POSITIVE REINFORCEMENT AND 3-DIMENSIONAL INFORMETRICS

by

Leo Egghe, Limburgs Universitair Centrum (LUC), Universitaire Campus, B-3590 Diepenbeek, Belgium[1]
and Universiteit Antwerpen (UIA), Universiteitsplein 1, B-2610 Wilrijk, Belgium
leo.egghe@luc.ac.be

ABSTRACT

We show that the composition of two information production processes (IPPs), where the items of the first IPP are the sources of the second, and where the ranks of the sources in the first IPP agree with the ranks of the sources in the second IPP, yields an IPP which is positively reinforced with respect to the first IPP. This means that the rank-frequency distribution of the composition is the composition of the rank-frequency distribution of the first IPP and an increasing function φ, which is explicitly calculable from the two IPPs' distributions.

From the rank-frequency distribution of the composition, we derive its size-frequency distribution in terms of the size-frequency distribution of the first IPP and of the function φ.

The paper also relates the concentration of the reinforced IPP to that of the original one.

This theory solves part of the problem of the determination of a third IPP from two given ones (so-called three-dimensional informetrics). In this paper we solved the "linear" case, i.e. where the third IPP is the composition of the other two IPPs.

I. Introduction

One-dimensional informetrics is that part of informetrics where one studies one object, one variable, ... as shown in the following examples:

(i) The number of books in a library
(ii) The number of circulations in a library
(iii) The total number of researchers in a field, e.g. mathematics, at a certain moment
(iv) The total number of publications in a field, e.g. mathematics, say in a year
(v) The total number of citations in a field, e.g. mathematics, say in a year.

Such studies can be interesting, certainly if they are made in connection with evolution in time.
However, in order to really understand information production, one needs to study two-dimensional informetrics, where one considers two objects, one called the sources and the other one called the items and where it is understood that the items are produced by the sources. Of course, here, one does

[1] Permanent address.

Keywords and phrases: positive reinforcement, linear 3-dimensional informetrics, rank-frequency distribution, size-frequency distribution, composition.

Acknowledgement. The author is grateful to R. Rousseau for interesting discussions on the topic of this paper and for drawing the author's attention Fellman's paper (1976).

not only perform two times a one-dimensional informetrics study (one for the sources and one for the items) but one studies the "higher" (more interesting) problem of the link between sources and items, i.e. which sources produce which items. One could say that such a study implies the two one-dimensional informetrics studies but not vice-versa. Two-dimensional informetrics is the one in which one has the celebrated (historic) informetric laws, describing this source-item relationship, such as the laws of Lotka, Zipf, Mandelbrot, Bradford, Leimkuhler and so on. It is also here that the notion of information production process (IPP) is defined as a general bibliography: an IPP is a triplet (S,I,F), where S is the set of sources, I is the set of items and $F : S \rightarrow I$ is a function describing which sources contain or produce which items. Two-dimensional informetrics is the most famous part of informetrics and it is even so that - in most studies - one does not even mention the adjective "two-dimensional". Examples abound (see e.g. Egghe and Rousseau (1990a)):

(i) Books (as sources) in a library and the circulations (as items) they generate (cf. examples (i) and (ii) in one-dimensional informetrics)
(ii) Researchers (as sources) in a field (e.g. mathematics) and their publications (as items) (cf. examples (iii) and (iv) in one-dimensional informetrics)
(iii) Publications (as sources) in a field (e.g. mathematics) and the citations (as items) they receive (or give) (cf. examples (iv) and (v) in one-dimensional informetrics).

We stress again that the above examples study more than the two one-dimensional examples that we indicated namely also, more importantly, the relationship between sources and items, e.g. the number of sources with 1,2,3,... items (size-frequency study such as Lotka's law) or the number of items in the source on rank r=1,2,3,... (where we rank sources decreasingly with respect to the number of items they have) (rank-frequency study such as Zipf's law or the law of Mandelbrot, Bradford or Leimkuhler). We can nowadays say that the majority of informetrics papers (theoretical as well as practical ones) deal with two-dimensional informetrics as a quick inspection of leading informetrics journals makes clear.

