The dependence of the height of a Lorenz curve of a Zipf function on the size of the system

Peer-reviewed author version

EGGHE, Leo (2006) The dependence of the height of a Lorenz curve of a Zipf function on the size of the system. In: MATHEMATICAL AND COMPUTER MODELLING, 43(7-8). p. 870-879.

DOI: 10.1016/j.mcm.2005.09.033
Handle: http://hdl.handle.net/1942/749

The dependence of the height of a Lorenz curve of a Zipf function on the size of the system

by

L. Egghe

Universiteit Hasselt (UHasselt), Agoralaan, B-3590 Diepenbeek, Belgium[1]
and
Universiteit Antwerpen (UA), Campus Drie Eiken, Universiteitsplein 1, B-2610 Wilrijk, Belgium

leo.egghe@uhasselt.be

ABSTRACT

The Lorenz curve of a Zipf function describes, graphically, the relation between the fraction of the items and the fraction of the sources producing these items. Hence it generalizes the so-called 80/20-rule to general fractions.

In this paper we examine the relation of such Lorenz curves with the size of the system (expressed by the total number of sources). We prove that the height of such a Lorenz curve is an increasing function of the total number of sources.

---

[1] Permanent address

Key words and phrases: Lorenz curve, Zipf, system size

In other words, we show that the share of items in function of the corresponding share of sources increases with increasing size of the system. This conclusion is opposite (but not in contradiction) to a conclusion of Aksnes (studied in an earlier paper of Egghe) but where "share of sources" is replaced by "number of sources".

I. Introduction

Two-dimensional informetrics (and sociometrics, econometrics,…) deal with the relation between items and sources (producing these items). Examples abound: journals have articles, authors write papers, papers are cited, cities have inhabitants, workers have salaries,… . In all these examples it is clear that the source-item relationship is very skew in the sense that few sources produce many items and many sources produce few items.

Basic to the measurement of this skewness are the underlying rank- and size-frequency functions. The rank-frequency function, denoted by g acts on elements $r \in \{1,2,...,T\}$ (T = total number of sources), where $g(r)$ denotes the number of items in (or produced by) the source on rank r (sources are ranked in decreasing order of number of items in these sources). The size-frequency function, denoted by f acts on elements $j \in \{1,2,...,\rho_m\}$ (ρ_m = the number of items in the source on rank 1, i.e. the maximal number of items in a source), where $f(j)$ denotes the number of sources with j items.

In Lotkaian informetrics (see e.g. [1]) the function f is a power law; i.e. the law of Lotka, [2]

$$f(j) = \frac{C}{j^{\alpha}} \qquad (1)$$

$C, \alpha > 0$. This already expresses how skew (unequal, concentrated) such systems are: e.g. the number of sources (authors in the historical formulation of Lotka's law in [2]) with 3 items is only (take e.g. $\alpha = 2$) $\frac{1}{9}$th of the number of sources with 1 item.

Also for the rank-frequency function g we can take a power law of the form

$$g(r) = \frac{E}{r^{\beta}} \tag{2}$$

$E, \beta > 0$. This form is called the law of Zipf and originates from linguistics. But, as contrasted with Lotka, Zipf did not invent his law but morely "promoted" it, e.g. in [3]; the law (2) itself was already appearing in [4] and (implicitely) even in [5]. That is why the law of Zipf sometimes is called the law of Estoup-Zipf. For more on this we refer to [6] or to [7].

In the next section we will show that the laws of Lotka and Zipf are equivalent, if we take continuous variables j and r (more details are given in the next section – these results and proofs are not new but are given for the sake of completeness).

Functions (1) and (2) are the basis for the description of inequality in these systems. Inequality can be described in several ways. One, rather simple, method is by looking at the most productive source and then checking which fraction of all the items it accounts for. One can also consider the 2,3,4,… most productive sources and check the same property. In this way different fields can be compared: if, say, the most productive source accounts for 5% of all items in one field and for 10% of all items in another field we can say (only based on the first source) that field 2 is skewer than field 1. This approach was followed in [8] where one found (experimentally) that, the smaller the field, the higher the share of the items that come from few (a fixed number: in [8] one took 1 and 5) highly productive sources. This property was (partially) explained in [9].

