of a single particle we have to multiply the frequency by the fraction of the particle namely $\phi_i n_i$. When the frequency increases the associate fraction of particles decreases exponentially with the frequency. The larger the β , the steeper is the decay. Since $P(\phi, \beta)$ is a linearly increasing function of ϕ_i multiplied by an exponentially decay function of ϕ_i , the distribution of particles in a box has a definite maximum.

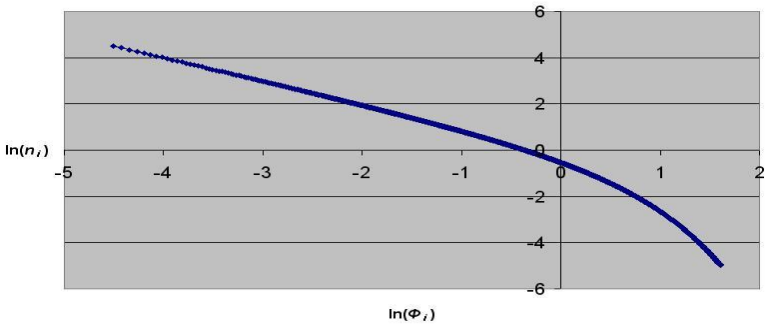


Graph 4. *The number of boxes and their probability to find a single particle for N=1000 and* $\beta = 1/50$

The maximum probability is obtained from $\frac{\partial P}{\partial \phi} = e^{-\beta\phi} - \beta\phi e^{-\beta\phi} = 0$ and is given by $\phi_{max} = 1/\beta$. In Graph (4) we see that the obtained curves is typical to the distributions of velocity of molecules, human age at death etc.

7. Discussion

The long tail distribution attracts a considerable attention because it is so ubiquitous [15]. Sometimes it is called a power law distribution and scale-free distribution. This is because a Log-Log presentation of the distribution yields a straight line as seen in Fig.2. When a power law fits are done, different slopes obtained for different statistics. For example, in Zipf law the ratio between the frequency of the 1st and the frequency of the 2nd is 2; in Pareto's rule and in Benford's law this ratio is about 1.7. Namely, in different regimes of P / N different "slopes" are obtained as is seen in Graph 5. Another notable point is that the normalized frequencies $f(n) f(n)$ for $P \gg N$ are not a function of β .This is with contradistinction to the case $P<<N$ in which the distribution is a function of β.



Graph 5. *A plot of* $\ln \phi$ *versus* $\ln n$*: for high values of n a "power law" decay is obtained, however for low values of n an exponential decay is obtained.*

The Lagrange multiplier β has a meaning. In ther modynamic the temperature is related to it via $T \propto \frac{1}{\beta}$.We see that in the case of Zipf law the frequency multiplied by the number of particles is proportional to the temperature. In the case of $n \ll 1$ the temperature is proportional to the frequency in which the probability to find a particle is the highest. This is the main difference between the long tail distribution and the "bell-like" distribution. In the long tail the temperature means the average wealth of a box. In the bell curve the temperature means the maximum probability.

8. Summary

The distribution of P non-interacting particles in N boxes iscalculated for a fair system. Since there is no preference to any configuration of particles and boxes, the entropy principle can be applied. It is shown that when the number of the particles is negligible as compared to the number of boxes, the "bell-like" distribution (which prefers the average) is obtained. However, when the number of particles is higher than the number of boxes, a long tail distribution is obtained. The obtained long tail distribution yields correctly Zipf law, Pareto's rule and Benford's law.

The Pareto's rule usually is conceived as an evolutionary law. Namely, the 20% of the drivers that cause 80% of the accidents are the bad drivers. Maybe the personality of these drivers is the reason for their excessive involvement in car accidents. Similarly, there might be good reasons for the fact that few people get rich and the majority remains poor. These kinds of questions cannot be answered by this kind of analysis. However, one should bear in mind that particles without personality, interactions or statistical bias are also distributed in the same way.

References

Bak, P. (1996). *How Nature Works: The Science of Self-Organized Criticality*, Springer-Verlag, New York.

Benford, F. (1938). The law of anomalous numbers, *Proceedings of the American Mathematical Society*, 78(4), 551-572.

Cohen, D., & Talbot K. (1984). Prime numbers and the first digit phenomenon, *Journal of Number Theory*, 18(3), 261-268. doi. http://dx.doi.org/10.1016/0022-314X%2884%2990061-1

Jurgan, J.M. (1951). *Quality Control Handbook*, McGraw-Hill Book Company.

Kafri, O. (2007). The second law as a cause for the evolution, *Unpublished Paper*, [Retrieved from].

Kafri, O. (2009). Entropy principle in direct derivation of Benford's law, *Unpublished Paper*, [Retrieved from].

Kafri, O. (2016). Sociological inequality and the second law, *Journal of Economic and Social Thought*, 3(4), xxx-xxx.

Kossovsky, A.E. (2012). Towards a better understanding of the leading digits phenomena, *Unpublished Paper*, [Retrieved from].

Miller, G.A. & Newman E.B. (1958). Tests of a statistical explanation of the rank-frequency relation for words in written English, *American Journal of Psychology*, 71(1), 209-218. doi. http://dx.doi.org/10.2307/1419208

Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers, *American Journal of Mathematics*, 4, 39-40.

Newman, M.E. (2006). Power-law, Pareto Distribution and Zipf's law, *Unpublished Paper*, [Retrieved from].

Nigrini, M. (1996). A taxpayer compliance application of Benford's law, Journal of the American Taxation Association, 18(1), 72-91.

Pareto, V. (1897). *Coursd' Economie Politique*. Droz, Geneva Switzerland, 1896, Rouge, Lausanne et Paris.

Planck, M. (1901). On the law of distribution of energy in the normal spectrum, *Annalen der Physik*, 4, 553-562.

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

SCAN ME