The mathematics of two-dimensional informetrics was developed in Egghe (1985, 1989, 1990) where the size and rank-frequency functions were studied and their interrelations proved. Exact results were obtained but only in the continuous setting, i.e. where the source set S (bijective with {1,2,...,T}) is replaced by the interval [0,T] and where the item set I (bijective with {1,2,...,A}) is replaced by the interval [0,A]. The reason for this is that interrelations, say between Lotka's law and the one of Mandelbrot can only be shown by evaluating integrals which is much easier than evaluating finite sums. Since we need this theory in the elaboration of 3-dimensional informetrics we give a brief overview of this theory in the next section.

So we arrived now at the challenge of studying three-dimensional informetrics. The attentive reader may already have thought of the following example of three-dimensional informetrics (based on the above mentioned cases):

(i) Researchers (as sources) in a field (e.g. mathematics), their publications as items produced by the researchers, but also considered as sources, producing citations (being items in the sources being publications) (cf. examples (iii), (iv) and (v) of one-dimensional informetrics or the examples (ii) and (iii) in two-dimensional informetrics).

Note again that three-dimensional informetrics is much more than three times the one-dimensional or two times the two-dimensional study. Now we face (as in the two-dimensional case) to describe the relations between the two types of sources and the two types of items. The above example is an example of what we could call <u>"linear three-dimensional informetrics"</u> since the items in the first IPP (publications) are the sources in the second IPP and three-dimensional informetrics takes the (symbolic) linear form (cf. Rousseau (1992))

$$\text{researchers} \rightarrow \text{publications} \rightarrow \text{citations}$$

Of the same type is the example

$$\text{journals} \rightarrow \text{articles} \rightarrow \text{citations}$$

meaning that journals publish articles (on a certain topic) and that articles give (or receive) citations.

In Egghe (1990), see also Egghe and Rousseau (1990a), other forms of (nonlinear) three-dimensional informetrics are defined (but no results obtained):

(ii) Journals produce articles and these articles are written by authors. Here we have two source sets (journals and authors) and one item set (articles), symbolically visualized by a triangle:

$$\begin{array}{ccc} \text{journals} & & \text{authors} \\ & \searrow \quad \swarrow & \\ & \text{articles} & \end{array}$$

(iii) Papers have references and receive citations. Here we have one source set (papers) and two item sets (references and citations), symbolically visualized by a triangle of another type

$$\begin{array}{ccc} & \text{papers} & \\ & \swarrow \quad \searrow & \\ \text{references} & & \text{citations} \end{array}$$

In all three examples of three-dimensional informetrics we have two IPPs (in the two-dimensional sense) but each example shows a different relation between these two IPPs.

On three-dimensional informetrics, almost no papers or results exist. There was a first (theoretical) attempt to model general three-dimensional informetrics in Egghe and Rousseau (1996), generalizing work of Qin (1995). Egghe and Rousseau (1996) also give further examples (found in the literature) of case (iii) above (examples of Lafouge (1995), Coleman (1992), Yitzhaki (1995) and Kyvik (1990)), where sources produce two sets of items. In general we can say that the theoretical remarks in Egghe and Rousseau (1996) are not capable of really modelling <u>general</u> three-dimensional informetrics (say cases (ii) and (iii) above).

In Rousseau (1990), Egghe (1990) and Egghe and Rousseau (1990a) the problem to model at least <u>linear</u> three-dimensional informetrics is mentioned. In Rousseau (1990) the author deals with the linear problem

$$\text{journals} \rightarrow \text{articles} \rightarrow \text{software programs}$$

meaning: journals have articles and these articles describe certain software programs. Linear three-dimensional informetrics problems were also (implicitely) mentioned in Rousseau (1992) and Fox (1983). In Fox (1983), one also uses the term "reinforcement" referring to the transitive effect of linear three-dimensional informetrics - here the context of high productivity of authors as a consequence of earlier recognition (heavy citations) is mentioned. Rousseau (1992) mentions positive reinforcement. This can (intuitively - exact definitions follow in the sequel) be described as: the item-producing sources increase their production. Rousseau (1992) shows that, under certain conditions (see also further), a positively reinforced IPP has a higher concentration, expressed by the fact that its Lorenz curve is above the original one (when ordering the production vectors decreasingly). It seems that Fellman (1976) was the first to notice these facts. Using the latter result, we will generalise Rousseau's result in section IV. Besides the example researchers $\rightarrow$publications$\rightarrow$citations, Rousseau (1992) also mentions the example of availability of CDs in certain music categories and its use (loans) in a public library as another example of linear three-dimensional informetrics.

