Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Mathematical and Computer Modelling 43 (2006) 1–8
MATHEMATICAL AND COMPUTER MODELLING
www.elsevier.com/locate/mcm

# On the relation between the Maximum Entropy Principle and the Principle of Least Effort

L. Egghe[a,b,*], T. Lafouge[c]

[a] Hasselt University, Campus Diepenbeek, Agoralaan, Gebouw D, 3590 Diepenbeek, Belgium
[b] Universiteit Antwerpen (UA), Universiteitsplein 1, B-2610 (Wilrijk), Belgium
[c] Laboratoire Docsi-URSIDOC, batiment Oméga, Université Claude Bernard Lyon 1, 43 Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France

Received 20 January 2004; accepted 20 January 2004

**Abstract**

The Maximum Entropy Principle (MEP) maximises the entropy subject to the constraint that the effort remains constant. The Principle of Least Effort (PLE) minimises the effort subject to the constraint that the entropy remains constant. The paper investigates the relation between these two principles. It is shown that (MEP) is equivalent with the principle "(PLE) or (PME)" where (PME) is (introduced in this paper) the Principle of Most Effort, meaning that the effort is maximised subject to the constraint that the entropy remains constant.


*Keywords:* Maximum Entropy Principle; Principle of Least Effort; Principle of Most Effort; Equivalence

## 1. Introduction

Suppose we have $m \in N$ probabilities $p_1, p_2, \ldots, p_m$ such that

$$\sum_{r=1}^{m} p_r = 1. \tag{1}$$

One could think of an $m$-letter alphabet e.g. $(a, b, \ldots, z, 1, \ldots, +)$, where the numbers $p_1, p_2, \ldots, p_m$ denote the probabilities of occurrence of these letters. Hence all probabilities are strictly positive. Of course many other interpretations are possible.

The average information content, called entropy, of such a system is defined as (cf. [16])

$$\overline{H} = -\sum_{r=1}^{m} p_r . \ln(p_r). \tag{2}$$

* Corresponding author at: Hasselt University, Campus Diepenbeek, Agoralaan, Gebouw D, 3590 Diepenbeek, Belgium.
*E-mail addresses:* leo.egghe@uhasselt.be (L. Egghe), Thierry.Lafouge@univ-lyon1.fr (T. Lafouge).

0895-7177/$ - see front matter 
doi:10.1016/j.mcm.2004.01.017

Note that any logarithm can be used; in information theory it is common to use $\log_2$ but, for reasons of simplification of the calculations we will use here $\ln = \log_e$. Of course any result proved in this paper on (2) will also be valid when using another logarithm. The measure $\overline{H}$ is fundamental in the sciences. In information theory it is the defining formula for bits (binary digits) and for the optimal text-length when using non-fixed length coding (cf. [5]).

Another important aspect of the numbers $p_1, p_2, \ldots, p_m$ can be described as effort (of using the $m$ attributes, e.g. symbols). Let $E_r > 0$ denote the effort of using attribute $r$ ($r = 1, \ldots, m$). We suppose that the order is such that $(E_r)_{r=1,\ldots,m}$ increases. The average effort, denoted by $\overline{E}$, is:

$$\overline{E} = \sum_{r=1}^{m} E_r . p_r . \tag{3}$$

Note that, supposing $(E_r)_{r=1,..,m}$ to be increasing, is not a mathematical restriction but for the applications it is convenient to order our attributes this way. Indeed consider the application in linguistics where the attributes $r = 1, \ldots, m$ stand for words in a text: $r$ then denotes the rank of a word (type) according to the number of times it is used (tokens) in the text (where the higher the rank, the less it is used in the text). In this application it is clear that, the higher the rank, the more "exotic" the word is (and, usually, because of this, the word is longer), hence the higher cost (or effort or energy) in using it.

Let us now define two important optimization principles. The Maximum Entropy Principle (MEP) requires $\overline{H}$ (formula (2)) to be maximised subject to certain "energy" conditions which have prescribed average values. The historic definition was given in [6] in the field of physics. Stated exactly, in this paper we define the Maximum Entropy Principle (MEP) as: maximise $\overline{H}$ in (2) subject to a fixed value of (3) and also subject to (1) (see also [17,4, 13] for a definition using continuous variables).

