Principle of least effort vs. maximum efficiency: derivingZipf-Pareto's laws

Qiuping A. Wang

Invited talk at the 6th SPMCS conference, 15-19 January 2020, Huaqiao University, Xiamen, Chine

Principle of least effort vs. maximum efficiency: deriving Zipf-Pareto's laws

Qiuping A. Wang

Laboratoire Systèmes Complexes et Information Quantique

ESIEA, 9 Rue Vésale, 75005 Paris

IMMM, CNRS UMR 6283, Le Mans Université, Ave. O. Messiaen,

72085 Le Mans Cedex, France

Abstract

This paper provides a derivation of Zipf-Pareto laws directly from the principle of least effort. A probabilistic functional of efficiency is introduced as the consequence of an extension of the nonadditivity of the efficiency of thermodynamic engine to a large number of living agents assimilated to engines, all randomly distributed over their output. Application of the maximum calculus to this efficiency yields the Zipf's and Pareto's laws.

Keywords: Least effort, Maximum efficiency, Zipf's law, Pareto law, calculus of variation

PACS codes: 89.75.−k Complex systems; 89.75.Da Systems obeying scaling laws;
02.70.Rr General statistical methods; 02.30.Xx Calculus of variation

1) Introduction

Zipf's law is an empirical law describing the discrete distribution of some measured values *x* as a function of their rank *r*. If all the measured values are binned into *W* rank r in a decreasing order $(x_1 > x_2 > x_3 \ldots . > x_r \ldots > x_W)$, Zipf's law is given by [1][2]

$$x_r = \frac{x_1}{r^{\alpha}} \tag{1}$$

where α is a parameter to be determined experimentally. For many systems, language for example, $\alpha \approx 1$. For others it is quite different from unity. This regularity was first discovered by Jean-Baptiste Estoup [3] and then popularized by Zipf 30 years later.

Zipf gave a first interpretation of this law [2] by postulating that all human being minimizes effort in his activities to get some fulfillment. This rule was first formulated in 1894 by Guillaume Ferrero in his paper discussing the mental inertia of human being [4]. But Zipf was the first to explore its possible application to quantitative study. He wrote : "*The **power laws** in linguistics and in other human systems reflect an economical rule: everything carried out by human being and other biological entities must be done with least effort* (*at least statistically*)*".* This rule was obviously an intuition from the observation of the behaviors of human being himself and probably of other animals, always trying to get more done by doing less. A summary is given in [5] about several experimental works checking the rule of least effort.

The idea of least effort for human and animal systems is very appealing and even fascinating, especially in a perspective of using it for quantitative and analytic methods in the same way as many variational principles in physics (stationary action, least time, maximum entropy etc.). However, no such implementation has been realized to date to our knowledge. The relationship between this beautiful principle and the Zipf's law remains a sort of speculation without direct mathematical proof. In the past several decades, much attempt was made to interpret or derive this power law with different mechanisms and models. There are almost as many models proposed as the systems in which Zipf's law and near-Zipf's laws have been observed. The reader can easily find relative information. The most recent model is, to our knowledge, about the origin of the Zipf's law in language by considering the interaction between syntax and semantics [6].

Zipf's law is closely related to Pareto distribution, a power law originally describing the wealth distribution of a population in a given society [7]:

$$P(X > x) = \left(\frac{x_m}{x}\right)^{\beta} \tag{2}$$

where X is a random variable representing the income, $P(X > x)$ the probability of finding a person with income larger than a value x, x_m the smallest income and β a constant characterizing the distribution. This distribution law is the origin of the famous 20-80 rule of Pareto. It is believed that there is an intrinsic link between Zipf's and Pareto's laws. If one is observed in a system, another must exist simultaneously [5].

The aim of this work is to derive these laws in a generic way from the idea of least effort. For this purpose, a universal functional of effort is necessary for the minimum calculus of variation. However, an effort is a cost whose nature differs in different domains. It can be an expenditure of energy, time, information, an amount of money and so on. It is difficult to define and quantify an effort in a general manner. In this work, I focus on another relative quantity instead: the efficiency, often defined by useful output divided by the input, or effort to get that output. We can maximize the efficiency instead of minimizing effort. The maximum efficiency (MAXEFF) has double advantages: the first being minimizing effort for a given output, and the second being to maximize the output with given effort. MAXEFF seems to be a more general rule than least effort.

The idea of MAXEFF in science and engineering is not new. A good example is the derivation of the Betz limit of the efficiency of wind turbine from fluid mechanics[1] [8]. The essential of the application of MAXEFF is to use an expression of efficiency as a functional. I adopt this method and introduce a functional of efficiency by considering the nonadditive property of the thermodynamic efficiency and the fact that we are tackling a large number of engines (living agents), all distributed over some output. Then the application of the maximum calculus to this functional of probability distribution yields the Zipf's and Pareto laws.