A far more popular way of expressing inequality in source productivity is by expressing the so-called 80/20-rule or, more generally, by describing the Lorenz curve of the system. In this way one does not consider a fixed number of highly productive sources, that produce a certain share of all the items, but one considers a certain (small) share (fraction) of highly productive sources that produce a certain share of all the items. The 80/20-rule then states (historically) that only 20% of the most productive sources account for 80% of all the items. Of course this is only an expression and the number corresponding with 20 can be different from 80: this is determined by the frequency functions f and g. Also, we are not only interested in a share of 20% of the sources but – in principle – in any share, say 100y%, $0 < y < 1$.

This is accomplished by constructing the Lorenz curve of the system, which we repeat now for the sake of completeness (for more information, see [10], [11] and references therein). The fraction of most productive sources is expressed by looking at $\frac{r}{T}$, for $r = 1, 2, ..., T$ and the fraction of the items produced by these sources is expressed by calculating

$$\frac{\sum_{s=1}^{r} g(s)}{A} \tag{3}$$

where

$$A = \sum_{s=1}^{T} g(s) \tag{4}$$

denotes the total number of items (in all sources $r = 1, 2, ..., T$). One then connects the points $(0,0)$ and $(y, L(g)(y))$ where $y = \frac{r}{T}$ and $L(g)(y)$ is given by (3) for $r = yT$. For every given fraction y of most productive sources we then know the corresponding fraction $L(g)(y)$ of items: this functional relationship is called the Lorenz curve (of g).

In the next section we will repeat results in [1] and [12] which describe the construction of Lorenz curves for continuous functions and we will give explicite formulae for $L(g)$ where g is a continuous Zipf function.

With these preparatory results we then treat, in Section III, the following problem which is the analogue of the problem treated in [9] but there for the Aksnes-Sivertsen type of formulation.

Problem:

Establish the relation between the Lorenz curve $L(g)$ and the size of the field as e.g. expressed by the total number of sources T.

If the Aksnes-Sivertsen conjecture would be true for Lorenz curves this would mean that, the smaller the field, the higher the share of the items that come from a (small) share (instead of number in the Aksnes-Sivertsen conjecture) of highly productive sources. Formulated more easily and mathematically more correctly this would mean that $L(g)$ increases with decreasing T. Surprisingly, however, in Section III, we will show that the opposite is a valid theorem ! We will show that $L(g)$ increases with increasing T, that is, if g is a Zipfian function.

II. Results from continuous informetrics and continuous concentration theory

II.1 The laws of Lotka and Zipf (see [1])

Continuous Lotkaian informetrics deals with the Lotka function (size-frequency function)

$$f(j) = \frac{C}{j^{\alpha}} \tag{5}$$

where C, $\alpha > 0$, $j \in [1, \rho_m]$. Now $f(j)$ denotes the density of sources in the item density j. The general relation between the size-frequency function f and rank-frequency function g is (using g^{-1}, the inverse function of g)

$$r = g^{-1}(j) = \int_j^{\rho_m} f(j')dj' \tag{6}$$

Formula (6) can be taken as the definition of g but it is clear that (6) expresses g as a rank-frequency function: if r and j match then both sides of (6) yield the number of sources with item-density larger than or equal to j. We can prove that function (5) is equivalent with the following rank-frequency function g:

$$g(r) = \frac{E}{(1+r)^{\beta}} \tag{7}$$

with $r \in [0,T]$ with certain relations between the parameters C, α, E and β. This is given in the next theorem.