The Principle of Least Effort, attributed to Zipf in linguistics [18], requires the average effort $\overline{E}$ (formula (3)) to be minimal subject to certain fixed conditions on average information content, i.e. entropy. Stated exactly, in this paper we define the Principle of Least Effort (PLE) as: minimise $\overline{E}$ (in (3)) subject to a fixed value of (2) and of course subject to (1).

When we interpret (intuitively) both principles in the same field, e.g. linguistics, we can say that (MEP) maximises the information content of a text, among all texts with a fixed average effort value. The (PLE) keeps the information content fixed and determines the text with this information content, requiring the least average effort in producing it. Even more intuitively we can say that (MEP) gives maximal profit among different situations requiring the same effort in producing it, while (PLE) requires a minimum effort among different situations that yield the same profit.

Formulated this way both principles look very similar and one can pose the problem: are (MEP) and (PLE) equivalent or does one principle only imply the other? In other words: establish the mathematical relation between (MEP) and (PLE). Before we go into this problem (which will be solved in this paper) we remark that the "state of the art" of these principles is very different. A search in April 2003 in Mathscinet yielded as much as 177 documents on the phrase "maximum entropy principle" while there were only 9 documents on the phrase "principle of least effort" (phrases to appear in the review text). No paper was in the intersection of both sets. Similar results were found in the database Zentralblatt Math. Hence no paper apparently deals with the relation between (MEP) and (PLE). Although this is surprising at first glance, the reason, most probably, lies in the fact that both principles have completely different origins. The (MEP), being introduced in 1957 by Jaynes in physics, is well known in exact sciences, such as, apart from physics: information science, chemistry, biology, artificial intelligence and of course in probability theory and analysis. For these applications see e.g. [9,8]. The (PLE), being attributed to [18], was not formulated within the exact sciences but in linguistics where, in addition, the formulation was not given in mathematical terms. The (PLE), in its historical formulation is, hence, to be understood as a sociological behaviour of persons and groups of persons, see also [14].

We even noted that some papers confuse (MEP) and (PLE): in [15] the correct definition of (PLE) is given but, in the subsequent mathematical calculations of it, one uses (MEP) instead. In [10,11] there is no mistake except for the fact that the (MEP) is called "Principle of Least Effort".

It is now clear from the above that a mathematical proof of a relation between (MEP) and (PLE) would be interesting. This will be given in the next section. The development of the proof makes it clear that, in order to describe the complete relation between (MEP) and (PLE), we are in need of another optimization principle, the Principle of Most Effort (PME). Its definition is clear from the one of (PLE): (PME) requires to maximise the average effort $\overline{E}$

(formula (3)) subject to a fixed value of $\overline{H}$ (formula (2)) and of course subject to (1). The practical usability of (PME) is of no importance here: (PME) is the (mathematically) missing link between (MEP) and (PLE) as will become clear in the next section. It also turns out that there is no such principle as the "minimal entropy principle": in the next section we will show that any constrained extremal problem for $\overline{H}$ (subject to constant $\overline{E}$ (formula (3)) and subject to (1)) yields a maximum for $\overline{H}$; this is not the case for the extremal problem for $\overline{E}$ subject to constant $\overline{H}$ and (1), hence leading to a possible maximum (hence (PME)) as introduced above.

## 2. Characterisation of the Maximum Entropy Principle in terms of the Principles of Least and Most Effort

In this section we will prove the following (surprising) result.

**Theorem.** *The following assertions are equivalent:*

(i) *The Maximum Entropy Principle*
(ii) *The Principle of Least Effort or the Principle of Most Effort.*

The proof is split up in several propositions: the first three giving necessary conditions for the $m$-tuple $p_1, p_2, \ldots, p_m$ to satisfy the (MEP), (PLE) and (PME) respectively; the fourth proposition gives in one proof that the necessary conditions are also sufficient. So we obtain characterisations (in terms of the values of the probabilities $p_1, p_2, \ldots, p_m$) of all three principles (MEP), (PLE) and (PME) from which the main theorem will follow. For the notation we refer to formulae (1)–(3).

