



**HAL**  
open science

# Uniqueness of the efficiency functional for deriving the Zipf and the Pareto laws from the principle of least effort

Aziz El Kaabouchi, François-Xavier Machu, Jeremy Cocks, Qiuping A. Wang

## ► To cite this version:

Aziz El Kaabouchi, François-Xavier Machu, Jeremy Cocks, Qiuping A. Wang. Uniqueness of the efficiency functional for deriving the Zipf and the Pareto laws from the principle of least effort. 2022. hal-03843384

**HAL Id: hal-03843384**

**<https://hal.science/hal-03843384>**

Preprint submitted on 8 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Uniqueness of the efficiency functional for deriving the Zipf and the Pareto laws from the principle of least effort

A. EL Kaabouchi<sup>3</sup>, F.X. Machu<sup>1</sup>, J. Cocks<sup>1</sup>, Q.A. Wang<sup>1,2\*</sup>

<sup>1</sup>Laboratoire SCIQ, ESIEA, 9 Rue Vésale, 75005 Paris, France

<sup>2</sup>IMMM, CNRS UMR 6283, Le Mans Université, Ave. O. Messiaen, Le Mans, France

<sup>3</sup>ESTACA, Parc universitaire Laval-Changé, Rue Georges Charpak, Laval, France

## Abstract

In a previous work, we applied the principle of least effort to derive the Zipf and the Pareto power law distributions using a calculus of variation and an efficiency functional. This functional was arrived at by considering living systems containing a great number of agents all trying to achieve something with effort, similarly to thermal engines producing work from source energy, and a nonadditive relationship of efficiency in thermodynamics. In the present work, we provide a complete proof of the uniqueness of this efficiency functional, thus confirms the intrinsic link between the power laws and the principle of least effort.

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

\*Corresponding author: alexandre.wang@esiea.fr

## 1) Introduction

Achieving more by doing less is a common rule in most if not all living systems<sup>1</sup>. Nowadays we refer to this rule as the principle of least effort (PLE) coined by Ferrero [1]. The connection of PLE to the power law distributions widely observed in living systems goes back to a remark of Zipf : *The power laws in linguistics and in other human systems reflect an economical rule: everything carried out by human beings and other biological entities must be done with least effort (at least statistically)* [2][3]. This connection has inspired a lot of work towards deriving mathematically the Zipf and the Pareto power laws [2][3][4] from PLE [5][6]. Nevertheless, this derivation has been slow in coming. The main obstacle has been the mathematical definition of the key quantities such as achievement and effort, both being process-dependent and system-specific, and impossible to define in a general way [7][8].

In a recent work [7][8], we proposed to derive the Zipf and the Pareto power laws using the principle of maximum efficiency (PME) which can be considered another side of PLE but is easier to implement mathematically with the calculus of variation. Here the efficiency is defined as the ratio of achievement to effort, similar to the efficiency  $\eta = \frac{W}{Q}$  of a thermal engine defined by the ratio of the output work  $W$  (achievement) to the input heat  $Q$  (effort). For a system containing a large number of agents all distributed randomly over a given number  $W$  of states, each state having a given efficiency  $\eta_i$ , the overall efficiency  $\eta$  of the system can be described by the statistical average of all  $\eta_i$ , i.e.  $\eta = \sum_{i=1}^W p_i \eta_i$ , where  $p_i$  is the probability to find an agent at the state  $i$  (a natural number) [7]. The general formula for this efficiency is given by [7]

$$\eta = \frac{\sum_{i=1}^W p_i^{1-a} - 1}{a}. \quad (1)$$

where  $a$  is a real parameter in the following nonadditive property of thermodynamic efficiency

$$\eta_{kj}(C) = \eta_k(A) + \eta_j(B) + a\eta_k(A)\eta_j(B). \quad (2)$$

---

<sup>1</sup> We use the term “living systems” to represent the living agents such as human beings, animals, insects and the systems used or driven by these living agents, such as linguistic, social, economic, educational, communicational systems etc.

Eq.(2) relates the joint efficiency  $\eta_{kj}(C)$  of a composite system  $C$  to the efficiencies of its subsystems  $A$  and  $B$ , where  $\eta_k(A)$  is the efficiency of an agent at a state  $k$  in  $A$ ,  $\eta_j(B)$  is the efficiency of an agent at a state  $j$  in  $B$ , and  $\eta_{kj}(C)$  is the joint efficiency of an agent at the joint state  $kj$  in  $C$ .