The key result in this paper will be that linear three-dimensional informetrics, in which the ranks of the sources in the first IPP agree with the ranks of the sources in the second IPP, is an example of positive reinforcement of the first IPP. We will present an explicit relation between the reinforcement function φ (to be defined later) and the size- and rank-frequency functions of the two IPPs in the composition.

A consequence of this result is that the positive reinforcement of a Lotkaian IPP with a reinforcement function φ that is a power law, yields a Lotkaian IPP.

As mentioned above the paper closes with an application of Fellman's concentration result (1976) (partially reproved in Rousseau (1992)). We give necessary and sufficient conditions for the positively reinforced IPP to have a lower, equal or higher concentration (i.e. a lower, equal or higher Lorenz curve) than the one of the original IPP, hereby generalizing Rousseau (1992)

II. Aspects of two-dimensional informetrics.

Here we have an IPP consisting of a set of sources S=[0,T] (the continuous extension of the counting of T sources in S, namely {1,2,...,T}) and a set of items I=[0,A] (the continuous extension of {1,2,...,A}). We also have a strictly increasing differentiable function V : [0,T] $\rightarrow$ [0,A] where, for every $r \in [0,T]$, V(r) denotes the cumulative number of items in the sources in the interval [T-r,T]. If sources are arranged in decreasing order of number of items they have, this definition guarantees that the function

$$\rho(r) = V'(r) \tag{1}$$

is not only positive but that it also increases: the function ρ can be interpreted as the density of the items at the source T-r. With an abuse of notation we define, for every $i \in [0,A]$,

$$\rho(i) = V'\left(V^{-1}(i)\right) \tag{2}$$

meaning that we put ρ (i)=ρ (r) iff i=V(r) (the relation $r \leftrightarrow i$ being unique since V is injective). Although it is not necessary, we will assume that ρ (0)=1, the minimal item-density of a source. The value ρ (A)=ρ (T) is the maximal item-density of a source. In discrete terms it is the maximum production of a source.

For this IPP we define the <u>size-frequency function f</u> as a positive function on the interval [1,ρ (A)], where f(j) (for each $j \in [1,\rho(A)]$) is the density function of sources as a function of j, i.e. for every $i \in [0,A]$

$$\int_1^{\rho(i)} f(j)\,dj \tag{3}$$

is the cumulative number of sources with item-density $j \in [1, \rho(i)]$. To define the <u>rank-frequency function g</u> we express that, for every $j \in [1, \rho(A)]$

$$\int_j^{\rho(A)} f(j')\,dj' \tag{4}$$

denotes the cumulative number of sources with item-density j or higher. Hence, by definition

$$r = r(j) = \int_j^{\rho(A)} f(j')\,dj' \tag{5}$$

and finally, g:[0,T] $\rightarrow$ [1, ρ (A)] is defined as the inverse of the function r=r(j). Thus

$$g^{-1}(j) = \int_j^{\rho(A)} f(j')\,dj' \tag{6}$$

defines g^{-1}, hence g. So j=g(r) is the item-density in the source r. From (6) it follows trivially that, for all $j \in [1, \rho(A)]$:

$$f(j) = -\frac{1}{g'\left(g^{-1}(j)\right)}, \tag{7}$$

an identity that we will use frequently in the next section. For more on two-dimensional informetric theory, we refer the reader to Egghe (1989, 1990) and to Egghe and Rousseau (1990a, 2003).

Note that, in Lotkaian informetrics, f corresponds to the law of Lotka and g to the law of Mandelbrot which is (by (6) and (7)) equivalent with Lotka's law (see again Egghe and Rousseau (1990a)).