**Proposition 2.1.** *(MEP) implies*

$$p_r = c\rho^{-E_r} \quad r = 1, \ldots, m, \exists c > 0, \exists \rho > 0.$$

**Proof.** (MEP) requires $\overline{H}$ to be maximal subject to a fixed value of $\overline{E}$ and subject to (1). The method of the multiplicators of Lagrange (see e.g. [1]) yields a necessary condition for this problem. So we form the function $G$ (in the variables $p_1, p_2, \ldots, p_m, \lambda, \mu$):

$$G = -\sum_{r=1}^{m} p_r . \ln(p_r) + \lambda\left(\overline{E} - \sum_{r=1}^{m} E_r . p_r\right) + \mu\left(1 - \sum_{r=1}^{m} p_r\right) \tag{4}$$

for which we require

$$\frac{\partial G}{\partial p_r} = 0 \tag{5}$$

for all $r = 1, 2, \ldots, m$ as a necessary condition for (MEP).

(At this moment we obtain a necessary condition for the constraint $\overline{H}$ to be extremal, hence also for $\overline{H}$ to be maximal but the proof of the sufficient condition will show that $\overline{H}$ can only be maximal and it will also be shown that this is not the case for the constraint extremum of the average effort $\overline{E}$).

Formulae (4) and (5) yield:

$$\frac{\partial G}{\partial p_r} = -1 - \ln(p_r) - \lambda E_r - \mu = 0$$

hence

$$p_r = e^{-1-\mu} . e^{-\lambda E_r}$$
$$p_r = c.\rho^{-E_r} \quad \text{with } c = e^{-1-\mu} > 0 \text{ and } \rho = e^{\lambda} > 0 \text{ for all } r = 1, 2, \ldots, m. \quad \square$$

**Proposition 2.2.** *(PLE) implies*

$$p_r = c\rho^{-E_r} \quad r = 1, \ldots, m, \exists c > 0, \exists \rho \geq 1. \tag{6}$$

**Proof.** (PLE) requires $\overline{E}$ to be minimal subject to a fixed value of $\overline{H}$ and subject to (1). The same method as in Proposition 2.1 gives (for the moment we only use the constrained extremality of $\overline{E}$ but later in this proof the obtained condition will be sharpened using the full (PLE)): we define the function $G^*$ as follows:

$$G^* = \sum_{r=1}^{m} p_r.E_r + \lambda\left(\overline{H} + \sum_{r=1}^{m} p_r.\ln(p_r)\right) + \mu\left(1 - \sum_{r=1}^{m} p_r\right). \tag{7}$$

A necessary condition for (PLE) is given by:

$$\frac{\partial G^*}{\partial p_r} = 0 \tag{8}$$

for all $r = 1, 2, \ldots, m$. This gives

$$\frac{\partial G^*}{\partial p_r} = E_r + \lambda(1 + \ln(p_r)) - \mu = 0$$

hence

$$p_r = e^{\frac{\mu}{\lambda}-1}.e^{-\frac{E_r}{\lambda}}$$

hence

$$p_r = c.\rho^{-E_r} \quad \text{with } c = e^{\frac{\mu}{\lambda}-1} > 0 \text{ and } \rho = e^{\frac{1}{\lambda}} > 0 \text{ for all } r = 1, 2, \ldots, m.$$

Note that $\lambda \neq 0$ since $\lambda = 0$ implies $\rho = \infty$ and hence $p_r = 0$ for all $r = 1, 2, \ldots, m$ contradicting (1) (and the fact that all probabilities are strictly positive).

Note that we proved (6) but only for $\rho > 0$. That $\rho \geq 1$ will follow from (PLE), i.e. that the average effort $\overline{E}$ is minimal. This will be proved now. Since we have (6), we will have proved $\rho \geq 1$ if we can show that the $p_1, p_2, \ldots, p_m$ decrease, since we assume the $E_1, E_2, \ldots, E_m$ to be positive and to increase (Note: our proof does not depend on the increasing order of the $E_1, E_2, \ldots, E_m$: for $E_1, E_2, \ldots, E_m$ in any order we then have to show that the $p_1, p_2, \ldots, p_m$ are in inverse order with respect to the $E_1, E_2, \ldots, E_m$, hence again $\rho \geq 1$; using increasing $E_1, E_2, \ldots, E_m$, however, better fixes the ideas).

