On the relation between the Maximum Entropy Principle and the principle of Least Effort: The continuous case

Abdelatif Agouzal [a,b], Thierry Lafouge [a,c,*]

[a] *Université de Lyon, France*
[b] *Université Lyon 1, Institut Camille Jordan, CNRS U.M.R. 5208, Villeurbanne F-69622, France*
[c] *Université Lyon 1, ELICO, Villeurbanne F-69622, France*

Abstract

The Maximum Entropy Principle (MEP) maximizes the entropy provided that the effort remains constant. The Principle of Least Effort (PLE) minimizes the effort provided that the entropy remains constant. The paper investigates the relation between these two principles. In some kinds of effort functions, called admissible, it is shown that these two principles are equivalent. The results are illustrated by the size–frequency statistical distribution met in infometry in Information Production Processes.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Maximum Entropy Principle; Principle of Least Effort; Effort function; Inverse power function

1. Introduction

The article is divided into two parts. In the first, through the signal theory, we mathematically formulate the equivalence of the two principles, Maximum Entropy Principle with constant effort, and the Principle of Least Effort with constant entropy. Then we situate these principles within the processes of producing and using information in infometry. In the second part, we mathematically demonstrate equivalence by introducing the concept of admissible effort function. Here, the mathematical proof is harder than in the discrete and finite case (Egghe & Lafouge, 2006), because the entropy (or the amount of effort) is not necessarily finished.

To conclude, we return to the information production processes to clarify the properties of the usual functions that model the information production processes.

2. Definitions and context

2.1. Information theory

In 1948 Shannon (Shannon, 1993) worked out a statistical theory on the transmission of electrical signals. This statistical theory of information stipulates that the more the states of a system are equiprobable, the more the process produces information. The theoretical bases used in this article come from this theory. Here, we uniquely consider

---

density functions of probability μ defined from $[1, \ldots, \infty[$ with real positive values and verifying:

$$\int_1^\infty \mu(x)\,dx = 1$$

We call effort function any continuous *f* function defined from $[1, \ldots, \infty[$ with real positive values. For each couple (μ, f), we define two quantities, $\bar{H}$[1] average information content, called entropy and $\bar{F}$ average effort:

$$\bar{H} = -\int_1^\infty \mu(x)\,\text{Ln}(\mu(x))\,dx, \qquad \bar{F} = \int_1^\infty f(x)\mu(x)\,dx,$$

$\bar{F}$ is necessarily positive. Let us now define two important optimization principles: Maximum Entropy Principle (MEP) and Principle of Least Effort (PLE).

Maximum Entropy Principle (MEP). The MEP maximizes the entropy provided that the effort remains constant.
Principle of Least Effort (PLE). The PLE minimizes the effort provided that the entropy remains constant.

2.1.1. Mathematical formulation of the two principles

With $\bar{H}$, $\bar{F}$ positive, two real numbers, and *f* an effort function, we define the sets of functions:

$$A(\bar{H}) = \left\langle \mu \geq 0, \int_1^\infty \mu(x)\,dx = 1,\ \bar{H} = -\int_1^\infty \mu(x)\,\text{Ln}(\mu(x)\,dx \right\rangle$$

$$C(\bar{F}) = \left\langle \mu \geq 0, \int_1^\infty \mu(x)\,dx = 1,\ \bar{F} = \int_1^\infty \mu(x)f(x)\,dx \right\rangle$$

With these notations, MEP and PLE are written:

$$\text{MEP} \Leftrightarrow \max_{\mu \in C(\bar{F})} \left(-\int_1^\infty \mu(x)\,\text{Ln}(\mu(x))\,dx \right)$$

$$\text{LEP} \Leftrightarrow \min_{\mu \in A(\bar{H})} \left(\int_1^\infty \mu(x)f(x)\,dx \right)$$

Our aim is to show, under certain conditions, that these two principles are equivalent. We will situate the problem within the sphere of infometry and recall results obtained in earlier work.

2.2. Informetrics and entropy aspects

2.2.1. Information production process and inverse power law

Statistical regularities observed in producing or using information have been studied for a long time (Bradford, 1934; Lotka, 1926). In practice, the production function has similar characteristics in very diverse situations of production or use of information.

These models can be represented by the diagram in Fig. 1, introduced into informetric systems by Egghe (1990) and called "Information Production Process" (IPP). An IPP is a triplet made up of a source, a production function, and all the elements (items) produced.

In practice, the production function of an IPP has similar characteristics in very diverse situations:

- Authors (sources) write articles (items).
- Words in a text (sources) produce occurrences of words in the text (items).
- Web pages (sources) contain links (items).
- Web sites (sources) are visited (items).
- Requests (sources) through a search engine are sent by users (items).

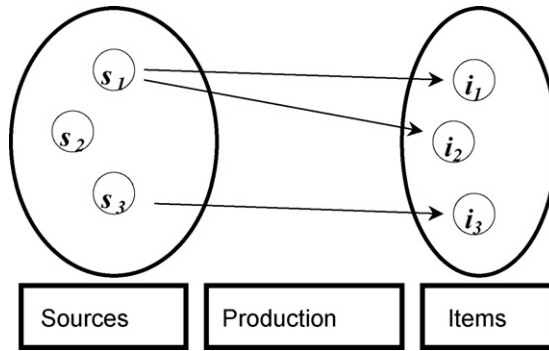[1] Unlike the discrete case, the entropy may be negative.