III. Linear three-dimensional informetrics

In this section we intend to give a mathematical description of linear three-dimensional informetrics (cf. the first example on three-dimensional informetrics, given in the previous section). Here we have a first IPP with source set $S_1=[0,T_1]$ and item-set $I_1=[0,A_1]$ and with f_1 and g_1 as size-frequency, respectively, rank-frequency function as defined in section II. Then we have a second IPP for which $S_2=[0,T_2]=I_1=[0,A_1]$ by definition (hence the items in the first IPP are the sources of the second IPP), defining a new item-set $I_2=[0,A_2]$ for this second IPP. In this second IPP we have a size-frequency function f_2 and a rank-frequency function g_2.

We can now state the following problem :

Problem III.1 : Give a mathematical description of the "composed" IPP with source-set $S=S_1=[0,T_1]$ and item-set $I=I_2=[0,A_2]$: describe its size-frequency function f and its rank-frequency function g. This composed IPP has sources in S_1 which produce items (in the first IPP) in I_1 which is considered as the source-set in the second IPP, which sources produce items in I_2.

We will give a complete solution in the following special (but most important) case.

Restriction III.2 : In this paper we will restrict ourselves to the case that the source-rankings r_2 in the second IPP are the same as the rankings in I_1, induced by the source-rankings r_1 in the first IPP. We will express this restriction in an exact mathematical way but first we give an intuitive interpretation: items in I_1 are ranked according to the ranking we have in S_1 (as always, sources are ranked in decreasing order of production). Considering $I_1=S_2$ as sources in the second IPP, we require that their productivity in the second IPP is such that the source-ranking r_2 in this second IPP is the same as the one we have already in I_1. An exact formulation of this restriction is

$$r_2 = \int_0^{r_1} g_1(r)\,dr \tag{8}$$

This situation will be proved to be a case of positive reinforcement of production of the first IPP. We will first define "positive reinforcement".

Definition III.3 : Let S=[0,T], I=[0,A] be an IPP with rank-frequency function g. Let $I^*=[0,A^*]$. We say that the IPP (S,I^*) is a positive reinforcement of the IPP (S,I) if its rank-frequency function g^* is given by

$$g^* = \varphi \circ g, \tag{9}$$

where φ is a strictly increasing function.

The interpretation of the above definition is clear: the positively reinforced IPP (S,I^*) has the same sources but with an increased productivity, given by (9).

We have the following basic result on linear three-dimensional informetrics. The used elementary calculus can e.g. be found in Apostol (1957).

Theorem III.4 : Let the first IPP have source-set $S_1=[0,T_1]$, item-set $I_1=[0,A_1]$, size-frequency function f_1 and rank-frequency function g_1 and let the second IPP have source-set $S_2=[0,T_2]=I_1=[0,A_1]$, item-set $I_2=[0,A_2]$, size-frequency function f_2 and rank-frequency function g_2. Then the composed IPP (under restriction III.2) has source-set $S=S_1$, item-set $I=I_2$, rank-frequency function

$$g = \varphi \circ g_1 , \tag{10}$$

where

$$\varphi(j_1) = g_2\left(\int_{j_1}^{\rho_1} \frac{-j\,dj}{g_1'\left(g_1^{-1}(j)\right)}\right) \tag{11}$$

$$\varphi(j_1) = g_2\left(\int_{j_1}^{\rho_1} j f_1(j)\,dj\right) \tag{12}$$

is a strictly increasing function. Hence the composed IPP is a positive reinforcement of the first IPP. Its size-frequency function f has the form

$$f(j) = \frac{f_1\left(\varphi^{-1}(j)\right)}{\varphi'\left(\varphi^{-1}(j)\right)} \tag{13}$$

$j \in [1, \varphi(\rho_1)]$, where $\rho_1=\rho(A_1)=\rho(T_1)$, the maximal item-density in the first IPP.

Proof : Since g_2 acts on the ranks r_2 in the second IPP and since the restriction (8) defines the relation between r_1 and r_2, we have that the composed IPP has the following rank-frequency function g:

$$g(r_1) = g_2\left(\int_0^{r_1} g_1(r)\,dr\right) \tag{14}$$

Now we are looking for a function φ such that

$$g(r_1) = (\varphi \circ g_1)(r_1) \tag{15}$$

We will prove its existence via (14), show that φ strictly increases and hence that the composed IPP is a positive reinforcement of the first one.