Let $\rho \neq 1$. Suppose that the $p_1, p_2, \ldots, p_m$ satisfy (PLE), hence are of the form (6) and that they do not decrease. Let then $i < j, i, j \in \{1, \ldots, m\}$ be such that $p_i < p_j$. Define $\pi$ to be the elementary permutation of $\{1, \ldots, m\}$ defined as:

$$(\pi(1), \ldots, \pi(m)) = (1, \ldots, i-1, j, i+1, \ldots, j-1, i, j+1, \ldots, m). \tag{9}$$

Then we have that

$$-\sum_{r=1}^{m} p_r.\ln(p_r) = \overline{H} = -\sum_{r=1}^{m} p_{\pi(r)}.\ln(p_{\pi(r)}) \tag{10}$$

and

$$\sum_{r=1}^{m} p_r = 1 = \sum_{r=1}^{m} p_{\pi(r)}. \tag{11}$$

But

$$\begin{aligned}\sum_{r=1}^{m} E_r.p_{\pi(r)} &= E_i p_{\pi(i)} + E_j p_{\pi(j)} + \sum_{r\neq i,j} E_r.p_{\pi(r)} \\ &= E_i p_i + E_j p_j + \sum_{r\neq i,j} E_r.p_{\pi(r)} + E_i(p_{\pi(i)} - p_i) + E_j(p_{\pi(j)} - p_j) \\ &= \sum_{r=1}^{m} E_r.p_r + (E_j - E_i)(p_{\pi(j)} - p_{\pi(i)})\end{aligned} \tag{12}$$

since $\pi(i) = j$ and $\pi(j) = i$ and $\pi(r) = r, \forall r \in \{1, \dots, m\} \setminus \{i, j\}$. But $i < j$ implies $E_i < E_j$ (since the $E_r$ are increasing and since $E_i = E_j$ implies $p_i = p_j$ by the already proved formula (6), contradicting $p_i < p_j$) and $p_{\pi(j)} - p_{\pi(i)} = p_i - p_j < 0$.

Hence

$$(E_j - E_i)(p_{\pi(j)} - p_{\pi(i)}) < 0.$$

Consequently

$$\sum_{r=1}^{m} E_r p_{\pi(r)} < \sum_{r=1}^{m} E_r p_r. \tag{13}$$

Now (10), (11) and (13) contradict the fact that the $p_1, p_2, \dots, p_m$ satisfy (PLE). So the $p_1, p_2, \dots, p_m$ decrease, hence $\rho > 1$ (if $\rho \neq 1$). Consequently $\rho \geq 1$. □

That (PME) is "complementary" with respect to (PLE), in the connection of (MEP) follows from the next proposition.

**Proposition 2.3.** *(PME) implies*

$$p_r = c\rho^{-E_r} \quad r = 1, \dots, m, \exists c > 0, \exists 0 < \rho \leq 1.$$

**Proof.** Using exactly the same function as in Proposition 2.2 we again find that

$$p_r = c\rho^{-E_r} \qquad r = 1, \dots, m, \exists c > 0, \exists \rho > 0. \tag{14}$$

(In fact at this stage we only used the constrained extremality of $\overline{E}$ which also contains (PME).) We now have to show that $0 < \rho \leq 1$. In other words, by (14) and the fact that the $E_1, E_2, \dots, E_m$ increase, we have to prove that the $p_1, p_2, \dots, p_m$ increase. This proof goes along the lines of the similar proof in Proposition 2.2. Let $\rho \neq 1$. Suppose that the $p_1, p_2, \dots, p_m$ are not increasing. Let then $i < j, i, j \in \{1, \dots, m\}$ be such that $p_i > p_j$. Define $\pi$ to be the elementary permutation of $\{1, \dots, m\}$ defined as in Proposition 2.2. Hence also (12) is valid. But now $i < j$ implies $E_i < E_j$ and $p_{\pi(j)} > p_{\pi(i)}$, hence by (12), $\sum_{r=1}^{m} E_r p_{\pi(r)} > \sum_{r=1}^{m} E_r p_r$ contradicting that the $p_1, p_2, \dots, p_m$ satisfy (PME). So the $p_1, p_2, \dots, p_m$ increase and hence $0 < \rho \leq 1$. □

We will now show, in one proof, that all the necessary conditions proved in Propositions 2.1–2.3 are also sufficient.