Theorem II.1 (see [1]) :

The following assertions are equivalent for $\alpha > 0$, $\alpha \neq 1$

(i)

$$f(j) = \frac{C}{j^{\alpha}} \tag{8}$$

$j \in [1, \rho_m]$, $C > 0$, i.e. Lotka's law with exponent α and where

$$\rho_m = \left(\frac{C}{\alpha - 1}\right)^{\frac{1}{\alpha - 1}} \tag{9}$$

(ii)

$$g(r) = \frac{E}{(1+r)^{\beta}} \tag{10}$$

$r \in [0,T]$, $E = \rho_m$, i.e. the law of Zipf with exponent β, where

$$\beta = \frac{1}{\alpha - 1} \tag{11}$$

Proof :

$(i) \Rightarrow (ii)$

The basic formula (6) yields

$$r(j) = g^{-1}(j) = \int_j^{\rho_m} \frac{C}{j'^{\alpha}} dj'$$

$$= \frac{C}{1-\alpha}\left(\rho_m^{1-\alpha} - j^{1-\alpha}\right)$$

hence, since $j = g(r)$, $r \in [0,T]$

$$g(r) = \frac{\rho_m}{(1+r)^{\frac{1}{\alpha-1}}}$$

using (9). Hence we have proved (10), (11).

$(ii) \Rightarrow (i)$

The basic relation (6) yields

$$f(j) = -\frac{1}{g'\left(g^{-1}(j)\right)} \tag{12}$$

for $j \in [1, \rho_m]$. But, using (10) we have

$$g'(r) = \frac{-\beta E}{(1+r)^{\beta+1}} \tag{13}$$

and

$$j = \frac{E}{\left(1+g^{-1}(j)\right)^{\beta}} \tag{14}$$

using that $j = g(r)$ and hence also that $r = g^{-1}(j)$ (see also (6)). So (14) gives

$$g^{-1}(j) = \left(\frac{E}{j}\right)^{\frac{1}{\beta}} - 1 \tag{15}$$

Putting (13) and (15) in (12) yields

$$f(j) = \frac{E^{\frac{1}{\beta}}}{\beta} \frac{1}{j^{1+\frac{1}{\beta}}}$$

hence (8) and where (9) follows from (11) and the fact that $E = \rho_m$. W

The above theorem shows that the law of Zipf is equivalent with the law of Lotka in which (9) is valid. Hence the law of Zipf is fully covered in Lotkaian informetrics (what we assume in this paper). We will now calculate the Lorenz curve for the Zipf function.

II.2 The Lorenz curve of the Zipf function

The results of this section can be found in [1] and [12]. The continuous analogue of the discrete Lorenz curve described in Section I is the curve $L(g)$ given by the set of points

$$\left(\frac{r}{T}, \frac{\int_0^r g(r')dr'}{\int_0^T g(r')dr'}\right) \tag{16}$$

In other words, putting $y = \frac{r}{T} \in [0,1]$, the Lorenz curve $L(g)$ of g is the function

$$L(g)(y) = \frac{\int_0^{yT} g(r')dr'}{\int_0^T g(r')dr'} \tag{17}$$

The rationale for this extension is clear: $\frac{r}{T}$ obviously measures the fraction of sources (up to r) and

$$\frac{\int_0^r g(r')dr'}{\int_0^T g(r')dr'}$$

measures the cumulative fraction of items in these sources.

For the Zipf function (10) we easily find for $\alpha > 0$, $\alpha \neq 1$, $\alpha \neq 2$, using (17):

$$L(g)(y) = \frac{(yT+1)^{1-\beta} - 1}{(T+1)^{1-\beta} - 1} \tag{18}$$

Note that, in terms of Lotka's α we have, by (11), for $\alpha > 0$, $\alpha \neq 1$, $\alpha \neq 2$ that

$$1 - \beta = 1 - \frac{1}{\alpha - 1} = \frac{\alpha - 2}{\alpha - 1} \tag{19}$$

For $\alpha = 2$ we have, by (11), that $\beta = 1$. We now have, from (17) that

$$L(g)(y) = \frac{\ln(yT+1)}{\ln(T+1)} \tag{20}$$

For the sake of simplicity we do not deal with the case $\alpha = 1$. This case can be recovered from [1] (see also [13]) and similar calculations as performed above. We henceforth suppose $\alpha > 1$ which is almost always encountered is practise.