2) The Nonadditivity of efficiency

The definition of efficiency of a thermal engine in thermodynamics differs from one type of engines to another. For example, suppose an work engine absorbs an energy Q_1, produces a useful work W, and rejects an energy Q_2. In the ideal case without energy loss where all heat cost $Q_1 - Q_2$ is converted into work W, we have $W = Q_1 - Q_2$. The efficiency of this engine is defined by

[1] https://www.youtube.com/watch?v=U_195_Uq3dU, https://en.wikipedia.org/wiki/Betz%27s_law

$$E = \frac{W}{Q_1} = 1 - \frac{Q_2}{Q_1} \tag{3}$$

For a heat pump engine (conditioner for heating for example) which absorbs a heat Q_1, consumes a work W, and produces a heat Q_2 for heating. We have $W = Q_2 - Q_1$ if all work is converted into heat. Its efficiency is defined by$E = \frac{Q_2}{W} = \frac{1}{1-\frac{Q_1}{Q_2}}$ or

$$\frac{1}{E} = \frac{W}{Q_2} = 1 - \frac{Q_1}{Q_2} \tag{4}$$

For a refrigerator (conditioner for cooling for example) which absorbs a heat Q_1, consumes a work W, and rejects a heat Q_2 for cooling, we have $W = Q_2 - Q_1$ if all work is converted into heat. Its efficiency is defined by $E = \frac{Q_1}{W} = \frac{1}{\frac{Q_2}{Q_1}-1}$ or

$$\frac{1}{E} = \frac{W}{Q_1} = \frac{Q_2}{Q_1} - 1 \tag{5}$$

The nonadditivity relationships of E and of $\frac{1}{E}$ are similar (see calculation in the Appendix). In what follows, I only give a summary of the nonadditivity of E for two working engines. Suppose two engines are connected in such a way that the first engine absorbs an energy Q_1, does a work W_1, and rejects an energy Q_2, and the second engine absorbs an energy Q_2, does a work W_2, and rejects an energy Q_3, one has $E_1 = \frac{W_1}{Q_1} = 1 - \frac{Q_2}{Q_1}$, $E_2 = \frac{W_2}{Q_2} = 1 - \frac{Q_3}{Q_2}$. The overall efficiency E of the ensemble of two engines is defined by $E = \frac{W_1+W_2}{Q_1} = 1 - \frac{Q_3}{Q_1}$. It is straightforward to calculate

$$E = E_1 + E_2 - E_1E_2. \tag{6}$$

If the engines cannot transform all the heat cost ($Q_1 - Q_2 = W_1$ for the first engine for example) due to energy loss (friction, vibration, heat conduction, heat radiation and so on), we can introduce a loss coefficient a in such a way to write (see Appendix)

$$E = E_1 + E_2 + aE_1E_2. \tag{7}$$

Formally, $a = -1$ corresponds to the case where the energy cost is totally converted into work. If $a < -1$, this is the case where the engines cannot transform all the heat cost into useful work, leading to a reduction of efficiency with respect to the case of $a = -1$. On the contrary, if $a > -1$, there is an enhancement of efficiency as if the collaboration created energy with respect to

the $a = -1$ case. This is possible only for processes in which the input and output are not energy or energy proportional quantities. This point will be discussed in the following section.

More complicated nonadditivity forms are possible for the more general case where, for example, the second engine does not consume all the heat Q_2 rejected by the first engine. A parameter can also be used for describing this partial collaboration (see Appendix). In any case, we can write $E = E_1 + E_2 + f(E_1, E_2)$ as a general nonadditive relationship.

3) Modeling an ensemble of thermal engines

Suppose an ensemble of a large number N of engines (or agents). The functioning of each engine is independent. They can however collaborate in the sense that the output of one engine can be, at least partially, the input of others. We do not consider the case where engines have no communication or collaboration between them. Each engine has a certain efficiency E_n with $n = 1,2 \dots N$. The total efficiency E of the ensemble should be a function of all E_n: $E = f(E_1, E_2 \dots E_N)$.

Efficiency is in general nonadditive, hence the total one $E = f(E_1, E_2 \dots E_N)$ cannot be a simple sum of the individual efficiencies. On the basis of the above analysis of the nonadditivity relationships for different type of engines (doing work, heating, cooling) and different collaboration type, we can model the whole system by using a simple nonadditivity given by Eq.(7) which also reads $(1 + aE) = (1 + aE_1)(1 + aE_2)$. This equation can be written as, for the whole system of *N* engines:

$$(1 + aE) = \prod_{n=1}^{N} (1 + aE_n) \tag{8}$$

where a is a parameter that characterizes collaboration between the engines as well as the energy loss during the processes from input to output. For many systems, a can be free from energy connection for non-thermodynamic processes for which the input and the output quantities are not energy connected and there is no necessarily energy conservation condition. For example, for agents trying to get connected to some objects (sites, friends, cities, richness etc.), the output can be frequency of connection, the population or the agents' richness. There is no energy conservation between these quantities and the inputs which can be energy cost, expenditure of time or money, used materials and so forth. Another example is the economic process of investment. This process is similar to the process of heating engine. The invested amount of money can be assimilated to the input heat Q_1, the consumed input work W is the

effort to increase the profit, and the total turnover can be considered as the heat production Q_2. In thermodynamics there would be a conservation condition: $Q_2 - Q_1 = W$. But this relationship does not exist for the economic process of investment because there is no quantitative measure of the effort W and of its conversion to Q_2.

Eq.(8) is obviously the simplest model for the efficiency as a nonadditive quantity. The mathematical advantage of this model will be shown later. A little bit more complicated model can be

$$(1 + aE) = \prod_{n=1}^{N}(1 + a_n E_n) \tag{9}$$

where a is the parameter characterizing the whole ensemble and a_n is the parameter of the nth subsystem or agent. This composite model can be used when it is necessary to consider the composite effect of subsystems. In what follows, I focus on the ensemble as a whole, the one parameter model Eq.(7) or (8) will be used.

