HAL
open science

On the relation between the maximum entropy Principle and the principle of Least effort : The continuous case

Abdelatif Agouzal, Thierry Lafouge

To cite this version:

Abdelatif Agouzal, Thierry Lafouge. On the relation between the maximum entropy Principle and the principle of Least effort : The continuous case. Journal of Informetrics, 2008, 2 (1), pp.75-88. hal-00606900

HAL Id: hal-00606900

https://hal.science/hal-00606900v1

Submitted on 7 Jul 2011

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.

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

Abdelatif, AGOUZAL[1]; Thierry LAFOUGE[2]

[1]Université de Lyon, France; Université Lyon 1, Institut Camille Jordan, CNRS U.M.R. 5208, Villeurbanne, F-69622, France

[2]Université de Lyon, France; Université Lyon 1, ELICO, Villeurbanne, F-69622, France

Abstract

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

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

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

1 Definitions and context

1.1 Information theory

In 1948 Shannon (Shanon, 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 $[1,..\infty[$ with real positive values and verifying:

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

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

$$\overline{H} = -\int_1^\infty f(x) Ln(\mu(x)) dx, \quad \overline{F} = \int_1^\infty f(x)\mu(x) dx,$$

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

<u>Maximum Entropy Principle (*MEP*)</u>

The *MEP* maximises the entropy provided that the effort remains constant.

<u>Principle of Least Effort (*PLE*)</u>

The *PLE* minimises the effort provided that the entropy remains constant.

<u>Mathematical formulation of the two principles</u>

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

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

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

With these notations, *MEP* and *PLE* are written:

$$MEP \Leftrightarrow Max_{\mu \in C(\overline{F})} \left(-\int_1^\infty \mu(x) Ln(\mu(x)) dx\right)$$

$$LEP \Leftrightarrow Min_{\mu \in A(\overline{H})} \left(\int_1^\infty \mu(x) f(x) dx\right)$$

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.

1.2 Informetrics and entropy aspects

1.2.1 Information Production Process and Inverse power law

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

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

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

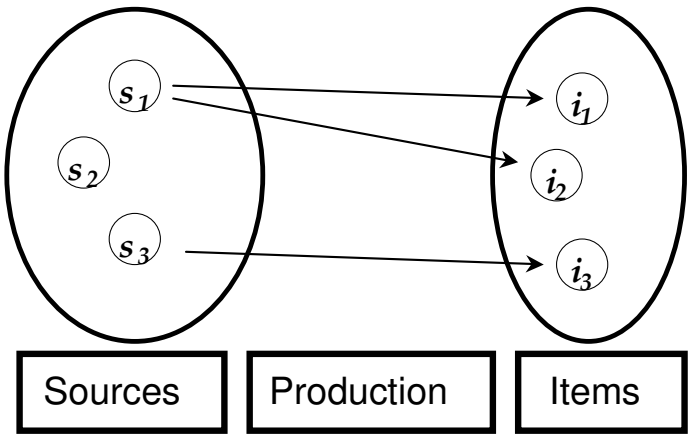

Figure 1: Schematic representation of an Information Production Process

- 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).

In all quoted examples, if we quantify the production of the items by the sources with a size frequency function, this one is decreasing with a long tail and a gap between a high number of sources producing few items and a small number of sources producing a lot. Their most current mathematical formulation is that of an inverse power function, usually called a lotkaian informetric distribution (Egghe, 2005), the corresponding density probability ν_λ is:

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

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

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

It is easy to show that the entropy is a decreasing function of λ. The classical interpretation of Lotka’s law (Lotka, 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?

1.2.2 Previous results

1.2.2.1 Finite discrete case

Suppose μ is a finite discrete distribution, so we have $m \in N$ probabilities $p_1, p_2,.., p_m$ so that $\sum_{r=1}^{m} p_r = 1$, the entropy of such a system is defined as $\overline{H} = -\sum_{r=1}^{m} p_r Ln(p_r)$. Let $F_r > 0$ denote the effort function $(r = 1,..,m)$ the quantity of effort is defined as $\overline{F} = \sum_{r=1}^{m} E_{r.} p_r Ln(p_r)$. In () we prove the following result, if we suppose $p_1, p_2,.., p_m$ 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 and 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 paragraph 1.2.2.2). The aim of this article is to generalize this result.

1.2.2.2 Exponential Informetric Process

In Lafouge and Prime Claverie (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.