Fig. 1. Schematic representation of an Information Production Process.

By using a size–frequency distribution on the examples above, we can observe that a high number of sources produce few items and a small number of sources produce a lot. The size–frequency function mostly used in Informetrics is the inverse power function, also called Lotkain informetric distribution (Egghe, 2005), the corresponding density probability ν_λ is:

$$x \in [1, \ldots \infty[, \ \nu_\lambda(x) = \frac{\lambda - 1}{x^\lambda}, \quad \lambda > 1$$

The calculation of the entropy (Yablonsky, 1981) $H(\lambda)$ according to λ gives:

$$H(\lambda) = \mathrm{Ln}\left(\frac{1}{\lambda - 1}\right) + \frac{\lambda}{\lambda - 1}, \quad \lambda > 1$$

It is easy to show that the entropy is a decreasing function of λ. The classical interpretation of Lotka's law (1926) is found, e.g. the higher λ is, the bigger is the gap between the number of scientists who produce a lot compared to the number of scientists who produce a little, knowing that there are few scientists who produce a lot compared to the number of scientists who produce a little.

Thus, the two following questions arise:

(1) For a given quantity of effort, what is the connection between the random distribution of the sources and the effort function when the quantity of information produced by the sources is maximized?
(2) For a given quantity of information, what is the connection between the random distribution of the sources and the effort function when the quantity of effort produced by the source is maximized?

2.2.2. Previous results

2.2.2.1. Finite discrete case. Suppose μ is a finite discrete distribution, so we have $m \in N$ probabilities $p_1, p_2, \ldots, p_m$ so that $\sum_{r=1}^{m} p_r = 1$, the entropy of such a system is defined as $\bar{H} = -\sum_{r=1}^{m} p_r \mathrm{Ln}(p_r)$. Let $F_r > 0$ denote the effort function ($r = 1, \ldots, m$) the quantity of effort is defined as $\bar{F} = \sum_{r=1}^{m} E_r \cdot p_r \mathrm{Ln}(p_r)$. In Egghe and Lafouge (2006) we prove the following result, if we suppose $p_1, p_2, \ldots, p_m$ is a decreasing sequence MEP and PLE are equivalent. In the conclusion to this article we provided a complete explanation of Zip's law (Egghe & Lafouge, 2006) or Lotka's law as a decreasing function.

An incomplete demonstration of the result of equivalence in the continuous case was subsequently proposed (see Section 2.2.2.2). The aim of this article is to generalize this result.

2.2.2.2. Exponential informetric process. In Lafouge and Prime Claverie (2005) we assume that an item produced requires a certain amount of effort and therefore we define the exponential informetric process by introducing the effort function (see Fig. 2). The effort function denotes the amount of effort from a source to produce items.

With f a strictly increasing unbounded effort function, λ a positive number where we have the condition:

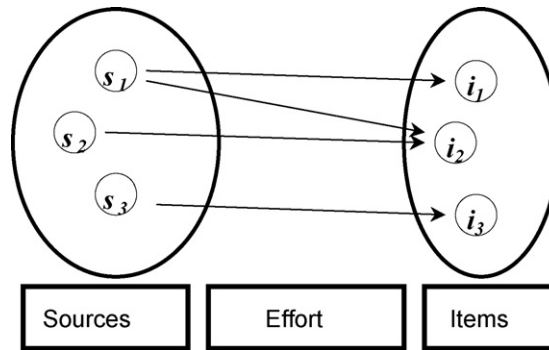$$\int_1^\infty f(x) \exp(-\lambda f(x))\, dx < \infty \tag{1}$$

Fig. 2. Schematic representation of an informetric process using the effort function.

Since *f* is strictly increasing, unbounded, we can easily show the condition:

$$\int_1^\infty \exp(-\lambda f(x))\, dx < \infty$$

Therefore, we call the following density function Exponential Informetric Process:

$$\nu_\lambda(x) = \frac{\exp(-\lambda f(x))}{\int_1^\infty \exp(-\lambda f(x))\, dx}$$

$\int_1^\infty f(x)\nu_\lambda(x)\, dx$ corresponds to the average $\bar{F}$ of effort produced by the Exponential Informetric Process. The following results explain the relationship between average amount of effort and entropy.

We have shown (Lafouge & Prime Claverie, 2005) the following results:

 (i) ν_λ is decreasing,
 (ii) the two principles maximum MEP and PLE are verified simultaneously,
(iii) $\bar{H} = \mathrm{Ln}(k) + \lambda \bar{F}$,

where $k = \int_1^\infty \exp(-\lambda f(x))\, dx$ and $\bar{H}$ corresponds to the entropy.

This result partly resolves the problem of the equivalence of the two principles given in Section 2.1. In fact, the condition is sufficient for these two principles to be equivalent. In the two preceding demonstrations, finite and discrete cases (equivalence) and in the continuous case (sufficient condition), we calculate a solution for the MEP that plays a central role in mechanical statistic,

$$\nu(x) = C \exp(-\lambda f(x))$$

where C is a constant, λ a positive number and f an effort function. This distribution, constructed with the MEP, is known under the name of Boltzmann distribution. To our knowledge, there is no mathematical demonstration of the equivalence of these two principles (MEP and PLE) in the continuous case.