If such a φ exists then, (14) and (15) combined yield

$$\varphi\left(g_1(r_1)\right) = g_2\left(\int_0^{r_1} g_1(r)\,dr\right).$$

Substituting $r = g_1^{-1}(j)$ gives

$$\int_0^{r_1} g_1(r)\,dr = \int_{\rho_1}^{j_1} \frac{j\,dj}{g_1'\left(g_1^{-1}(j)\right)}$$

since $g_1(0)=\rho_1$ (see (6)) and where $r_1=g_1^{-1}(j_1)$, the existing relation between j_1 and r_1 in the first IPP. Hence

$$\varphi(j_1)=g_2\left(\int_{\rho_1}^{j_1}\frac{j\,dj}{g_1'\left(g_1^{-1}(j)\right)}\right)=g_2\left(\int_{j_1}^{\rho_1} j f_1(j)\,dj\right)$$

using (7), hence proving (11) and (12). That φ strictly increases follows easily from (11):

$$\varphi'(j_1)=g'_2\left(\int_{\rho_1}^{j_1}\frac{j\,dj}{g_1'\left(g_1^{-1}(j)\right)}\right).\frac{j_1}{g_1'\left(g_1^{-1}(j)\right)}$$

since $j_1 \geq 1>0$ and since g_2' and g_1' are negative (by (6)), $\varphi'>0$. Hence linear three-dimensional informetrics is an example of positive reinforcement.

Let us now determine the size-frequency function f of this positively reinforced IPP. Invoke (7):

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

and then (10) gives $r_1 \in [0,T_1]$, $j \in [1, \varphi(\rho_1)]$, $g(r_1)=j$:

$$g'(r_1)=\varphi'\left(g_1(r_1)\right)g_1'(r_1)$$

$$g'\left(g^{-1}(j)\right)=\varphi'\left(g_1\left(g^{-1}(j)\right)\right)g_1'\left(g^{-1}(j)\right)$$

But $g=\varphi \circ g_1$, hence $g^{-1}=g_1^{-1} \circ \varphi^{-1}$ hence

$$g'\left(g^{-1}(j)\right)=\varphi'\left(\varphi^{-1}(j)\right)g_1'\left(g_1^{-1}\left(\varphi^{-1}(j)\right)\right)$$

So, by (7),

$$f(j)=-\frac{1}{\varphi'\left(\varphi^{-1}(j)\right)g_1'\left(g_1^{-1}\left(\varphi^{-1}(j)\right)\right)}=\frac{f_1\left(\varphi^{-1}(j)\right)}{\varphi'\left(\varphi^{-1}(j)\right)}$$

for all $j \in [1, \varphi(\rho_1)]$, again using (7) in the first IPP. □

Corollary III.5 : Let μ_1 be the average number of items per source (i.e. $\mu_1=A_1/T_1$) in the first IPP and μ be the average number of items per source in the positively reinforced IPP. Then

$$\mu \geq \mu_1 \tag{16}$$

Proof : Since the number of sources in both the first and the reinforced IPP are the same (being T_1) we have, by (6) and (10)

$$\mu=\frac{1}{T_1}\int_0^{T_1} g(r_1)\,dr_1$$

$$=\frac{1}{T_1}\int_0^{T_1} \varphi\left(g_1(r_1)\right)dr_1$$

$$\geq \frac{1}{T_1}\int_0^{T_1} g_1(r_1)\,dr_1=\mu_1$$

since $\varphi \geq 1$. This follows from the fact that $g \geq 1$ and by (10). □

Corollary III.6 : If the first IPP is Lotkaian (i.e. if f_1 is a power law) and if φ is a power law then the positively reinforced IPP is also Lotkaian.