4) Efficiency as a functional of probability

Suppose that all agents in the ensemble are making effort to achieve as much as possible a measurable quantity represented by a variable X having w discrete values x_i with $i = 1,2, \dots w$. More they get that quantity, larger is x_i. This quantity can be income, wealth, city population, firm size, frequency of events, and so forth. At equilibrium (or stationary) states of the whole systems, all agents are distributed over the range of X with n_i agents at the value x_i. We have $\sum_{n=1}^{w} n_i = N$. The probability p_i of finding an agents at the value x_i is $p_i = \frac{n_i}{N}$. The normalization condition is $\sum_{i=1}^{w} p_i = 1$.

Due to the statistical nature of the model with a large number of agents distributed over all the values of X, it is reasonable to suppose that the total efficiency E_i of the agents that have the value x_i depends on the number n_i with $E_i = f(n_i)$ or on the probability distribution $E_i = f(p_i)$. The average efficiency E of the whole system reads $E = \sum_{i=1}^{w} p_i E_i$.

Now let us separate the whole ensemble of agents into two *independent* subsystems A and B, with efficiency $E_k(A)$ and $E_j(B)$, respectively. The probability distribution of the agents in A is $p_k(A)$ and that in B is $p_j(B)$. The probability distribution of the whole ensemble can be written as

$$p_i = p_k(A)p_j(B) \text{ with } i = kj \tag{10}$$

We choose Eq.(7) as the efficiency nonadditivity. This implies a total efficiency given by

$$E_i = E_k(A) + E_j(B) + aE_k(A)E_j(B) \tag{11}$$

or $(1 + aE_i) = [1 + aE_k(A)]\left[1 + aE_j(B)\right]$. It can be proved that Eq.(10) and Eq.(11) together lead uniquely to $(1 + aE_i) = p_i^b$ or

$$E_i = \frac{p_i^b - 1}{a}. \tag{12}$$

Obviouly E_k and E_j have the same functional of p_k and p_j, respectively. The proof of the uniqueness of Eq.(12) will be given in another paper [9].

The parameter b is related to a by the following considerations. First, due to the fact that the efficiency E_i is positive and that p_i is smaller than unity, b should have opposite sign of a. Secondly, from Eq.(11), if a goes to the zero limit $a \to 0$, the efficiency tends to additive limit $E_i \to E_k(A) + E_j(B)$. Taking into account Eq.(10), the only possible relationship is $b = -a$, i.e.

$$E_i = \frac{p_i^{-a} - 1}{a} \tag{13}$$

which tends to $E_i = -\ln p_i$ as a tends to zero, allowing the additive efficiency relationship $E_i = -\ln p_k p_j = -\ln p_k - \ln p_j = E_k(A) + E_j(B)$. Finally, the average efficiency of the whole ensemble of N agents reads

$$E = \sum_{k=1}^{w} p_i E_i = \frac{\sum p_i^{1-a} - 1}{a} \tag{14}$$

in which the normalization $\sum p_i = 1$ is considered. From now on, if not specified, the summation is over all the w possible values (states) of X for the whole ensemble. It is noteworthy that the above discussion was made for the discrete case. The continuous case will introduce other conditions on the positivity of E [9].

5) Maximization of efficiency

As mentioned in the introduction, following the idea of least effort, I propose to maximize the efficiency, meaning that a calculus of variation applied to the functional of the average efficiency in Eq.(14) with respect to the probability p_i ($i = 1,2 \ldots w$). However, it is easy to verify that the maximization of E alone cannot lead to correct probability distribution. Indeed, $\delta E = 0$ means $\frac{\partial E}{\partial p_i} = 0$ for all i, leading to $p_i = \infty$ for $a > 0$ and $p_i = 0$ for $a > 0$. If we introduce the normalization as a constraint of the variation, $\delta(E + \sum p_i) = 0$ will lead to uniform distribution $p_i = 1/w$ for all i, which is of course not what we are looking for.

This is because the maximum efficiency is not an isolated property. The efficiency is at maximum when fulfillment is the best due to the effort of the agents. Hence there must a connection between the efficiency and the fulfillment or the output. It is quite reasonable to associate the maximum of efficiency to the maximum output. The output is represented by the variable *X*. As a consequence, its average, $\bar{x} = \sum p_i x_i$, should be maximized at the same time as the efficiency. As mentioned above, X can be the income, the frequency of events, the population of cities, the size of companies representing their wealth and so forth. In the case of income for example, if the maximum total efficiency is achieved, then the total income of the population should reach its maximum as well. In other words, the two maximums are mutually conditioned. The functional to be maximized should be the sum $(E + c\bar{X})$. We write now

$$\delta(E + c\bar{X}) = 0 \tag{15}$$

where c is a constant multiplier characterizing the balance between the two maximums.

6) Deriving Pareto law

Eqs.(14) and (15) means $\frac{\partial}{\partial p_i}\left(\frac{p_i^{1-a}-1}{a} + c p_i x_i\right) = 0$ for all i. After the normalization, the result is

$$p_i = C x_i^{-\frac{1}{a}} \tag{16}$$

where the normalization constant $C = 1/\sum x_i^{-\frac{1}{a}}$. Remember that p_i is the probability of finding an agent at the value x_i of *X*.