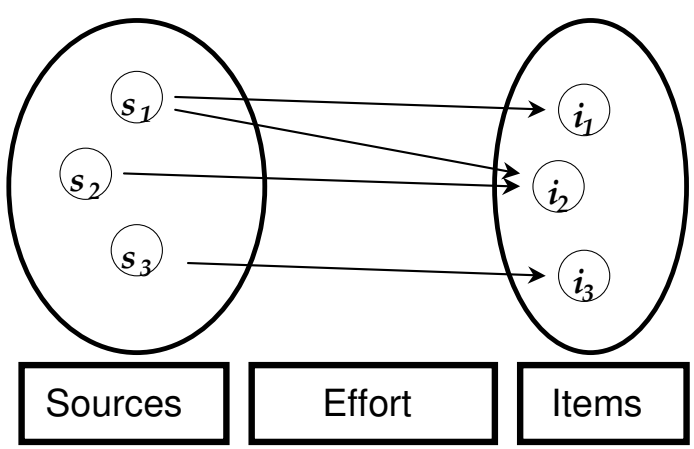

Figure 2: Schematic representation of an informetric process using the effort function

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

$$\int_1^\infty f(x)\exp(-\lambda f(x))dx < \infty \quad [1]$$

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 $\overline{F}$ 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 and Prime Claverie, 2005) the following results:

(i) ν_λ is decreasing,

(ii) The two principles maximum *MEP* and *PLE* are verified simultaneously,

(iii) $\overline{H} = Ln(k) + \lambda\overline{F}$

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

This result partly resolves the problem of the equivalence of the two principles given in paragraph 1.1. 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 Bolzmann 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 $\overline{F} = \int_1^{\infty} f(x)\exp(-\lambda f(x))dx$, the existence of a unique λ verifying this inequality. The demonstration given in paragraph 2 will clarify this point.

1.2.2.3 Effort function and inverse power law

With the preceding formalism, the density function corresponding to an inverse power is written:

$$x \in [1,..\infty[,\ \nu_\lambda(x) = (\lambda - 1)\exp(-(\lambda - 1)Ln(x)),\ \lambda > 1$$

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

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

The logarithmic function is the effort function corresponding to the Lotka distributions. When we are interested for example in the production of words in a text and when we express this regularity using the formalism of Zipf's law (see Conclusion). In (Egghe and Lafouge, 2006, p 8), this hypothesis is formulated thus: *the cost (effort) of using a word with* $i, (i = 1,2,3,....)$ *letters is proportional to* i *hence to* $Log_N(r)$ *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,

$$Limit_{\ x \to \infty} \frac{f(x)}{Ln(x)} = \lambda,\ \lambda > 1$$

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 [1].

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

2.1 Main results

2.1.1 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 reality $\sigma(f)\geq 0$ 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,

$$\int_1^\infty \nu(x)dx=1$$

We define the entropy of μ,

$$H(\mu)=-\int_1^\infty \mu(x)Ln(\mu(x))dx$$

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

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

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

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

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

$$H(\lambda)=-\int_1^\infty \nu_\lambda(x)Ln(\nu_\lambda(x))dx$$

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

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

$$H(\lambda)=-\int_1^\infty \frac{1}{k}\exp(-\lambda f(x))Ln(\frac{\exp(-\lambda f(x))}{k})dx=\frac{1}{k}\int_1^\infty \exp(-\lambda f(x))(\lambda f(x)+Ln(k))dx$$

$$=\lambda\overline{F}+Ln(k)$$

Also, the two functions are defined:

$$Z(\lambda)=-Ln(\int_1^\infty \exp(-\lambda(f(x)-\overline{F}))dx$$

$$R(\lambda)=-\frac{1}{\lambda}Ln(\int_1^\infty \exp(-\lambda f(x)-\overline{H})dx)$$

Lemma 2.2

With $\mu\in C(\overline{F})$, $\lambda>\sigma(f)$, we have: $\int_1^\infty \mu(x)Ln(\nu_\lambda(x))dx=Z(\lambda)$

Proof

$$\int_1^\infty \mu(x)Ln(\nu_\lambda(x))dx=\int_1^\infty \mu(x)(-\lambda f(x)-Ln(\int_1^\infty \exp(-\lambda f(x))dx)dx$$

$$=-\lambda.\overline{F}-Ln(\int_1^\infty \exp(-\lambda f(x))dx)=-Ln(\exp(\lambda\overline{F})\int_1^\infty \exp(-\lambda f(x)dx)$$

$$=-Ln(\int_1^\infty \exp(-\lambda(f(x)-\overline{F}))dx$$

Lemma 2.3

Let $\nu_\lambda\in C(\overline{F})$, we have $\int_1^\infty (f(x)-\overline{F})\exp(-\lambda f(x))dx=0$