We can now prove the dependence of the height of $L(g)$ on T, a variant of the Aksnes-Sivertsen conjecture.

III. Dependence of the height of L(g) on T

We have the following result.

Theorem III.1 :

For fixed $\alpha > 1$ we have that $L(g)(y)$ is a strictly increasing function of T.

Proof :

(1) Let $\alpha > 1$ and $\alpha \neq 2$. Then

$$L(g)(y) = \frac{(yT+1)^{\frac{\alpha-2}{\alpha-1}} - 1}{(T+1)^{\frac{\alpha-2}{\alpha-1}} - 1} \tag{21}$$

for all $y \in [0,1]$. We have that

$$\frac{dL(g)(y)}{dT} = \frac{(*)}{\left[(T+1)^{\frac{\alpha-2}{\alpha-1}} - 1\right]^2} \tag{22}$$

where

$$(*) = \left[(T+1)^{\frac{\alpha-2}{\alpha-1}} - 1\right]\frac{\alpha-2}{\alpha-1}(yT+1)^{\frac{\alpha-2}{\alpha-1}-1} y - \left[(yT+1)^{\frac{\alpha-2}{\alpha-1}} - 1\right]\frac{\alpha-2}{\alpha-1}(T+1)^{\frac{\alpha-2}{\alpha-1}-1} \tag{23}$$

So in order to prove that $L(g)(y)$ strictly increases in T it is sufficient to prove that $(*) > 0$.

(1.1) Suppose first that

$$\frac{\alpha-2}{\alpha-1} > 0 \tag{24}$$

(i.e. $\alpha > 2$). By $(*)$ it suffices to prove that

$$y\left[(T+1)^{\frac{\alpha-2}{\alpha-1}} - 1\right](T+1)^{\frac{1}{\alpha-1}} > \left[(yT+1)^{\frac{\alpha-2}{\alpha-1}} - 1\right](yT+1)^{\frac{1}{\alpha-1}} \tag{25}$$

This will be proved (as is clear from Fig. 1) if we can show that the function (on $x \geq 1$)

$$\varphi(x) = \left(x^{\frac{\alpha-2}{\alpha-1}} - 1\right)x^{\frac{1}{\alpha-1}} \tag{26}$$

is strictly increasing and convex such that $\varphi(1) = 0$.

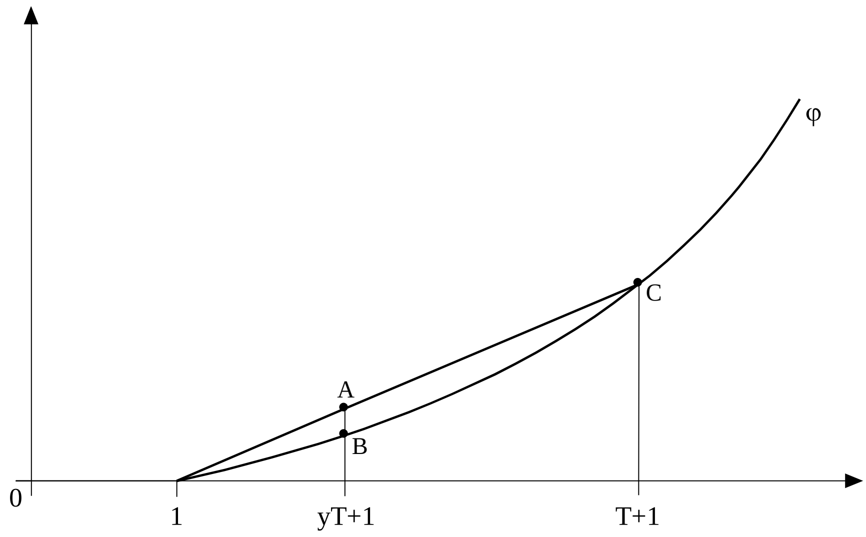

Fig. 1

Indeed, Fig. 1 makes clear that ordinate $(C) = \varphi(T+1)$, hence, by the theorem of Thales, ordinate $(A) = y\varphi(T+1)$ which is larger than ordinate (B) $= \varphi(yT+1)$, proving (25). Now $\varphi(1) = 0$ is clear. Further