The continuous version of the discrete distribution Eq.(16) is the probability density function $\rho(x) = Cx^{-\frac{1}{a}}$ with $dp(x) = \rho(x)dx = Cx^{-\frac{1}{a}}dx$ the probability of finding an agent in the interval from x to $x + dx$. The Pareto law follows from the integral of *dp* from *x* to the maximum value of *X* or infinity for simplicity: $p(X > x) = \int_x^\infty Cx^{-\frac{1}{a}}dx = \frac{C}{\frac{1}{a}-1}x^{-(\frac{1}{a}-1)}$. Since $p(X > x_{min}) = 1$ with x_{min} the minimum value of *X*, one gets

$$p(X > x) = \left(\frac{x_{min}}{x}\right)^{\frac{1}{a}-1} \tag{17}$$

which is the Pareto distribution Eq.(2) with $\beta = \frac{1}{a} - 1$. $p(X > x)$ is a decreasing distribution, it follows that $0 < a < 1$ and $0 < \beta < \infty$.

7) Deriving Zipf's law

Now let us put the values of X into W bins. These bins are ranked in a decreasing order in the magnitude of x. Let x_r be the benchmark value of the bin of rank r, we have $x_1 > x_2 \ldots > x_r > \ldots > x_W$. The Zipf's law Eq.(1) describes the relationship between x_r and r. As mentioned above, the Zipf's law and the Pareto law are regarded as two sides of the same thing. In the literature [5], they are connected one to another by the hypothesis that the probability $p(X > x)$ is proportional to the rank r.

By definition, the population (number of agents) having more income than x_r increases with increasing r. In other words, $p(X > x_r)$ increases when r increases until the maximum W at which X reaches its minimum value x_W and $p(X > x_W) = 1$. But saying that the probability $p(X > x)$ is always proportional to the rank r seems to be just an observation from some empirical results. It would hard to say it is a general rule. In what follows, I will use MAXEFF to derive a general relationship of $p(X > x_r)$ to r leading to Zipf's law from Pareto law Eq.(17).

Let us now use the average of the rank in the MAXEFF. Notice that if an agent increases its income X, its rank value decreases. Hence whenever $\bar{x} = \sum p_i x_i$ has a maximum, the average rank $\bar{r} = \sum p_r r$ should reach its minimum where $p_r = C x_r^{-\frac{1}{a}}$ given by Eq.(16). The calculus of variation applies with $\delta(E - c'\bar{r}) = 0$ which should be maximum because $(E - c\bar{r})$ is a difference between the maximum $E = \frac{\sum_r p_r^{1-b} - 1}{b}$ (b is not necessarily equal to a) and the minimum $\bar{r}$. This leads to $\frac{\partial}{\partial p_r}\left(\frac{p_r^{1-b}-1}{b} - c' p_r r\right) = 0$ and $p_r = C' r^{-\frac{1}{b}}$. By definition of rank distribution, p_r must be increasing function of r, hence b must be negative. For simplicity, let $\gamma = -\frac{1}{b} > 0$, we have

$$p_r = C' r^{\gamma} \tag{18}$$

Substituting this equation into $p_r = C x_r^{-\frac{1}{a}}$ gives $C' r^{\gamma} = C x_r^{-\frac{1}{a}}$ and $x_r = \left(\frac{C'}{C}\right)^{-a} r^{-a\gamma}$. Notice that $x_1 = \left(\frac{C'}{C}\right)^{-a}$, Zipf's law reads

$$x_r = \frac{x_1}{r^{\alpha}} \tag{19}$$

with $\alpha = a\gamma$. An example of this relationship comes from the Zipf-Pareto distributions of American city populations [5]. The Pareto distribution Eq.(17) shows $\beta = 1.366$, leading to

$a = 0.423$. While the Zipf's distribution shows $\alpha = 0.823$, meaning that $\gamma = 1.95$ or $p_r = C' r^{1.95}$ for the system of city population.

8) Zipf-Pareto efficiency as a measure of performance

The efficiency given by Eq.(14) (from now on we refer to it as Zipf-Pareto or ZP efficiency) provides a possible measure of performance of the ensemble of agents as a whole all making effort for some fulfillment. Using the probability density function $\rho(x) = \frac{1}{Z} x^{-\frac{1}{a}}$, the efficiency reads:

$$E = \frac{\int_{x_{min}}^{\infty} \rho^{1-a}\, dx - 1}{a} \tag{20}$$

The partition function $Z = \int_{x_m}^{\infty} x^{-\frac{1}{a}} dx = \frac{1}{\beta} x_{min}^{-\beta}$. Choosing $x_{min} = 1$, we get $Z = \frac{1}{\beta} = \frac{a}{1-a}$. Since $0 < a < 1$ and $0 < \beta < \infty$, the partition function is increasing function of a in the interval $0 \leq Z < \infty$. The average of x is given by $\bar{x} = \int_{x_{min}}^{\infty} x\rho\, dx = \frac{\beta}{\beta - 1} = \frac{1-a}{1-2a}$ which increases with increasing a up to infinity for $a = 0.5$ and becomes negative for $0.5 < a \leq 1$. If we impose the condition of $\bar{x} \geq 0$, then $0 < a \leq 0.5$ and $1 < \beta < \infty$.