Note. Demonstration (ii) of the preceding result supposes, when we write the equality $\bar{F} = \int_1^\infty f(x)\exp(-\lambda f(x))\, dx$, the existence of a unique λ verifying this equality. The demonstration given in Section 3 will clarify this point.

2.2.2.3. Effort function and inverse power law. With the preceding formalism, the density function corresponding to an inverse power is written:

$$x \in [1, \ldots \infty[,\ \nu_\lambda(x) = (\lambda - 1)\exp(-(\lambda - 1)\mathrm{Ln}(x)),\quad \lambda > 1$$

The calculation of the amount of effort according to λ gives:

$$F(\lambda) = \frac{1}{\lambda - 1},\quad \lambda > 1$$

The logarithmic function is the effort function corresponding to the Lotka distributions. In the production of words in a text, we express this regularity using the formalism of Zipf's law (see Section 4). In Egghe and Lafouge (2006) page 8, this hypothesis is formulated thus:

the cost (effort) of using a word with i ($i = 1, 2, 3 \ldots$) letters is proportional to i hence to $\mathrm{Log}_N(r)$ where N is the number of different letters and r is the rank of a word with i letters.

Moreover, the results presented in Lafouge and Smolczeswska (2006) show that if an effort function f verifies the condition,

$$\mathrm{Limit}_{x \to \infty} \frac{f(x)}{\mathrm{Ln}(x)} = \lambda, \quad \lambda > 1$$

we have an Exponential Informetric Process.

We will now directly demonstrate the equivalence of the two principles using the continuous case with effort functions, strictly increasing, unbounded and verifying the condition (i).

3. Characterising the Maximum Entropy Principle in terms of the Principle of Least Effort

3.1. Main results

3.1.1. Preliminary and notation

In all that follows f denotes a strictly increasing unbounded effort function, f will be called admissible if there is also a real number λ_0 verifying the condition

$$\int_1^\infty f(x) \exp(-\lambda_0 f(x)) \, dx < \infty$$

Lemma 2.0. *With f an admissible effort function, there exists a real number $\sigma(f) \geq 0$ so that*

$$\forall \lambda > \sigma(f), 0 \leq \int_1^\infty f(x) \exp(-\lambda f(x)) \, dx < \infty$$

Proof. f being admissible,

there exists $\lambda_0 \in \Re$ so that $0 \leq \int_1^\infty f(x) \exp(-\lambda_0 f(x)) \, dx < \infty$

f being increasing $\forall \lambda > \lambda_0$ we have,

$$0 \leq \int_1^\infty f(x) \exp(-\lambda f(x)) \, dx < \int_1^\infty f(x) \exp(-\lambda_0 f(x)) \, dx < \infty$$

Moreover, it is clear that $\int_1^\infty f(x) \, dx = +\infty$, consequently λ_0 is strictly positive. We can then define the following real number:

$$\sigma(f) = \inf \left\langle \lambda > 0, 0 < \int_1^\infty f(x) \exp(-\lambda f(x)) \, dx < \infty \right\rangle$$

We will now define two functions H and F quantifying the entropy and effort of a density function μ,

$$\int_1^\infty v(x) \, dx = 1$$

We define the entropy of μ,

$$H(\mu) = - \int_1^\infty \mu(x) \, \mathrm{Ln}(\mu(x)) \, dx$$

which can be infinite and the quantity of effort of μ,

$$F(\mu) = \int_1^\infty \mu(x) f(x) \, dx$$

Let $\bar{H}$ and $\bar{F}$ two constants, $A(\bar{H})$ and $C(\bar{F})$ the set of functions defined in the introduction, $\forall \lambda > \sigma(f)$ we put:

$$k = \int_1^\infty \exp(-\lambda f(x)) \, dx$$

$$\nu_\lambda(x) = \frac{\exp(-\lambda f(x))}{k}$$

$$H(\lambda) = -\int_1^\infty \nu_\lambda(x) \, \mathrm{Ln}(\nu_\lambda(x)) \, dx$$

$$F(\lambda) = \int_1^\infty f(x) \nu_\lambda(x) \, dx \quad \square$$

The following results are simple consequences of the preceding definitions.

Lemma 2.1. $\forall \lambda > \sigma(f)$, *the entropy $H(\lambda)$ of ν_λ is finite.*

Proof.

$$H(\lambda) = -\int_1^\infty \frac{1}{k} \exp(-\lambda f(x)) \, \mathrm{Ln}\left(\frac{\exp(-\lambda f(x))}{k}\right) \, dx = \frac{1}{k} \int_1^\infty \exp(-\lambda f(x))(\lambda f(x) + \mathrm{Ln}(k)) \, dx$$
$$= \lambda \bar{F} + \mathrm{Ln}(k)$$

Also, the two functions are defined:

$$Z(\lambda) = -\mathrm{Ln}\left(\int_1^\infty \exp(-\lambda(f(x) - \bar{F})) \, dx\right)$$

$$R(\lambda) = -\frac{1}{\lambda} \mathrm{Ln}\left(\int_1^\infty \exp(-\lambda f(x) - \bar{H}) \, dx\right) \quad \square$$