Eq.(2) must be satisfied by any two agents at the state  $k$  in  $A$  and the state  $j$  in  $B$  when they are connected in series to form the joint agent in  $C$  [7]. The origin of this nonadditivity is in thermodynamics where Eq.(2) with  $a = -1$  must be satisfied by any two engines connected in series to form a joint engine [9]. For living systems where the involved quantities (achievement and effort) are in general non-energy connected and non-conservative (such as money, time, physical effort, mental effort and so on),  $a$  is allowed to be different from -1 [7][8].

In our previous work, the two subsystems  $A$  and  $B$  were supposed to be statistically independent, i.e.

$$p_{kj}(C) = p_k(A)p_j(B) \quad (3)$$

relating the joint probability  $p_{kj}(C)$  of finding an joint agent in  $C$  at the joint state  $kj$  to the probabilities  $p_k(A)$  and  $p_j(B)$  of finding agents in the two subsystems  $A$  and  $B$  at the states  $k$  and  $j$ , respectively.

The efficiency functional in the form of Eq.(1) is crucial for the successful derivation of the Zipf and the Pareto laws as well as the Pareto rule of 20-80 from PLE and PME using variational calculus [7]. It is hence important to prove the uniqueness of the functional using Eq.(2) and Eq.(3) as basic axioms. However, in the previous work [8], we have only proved that Eq.(1) is the simplest form of all the possible functionals satisfying Eq.(2) and Eq.(3), meaning that other more complicated forms of the efficiency functional may exist and not necessarily lead to the power laws we want to derive using PLE and PME.

In this work, we provide a complete proof of the uniqueness of Eq.(1) given Eq.(2) and Eq.(3). The aim is to demonstrate that, if living agents can be considered as thermal engines doing work (achievement) using source energy (effort), then the ubiquitous power laws and the Pareto rule of 20-80 are necessarily a consequence of PLE and PME.

## 2) Uniqueness of the efficiency functional

We propose the following axioms for the efficiency functional.

- I) The efficiency  $\eta_i$  of an agent at the state  $i$  is a continuous function with respect to its

argument  $p_i$ , with  $\eta_i = 0$  if  $p_i = 0$  (no agent at the state).

II)  $\eta_{kj}(C) = \eta_k(A) + \eta_{kj}(B) + a\eta_k(A)\eta_{kj}(B)$  or equivalently

$$1 + a\eta_{kj}(C) = [1 + a\eta_k(A)][1 + a\eta_{kj}(B)]$$

In what follows, for simplicity, let  $z$  be a value of the joint probability of  $C$ ,  $x$  be a value of the probability of  $A$ , and  $y$  a value of the conditional probability of  $B$  given  $A$ , then Eq.(3), the statistical independence between the two subsystems  $A$  and  $B$ , can be written as  $z = xy$ . If now we define  $f(x) = 1 + a\eta(x)$ , then the axiom II reads  $f(z) = f(x)f(y)$  or  $f(xy) = f(x)f(y)$ .

**Theorem 1:**

The functional satisfying the axioms I and II is uniquely in the form  $\eta_i = \frac{p_i^{-a}-1}{a}$ .

Theorem 1 is to be proven through the following lemmas.

**Lemma 1:**

If  $f$  is a continuous function of a single variable ( $x$  or  $y$ ) on the  $\sigma$ -algebra  $B(\mathbb{R}^+)$  satisfying the condition

$\forall x \in \mathbb{R}^+, \forall y \in \mathbb{R}^+, f(xy) = f(x)f(y)$ , then there exists  $b \in \mathbb{R}$  such that

$$f(x) = \begin{cases} x^b & \text{if } x > 0 \\ 1 & \text{if } x = 0 \end{cases} .$$

**Proof**

$\forall x > 0, f(x) = f(\sqrt{x}\sqrt{x}) = f(\sqrt{x})^2$ , thus  $f(x) \geq 0$ .

Moreover if there exists a  $y$  in  $\mathbb{R}^+$  such that  $f(y) = 0$ , we have  $f(x) = f\left(y\frac{x}{y}\right) = f(y)f\left(\frac{x}{y}\right) = 0$ , leading to  $f(x) = 0$  everywhere, which is contradictory to the definition of  $f(x)$  being a function of  $x$ . Thus  $f(x) > 0$ .

If  $x = 0$ , we have  $f(0) = f(00) = f(0)f(0) = (f(0))^2$ , implying  $f(0) = 1$ , which is in accordance with the definition  $f(x) = 1 + a\eta(x)$  since  $\eta(0) = 0$  according to Axiom I.