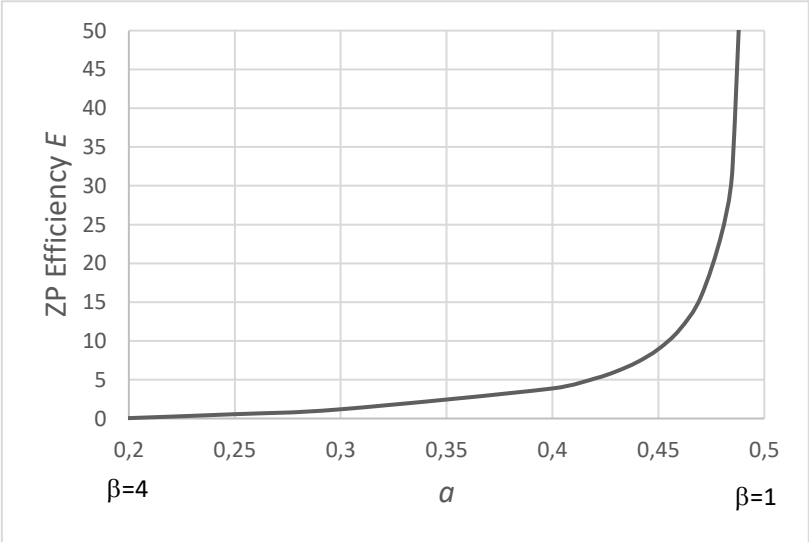

Figure 1: Evolution of the efficiency of Eq.(20) as a function of the parameter a in the interval $0.2 < a < 0.5$ or $4 < \beta < 1$. E diverges for a=0.5 or $\beta = 1$.

Finally, the integral in Eq.(20) gives $E = \frac{Z^a \bar{x} - 1}{a} = \frac{1}{a}\left[\left(\frac{a}{1-a}\right)^a \frac{1-a}{1-2a} - 1\right]$ which increases from zero to infinity with increasing a in the interval $0 \leq a < 0.5$. . It is worth noticing that the ZP efficiency increases when $\beta = \frac{1}{a} - 1$ decreases from infinity (a=0) to 1 (a=0.5). This

evolution of the ZP efficiency is plotted in **Figure 1** where only positive efficiency is shown for $\sim 0.2 \leq a$.

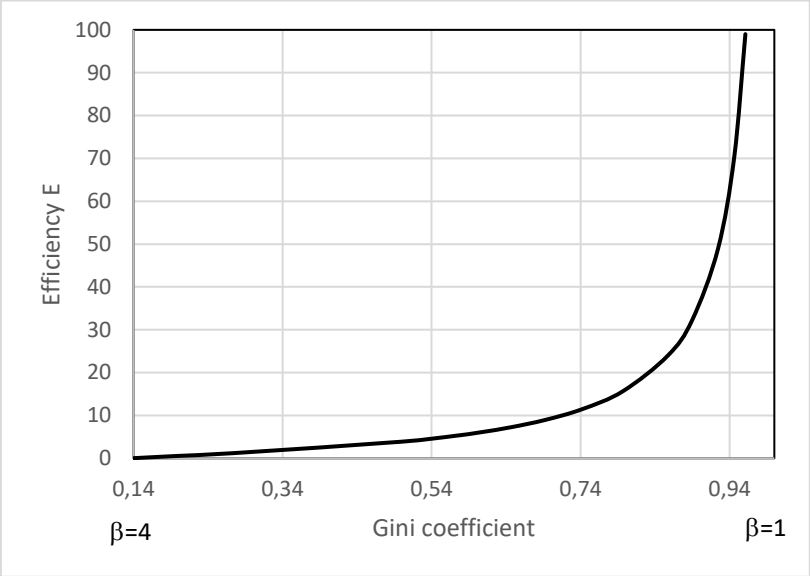

Figure 2: Relationship between the ZP efficiency and the Gini coefficient G in the interval of $4 < \beta < 1$. E diverges for G=1 or $\beta = 1$.

In order to help the understanding the ZP efficiency, let's compare it to another known properties of Zipf-Pareto distributions, the Gini coefficient G, an indicator of inequality for given population in many fields such as economy, sociology, biology, education, health science, ecology and so on[2]. For Pareto distribution, $G = \frac{1}{2\beta - 1}$, meaning that $G = 0$ for $\beta = \infty$ (a=0), and $G = 1$ for $\beta = 1$ (a=0.5). $G = 0$ implies absolute equality (all agents have the same income for example), and $G = 1$ indicates absolute inequality (one person has all the income and all others have zero income). Clearly, the ZP efficiency increases with increasing G and is very large when G approaches its maximum (see **Figure 2**). This mathematical behavior of the ZP efficiency coincides with the classical point of view in economics [10]. Of course, in a real system, growth may deviate from this ideal behavior and undergoes the impacts of different factors. It is convincing to say that an economic system is most efficient when only one hyper-trillionaire has all the wealth and the rest of the population has nothing, or that an educational system has maximal efficiency with only one person very educates and the rest of the population without any education (G=1). The relationship between wealth inequality and economic growth is an important issue and widely debated until nowadays[3]. In order to appreciate fully the

[2] https://en.wikipedia.org/wiki/Gini_coefficient
[3] https://en.wikipedia.org/wiki/Economic_growth

properties of the ZP efficiency, it is necessary to make further investigation of its relationship with other properties such as inequality, 80-20 rule, 20/20 ratio, performance, opportunity etc.) by either empirical study or numerical modeling of real systems [11].