$$\varphi'(x) = \frac{\alpha - 2}{\alpha - 1} + \left(1 - x^{-\frac{\alpha-2}{\alpha-1}}\right)\frac{1}{\alpha - 1} > 0$$

by (24) and since $x > 1$. Also

$$\varphi''(x) = \frac{\alpha - 2}{(\alpha - 1)^2}\, x^{\frac{-2\alpha+3}{\alpha-1}} > 0$$

Hence φ is strictly increasing and convex on $[1, +\infty[$ which shows (25).

(1.2) Suppose now that

$$\frac{\alpha - 2}{\alpha - 1} < 0 \qquad (27)$$

(hence $1 < \alpha < 2$). Hence now $(*) > 0$ if

$$y\left[(T+1)^{\frac{\alpha-2}{\alpha-1}} - 1\right](T+1)^{\frac{1}{\alpha-1}} < \left[(yT+1)^{\frac{\alpha-2}{\alpha-1}} - 1\right](yT+1)^{\frac{1}{\alpha-1}} \qquad (28)$$

This will be proved (as is clear from Fig.2) if we can show that the function (26) (but now with (27) valid) is strictly decreasing and concave such that $\varphi(1) = 0$.

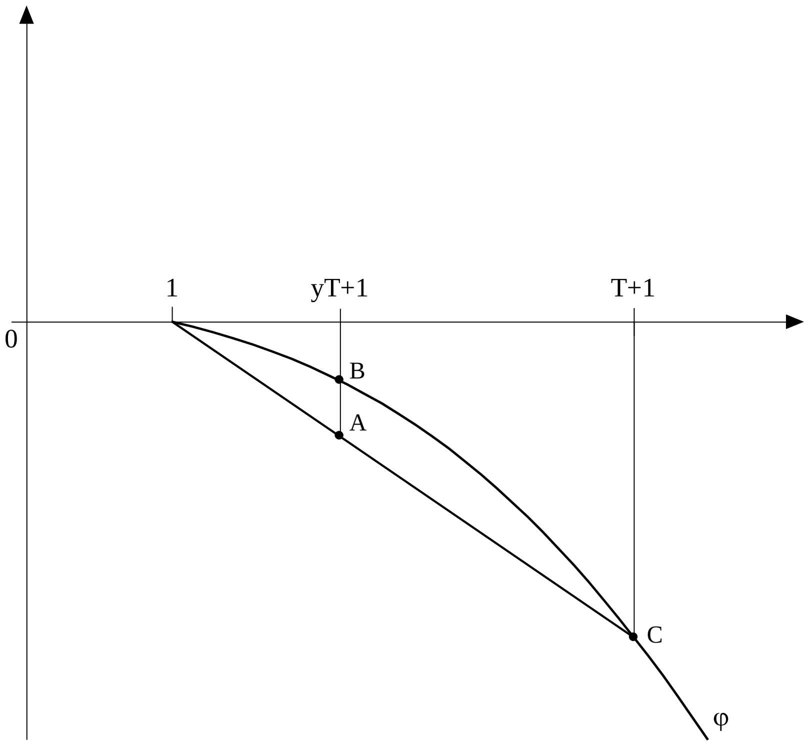

Fig.2

Indeed, Fig.2 makes it clear that ordinate $(C) = \varphi(T+1)$, hence by the theorem of Thales, ordinate $(A) = y\varphi(T+1)$ which is smaller than ordinate $(B) = \varphi(yT+1)$, proving (28). Again $\varphi(1) = 0$ is clear. Further