Lemma 2.2. *With $\mu \in C(\bar{F})$, $\lambda > \sigma(f)$, we have*:

$$\int_1^\infty \mu(x) \, \mathrm{Ln}(\nu_\lambda(x)) \, dx = Z(\lambda)$$

Proof.

$$\int_1^\infty \mu(x) \, \mathrm{Ln}(\nu_\lambda(x)) \, dx = \int_1^\infty \mu(x)(-\lambda f(x) - \mathrm{Ln}\left(\int_1^\infty \exp(-\lambda f(x)) \, dx\right)$$
$$= -\lambda \bar{F} - \mathrm{Ln}\left(\int_1^\infty \exp(-\lambda f(x)) \, dx\right) = -\mathrm{Ln}\left(\exp(\lambda \bar{F}) \int_1^\infty \exp(-\lambda f(x)) \, dx\right)$$
$$= -\mathrm{Ln}\left(\int_1^\infty \exp(-\lambda(f(x) - \bar{F})) \, dx \quad \square$$

Lemma 2.3. *Let $v_\lambda \in C(\bar{F})$, we have*

$$\int_1^\infty (f(x) - \bar{F}) \exp(-\lambda f(x)) \, dx = 0$$

Proof. $v_\lambda \in C(\bar{F})$, hence

$$\int_1^\infty f(x) \exp(-\lambda f(x)) \, dx = k\bar{F}$$

$$\int_1^\infty (f(x) - \bar{F}) \exp(-\lambda f(x)) \, dx = \int_1^\infty f(x) \exp(-\lambda f(x)) \, dx - k\bar{F} = k\bar{F} - k\bar{F} = 0 \quad \square$$

3.1.2. Equivalence theorem

We can now state the equivalence theorem of the two principles by characterizing the solutions of the MEP and the PLE using the functions $z(\lambda)$ and $R(\lambda)$.

Theorem. *With f an admissible effort function, we have the following results*:

(i) $-\max H(\mu)_{\mu \in C(\bar{F})} \geq \sup(Z(\lambda))_{\lambda > \sigma(f)}$
   Also, if there is a real number λ_0 so that $v_{\lambda_0} \in C(\bar{F})$ then this real number is unique and we have:

$$H(\lambda_0) = -\max H(\mu)_{\mu \in C(\bar{F})} = \sup(Z(\lambda))_{\lambda > \sigma(f)} = Z(\lambda_0)$$

(ii) $0 \geq \min F(\mu)_{\mu \in A(\bar{H})} \geq \sup(R(\lambda))_{\lambda > \sigma(f)}$
   Also, if there is a real number λ_1 so that $v_{\lambda_1} \in A(\bar{H})$ then this real number is unique and we have:

$$F(\lambda_1) = \min F(\mu)_{\mu \in A(\bar{H})} = \sup(R(\lambda))_{\lambda > \sigma(f)} = R(\lambda_1)$$

(iii) *If $\bar{H} = -Z(\lambda_1)$ or $\bar{F} = R(\lambda_0)$ then $v_{\lambda_0} = v_{\lambda_1}$.*

(i) characterizes the solution of the MEP, in other words maximizing the entropy subject to a constant effort, (ii) characterizes the solution of the PLE, in other words minimizing the effort subject to constant entropy. Finally, if the information content $\bar{H}$ and quantity of effort $\bar{F}$ are linked, in other words if $\bar{H} = -Z(\lambda_1)$ or if $\bar{F} = R(\lambda_0)$, the two principles are equivalent. We note that in this case we find the result (iii) of Section 2.2.2.2, $\bar{H} = \lambda_1 \bar{F} + \text{Ln}(k)$ or $\bar{H} = \lambda_0 \bar{F} + \text{Ln}(k)$.

Before solving this problem, we will give arguments by a formal result to demonstrate (i) as follows.

We have,

$$\frac{dZ}{d\lambda} = \frac{\int_1^\infty (f - \bar{F}) \exp(-\lambda(f(x) - \bar{F})) \, dx}{\int_1^\infty \exp(-\lambda(f(x) - \bar{F})) \, dx}$$

With $v_{\lambda_0} \in C(\bar{F})$, according to Lemma 2.3, we have

$$\frac{dZ}{d\lambda}(\lambda_0) = 0,$$

which means that $Z(\lambda_0)$ is an extremum of function Z.

3.2. Demonstration of the equivalence theorem

3.2.1. Preliminary

For this problem to be of mathematical interest, we need to show that $A(\bar{H})$ and $C(\bar{F})$ are not empty and to suppose additional hypotheses regarding the effort function f so that the maximum entropy is finite. Readers should refer to points (1)–(3) in Appendix A for these demonstrations.

We will solve this problem uniquely through the use of elementary analytical arguments.

Lemma 2.4. *Let two real numbers $y \geq 0$ and $z > 0$, we have*:

(i) $y \operatorname{Ln}(y) \geq y - z + y \operatorname{Ln}(z)$, *and strict inequality if, and only if, $y \neq z$.*

(ii) $y \operatorname{Ln}(y) - z \operatorname{Ln}(z) \geq (1 + \operatorname{Ln}(z))(y - z)$.

Proof. (ii) is immediately deduced from (i).