Proof

$\nu_\lambda\in C(\overline{F})$, hence

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

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

$$= k\overline{F} - k\overline{F} = 0$$

2.1.1 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) $0 \geq Max\, H(\mu)_{\mu \in C(\overline{F})} \geq Sup(Z(\lambda))_{\lambda > \sigma(f)}$

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

$$H(\lambda_0) = Max\, H(\mu)_{\mu \in C(\overline{F})} = Sup(Z(\lambda))_{\lambda > \sigma(f)} = Z(\lambda_0)$$

(ii) $0 \geq Min\, F(\mu)_{\mu \in A(\overline{H})} \geq Sup(R(\lambda))_{\lambda > \sigma(f)}$

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

$$F(\lambda_1) = Min F(\mu)_{\mu \in A(\overline{H})} = Sup(R(\lambda))_{\lambda > \sigma(f)} = R(\lambda_1)$$

(iii) If $\overline{H} = -Z(\lambda_1)$ or $\overline{F} = R(\lambda_0)$ then $\nu_{\lambda_0} = \nu_{\lambda_1}$

(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 $\overline{H}$ and quantity of effort $\overline{F}$ are linked, in other words if $\overline{H} = -Z(\lambda_1)$ or if $\overline{F} = R(\lambda_0)$, the two principles are equivalent. We note that in this case we find the result (iii) of paragraph 1.2.2.2, $\overline{H} = \lambda_1 \overline{F} + Ln(k)$ or $\overline{H} = \lambda_0 \overline{F} + Ln(k)$.

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

We have,

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

With $\nu_{\lambda_0} \in C(\overline{F})$, 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.

2.2 Demonstration of the equivalence theorem

Preliminary

For this problem to be of mathematical interest, we need to show that $A(\overline{H})$ and $C(\overline{F})$ 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 1,2 and 3 in the appendix 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) $yLn(y) \geq y - z + yLn(z)$, and strict inequality if, and only if, $y \neq z$

(ii) $yLn(y) - zLn(z) \geq (1 + Ln(z))(y - z)$

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 $t = \frac{z}{y}$, we have

$\forall t > 0, t \neq 1,\ Ln(t) < t - 1$, hence $Ln(z) - Ln(y) < \frac{z-y}{y}$.

Proposition 2.1

For any admissible function, we have:

$$0 \geq Max\, H(\mu)_{\mu \in C(\overline{F})} \geq Sup\, Z(\lambda)_{\lambda > \sigma(f)}$$

Proof

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

$$\int_1^\infty \mu(x) Ln(\mu(x))dx \geq \int_1^\infty (\mu(x) - v_\lambda(x))dx + \int_1^\infty \mu(x) Ln(\nu_\lambda(x))dx$$

we know,

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

consequently

$$-H(\mu) \geq \int_1^\infty \mu(x) Ln(\nu_\lambda(x)) dx$$

thus according to lemma 2.2

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

hence

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

Also, according to point 2 in the appendix, there is a function $\mu_z \in C(\overline{F})$ verifying $\mu_z \leq 1$, consequently $\mu_z Ln(\mu_z) \leq 0$ and thus,

$$0 \geq Max\, H(\mu)_{\mu \in C(\overline{F})} \geq Sup\, Z(\lambda)_{\lambda > \sigma(f)}$$

Proposition 2.2

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

$$\int_1^\infty \mu(x) Ln(\mu(x)) dx - \int_1^\infty \nu_\lambda(x) Ln(\nu_\lambda(x)) dx > \lambda \int_1^\infty (\nu_\lambda(x) - \mu(x)) f(x) dx$$

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:

$$\mu(x) Ln(\mu(x)) - \nu_\lambda(x) Ln(\nu_\lambda(x)) \geq (\mu(x) - \nu_\lambda(x))(1 + Ln(\nu_\lambda(x))$$

By integrating the right member, we obtain:

$$\int_1^\infty (\mu(x) - \nu_\lambda(x)) dx + \int_1^\infty (\mu(x) - \nu_\lambda(x)) Ln(\nu_\lambda(x)) dx$$

However

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

Consequently,

$$\int_1^\infty (\mu(x)-\nu_\lambda(x))Ln(\nu_\lambda(x))dx=\int_1^\infty (\mu(x)-\nu_\lambda(x))(-\lambda f(x)-Ln(k))dx$$

Hence

$$=\int_1^\infty (\mu(x)-\nu_\lambda(x))(-\lambda f(x))dx$$

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 $C(\overline{F})$ are sets of uniqueness for the functions ν_{λ_0}. More precisely, we demonstrate the result below.