We now define:  $\forall x \in \mathbb{R}_+^*, g(x) = \ln(f(e^x))$ ,  $g$  being continuous on  $\mathbb{R}$ .

We have, for every  $x \in \mathbb{R}_+^*$  and  $y \in \mathbb{R}_+^*$ , the following relationship:

$g(x + y) = \ln(f(e^{x+y})) = \ln(f(e^x e^y)) = \ln(f(e^x)f(e^y))$ . According to the condition in the lemma, we get  $g(x + y) = g(x) + g(y)$ .

Consequently,  $\forall n \in \mathbb{N}^*$ ,  $g(n) = ng(1)$ ,  $g(1) = g\left(\overbrace{\frac{1}{n} + \dots + \frac{1}{n}}^n\right) = ng\left(\frac{1}{n}\right)$ , and  $g\left(\frac{1}{n}\right) = \frac{g(1)}{n}$ .

Thus,  $\forall p \in \mathbb{N}^*$ ,  $\forall q \in \mathbb{N}^*$ ,  $g\left(\frac{p}{q}\right) = g\left(\overbrace{\frac{1}{q} + \dots + \frac{1}{q}}^p\right) = pg\left(\frac{1}{q}\right) = p\frac{g(1)}{q}$ .

$\forall x \in \mathbb{R}^+$ , there is a sequence  $(p_n, q_n)$  in  $\mathbb{N}^* \times \mathbb{N}^*$ , such that  $\lim_{n \rightarrow \infty} \frac{p_n}{q_n} = x$ . So since  $g$  is continuous in  $x$ , we can write

$$\lim_{n \rightarrow \infty} g\left(\frac{p_n}{q_n}\right) = g(x) = \lim_{n \rightarrow \infty} p_n \left(\frac{g(1)}{q_n}\right) = xg(1)$$

Now,  $\forall x \in \mathbb{R}_+^*$ , let  $X = \ln x$ . This gives

$f(x) = f(e^X) = e^{g(X)} = e^{Xg(1)} = (e^X)^{g(1)} = x^b$ , where  $b = g(1)$ . *Lemma 1* has thus been proven.

From the definition  $f(p_i) = 1 + a\eta_i(p_i)$ , we can write  $1 + a\eta_i(p_i) = p_i^b$  or  $\eta_i = \frac{p_i^{b-1}}{a}$ .

The average of  $\eta_i$  is given by  $\eta = \sum_{k=1}^W p_k \eta_k = \frac{\sum p_i^{1+b-1}}{a}$ .

**Lemma 2 :** The maximum or the minimum of  $\eta$  correspond to uniform distribution

$$(p_i)_{1 \leq i \leq W} = \left(\frac{1}{W}, \dots, \frac{1}{W}\right)$$

**Proof :**

Let  $\eta((p_i)_{1 \leq i \leq W}) = \sum_{i=1}^W p_i \frac{p_i^{b-1}}{a} = \sum_{i=1}^W \frac{p_i^{1+b-p_i}}{a}$ , we consider the function  $\varphi: x \mapsto \frac{x^{1+b-x}}{a}$  defined on  $]0, +\infty[$ .

The function  $\varphi$  is of class  $C^2$  on  $]0, +\infty[$ , with a second derivative  $\varphi''(x) = \frac{(b+1)b}{a} x^{b-1}$ .

It is obvious that  $\varphi$  is convex if  $\frac{(b+1)b}{a} \geq 0$ , and  $\varphi$  is concave if  $\frac{(b+1)b}{a} \leq 0$ .

As a consequence, if  $\frac{(b+1)b}{a} \geq 0$ , we have

$$\left(\sum_{i=1}^W \frac{1}{W} \varphi(p_i)\right) \geq \varphi\left(\sum_{i=1}^W \frac{1}{W} p_i\right) = \varphi\left(\frac{1}{W} \sum_{i=1}^W p_i\right) = \varphi\left(\frac{1}{W}\right).$$

And if  $\frac{(b+1)b}{a} \leq 0$ ,

$$\left(\sum_{i=1}^W \frac{1}{W} \varphi(p_i)\right) \leq \varphi\left(\sum_{i=1}^W \frac{1}{W} p_i\right) = \varphi\left(\frac{1}{W} \sum_{i=1}^W p_i\right) = \varphi\left(\frac{1}{W}\right).$$