For $y = 0$ and $y = z$, the result is trivial. With $y > 0$, $z > 0$, $y \neq z$, we put $t = z/y$, we have $\forall t > 0$, $t \neq 1$, $\operatorname{Ln}(t) < t - 1$, hence $\operatorname{Ln}(z) - \operatorname{Ln}(y) < z - y/y$. $\square$

Proposition 2.1. *For any admissible function, we have*:

$$-\max H(\mu)_{\mu \in C(\bar{F})} \geq \sup Z(\lambda)_{\lambda > \sigma(f)}$$

Proof. With $\mu \in C(\bar{F})$, $\lambda > \sigma(f)$ using the result (i) of Lemma 2.4 with $y = \mu$ and $z = v_\lambda$, and by integrating the functions, we have:

$$\int_1^\infty \mu(x) \operatorname{Ln}(\mu(x)) \, dx \geq \int_1^\infty (\mu(x) - v_\lambda(x)) \, dx + \int_1^\infty \mu(x) \operatorname{Ln}(v_\lambda(x)) \, dx$$

we know,

$$\int_1^\infty (\mu(x) - v_\lambda(x)) \, dx = 1 - 1 = 0,$$

consequently

$$-H(\mu) \geq \int_1^\infty \mu(x) \operatorname{Ln}(v_\lambda(x)) \, dx$$

thus according to Lemma 2.2

$$-H(\mu) \geq Z(\lambda)$$

hence

$$\forall \mu \in C(\bar{F}), \forall \lambda > \sigma(f), -H(\mu) \geq Z(\lambda)$$

$$-\max H(\mu)_{\mu \in C(\bar{F})} \geq \sup Z(\lambda)_{\lambda > \sigma(f)} \quad \square$$

Proposition 2.2. *With f an admissible effort function, for any density function μ we have $\forall \lambda > \sigma(f)$, $v_\lambda \neq \mu$, the following inequality*:

$$\int_1^\infty \mu(x) \operatorname{Ln}(\mu(x)) \, dx - \int_1^\infty v_\lambda(x) \operatorname{Ln}(v_\lambda(x)) \, dx > \lambda \int_1^\infty (v_\lambda(x) - \mu(x)) f(x) \, dx$$

Proof. With $\lambda > \sigma(f)$ so that $\nu_\lambda \neq \mu$, using Lemma 2.4 (ii) with $y = \mu$ and $z = \nu_\lambda$, we have the inequality:

$$\mu(x)\,\mathrm{Ln}(\mu(x)) - \nu_\lambda(x)\,\mathrm{Ln}(\nu_\lambda(x)) \geq (\mu(x) - \nu_\lambda(x))(1 + \mathrm{Ln}(\nu_\lambda(x))$$

By integrating the right member, we obtain:

$$\int_1^\infty (\mu(x) - \nu_\lambda(x))\,\mathrm{d}x + \int_1^\infty (\mu(x) - \nu_\lambda(x))\,\mathrm{Ln}(\nu_\lambda(x))\,\mathrm{d}x$$

However,

$$\int_1^\infty (\mu(x) - \nu_\lambda(x))\,\mathrm{d}x = 1 - 1 = 0$$

Consequently,

$$\int_1^\infty (\mu(x) - \nu_\lambda(x))\,\mathrm{Ln}(\nu_\lambda(x))\,\mathrm{d}x = \int_1^\infty (\mu(x) - \nu_\lambda(x))(-\lambda f(x) - \mathrm{Ln}(k))\,\mathrm{d}x$$

Hence

$$= \int_1^\infty (\mu(x) - \nu_\lambda(x))(-\lambda f(x))\,\mathrm{d}x \quad \square$$

Lemma 2.5. *For any couple of real numbers z_1, z_2, we have the following inequality*:

$$\exp(-z_2) - \exp(-z_1) \geq (z_1 - z_2)\exp(-z_1)$$

Proof. We have the strict inequality if $z_1 \neq z_2$. Let us suppose $z_1 < z_2$, according to the mean theorem, there is z, $z_1 < z < z_2$, verifying the equality:

$$\exp(z_2) - \exp(z_1) = -(z_2 - z_1)\exp(-z)$$

The inequality is then written:

$$(z_1 - z_2)\exp(-z) \geq (z_1 - z_2)\exp(-z_1)$$

with

$$\exp(-z) \leq \exp(-z_1)$$

We do the same if $z_2 < z_1$.

We will show the most important intermediate result, namely the sets $C(\bar{F})$ are sets of uniqueness for the functions ν_{λ_0}. More precisely, we demonstrate the result below. $\quad \square$

Proposition 2.3. *With f an admissible function, there exists at most a function ν_{λ_0} belonging to $C(\bar{F})$ verifying the inequality*: $\sup Z(\lambda)_{\lambda > \sigma(f)} = Z(\lambda_0)$.

Proof.