Proof : Let

$$\varphi(x) = Bx^{\beta} \tag{17}$$

$\beta > 0$ (hence $\varphi^{-1}(j) = \left(\frac{j}{B}\right)^{\frac{1}{\beta}}$) and

$$f_1(j_1) = \frac{C}{j_1^{\alpha_1}} \tag{18}$$

$(\alpha > 0)$. Then, by (13)

$$f(j) = \frac{C}{\left(\frac{j}{B}\right)^{\frac{\alpha_1}{\beta}}} \frac{1}{B\beta\left(\frac{j}{B}\right)^{\frac{\beta-1}{\beta}}}$$

$$f(j) = \frac{C}{\beta} B^{\frac{\alpha_1-1}{\beta}} \frac{1}{j^{\frac{\alpha_1+\beta-1}{\beta}}} \tag{19}$$

which is also a power function, hence a Lotka law with exponent

$$\alpha = \frac{\alpha_1 + \beta - 1}{\beta}. \tag{20}$$

□

Examples III.7 :

(i) If $\beta = 1$, i.e. φ is linear, then $\alpha = \alpha_1$ as could be expected.

(ii) For $\beta = 2$ we have that

$$\alpha = \frac{\alpha_1 + 1}{2}. \tag{21}$$

This is the case that $\varphi(x)$ is proportional to x^2. This case has links with the "Type/Token-Taken" theory of Egghe (2002) in which one does not measure the source-item (=Type/Token)-relationship but the source/used-item relationship. Examples of this are

- books (sources) are borrowed (a loan being an item) but the more a book is borrowed, the more it is encountered (when one wants to borrow it) as "in use"; i.e. sources with a high number of items are encountered more in use,
- search keys (e.g. N-grams) occur in a database (via books) but frequently occurring search keys are also encountered more often than less-frequently occurring search keys.

(iii) If $\beta > 1$ and $\alpha_1 > 1$, then $\alpha < \alpha_1$: indeed, $\alpha_1 > 1$ and $\beta > 1$ implies $\alpha_1(\beta - 1) > \beta - 1$ hence $\alpha_1 + \beta - 1 < \alpha_1\beta$, hence, by (20), $\alpha < \alpha_1$.

IV. Concentration aspects

In the previous section we showed that linear three-dimensional informetrics yields a positive reinforcement of the first IPP: in terms of the rank-frequency distribution g_1 of the first IPP, the rank-frequency distribution g of the reinforced IPP has the form $g = \varphi \circ g_1$ (cf.(10)), where φ is given by (11) or (12).

Each IPP represents an unequal situation expressed by the values of the rank-frequency function: if g represents the rank-frequency function of an IPP with T items, then we calculate the inequality in the numbers g(r), $r \in [0,T]$ by considering its Lorenz curve.

This curve is constructed (see Egghe (2003)) by the following formulae

$$L(g_1)(y) = \frac{\int_1^{yT_1+1} g_1(r)\,dr}{\int_1^{T_1+1} g_1(r)\,dr} \tag{22}$$

$$L(g)(y) = L(\varphi^\circ g_1)(y) = \frac{\int_1^{yT_1+1} \varphi(g_1(r))\,dr}{\int_1^{T_1+1} \varphi(g_1(r))\,dr} \tag{23}$$

where $y \in [0,1]$ and where, in g_1, the argument r is replaced by (substitution) r+1.

We now invoke the following Theorem of Fellman (1976) on the Lorenz curves of g_1 and of $g = \varphi \circ g_1$

Theorem IV.1 (Fellman): Let $L(g_1)$ respectively $L(g)$ denote the Lorenz curves of g_1 and $g = \varphi \circ g_1$.

(i) $L(g) \geq L(g_1)$ iff $\frac{\varphi(x)}{x}$ is increasing

(ii) $L(g)=L(g_1)$ iff $\frac{\varphi(x)}{x}$ is constant

(iii) $L(g) \leq L(g_1)$ iff $\frac{\varphi(x)}{x}$ is decreasing.

We do not give the proof: it can be found in Fellman (1976) where a different formalism is used to define the Lorenz curve. It is, however, easy to adapt this proof to the formalism (22)-(23). The proof is left to the reader.