9) Conclusion and remarks

I have implemented the idea of least effort via a calculus of variation MAXEFF in order to derive Zipf-Pareto laws. I introduce a functional of the efficiency from the consideration of a nonadditive relationship of efficiency of thermal engines in thermodynamics. The Zipf's and Pareto's laws come out naturally from this maximum calculus. This efficiency functional also provides a possible measure of the performance of real systems.

One of the underlying meaning of this approach is that Zipf-Pareto laws are ubiquitous for all systems composed of a large number of agents, no matter what are their nature and behavior, human beings, animals or objects recipient of effort and representing fulfillment, whenever they try to achieve something or become objects of effort, Zipf-Pareto laws take place. As objects recipient of effort or fulfillment, one can think about words, webpages, cities, firms, books, phone numbers and so on. From this universality of power laws, we understand why the Pareto's 80-20 rule or similar distributions happen so often everywhere whenever an effort is involved to achieve or to produce some quantities.

I used the nonadditivity Eq.(7) or (8) of the efficiency of the heat engines. The reader can notice in the Appendix that, for heating or cooling engine, similar additivity occurs for the inverse of the efficiency $1/E$. This means that $1/E$, instead of E, has a functional similar to Eq.(14). This modification of nonadditivity does not affect the calculus of variation of least effort and maximum efficiency, because if E has a maximum, $1/E$ should have a minimum, and the calculus of variation $\delta\left(\frac{1}{E}+c\bar{X}\right)=0$ will generate the same probability distribution if $1/E$ has the same functional as E. Therefore, the result of the MAXEFF is independent of the type of the engines in the ensemble.

It is worth noticing that the result of MAXEFF in the present work is a single power law, showing a straight-line in the log-log plot. In practice, most systems claimed to have Zipf-Pareto distributions only show near-Zipf's laws with different (curved down or up) tails [12]. It is possible to account for this behavior within this framework by separating the ensemble of agents into two or more subsystems which differ in behavior as well as in the parameter a. This composite approach is to be developed in another work by using the technique of incomplete

statistics [14]. It is also possible to explain the exponential distributions of certain complex systems [13] since the efficiency functional approaches logarithm form for small *a*. The MAXEFF can generate near-exponential probability distribution in this case.

The resemblance of the ZP efficiency to Tsallis entropy[4] [14][15] is noteworthy. This comes from the fact that the ZP efficiency has the same nonadditivity Eq.(7) as Tsallis entropy. However, for Tsallis entropy, this nonadditivity is a consequence of the postulated entropy [15]. But here, the nonadditivity, as suggested by the efficiency of thermal engines, is the starting point which uniquely leads to ZP efficiency. The parameter *a* has a concrete physical meaning as shown in the Appendix. The maximization of Tsallis entropy has its origin in the Jaynes principle[5] [16] stipulating the maximization of Boltzmann entropy for thermodynamic systems, while the maximum efficiency arises from the principle of least effort for living systems which is well illustrated by the dictum "achieving more by doing less" that each of us is applying in every detail of our life.

Acknowledgement

The author thanks Dr Aziz El Kaabouchi, Dr Congjie Ou, Dr Ru Wang, and Dr François-Xavier Machu for valuable discussion and help in the calculation.

[4] https://en.wikipedia.org/wiki/Tsallis_entropy
[5] https://en.wikipedia.org/wiki/Principle_of_maximum_entropy

References

[1] G.K. Zipf, Selected Studies of the Principle of Relative Frequency in Language. Cambridge, MA: Harvard University Press (1932).

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

[3] J.B. Estoup, Gammes sténographiques : méthode et exercices pour l'acquisition de la vitesse, 4e édition rev. et aug., 151p., Paris (1916).

[4] G. Ferrero, L'inertie Mentale et la Loi du Moindre Effort, Philosophical Review **3** (1894)362 ; Revue Philosophique de la France Et de l'Etranger **37**(1894)169

[5] Y.Y. Zhu, Q.A. Wang, W. Li, X. Cai, The principle of least effort and Zipf distribution, J. Phys.: Conf. Ser. **1113**(2018)012007

[6] S. Lestrade, Unzipping Zipf's law, PLOS ONE, **12**(2017)e0181987.

[7] V. Pareto, Cours d'economie politique, Journal of Political Economy. **6** (1898)

[8] A. Betz, Introduction to the Theory of Flow Machines. (D. G. Randall, Trans.) Oxford: Pergamon Press (1966)

[9] A. El Kaabouchi, F.X. Machu, R. Wang, Q.A. Wang, About the efficiency function implementing the principle of least effort, to be submitted

[10] A. Berg; J. Ostry, Inequality and Unsustainable Growth: Two Sides of the Same Coin". IMF Economic Review. **65**(2017)792–815

[11] D. Dollar, D. T. Kleineberg and A. Kraay, Growth, Inequality and Social Welfare: Cross-Country Evidence, Economic Policy, **30**(2015)335-377

[12] M. Cristellii, M. Battyy, L. Pietroneroo, There is More than a Power Law in Zipf, Scientific Reports, **2**(2012)812

[13] W.B. Deng, W. Li, X. Cai, Q.A. Wang, The exponential degree distribution in complex networks: non-equilibrium network theory, numerical simulation and empirical data, Physica A, **390**(2011)1481-1485

[14] Q.A. Wang, Nonextensive statistics and incomplete information, Euro. Phys. J. B, **26**(2002)357

[15] C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics. Journal of Statistical Physics, **52**(1988)479–487

[16] E.T. Jaynes, Information Theory and Statistical Mechanics. *Physical Review*. Series II, **106**(1957)620–630

Appendix

Analysis of the nonadditivity of the efficiency of thermal engines.