(1) *Uniqueness*. Let us suppose $\lambda_1 \neq \lambda_2$ so that ν_{λ_1}, $\nu_{\lambda_2} \in C(\bar{F})$, according to Lemma 2.1 we know that the entropy of ν_λ is finite, by applying Proposition 2.2 twice, we obtain

$$\int_1^\infty \nu_{\lambda_2}(x)\,\mathrm{Ln}(\nu_{\lambda_2}(x))\,\mathrm{d}x - \int_1^\infty \nu_{\lambda_1}(x)\,\mathrm{Ln}(\nu_{\lambda_1}(x))\,\mathrm{d}x > \lambda_1 \int_1^\infty (\nu_{\lambda_1}(x) - \nu_{\lambda_2}(x))f(x)\,\mathrm{d}x = 0$$

And in the same way

$$\int_1^\infty \nu_{\lambda_1}(x)\,\mathrm{Ln}(\nu_{\lambda_1}(x))\,\mathrm{d}x - \int_1^\infty \nu_{\lambda_2}(x)\,\mathrm{Ln}(\nu_{\lambda_2}(x))\,\mathrm{d}x > \lambda_2 \int_1^\infty (\nu_{\lambda_2}(x) - \nu_{\lambda_1}(x))f(x)\,\mathrm{d}x = 0$$

which is contradictory and consequently there is at most one item belonging to $C(\bar{F})$.

(2) *Existence*. Let a function $v_\lambda \in C(\bar{F})$, according to Lemma 2.5, we can write

$$z_2 = \lambda y, \quad z_1 = \lambda_0 y, \quad \exp(-\lambda y) \geq \exp(-\lambda_0 y) - y(\lambda - \lambda_0) \exp(-\lambda_0 y)$$

By putting $y = f(x) - \bar{F}$ and integrating,

$$\int_1^\infty \exp(-\lambda(f(x) - \bar{F})) \, dx \geq \int_1^\infty \exp(-\lambda_0(f(x) - \bar{F})) \, dx - (\lambda - \lambda_0) \int_1^\infty (f(x) - \bar{F})$$
$$\exp(-\lambda_0(f(x) - \bar{F})) \, dx$$

$v_{\lambda_0} \in C(\bar{F})$ we know according to Lemma 2.3,

$$\int_1^\infty (f(x) - \bar{F}) \exp(-\lambda_0(f(x) - \bar{F})) \, dx = 0$$

consequently,

$$\int_1^\infty \exp(-\lambda(f(x) - \bar{F})) \, dx \geq \int_1^\infty \exp(-\lambda_0(f(x) - \bar{F})) \, dx$$

hence

$$Z(\lambda) \leq Z(\lambda_0)$$

and thus

$$\sup Z(\lambda)_{\lambda > \sigma(f)} \leq Z(\lambda_0) \quad \square$$

We can now demonstrate the result (i) of the equivalence theorem.

Theorem 2.1. *We suppose that there is a real number $\lambda_0 > \sigma(f)$ so that $v_{\lambda_0} \in C(\bar{F})$, then this real number is unique and moreover*:

$$\sup(Z(\lambda)_{\lambda > \sigma(f)} = Z(\lambda_0) = -\max H(\mu)_{\mu \in C(\bar{F})} = -\int_1^\infty v_{\lambda_0}(x) \operatorname{Ln}(v_{\lambda_0}(x)) \, dx$$

Proof. Uniqueness and characterization by the function Z are consequences of Proposition 2.3.
With $\mu \in C(\bar{F})$, $\mu \neq v_{\lambda_0}$ using the inequality of Proposition 2.2,

$$\int_1^\infty \mu(x) \operatorname{Ln}(\mu(x)) \, dx - \int_1^\infty v_{\lambda_0}(x) \operatorname{Ln}(v_\lambda(x)) \, dx > \lambda \int_1^\infty (v_{\lambda_0}(x) - \mu(x)) f(x) \, dx$$

However,

$$\int_1^\infty (v_{\lambda_0}(x) - \mu(x)) f(x) \, dx = \bar{F} - \bar{F} = 0$$

Thus,

$$-\int_1^\infty \mu(x)\,\mathrm{Ln}(\mu(x))\,\mathrm{d}x < -\int_1^\infty \nu_{\lambda_0}(x)\,\mathrm{Ln}(\nu_{\lambda_0}(x))\,\mathrm{d}x \quad \square$$

Note. It is essential to suppose that the function ν_{λ_0} is well defined because the whole definition of the function Z is greater than $]\sigma(f), \dots \infty[$, as the following example shows. With the effort function $f(x) = \mathrm{Ln}(x+1) + \mathrm{Ln}(\mathrm{Ln}^2(x+1))$, we show $\int_1^\infty \exp(-f(x))\,\mathrm{d}x < \infty$, therefore $Z(1)$ is well defined whereas $\int_1^\infty f(x)\exp(-f(x))\,\mathrm{d}x = \infty$.

In the same way, with $\bar{H}$ a real number we consider the function $R(\lambda)$, we demonstrate Proposition 2.4 using the same techniques as previously.

Proposition 2.4. *With f an admissible function, there is at most a function ν_{λ_1} belonging to $A(\bar{H})$ verifying the equality*: $\sup R(\lambda)_{\lambda > \sigma(f)} = R(\lambda_1)$

We can now demonstrate the result (ii) of the equivalence theorem.