**Proposition 2.4.**

(i)

$$p_r = c\rho^{-E_r} \quad r = 1, \dots, m, \exists c > 0, \exists \rho > 0 \tag{15}$$

*implies (MEP).*

(ii)

$$p_r = c\rho^{-E_r} \quad r = 1, \dots, m, \exists c > 0, \exists \rho \geq 1 \tag{16}$$

*implies (PLE).*

(iii)

$$p_r = c\rho^{-E_r} \quad r = 1, \dots, m, \exists c > 0, \exists 0 < \rho \leq 1 \tag{17}$$

*implies (PME).*

**Proof.** Given one of the situations (15), (16) or (17), define the following function $f_r$ of the variable $x_r > 0, r = 1, \dots, m$:

$$f_r(x_r) = x_r . \ln(x_r) - (1 + \ln c)x_r + (\ln \rho)E_r .x_r. \tag{18}$$

Then we have that

$$f_r'(x_r) = 1 + \ln(x_r) - 1 - \ln(c) + \ln(\rho)E_r = 0$$

implies that

$$x_r = e^{\ln(c)-\ln(\rho)E_r}$$
$$x_r = c\rho^{-E_r}$$

$r = 1, \ldots, m$, hence the given functions (15), (16) or (17) respectively. Furthermore

$$f''(x_r) = \frac{1}{x_r} > 0$$

for all $x_r$. Hence the given functions (15), (16) or (17) satisfy $f_r'(p_r) = 0$ and $f_r''(p_r) > 0$, hence $f_r$ has a minimum in $x_r = p_r$ with $p_r$ as in (15), (16) or (17), respectively, $r = 1, \ldots, m$. So for all $x_r > 0$, $r = 1, \ldots, m$:

$$f_r(x_r) \geq f_r(p_r)$$

i.e.

$$x_r.\ln(x_r) - (1+\ln c)x_r + (\ln\rho)E_r.x_r \geq p_r.\ln(p_r) - (1+\ln c)p_r + (\ln\rho)E_r.p_r. \tag{19}$$

Hence

$$\sum_{r=1}^m x_r.\ln(x_r) - (1+\ln c)\sum_{r=1}^m x_r + (\ln\rho)\sum_{r=1}^m E_r.x_r \geq \sum_{r=1}^m p_r.\ln(p_r) - (1+\ln c)\sum_{r=1}^m p_r + (\ln\rho)\sum_{r=1}^m E_r.p_r. \tag{20}$$

The proof is now split into four parts: (a) represents the proof of (i), (b) represents the proof of (ii) for $\rho \neq 1$, (c) represents the proof of (iii) for $\rho \neq 1$, and (d) represents the proof of (ii) and (iii) for $\rho = 1$.

(a) Let now (15) be given: $\rho > 0$ and require:

$$\sum_{r=1}^m x_r = \sum_{r=1}^m p_r = 1, \quad \sum_{r=1}^m E_r x_r = \sum_{r=1}^m E_r p_r = \overline{E}$$

be constants. Then (20) implies, for any $\rho > 0$: $-\sum_{r=1}^m x_r.\ln(x_r) \leq -\sum_{r=1}^m p_r.\ln(p_r)$.

Hence the $p_1, p_2, \ldots, p_m$ of the form (15) (for any $\rho > 0$) satisfy the (MEP). This completes the proof of (a).

(b) Let now (16) given and let $\rho \neq 1$. So $\rho > 1$ and

$$\sum_{r=1}^m x_r = \sum_{r=1}^m p_r = 1$$
$$-\sum_{r=1}^m x_r.\ln(x_r) = -\sum_{r=1}^m p_r.\ln(p_r) = \overline{H}$$

be constants. Then (20) implies, for $\rho > 1$ (hence $\ln(\rho) > 0$):

$$\sum_{r=1}^m E_r x_r \geq \sum_{r=1}^m E_r p_r.$$

Hence the $p_1, p_2, \ldots, p_m$ of the form (16) ($\rho > 1$) satisfy the (PLE).