A) Nonadditivity of the efficiency of heat engine doing work

The definition of efficiency of a thermal engine in thermodynamics differs for one type of engines to another. For example, suppose an engine absorbs an energy Q_1, does a useful positive work *W*, and rejects an energy Q_2. In the ideal case without energy loss where all heat cost $Q_1 - Q_2$ is converted into work *W*, we have $W = Q_1 - Q_2$. the efficiency of this engine is defined by

$$E = \frac{W}{Q_1} = 1 - \frac{Q_2}{Q_1}$$

A1) If two engines are connected in such a way that the first engine absorbs an energy Q_1, does a work W_1, and rejects an energy Q_2, and the second engine absorbs the energy Q_2, does a work W_2, and rejects an energy Q_3, one has $E_1 = 1 - \frac{Q_2}{Q_1}$, $E_2 = 1 - \frac{Q_3}{Q_2}$, and the efficiency *E* of the ensemble of two engines is given by

$$E = \frac{W_1 + W_2}{Q_1} = 1 - \frac{Q_3}{Q_1} = 1 - \frac{Q_2}{Q_1}\frac{Q_3}{Q_2} = 1 - (1 - E_1)(1 - E_2) = E_1 + E_2 - E_1E_2$$

A2) Now if the second engine only absorbs a part of the energy Q_2, say, bQ_2 (b<1), does a work W_2., and rejects an energy Q_3, one has $E_1 = 1 - \frac{Q_2}{Q_1}$, $E_2 = \frac{W_2}{bQ_2} = \frac{bQ_2 - Q_3}{bQ_2} = 1 - \frac{Q_3}{bQ_2}$, and the overall efficiency *E* of the ensemble of two engines:

$$E = \frac{W_1 + W_2}{Q_1} = \frac{Q_1 - Q_2 + bQ_2 - Q_3}{Q_1} = \frac{(b-1)Q_2 + Q_1 - Q_3}{Q_1} = \frac{(b-1)Q_2}{Q_1} + 1 - \frac{Q_3}{Q_1}$$

$$= (b-1)(1 - E_1) + 1 - b(1 - E_1)(1 - E_2)$$

$$= E_1 + bE_2 - bE_1E_2$$

A3) Now suppose not all heat $Q_1 - Q_2$ is converted into work *W* due to some energy loss (friction, thermal radiation, vibration etc.), we can write

$W = \frac{1}{a}(Q_1 - Q_2)$ and $E_1 = \frac{1}{a}\left(1 - \frac{Q_2}{Q_1}\right)$, $E_2 = \frac{1}{a}\left(1 - \frac{Q_3}{Q_2}\right)$,

where a>1 characterizes the loss of energy of the engines,

$$E = \frac{W_1 + W_2}{Q_1} = \frac{1}{a}\left(1 - \frac{Q_3}{Q_1}\right) = \frac{1}{a}\left(1 - \frac{Q_2}{Q_1}\frac{Q_3}{Q_2}\right)$$

$$= \frac{1}{a}[1 - (1 - aE_1)(1 - aE_2)]$$

$$= E_1 + E_2 - aE_1E_2$$

or

$$(1-aE)=(1-aE_1)(1-aE_2)$$

A4) If the two engines have different loss coefficients, say, a_1 and a_2, we have $E_1=\frac{1}{a_1}\left(1-\frac{Q_2}{Q_1}\right)$, $E_2=\frac{1}{a_2}\left(1-\frac{Q_3}{Q_2}\right)$, then

$$E=\frac{W_1+W_2}{Q_1}=\frac{1}{a}\left(1-\frac{Q_3}{Q_1}\right)=\frac{1}{a}\left(1-\frac{Q_2}{Q_1}\frac{Q_3}{Q_2}\right)$$

$$=\frac{1}{a}[1-(1-a_1E_1)(1-a_2E_2)]$$

One gets

$$(1-aE)=(1-a_1E_1)(1-a_2E_2)$$

B) Nonadditivity of the efficiency of heat pump

The definition of efficiency of a heat pump (heating engine) in thermodynamics is the following. Suppose heat pump absorbs a heat Q_1, consumes a work *W*, and provides a heat Q_2 for heating. We have $W=Q_2-Q_1$ if all work is converted into heat. Its efficiency is defined by

$$E=\frac{Q_2}{W}=\frac{1}{1-\frac{Q_1}{Q_2}}$$

$$\frac{1}{E}=\frac{W}{Q_2}=1-\frac{Q_1}{Q_2}$$

B1) If two pumps are connected in series in such a way that the first pump absorbs an energy Q_1, uses a work W_1, and supplies Q_2, and the second engine absorbs Q_2, consumes a work W_2, and supplies Q_3, one has $\frac{1}{E_1}=1-\frac{Q_1}{Q_2}$, $\frac{1}{E_2}=1-\frac{Q_2}{Q_3}$, and the overall efficiency *E* of the ensemble of two engines:

$$\frac{1}{E}=\frac{W_1+W_2}{Q_3}=1-\frac{Q_1}{Q_3}=1-\frac{Q_1}{Q_2}\frac{Q_2}{Q_3}$$

$$=1-\left(1-\frac{1}{E_1}\right)\left(1-\frac{1}{E_2}\right)$$

$$=\frac{1}{E_1}+\frac{1}{E_2}-\frac{1}{E_1}\frac{1}{E_2}$$

or

$$\left(1-\frac{1}{E}\right)=\left(1-\frac{1}{E_1}\right)\left(1-\frac{1}{E_2}\right)$$