$$\varphi'(x) = \frac{\alpha-2}{\alpha-1} + \left(1 - x^{-\frac{\alpha-2}{\alpha-1}}\right)\frac{1}{\alpha-1} < 0$$

since we have (27), $\alpha > 1$ and $x > 1$. Finally

$$\varphi''(x) = \frac{\alpha-2}{(\alpha-1)^2} x^{\frac{-2\alpha+3}{\alpha-1}} < 0$$

since, by (27), $1 < \alpha < 2$. Hence φ is strictly decreasing and concave on $[1,+\infty[$ which shows (28).

Finally we treat the case $\alpha = 2$.

(2) Let $\alpha = 2$. Now we use (20) and hence we have that

$$\frac{dL(g)(y)}{dT} = \frac{(**)}{\ln^2(T+1)}$$

with

$$(**) = \frac{y}{yT+1}\ln(T+1) - \frac{1}{T+1}\ln(yT+1).$$

So, in order to prove that $L(g)(y)$ strictly increases in T it is sufficient to show that

$$(**) = y(T+1)\ln(T+1) > (yT+1)\ln(yT+1) \tag{29}$$

Again as in part (1.1) it suffices to prove that

$$\psi(x) = x\ln x$$

is strictly increasing, convex on $[1,+\infty[$ such that $\varphi(1) = 0$. That $\varphi(1) = 0$ is clear. Further

$$\psi'(x) = \ln x + 1 > 0$$

$$\psi''(x) = \frac{1}{x} > 0$$

on $[1,+\infty[$ which shows that we are in a situation as in Fig.1, which now proves (29).

W

Corollary III.2:

Let us have two Lotkaian sytems with the same Lotka exponent $\alpha > 1$. Let the first system have T_1 sources and the second one T_2 sources such that $T_1 < T_2$. Then, denoting by $L(g_i)(y)$ $(y \in [0,1])$ the Lorenz curves of system $i (i = 1,2)$ we have that

$$L(g_1)(y) < L(g_2)(y) \tag{30}$$

for every $y \in]0,1[$.

In words: “The smaller the field, the smaller the share of the items that come from an equal share of the sources”. Looking only at a small share of highly productive sources this yields: “The smaller the field, the smaller the share of the items that come from a small equal share of highly productive sources”.

This result is opposite to the conjecture of [8] but with “share of sources” replaced by “number of sources” as follows (see [8] (there formulated in terms of papers and citations) and [9]): “The smaller the field, the higher the share of the items that come from a small equal number of highly productive sources”. This conjecture was studied in different versions in [9] where the above Aksnes and Sivertsen conjecture was proved in the following version (compare with Corollary III.2).

Theorem III.3 ([9]):

Let us have two Lotkaian systems with the same exponent α. Let the first system have T_1 sources and the second one T_2 sources such that $T_1 < T_2$. Then, for all $r \in]0,T_1]$

$$\varphi_1(r) = \frac{\int_0^r g_1(r')dr'}{\int_0^{T_1} g_1(r')dr'} > \varphi_2(r) = \frac{\int_0^r g_2(r')dr'}{\int_0^{T_2} g_2(r')dr'} \tag{31}$$

where g_i denote the rank-frequency functions of system i $(i = 1,2)$.

This can be compared with formula (17): in $L(g_i)(y)$ $(i = 1,2)$ we use an identical fraction yT_1, respectively yT_2 in both systems while in (31) we use an identical number r of sources in both systems. So these are different things and hence, of course, Theorem III.3 is not in contradiction with Corollary III.2, but we can at least say that these "opposite" results are surprising.

An intuitive "understanding" of the difference of both results can be given as follows. When $T_1 < T_2$ then, for every $y \in]0,1[$, $yT_1 < yT_2$ and hence, in (30) we have (31) but with r in φ_1 replaced by yT_1 and with r in φ_2 replaced by yT_2. So the left hand side of (31) is reduced w.r.t. the right hand side of (31) but this does not prove that the inequality is reversed! This is shown in Corollary III.2.