It is clear that linear three-dimensional informetrics comprises all cases mentioned in Theorem IV.1. Indeed, let e.g. $\varphi(x) = Bx^\beta$ as in (17). If $\beta < 1$, then $\varphi(x)/x$ is strictly decreasing and hence $L < L_1$. If $\beta = 1$ then $L = L_1$ and, finally, if $\beta > 1$, then $\varphi(x)/x$ strictly increases and hence $L > L_1$. These results confirm the one of Rousseau (1992) who defines R(x) to be proportional to $\varphi(x)/x$. Hence, in Rousseau (1992), a special case of positive reinforcement was studied namely the case where R increases and hence, where $L > L_1$, i.e. the concentration in the positively reinforced IPP is higher than in the original IPP. The author feels that the term “positive reinforcement” should be reserved for the general situation expressed by Theorem III.4: indeed, the first IPP’s production is increased by composing g_1 with φ (see (10)), a strictly increasing function. Hence all productions of the sources are increased, hence the name “positive reinforcement”. In this case, it can still be that the concentration decreases, remains the same or increases, according to Theorem IV.1. Direct examples of a decrease are easy to give: one can even add items to all sources so that they all have equal production, in which case the new Lorenz curve L is the first bisector of the unit square (the smallest one possible). Another, more realistic, case is given by adding to all productions, the same amount of items in which case L is below L_1 (this is called the principle of nominal increase and it is known that in this case inequality decreases (see Egghe and Rousseau (1990b)).

References

T.M. Apostol (1957). Mathematical Analysis. Addison-Wesley, Reading, Massachusetts, 1957.

S.R. Coleman (1992). The laboratory as a productivity and citation unit in the publications of an experimental-psychology specialty. Journal of the American Society for Information Science 43(9), 639-643, 1992.

L. Egghe (1985). Consequences of Lotka's law for the law of Bradford. Journal of Documentation 41(3), 173-189, 1985.

L. Egghe (1989). The Duality of informetric Systems with Applications to the empirical Laws. Ph.D. Thesis. The City University, London, UK, 1989.

L. Egghe (1990). The duality of informetric systems with applications to the empirical laws. Journal of Information Science 16(1), 17-27, 1990.

L. Egghe (2002). Type/Token-Taken informetrics. Journal of the American Society for Information Science and Technology, to appear.

L. Egghe (2003). Zipfian and Lotkaian continuous concentration theory. Preprint, 2003.

L. Egghe and R. Rousseau (1990a). Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science. Elsevier, Amsterdam, 1990.

L. Egghe and R. Rousseau (1990b). 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.

L. Egghe and R. Rousseau (1991). Transfer principles and a classification of concentration measures. Journal of the American Society for Information Science 42(7), 479-489, 1991.

L. Egghe and R. Rousseau (1996). Modelling multi-relational data with special attention to the average number of collaborators as a variable in informetric distributions. Information Processing and Management 32(5), 563-571, 1996.

L. Egghe and R. Rousseau (2003). Size-frequency and rank-frequency relations, power laws and exponential relations: a unified approach. Progress of Natural Science (to appear).

J. Fellman (1976). The effect of transformations on Lorenz curves. Econometrica 44, 823-824, 1976.

M.F. Fox (1983). Publication productivity among scientists: A critical review. Social Studies of Science 13, 285-305, 1983.

S. Kyvik (1990). Age and scientific productivity. Differences between fields of learning. Higher Education 19, 37-55, 1990.

T. Lafouge (1995). Stochastic information field. JISSI: The International Journal of Scientometrics and Informetrics 1(2), 57-64, 1995.

C.J. Mankin and J.D. Bastille (1981). An analysis of the differences between density-of-use ranking and raw-use ranking of library journal use. Journal of the American Society for Information Science 32, 224-228, 1981.

J. Qin (1995). Collaboration and publication productivity: an experiment with a new variable in Lotka's law. In: Proceedings of the fifth Conference of the International Society for Scientometrics and Informetrics (M. Koenig and A. Bookstein, eds.), Learned Information, Medford (NJ), USA, 445-454, 1995.

R. Rousseau (1990). A bibliometric study of Nieuwenhuysen's bibliography of microcomputer software for online information and documentation work. Journal of Information Science 16, 45-50, 1990.

R. Rousseau (1992). Concentration and diversity of availability and use in information systems: A positive reinforcement model. Journal of the American Society for Information Science 43(5), 391-395, 1992.

M. Yitzhaki (1995). Relation between number of references and length of journal article. In: Proceedings of the fifth Conference of the International Society for Scientometrics and Informetrics (M. Koenig and A. Bookstein, eds.), Learned Information, Medford (NJ), USA, 647-657, 1995.