B2) If not all the work W is converted into heat $Q_1 - Q_2$ due to some loss, let $Wa = (Q_1 - Q_2)$

$$E_1=\frac{Q_2}{W_1}=a\frac{Q_2}{Q_2-Q_1}=\frac{a}{\left(1-\frac{Q_1}{Q_2}\right)}$$

$$\frac{1}{E_1}=\frac{1}{a}\left(1-\frac{Q_1}{Q_2}\right)$$

and

$$\frac{1}{E_2}=\frac{1}{a}\left(1-\frac{Q_2}{Q_3}\right)$$

where a<1 characterizes the loss of heat energy of the engines.

$$\frac{1}{E}=\frac{W_1+W_2}{Q_3}=\frac{1}{a}\left(1-\frac{Q_1}{Q_3}\right)$$

$$=\frac{1}{a}\left(1-\frac{Q_2}{Q_1}\frac{Q_3}{Q_2}\right)=\frac{1}{a}\left[1-\left(1-a\frac{1}{E_1}\right)\left(1-a\frac{1}{E_2}\right)\right]$$

$$=\frac{1}{E_1}+\frac{1}{E_2}-a\frac{1}{E_1}\frac{1}{E_2}$$

Or

$$\left(1-a\frac{1}{E}\right)=\left(1-a\frac{1}{E_1}\right)\left(1-a\frac{1}{E_2}\right)$$

B3) If now the two pumps have different loss coefficients, say, a_1 and a_2, we have

$$\frac{1}{E_1}=\frac{1}{a_1}\left(1-\frac{Q_1}{Q_2}\right)$$

$$\frac{1}{E_2}=\frac{1}{a_2}\left(1-\frac{Q_2}{Q_3}\right)$$

then

$$\left(1-a\frac{1}{E}\right)=\left(1-a_1\frac{1}{E_1}\right)\left(1-a_2\frac{1}{E_2}\right)$$

C) Nonadditivity of the efficiency of refrigerator

The definition of efficiency of a refrigerator (cooling engine) is the following. Suppose a refrigerator absorbs a heat Q_1, consumes a work W, and rejects a heat Q_2 for cooling. We have $W = Q_2 - Q_1$ if all work is converted into heat. Its efficiency is defined by

$$E = \frac{Q_1}{W} = \frac{1}{\frac{Q_2}{Q_1} - 1}$$

$$\frac{1}{E} = \frac{W}{Q_1} = \frac{Q_2}{Q_1} - 1$$

C1) If two refrigerators are connected in such a way that the first one absorbs an energy Q_1, uses a work W_1, and rejects Q_2, and the second one absorbs Q_2, consumes a work W_2, and rejects Q_3, one has $\frac{1}{E_1} = \frac{Q_2}{Q_1} - 1$, $\frac{1}{E_2} = \frac{Q_3}{Q_2} - 1$, and the overall efficiency E of the ensemble of two engines reads:

$$\frac{1}{E} = \frac{W_1 + W_2}{Q_1} = \frac{Q_3}{Q_1} - 1 = \frac{Q_3}{Q_2}\frac{Q_2}{Q_1} - 1$$

$$= \left(\frac{1}{E_1} + 1\right)\left(\frac{1}{E_2} + 1\right) - 1$$

$$= \frac{1}{E_1} + \frac{1}{E_2} + \frac{1}{E_1}\frac{1}{E_2}$$

or

$$\left(\frac{1}{E} + 1\right) = \left(\frac{1}{E_1} + 1\right)\left(\frac{1}{E_2} + 1\right)$$

C2) In case of loss with a coefficient a, we have $E_1 = \frac{1}{a}\left(\frac{Q_2}{Q_1} - 1\right)$, $E_2 = \frac{1}{a}\left(\frac{Q_3}{Q_2} - 1\right)$,

$$\frac{1}{E} = \frac{1}{E_1} + \frac{1}{E_2} + a\frac{1}{E_1}\frac{1}{E_2}$$

$$\left(\frac{a}{E} + 1\right) = \left(\frac{a}{E_1} + 1\right)\left(\frac{a}{E_2} + 1\right)$$

C3) In case of loss with two different coefficients a_1 and a_2, we have $E_1 = \frac{1}{a_1}\left(\frac{Q_2}{Q_1} - 1\right)$, $E_2 = \frac{1}{a_2}\left(\frac{Q_3}{Q_2} - 1\right)$, one gets

$$\left(\frac{a}{E} + 1\right) = \left(\frac{a_1}{E_1} + 1\right)\left(\frac{a_2}{E_2} + 1\right)$$