A generalization of Theorem III.3 (also proved in [9]) consist of not taking the same Lotka exponent α in both systems but to only suppose that $\alpha_1 \leq \alpha_2$ where α_i is the Lotka exponent of system i $(i = 1,2)$. A similar (but different) extension of Corollary III.2 is as follows.

<u>Corollary III.3</u>:

Let us have two Lotkaian systems i $(i = 1,2)$ with number of sources T_i and Lotkaian exponent α_i such that $T_1 < T_2$ and such that $\alpha_1 > \alpha_2$. Then, denoting by $L(g_i)(y)$ $(y \in [0,1])$ the Lorenz curves of system i we have that

$$L(g_1)(y) < L(g_2)(y) \tag{32}$$

for all $y \in]0,1[$.

<u>Proof</u>:

In [1] and [12] it is proved that $L(g)$ (as in (17)) (when the number of sources T is kept fixed) is a decreasing function of α, the Lotka exponent in case we have a Lotkaian system (see e.g. Corollary IV.3.2.1.5 in [1] or see [12]). Hence and also using Corollary III.2 above, we have, for every $y \in]0,1[$:

$$L(g_1)(y) < L(g_1, T_2)(y) < L(g_2)(y)$$

where $L(g_1, T_2)$ means the Lorenz curve of the first system but with $T_2 > T_1$ sources. The first inequality hence follows from Corollary III.2 while the second one follows from the indicated decreasing dependence of $L(g)$ on α (keeping T, here T_2 fixed). $\square$

<u>Note</u> :
Different Lotkaian systems but with the same Lotka exponent α exist : just take different constants C in (8), changing also the constants in (10), so both functions f_i (size-frequency) and g_i (rank-frequency) $(i = 1, 2)$ are different which makes it also possible to have a different number of sources (and items).

Let us give some examples that illustrate the main result, Corollary III.2.

<u>Examples III.4</u>

1. Let us take $\alpha = 3$, $T_1 = 1{,}000 < T_2 = 10{,}000$: it follows from [1], Chapter II that these values can be taken (even with $\rho_m = \infty$): the constant C can be choosen such that these values occur. We have, by (18) and (19)

$$L(g_1)(y) = \frac{\sqrt{1{,}000y + 1} - 1}{\sqrt{1{,}001} - 1}$$

$$L(g_2)(y) = \frac{\sqrt{10{,}000y + 1} - 1}{\sqrt{10{,}001} - 1}$$

Let us take $y = 0.1$ (but any value $y \in \left]0, 1\right[$ can be taken) and verify that

$$L(g_1)(0.1) = 0.2954 < L(g_2)(0.1) = 0.3095$$

2. Let us take $\alpha = 1.5$. Again it follows from [1], Chapter II that $T_1 = 1{,}000 < T_2 = 10{,}000$ can be taken (but $\rho_{m,1}, \rho_{m,2} < \infty$ but that has no importance for $L(g_i)$). Now, by (18) and (19) we have

$$L(g_1)(y) = \frac{\dfrac{1}{1{,}000y + 1} - 1}{\dfrac{1}{1{,}001} - 1}$$

$$L(g_2)(y) = \frac{\dfrac{1}{10{,}000y + 1} - 1}{\dfrac{1}{10{,}001} - 1}$$

For (e.g.) $y = 0.01$ we now have

$$L(g_1)(0.01) = 0.9100 < L(g_2)(0.01) = 0.9902$$

3. Let us take $\alpha = 2$, $T_1 = 1{,}000$, $T_2 = 10{,}000$. By (20) we now have

$$L(g_1)(y) = \frac{\ln(1{,}000y + 1)}{\ln(1{,}001)}$$

$$L(g_2)(y) = \frac{\ln(10{,}000y + 1)}{\ln(10{,}001)}$$

For $y = 0.1$ we have

$$L(g_1)(0.1) = 0.6680 < L(g_2)(0.1) = 0.7501$$

If we take $\lim_{T \to \infty} L(g)(y)$ in (18) and (20) we know, from Theorem III.1 that we find the highest possible value of $L(g)(y)$ (given α fixed). It is easy to see that