(c) Let now (17) be given and let $\rho \neq 1$. So $0 < \rho < 1$ and

$$\sum_{r=1}^m x_r = \sum_{r=1}^m p_r = 1$$
$$-\sum_{r=1}^m x_r.\ln(x_r) = -\sum_{r=1}^m p_r.\ln(p_r) = \overline{H}$$

be constants. Then (20) implies, for $0 < \rho < 1$ (hence $\ln(\rho) < 0$):

$$\sum_{r=1}^{m} E_r x_r \leq \sum_{r=1}^{m} E_r p_r.$$

Hence the $p_1, p_2, \ldots, p_m$ of the form (17) ($0 < \rho < 1$) satisfy the (PME).

(d) Let now $\rho = 1$. Hence $p_1 = p_2 = \cdots = p_m = \frac{1}{m}$. Now (20) cannot be used anymore since $\ln(\rho) = 0$. We will now show directly that (PLE) and (PME) are valid. Let

$$\sum_{r=1}^{m} x_r = \sum_{r=1}^{m} p_r = 1$$

$$-\sum_{r=1}^{m} x_r . \ln(x_r) = -\sum_{r=1}^{m} p_r . \ln(p_r) = \overline{H} = \ln(m)$$

then $x_r = p_r$ for all $r = 1, \ldots, m$. This is well known since $\overline{H}$ attains its free maximum (apart from the requirement $\sum_{r=1}^{m} x_r = 1$) in only one point, namely

$$(x_1, \ldots, x_m) = \left(\frac{1}{m}, \ldots, \frac{1}{m}\right) = (p_1, \ldots, p_m).$$

We refer the reader to [7] or to [12]. Hence

$$\sum_{r=1}^{m} E_r x_r = \sum_{r=1}^{m} E_r p_r = \frac{1}{m} \sum_{r=1}^{m} E_r,$$

hence (PLE) as well as (PME). Note that this case is degenerate since only one point $(\frac{1}{m}, \ldots, \frac{1}{m})$ is involved here. This completes the proof of (ii) and (iii) and hence of the proposition. □

We hence have proved the theorem:

**Theorem 2.5.** (i) *(MEP) is equivalent with* $p_r = c\rho^{-E_r}$ $r = 1, \ldots, m, \exists c > 0, \exists \rho > 0$ (15)
(ii) *(PLE) is equivalent with* $p_r = c\rho^{-E_r}$ $r = 1, \ldots, m, \exists c > 0, \exists \rho \geq 1$ (16)
(iii) *(PME) is equivalent with* $p_r = c\rho^{-E_r}$ $r = 1, \ldots, m, \exists c > 0, \exists 0 < \rho \leq 1$ (17).

The case $\rho = 1$ is the only one in the intersection (PLE) and (PME) and corresponds to the "degenerate" case of (MEP) where an unconstrained (w.r.t. $\overline{E}$) maximum is obtained for $\overline{H}$.

Finally, from Theorem 2.5, we obtained the theorem announced in the beginning of this section:

**Theorem 2.6.** *The following assertions are equivalent:*

(i) *(MEP)*
(ii) *(PLE) or (PME).*

**Remark.** The use of the method of the multiplicators of Lagrange to solve the constraint extremal problems such as (MEP), (PLE) or (PME) is well known (see [1]) but yields only a necessary condition; only for free extrema one can prove necessary and sufficient conditions. As said above, Rapoport [15] obtains the function (15) for $\rho > 0$, using (MEP) but calls it (PLE). He then continues by assuming the value:

$$E_r = E . \ln(r) \tag{21}$$

for each $r = 1, \ldots, m$, with $E > 0$ a certain constant. This then leads to the function (cf. (15))

$$p_r = c.\rho^{-E\ln(r)} = cr^{-E.\ln(\rho)}$$

$$p_r = \frac{c}{r^\beta} \tag{22}$$

a power law, where $\beta > 0$ in case $\rho > 1$, hence the decreasing law of Zipf if $r$ denotes the rank. Based on our results we see this only follows from (PLE). (PME) implies an increasing sequence $p_r$ and (MEP) allows both increasing and decreasing sequences. In the same sense, the papers [17,4,13], which use (MEP) (but with continuous variables), do not provide a complete explanation of Zipf's law as a decreasing function. If we interpret $p_r$ as $p_n$ where $p_n$ is the fraction of sources (e.g. journals or authors) with $n$ articles, then (22) is known as the law of Lotka. So also this law follows from (PLE) in its decreasing version (which is the only acceptable one). For more on the law of Lotka and Zipf we refer to [3] or the recent [2].