On the other hand

$$\eta((p_i)_{1 \leq i \leq W}) = \sum_{i=1}^W p_i \frac{p_i^{b-1}}{a} = \sum_{i=1}^W \varphi(p_i) = W \left( \sum_{i=1}^W \frac{1}{W} \varphi(p_i) \right).$$

So if  $\frac{(b+1)b}{a} \geq 0$ , we obtain

$$\eta((p_i)_{1 \leq i \leq W}) = W \left( \sum_{i=1}^W \frac{1}{W} \varphi(p_i) \right) \geq W \varphi\left(\frac{1}{W}\right) = \sum_{i=1}^W \varphi\left(\frac{1}{W}\right) = \eta\left(\frac{1}{W}, \dots, \frac{1}{W}\right).$$

And if  $\frac{(b+1)b}{a} \leq 0$ , we have

$$\eta((p_i)_{1 \leq i \leq W}) = W \left( \sum_{i=1}^W \frac{1}{W} \varphi(p_i) \right) \leq W \varphi\left(\frac{1}{W}\right) = \sum_{i=1}^W \varphi\left(\frac{1}{W}\right) = \eta\left(\frac{1}{W}, \dots, \frac{1}{W}\right).$$

*Lemma 2* has thus been proven.

**Lemma 3** : The extremum of  $\eta = \frac{\sum p_i^{1+b}-1}{a}$  means  $b = -a$ , yielding  $\eta = \frac{\sum p_i^{1-a}-1}{a}$ .

**Proof:**

Consider the set  $Y = \{(1 \geq p_1 \geq 0, \dots, 1 \geq p_W \geq 0) | \sum_{i=1}^W p_i = 1\}$ , the function  $\eta = \frac{\sum p_i^{1+b}-1}{a}$  is continuous on a compact set  $Y$ , and therefore admits one or more extremums (maximum or minimum) in  $C = (C_1, \dots, C_W)$ . From Lemma 2, we know that  $C_i = \frac{1}{W}, \forall 1 \leq i \leq W$ . In addition, from physical consideration, we can exclude  $C_i = 1$  and  $C_i = 0 \forall 1 \leq i \leq W$  since they are non-probabilistic case and irrelevant to this work. We can apply Lagrange's multiplier theorem to  $\eta = \frac{\sum p_i^{1+b}-1}{a}$  and the constraint function  $g(p) = \sum_{i=1}^W p_i$  over the set  $X = \{(1 > p_1 > 0, \dots, 1 > p_W > 0) | g(p) = 1\}$  with the multiplier  $\lambda \in R | d\eta(C) = \lambda dg(C)$  and

$$\frac{\partial \eta}{\partial p_i}(C) = \frac{(1+b)p_i^b(C)}{a} = \frac{(1+b)C_i^b}{a} = \lambda \frac{\partial g}{\partial p_i}(C) = \lambda$$

Leading to

$$\lambda = \frac{(1+b)W^{-b}}{a}$$

In order to determine the relationship between  $b$  and  $a$ , we should determine how  $\lambda$  depends on  $a$  or  $b$ . Considering the efficiency nonadditivity  $\eta_{kj}(C) = \eta_k(A) + \eta_j(B) + a\eta_k(A)\eta_j(B)$ , we see that if  $a$  goes to the zero limit  $a \rightarrow 0$ , the efficiency tends to additive limit  $\eta_{kj}(C) \rightarrow \eta_k(A) + \eta_j(B)$ . Now taking into account  $p_{kj}(C) = p_k(A)p_j(B)$  and  $\eta_i = \frac{p_i^b-1}{a} = \frac{\exp(b \ln p_i)-1}{a}$ , we expect an asymptotic behavior  $\eta_i \propto \ln p_i$  when  $a \rightarrow 0$ . This requires  $b \rightarrow 0$ , or  $\eta_i = \frac{b \ln p_i}{a}$ ,

which implies  $b \propto a$  when  $a \rightarrow 0$ . If we require  $\eta_i \geq 0$  as expected in physics for thermodynamic efficiency, we can write  $b = -a$  without loss of generality. This gives  $\lambda = \frac{(1-a)W^a}{a}$  for  $a \rightarrow 0$ . Substituting this into  $\lambda = \frac{(1+b)W^{-b}}{a}$ , we obtain the general relation  $b = -a$  and

$$\eta_i = \frac{p_i^{-a} - 1}{a}$$

*Lemma 3* has thus been proven. This completes the proof of *Theorem 1*. The uniqueness of the efficiency functional Eq.(1) follows as the average of  $\eta_i$  given by