$$\lim_{T \to \infty} L(g)(y) = 1 \tag{33}$$

for all $\alpha \geq 2$ and

$$\lim_{T \to \infty} L(g)(y) = y^{\frac{\alpha - 2}{\alpha - 1}} = y^{1-\beta} \tag{34}$$

if $1 < \alpha < 2$ but this result was already known and appears in [1] and [12].

Of course, if the conditions of the most general result – Corollary III.3 – are not satisfied, we can produce counterexamples to inequality (32).

<u>Counterexample III.5</u> :

Let $\alpha_1 = 3 < \alpha_2 = 4$, $T_1 = 1{,}000 < T_2 = 10{,}000$. We have now, by (18), (19):

$$L(g_1)(y) = \frac{\sqrt{1{,}000y + 1} - 1}{\sqrt{1{,}001} - 1}$$

$$L(g_2)(y) = \frac{(10{,}000y + 1)^{\frac{2}{3}} - 1}{(10{,}001)^{\frac{2}{3}} - 1}$$

For $y = 0.1$ this gives

$$L(g_1)(0.1) = 0.2954 > L(g_2)(0.1) = 0.2139 .$$

IV. Conclusion

The main result of this paper is that, in Lotkaian informetrics, “the smaller the field, the smaller the share of the items that come from an equal share of the sources”.

In [9] we proved the opposite result when “share of the sources” is replaced by “number of sources”.

The difference between the two results lies in the fact that, if $r \in [0, T]$ is kept fixed (the “number of sources”) then $y = \frac{r}{T}$ (the “share of sources”) increases with decreasing total number of sources T; hence these sources account for a higher fraction of items.

Of course both results are exact and hence do not contradict each other.

The paper shows that, if the conditions of the proved theorems are not met, counterexamples to these properties can be given.

References

[1] L. Egghe. Power Laws in the Information Production Process: Lotkaian Informetrics. Elsevier, Oxford, 2005.

[2] A.J. Lotka. The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences 16(12), 317-323, 1926.

[3] G.K. Zipf. Human Behavior and the Principle of least Effort. Addison-Wesley, Cambridge, USA, 1949. Reprinted: Hafner, New York, USA, 1965.

[4] E.U. Condon. Statistics of vocabulary. Science 67(1733), 300, 1928.

[5] J.B. Estoup. Gammes Sténographiques. 4th edition. Institut Sténographique, Paris, 1916.

[6] R. Rousseau. George Kingsley Zipf: life, ideas, his law and informetrics. Glottometrics 3, 11-18, 2002.

[7] C.S. Wilson. Informetrics. Annual Review of Information Science and Technology (ARIST) 34 (M.E. Williams, ed.) 107-247, 1999.

[8] D.W. Aksnes and G. Sivertsen. The effect of highly cited papers on national citation indicators. Scientometrics 59(2), 213-224, 2004.

[9] L. Egghe. The share of items of highly productive sources in function of the size of the system. Scientometrics 65(3), 265-281, 2005.

[10] L. Egghe and R. Rousseau. Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science. Elsevier, Amsterdam, 1990.

[11] L. Egghe and R. Rousseau. Elements of concentration theory. In: Informetrics 89/90. Proceedings of the second international Conference on Bibliometrics, Scientometrics and Informetrics, London (Canada), (L. Egghe and R. Rousseau, eds.). Elsevier, Amsterdam, 97-137, 1990.

[12] L. Egghe. Zipfian and Lotkaian continuous concentration theory. Journal of the American Society for Information Science and Technology 56(9), 935-945, 2005.

[13] L. Egghe. The source-item coverage of the Lotka function. Scientometrics, 61(1), 103-115, 2004.