Proposition 2.3

With f an admissible function, there exists at most a function ν_{λ_0} belonging to $C(\overline{F})$ 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 $\nu_{\lambda_1}, \nu_{\lambda_2} \in C(\overline{F})$, according to lemma 2.1 we know that the entropy of ν_λ is finite, by applying proposition 2.2 twice, we obtain

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

And in the same way

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

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

2) Existence

Let a function $\nu_{\lambda_0} \in C(\overline{F})$, according to lemma 2.5, we can write

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

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

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

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

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

consequently,

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

hence

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

and thus

$$Sup\, Z(\lambda)_{\lambda > \sigma(f)} \leq Z(\lambda_0)$$

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 $\nu_{\lambda_0} \in C(\overline{F})$, then this real number is unique and moreover:

$$Sup(Z(\lambda)_{\lambda > \sigma(f)} = Z(\lambda_0) = MaxH(\mu)_{\mu \in C(\overline{F})} = -\int_1^\infty \nu_{\lambda_0}(x) Ln(\nu_{\lambda_0}(x))dx$$

Proof

Uniqueness and characterization by the function Z are consequences of proposition 2.3.

With $\mu \in C(\overline{F}), \mu \neq \nu_{\lambda_0}$ using the inequality of proposition 2.2,

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

However

$$\int_1^\infty (\nu_{\lambda_0}(x) - \mu(x)) f(x)dx = \overline{F} - \overline{F} = 0$$

Thus

$$-\int_1^\infty \mu(x) Ln(\mu(x))dx < -\int_1^\infty \nu_{\lambda_0}(x) Ln(\nu_{\lambda_0}(x))dx$$

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 $]\sigma(f),..\infty[$, as the following example shows. With the effort function $f(x) = Ln(x+1) + Ln(Ln^2(x+1))$, we show $\int_1^\infty \exp(-f(x)dx < \infty$, therefore $Z(1)$ is well defined whereas $\int_1^\infty f(x)\exp(-f(x))dx = \infty$.

In the same way, with $\overline{H}$ 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 $A(\overline{H})$ verifying the inequality: $Sup\, R(\lambda)_{\lambda > \sigma(f)} = R(\lambda_1)$

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

Theorem 2.2

We suppose that there is a reality $\lambda_1 > \sigma(f)$ so that $\nu_{\lambda_1} \in A(\overline{H})$, then this reality is unique and also:

$$Sup(Z(\lambda))_{\lambda > \sigma(f)} = R(\lambda_1) = Min\, F(\mu)_{\mu \in A(\overline{H})} = \int_1^\infty \nu_{\lambda_1}(x) f(x) dx$$

Proof

Uniqueness and characterization by the function R are consequences of proposition 2.3.

With $\mu \in A(\overline{H})$, using the inequality of proposition 2.2 we obtain,

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

$$0 = \overline{H} - \overline{H} > \lambda \int_1^\infty (\nu_{\lambda_1}(x) - \mu(x)) f(x) dx$$

consequently, since $\lambda > 0$,

$$\int_1^\infty f(x) \nu_{\lambda_1}(x) dx < \int_1^\infty f(x) \mu(x) dx$$

using theorems 2.1 and 2.2, we obtain the equivalence theorem.