Theorem 2.2. *We suppose that there is a real number $\lambda_1 > \sigma(f)$ so that $\nu_{\lambda_1} \in A(\bar{H})$, then this real number is unique and also*:

$$\sup(Z(\lambda))_{\lambda > \sigma(f)} = R(\lambda_1) = \min F(\mu)_{\mu \in A(\bar{H})} = \int_1^\infty \nu_{\lambda_1}(x)f(x)\,\mathrm{d}x$$

Proof. Uniqueness and characterization by the function R are consequences of Proposition 2.3.
With $\mu \in A(\bar{H})$, using the inequality of Proposition 2.2 we obtain,

$$\int_1^\infty \mu(x)\,\mathrm{Ln}(\mu(x))\,\mathrm{d}x - \int_1^\infty \nu_{\lambda_1}(x)\,\mathrm{Ln}(\nu_{\lambda_1}(x))\,\mathrm{d}x > \lambda \int_1^\infty (\nu_{\lambda_1}(x) - \mu(x))f(x)\,\mathrm{d}x$$

$$0 = \bar{H} - \bar{H} > \lambda \int_1^\infty (\nu_{\lambda_1}(x) - \mu(x))f(x)\,\mathrm{d}x$$

consequently, since $\lambda > 0$,

$$\int_1^\infty f(x)\nu_{\lambda_1}(x)\,\mathrm{d}x < \int_1^\infty f(x)\mu(x)\,\mathrm{d}x$$

using Theorems 2.1 and 2.2, we obtain the equivalence theorem. $\square$

Theorem 2.3. *Let a positive, unbounded, strictly increasing, continuous function, we suppose that there exists $\lambda_0 \in \Re$ so that*:

$$\int_1^\infty f(x)\exp(-\lambda_0 f(x))\,\mathrm{d}x < \infty$$

then, we have:

1. *There is a real number $\sigma(f)$ so that*:

$$\forall \lambda > \sigma(f), 0 \le \int_1^\infty f(x)\exp(-\lambda f(x))\,\mathrm{d}x < \infty$$

2. *We have*

$$0 \ge -\max H(\mu)_{\mu \in C(\bar{F})} \ge \sup Z(\lambda)_{\lambda > \sigma(f)}$$

Also, if there is a real number λ_0 so that $v_{\lambda_0} \in C(\bar{F})$ then this real number is unique and we have:

$$H(\lambda_0) = -\max H(\mu)_{\mu \in C(\bar{F})} = \sup Z(\lambda)_{\lambda > \sigma(f)} = Z(\lambda_0)$$

where

$$Z(\lambda) = -\text{Ln} \left(\int_1^\infty \exp(-\lambda(f(x) - \bar{F})) \, dx \right)$$

3. *We have*

$$0 \geq \min F(\mu)_{\mu \in A(\bar{H})} \geq \sup R(\lambda)_{\lambda > \sigma(f)}$$

Also, if there is a real number λ_1 so that $v_{\lambda_1} \in A(\bar{H})$ then this real number is unique and we have:

$$F(\lambda_1) = \min F(\mu)_{\mu \in A(\bar{H})} = \sup R(\lambda)_{\lambda > \sigma(f)} = R(\lambda_1)$$

where

$$R(\lambda) = \frac{1}{\lambda}\text{Ln} \left(\int_1^\infty \exp(-\lambda f(x) - \bar{H}) \, dx \right)$$

4. *Finally, if $\bar{H} = -Z(\lambda_1)$ or $\bar{F} = R(\lambda_0)$ then $v_{\lambda_0} = v_{\lambda_1}$, the two principles are equivalent.*

Examples.   The most common distributions in infometry are:

- linear effort functions $f(x) = x$, in other words exponential distributions (Lafouge, 2007), in this case $\sigma(f) = 0$.
- logarithmic effort functions $f(x) = \text{Ln}(x)$, in other words Lotka distributions (see Section 2.2.1), in this case $\sigma(f) = 1$.

4. Conclusion

One of the most intriguing phenomena in infometry, and widely studied in quantitative linguistics, is Zipf's law (Zipf, 1949). The name of Zipf's law has been given to the following approximation of the rank–frequency function:

$$g(r) = \frac{C}{r}$$

where r is the rank of a word type, $g(r)$ is the occurrence of the word type and C is a constant. The Principle of Least Effort is attributed to Zipf in linguistics. Zipf thinks that the reason for this regularity is linked to the behaviour of the individuals. More precisely, he defines this principle philosophically thus (comment of Zipf quoted in Wyllis, 1982):

The Principle of Least Effort means . . . that a person . . . will strive to solve his problem in such a way as to minimize the total work that he must expend in solving both his immediate problems and his probabilistic future problems . . ..

We noted in Egghe and Lafouge (2006) that several authors confuse MEP and PLE. In the first article on the subject, introducing another principle (PME: Principle of Most Effort), we give a complete explanation of Zipf's law as a decreasing distribution, in the discrete and finite case. We place ourselves here in the case where the informetric statistical distributions are formulated in the shape of frequency–size. We refer readers interested in the rank–frequency

formulation of Zipf's law given above in Egghe's work (Egghe, 2005) where the place within Lotka informetric distributions is studied with precision.

In the continuous case, the two principles are equivalent if the effort function is strictly increasing, unbounded and admissible. This is the case of the logarithmic function, effort function of inverse power distributions and the linear function, effort function of exponential distributions. In practice, the admissibility condition is not very restrictive (readers should refer to points 4 in Appendix A). To completely cover the problem of equivalence, we must now study the infinite discrete case.

Appendix A

(1) $A(\bar{H})$ is not empty.

Proof.

- Let $\bar{H}$ a real number and the function μ,

$$\mu(x) = \exp(-\bar{H}), \quad 1 \le x \le \exp(\bar{H})$$

$$\mu(x) = 0, \qquad \exp(\bar{H}) < x \quad \square$$

It is easy to verify, $\int_1^\infty \mu(x)\,dx = 1$ and $H(\mu) = -\int_1^\infty \mu(x)\,\text{Ln}(x)\,dx = \bar{H}$

(2) $C(\bar{F})$ is not empty.