Assuming (21) is very natural. Indeed, as explained in [15] and in terms of texts consisting of words with $i$ letters ($i = 1, 2, 3, \ldots$) we have that the cost (effort) of using a word with $i$ 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. In fact, this argument is the same as the one leading to the definition of entropy and bits where $\log_2(r)$ is the number of bits needed to the binary coding of $r$ symbols; for $N$-ary coding of $r$ symbols we need $\log_N(r)$ "$N$-ary bits" (e.g. $N = 10$: decimals).

In [11], other functions for $E_r$ than the one in (21) are used, leading to other functional relations for $p_r, r = 1, \ldots, m$.

## Acknowledgement

The first-named author is grateful to Prof. Dr. R. Rousseau for stimulating discussions on the topics of this paper.

## References

[1] T.M. Apostol, Mathematical Analysis, A Modern Approach to Advanced Calculus, Addison-Wesley, Reading, MA, USA, 1957.
[2] L. Egghe, Power Laws in the Information Production Process: Lotkaian Informetrics, Elsevier, Oxford, UK, 2005.
[3] L. Egghe, R. Rousseau, Introduction to Informetrics, Quantitative Methods in Library, Documentation and Information Science, Elsevier, Amsterdam, The Netherlands, 1990.
[4] S.D. Haitun, Stationary scientometric distributions. Part 3. The role of the Zipf distribution, Scientometrics 4 (3) (1982) 181–194.
[5] D.A. Huffman, A method for the construction of minimum redundancy codes, Proceedings of the IRE 40 (1952) 1098–1101.
[6] E.T. Jaynes, Information theory and statistical mechanics, Physical Review (2) 106 (1957) 620–630.
[7] D.S. Jones, Elementary Information Theory, Clarendon Press, Oxford, UK, 1979.
[8] J.N. Kapur, Application of generalised maximum-entropy principle to population dynamics, innovation diffusion models and chemical reactions, Journal of Mathematical and Physical Sciences 26 (2) (1992) 183–211.
[9] J.N. Kapur, H.K. Kesavan, Entropy Optimization Principles with Applications, Academic Press, Boston, MA, USA, 1992.
[10] T. Lafouge, Mathématiques du Document et de l'Information, Bibliométrie distributionnelle, Habilitation à diriger des Recherches, Université de Lyon 3, 1998.
[11] T. Lafouge, C. Michel, Links between information construction and information gain. Entropy and bibliometrics distributions, Journal of Information Science 27 (1) (2001) 39–49.
[12] M. Mansuripur, Introduction to Information Theory, Prentice-Hall, Englewood Cliffs, NJ, USA, 1987.
[13] G. Nicolis, C. Nicolis, J.S. Nicolis, Chaotic dynamics, Markov partitions, and Zipf's law, Journal of Statistical Physics 54 (3–4) (1989) 915–924.
[14] M. Petruszewycz, L'Histoire de la loi d'Estoup-Zipf: documents, Mathématiques des Sciences Humaines 11 (44) (1973) 41–56.
[15] A. Rapoport, Zipf's law re-visited, in: H. Guiter, M.V. Arapov (Eds.), Studies on Zipf's law, in: Quantitative Linguistics, vol. 16, Studien-Verlag Dr. N. Brockmeyer, Bochum, Germany, 1982, pp. 1–28.
[16] C.E. Shannon, W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, Urbana, IL, USA, 1975.
[17] A.I. Yablonsky, On fundamental regularities of the distribution of scientific productivity, Scientometrics 2 (1) (1980) 3–34.
[18] G.K. Zipf, Human Behavior and the Principle of Least Effort, Addison-Wesley, Cambridge, MA, USA, 1949. Reprinted: Hafner, NY, USA, 1965.