$$\eta = \sum_{i=1}^w p_i \eta_i = \frac{\sum p_i^{1-a} - 1}{a}.$$

### 3) Conclusion

We have provided a proof of the uniqueness of the efficiency functional Eq.(1) for living systems containing a great number of agents all trying to achieve something with effort, similarly to thermal engines producing work (achievement) from source energy (effort). Since the maximization of Eq.(1), as an application of PLE and PME, necessarily yields the Zipf and the Pareto power laws [7][8], this uniqueness strongly confirms the intrinsic link between the widely observed Zipf and Pareto power laws and the universal behavior of living agents to achieve more by doing less, following the principle of least effort.

We would like to mention that this work on the uniqueness of efficiency functional has been inspired by the previous works on the uniqueness theorem of the generalized entropies [10]-[16] as extension of the uniqueness theorem of the Shannon entropy [17][18]. The case of the efficiency functional is a little different from the case of entropy. On the one hand, unlike entropy, efficiency does not have the a priori maximum corresponding to uniform distribution. On the other hand, the additivity and nonadditivity of entropy are hypothetical properties of the entropy of the entire considered system, while the nonadditivity of efficiency is a general property of thermodynamic efficiency of single engines. It is worth mentioning also that in certain previous works [11][16], the uniqueness theorem of entropy has been formulated for interdependent sub-systems using conditional probability, while the uniqueness theorem in this work is formulated for independent sub-systems. This is one of the aims of our future work on efficiency. Another future challenge is the uniqueness theorem of entropy or efficiency functional with continuous probability distribution, an old question still open to date [19].

## References

- [1] 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.
- [2] G.K. Zipf, *Selected Studies of the Principle of Relative Frequency in Language* Cambridge, MA: Harvard University Press (1932).
- [3] G.K. Zipf, *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley (1949).
- [4] V. Pareto, *Cours d'économie politique*, *Journal of Political Economy* **6** (1898)
- [5] B. Mandelbrot, An informational theory of the statistical structure of language, *Communication Theory*, **84** (1953) 486
- [6] R.F. Cancho and R.V. Sole, *PNAS*, **100** (2002) 788
- [7] Q.A. Wang, Principle of least effort vs. Maximum efficiency : deriving Zipf-Pareto's laws, *Chaos, Solitons & Fractals*, **153**(2021)111489
- [8] A. El Kaabouchi, F.X. Machu, J. Cocks, R. Wang, Y.Y. Zhu and Q.A. Wang, Study of a measure of efficiency as a tool for applying the principle of least effort to the derivation of the Zipf and the Pareto laws, *Advances in Complex Systems*, **24** (2021) 2150013
- [9] H.B. Callen, *Thermodynamics and Introduction to Thermostatistics*, John Wiley & Sons, New York, 1985
- [10] R. J. V. dos Santos, *J. Math. Phys.* **38** (1997) 4104
- [11] S. Abe, "Axioms and uniqueness theorem for Tsallis entropy," *Phys. Lett. A* **271** (2000) 74
- [12] H. Suyari, Generalization of Shannon–Khinchin axioms to nonextensive systems and the uniqueness theorem for the nonextensive entropy, *IEEE Trans. Inf. Theory* **50** (2004) 1783
- [13] Petr Jizba, Toshihico Arimitsu, The world according to Rényi: thermodynamics of multifractal systems, *Ann. Physics* **312** (2004) 17
- [14] Petr Jizba, Toshihico Arimitsu, Generalized statistics: yet another generalization, *Physica A* **340** (2004) 110
- [15] V. M. Ilić, M. Stanković, and E. H. Mulalić, Comments on "Generalization of Shannon–Khinchin axioms to nonextensive systems and the uniqueness theorem for the nonextensive Entropy", *IEEE Trans. Inf. Theory* **59** (2013) 6950
- [16] V. M. Ilić, M. Stanković, Generalized Shannon–Khinchin axioms and uniqueness theorem for pseudo-additive entropies, *Physica A* **411** (2014) 138

- [17] C.E. Shannon, A Mathematical Theory of Communication. Bell System Technical Journal **27** (1948) 379.
- [18] A. I. Khinchin, Mathematical Foundations of Information Theory, Dover, New York, 1957.
- [19] R. Wang, F.X. Machu, A. El Kaabouchi and Q.A. Wang, A non-negative informational entropy for continuous probability distribution, preprint [hal-03841203](https://hal.archives-ouvertes.fr/hal-03841203)