Theorem 2.3

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

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

then, we have:

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

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

2. We have

$$0 \geq Max\, H(\mu)_{\mu \in C(\overline{F})} \geq Sup\, Z(\lambda)_{\lambda > \sigma(f)}$$

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

$$H(\lambda_0) = Max\, H(\mu)_{\mu \in C(\overline{F})} = Sup\, Z(\lambda)_{\lambda > \sigma(f)} = Z(\lambda_0)$$

where

$$Z(\lambda) = -Ln(\int_1^{\infty} \exp(-\lambda(f(x) - \overline{F}))dx$$

3. We have

$$0 \geq Min\, F(\mu)_{\mu \in A(\overline{H})} \geq Sup\, R(\lambda)_{\lambda > \sigma(f)}$$

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

$$F(\lambda_1) = Min F(\mu)_{\mu \in A(\overline{H})} = Sup\, R(\lambda)_{\lambda > \sigma(f)} = R(\lambda_1)$$

where

$$R(\lambda) = \frac{1}{\lambda} Ln(\int_1^{\infty} \exp(-\lambda f(x) - \overline{H})dx)$$

4. Finally, if $\overline{H} = -Z(\lambda_1)$ or $\overline{F} = R(\lambda_0)$ then $\nu_{\lambda_0} = \nu_{\lambda_1}$, 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 $f(x) = Ln(x)$, in other words Lotka distributions (see 1.2.1), in this case $\sigma(f) = 1$.

Conclusion

One of the most intriguing phenomena in infometry, and widely studied in quantitative linguistics, is Zipf's law. 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 (Ronald E. 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 infometric 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 infometric distributions is studied with precision.

In the continuous case, the two principles are equivalent if the effort function is strictly increasing, unbounded and admissible, which 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 the appendix). To completely cover the problem of equivalence, we must now study the infinite discrete case.

Appendix

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

Proof

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

$$\mu(x)=\exp(-\overline{H})\ ,\ 1\leq x\leq \exp(\overline{H})$$

$$\mu(x)=0, \qquad \exp(\overline{H})<x$$

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

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

Proof

- Let $\overline{F}$ a real number and f an unbounded, strictly increasing function, $C(\overline{F})$ is not empty if and only if, $\overline{F} > f(1)$.

With $\mu \in C(\overline{F})$, we have $\int_1^\infty \mu(x) f(x) dx = \overline{F} < f(1) \int_1^\infty \mu(x) dx = f(1)$

- Conversely, let us suppose $\overline{F} > f(1)$, the function f being unbounded, strictly increasing and continuous, there is $y > 1$, unique so that $\overline{F} = f(y)$. Let us fix $z > y$, with μ_z the density function defined by the three constants μ_1, μ_2, μ_3:

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

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

$$\mu_3 = 0 \qquad x > z$$

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

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

3) Additional conditions required for f.

Proof

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

$$\mu_\sigma(x) = \frac{1}{\sigma(x-1)\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\overline{F})^2\right), \; x > 1$$

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

The properties of the lognormal law imply:

$$\int_1^{\infty} \mu_\sigma(x)dx = 1 \text{ and } \int_1^{\infty} f(x)\mu_\sigma(x)dx = \overline{F}$$

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

4) Example of effort function, stricty increasing, unbounded et not admissible
Proof

We consider the function $g(x) = 1 + \int_1^x \frac{\exp(t^2)}{t} dt$, 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 $dy = \frac{y}{\exp(-y^2)} dx$ with,

$$\int_1^{\infty} f(x)\exp(-\lambda f(x))dx = \int_1^{\infty} \exp(y^2 - \lambda y)dy = +\infty$$

References

Egghe, L.(1990). On the duality of informetric systems with application to the empirical law. In *Journal of Information Science*, 16, 17-27.

Egghe, L. (2005). Power laws in the information production process: Lotkaian Informetrics, Elsevier 2005.

Egghe, L. and Lafouge, T. (2006). On the Relation Between the Maximum Entropy Principle and the Principle of Least Effort In *Mathematical and Computer Modelling*, 43,1-8.

Lafouge, T., Prime Claverie, C. (2005). Production and use of information characterization of informetric distributions using effort function and density function. In *Information Processing and Management*, 41, 1387-1394.

Lafouge, T., and Smolczeswska, A C. (2006). An interpretation of the effort function trough the mathematical formalism of Exponential Informetric Process. *In Information Processing and Management*, 42, 1142-1450.

Lafouge, T. (2007). The source-item coverage of the exponential distribution. In *Journal of Informetrics.* Article in press.

Lotka, A.J. (1926). The frequency distribution of scientific productivity. In *Journal of the Washington Academy of Science*, 16 317-323.

Ronald E. Wyllis (1982). Empirical and Theoretical Bases of Zipf's Law. In *Library Trends*, 53-64.

Shannon C. (1993) Collected papers edited by N.J.A. Sloane, Aaron D. Wyner.
New York: IEE Press c1993.

Yablonsky, A.L. (1981). On fundamental regularities of the distribution of scientific productivity. In *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.