Proof.

- Let $\bar{F}$ a real number and f an unbounded, strictly increasing function, $C(\bar{F})$ is not empty if and only if, $\bar{F} > f(1)$. With $\mu \in C(\bar{F})$, we have $\int_1^\infty \mu(x) f(x)\,dx = \bar{F} < f(1)\int_1^\infty \mu(x)\,dx = f(1)$
- Conversely, let us suppose $\bar{F} > f(1)$, the function f being unbounded, strictly increasing and continuous, there is $y > 1$, unique so that $\bar{F} = f(y)$. Let us fix $z > y$, with μ_z the density function defined by the three constants μ_1, μ_2, μ_3:

$$\mu_1 = \frac{\int_y^z f(t)\,dt - \bar{F}(z-y)}{(y-1)\int_y^z f(t)\,dt - (z-y)\int_1^y f(t)\,dt}, \quad 1 \le x \le y$$

$$\mu_2 = \frac{\bar{F}(y-1) - \int_1^y f(t)\,dt}{(y-1)\int_y^z f(t)\,dt - (z-y)\int_1^y f(t)\,dt}, \quad y < x \le z$$

$$\mu_3 = 0, \quad x > z$$

It is easy to verify that we have: $\int_1^\infty \mu_z(x)\,dx = 1$ and $\int_1^\infty \mu_z(x) f(x)\,dx = \bar{F}. \quad \square$

Also, the function μ_z is positive because f is strictly increasing and we have $f(1) < f(y) = \bar{F} < f(z)$. For each value of z we thus define a function μ_z.

(3) Additional conditions required for f.

Proof. With the effort function $f(x) = \mathrm{Ln}(x-1)$, for all $\sigma > 0$ we define the lognormal density function μ_σ with parameters, $1, \bar{F}, \sigma$

$$\mu_\sigma(x) = \frac{1}{\sigma(x-1)\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\mathrm{Ln}(x-\bar{F})^2)\right), \quad x > 1$$

$$\mu_\sigma(1) = 0$$

The properties of the lognormal law imply:

$$\int_1^\infty \mu_\sigma(x)\,\mathrm{d}x = 1 \quad \text{and} \quad \int_1^\infty f(x)\mu_\sigma(x)\,\mathrm{d}x = \bar{F}$$

Also, we know that a normal density function $H(0, \sigma)$ has an entropy equal to $\mathrm{Ln}(\sigma\sqrt{2\pi e})$, we deduce from it that the entropy of μ_σ is equal to $(1/2)(1 + \mathrm{Ln}(2) + \mathrm{Ln}(\sigma))$, with $\sup H(\mu)_{\mu \in C(\bar{F})} = +\infty$. $\square$

(4) Example of effort function, strictly increasing, unbounded et not admissible.

Proof. We consider the function $g(x) = 1 + \int_1^x (\exp(t^2)/t)\,\mathrm{d}t$, it is easy to see that this function is strictly increasing. We can then define its inverse f, we then put $f(x) = g^{-1}(x)$. g is also increasing, continuous, positive and unbounded, let us show that it is not admissible. Let λ any positive number, we put $y = f(x)$, we then have $\mathrm{d}y = (y/\exp(-y^2))\,\mathrm{d}x$ with,

$$\int_1^\infty f(x)\exp(-\lambda f(x))\,\mathrm{d}x = \int_1^\infty \exp(y^2 - \lambda y)\,\mathrm{d}y = +\infty \quad \square$$

References

Bradford, S. C. (1934). Sources of information on specific subject. *Engineering*, 85–86.
Egghe, L. (1990). On the duality of informetric systems with application to the empirical law. *Journal of Information Science*, *16*, 17–27.
Egghe, L. (2005). *Power laws in the information production process: Lotkaian Informetrics*. Elsevier.
Egghe, L., & Lafouge, T. (2006). On the relation between the maximum entropy principle and the principle of least effort. *Mathematical and Computer Modelling*, *43*, 1–8.
Lafouge, T. (2007). The source-item coverage of the exponential distribution. *Journal of Informetrics*, *1*(1), 59–67.
Lafouge, T., & Prime Claverie, C. (2005). Production and use of information characterization of informetric distributions using effort function and density function. *Information Processing and Management*, *41*, 1387–1394.
Lafouge, T., & Smolczeswska, A. C. (2006). An interpretation of the effort function trough the mathematical formalism of Exponential Informetric Process. *Information Processing and Management*, *42*, 1142–1450.
Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Science*, *16*, 317–323.
Wyllis, R. E. (1982). Empirical and theoretical bases of Zipf's Law. *Library Trends*, 53–64.
Shannon C. (1993). Collected papers edited by N.J.A. Sloane, & Aaron D. Wyner. New York: IEE Press.
Yablonsky, A. L. (1981). On fundamental regularities of the distribution of scientific productivity. *Scientometrics*, *2*(1), 3–34.
Zipf, G. K. (1949). *Human behaviour and the principle of least effort*. Cambridge, MA, USA: Addison-Wesley. (reprinted: Hafner, New York, USA, 1965).