

# THE SOURCE-ITEM COVERAGE OF THE LOTKA FUNCTION

by

L. Egghe, Limburgs Universitair Centrum (LUC), Universitaire Campus, B-3590  
Diepenbeek, Belgium<sup>1</sup>

and

Universiteit Antwerpen (UA), Campus Drie Eiken, Universiteitsplein 1, B-2610  
Wilrijk, Belgium

e-mail : leo.egghe@luc.ac.be

## **ABSTRACT**

The following problem has never been studied : Given  $A$ , the total number of items (e.g. articles) and  $T$ , the total number of sources (e.g. journals that contain these articles) (hence  $A > T$ ), when is there a Lotka function

$$f(j) = \frac{D}{j^\alpha}$$

that represents this situation (i.e. where  $f(j)$  denotes the density of the sources in the item-density  $j$ ) ? And, if it exists, what are the formulae for  $D$  and  $\alpha$ ? This problem is solved in both cases with  $j \in [1, \rho]$ : where (a)  $\rho = \infty$  and where (b)  $\rho < \infty$ . Note that  $\rho$  = the maximum density of the items. If  $\rho = \infty$ , then  $A$  and  $T$  determine uniquely  $D$  and  $\alpha$ . If

<sup>1</sup> Permanent address.

Key words and phrases : Lotka, source, item, existence theorem.

$\rho < \infty$ , then we have, for every  $\alpha \leq 2$ , a solution for  $D$  and  $\rho$ , hence for  $f$ . If  $\rho < \infty$  and  $\alpha > 2$  then we show that a solution exists if and only if

$$\mu = \frac{A}{T} < \frac{\alpha - 1}{\alpha - 2}.$$

This sheds some light on the source-item coverage power of Lotka's law.

## **I. Introduction**

Duality in informetrics is explicitly present in any "historical" informetric law such as these of Lotka, Mandelbrot, Zipf, Bradford and Leimkuhler (see e.g. Egghe and Rousseau (1990)). Basically, these laws all describe the source-item relationship such as journal-article, author-publication, article-citation and so on. The dual formalism was mathematically described in Egghe (1989, 1990) - see also Egghe and Rousseau (1990) - in a continuous setting as follows. Let  $S = [0, T]$  be the source set and  $I = [0, A]$  be the item set. An "Information Production Process" (IPP) is the triple  $(S, I, V)$ , where  $V: S \rightarrow I$  is a strictly increasing differentiable function such that  $V(0) = 0$  and  $V(T) = A$ . The dual IPP of the IPP  $(S, I, V)$  is defined to be  $(I, S, U)$ , where  $U: I \rightarrow S$  is the function

$$U(i) = T - V^{-1}(A - i), \quad (1)$$

,where  $V^{-1}$  denotes the inverse function of  $V$ . Note that the dual IPP of  $(I, S, U)$  is  $(S, I, V)$ , the original IPP. We define  $\sigma(i) = U'(i) = \frac{dU(i)}{di}$  and  $\rho(r) = V'(r) = \frac{dV(r)}{dr}$  for every  $i \in I$  and  $r \in S$ . Also denote  $\rho(i) = V'(V^{-1}(i))$  so that, with an abuse of notation, we have  $\rho(i) = \rho(r)$  if  $i = V(r)$ . For example,  $\rho(A) = \rho(T)$  which we will denote simply by  $\rho$ . The following function, describing the relation between the density of the items and the density of the sources, is defined :

$$\begin{aligned} f &: [1, \rho] \rightarrow \mathbb{R}^+ \\ j &\rightarrow f(j) \end{aligned} \quad (2)$$

where  $f(j)$  is the density function (w.r.t. the IPP  $(S, I, V)$ ) of the sources as a function of  $j \in [1, \rho]$ , the density of the items. In other words, for every  $i \in I$

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

is the cumulative number of sources for which  $j \in [1, \rho(i)]$ . In Egghe (1989) one proves (the item-analogue of (3)) that

$$\int_1^{\rho(i)} j f(j) dj = i \quad (4)$$

for every  $i \in I$ . This is the basis for (continuous) duality in informetrics. The function  $f$  is the size-frequency density function and is the continuous analogue of the frequency function  $f(n)$  representing the number of sources with  $n = 1, 2, 3, \dots$  items.

If  $f$  is a power function :

$$f(j) = \frac{D}{j^\alpha} \quad (5)$$

$D, \alpha$  constants,  $\alpha > 1$ , then we say that we have the Lotka formalism (or ‘‘Lotkaian’’ informetrics). This is, of course, based on the historical paper, Lotka (1926), where Alfred Lotka describes his size-frequency distributions (functions) using power functions.

Since  $A, T, \rho, f$  appear in any IPP but  $f$  not necessarily under the form (5) we hence formulate the following problem :

**Problem** : What are the conditions, given  $A$  and  $T$  ( $A > T$ ), for the existence of a “Lotka”  $f$  as in (5) that yields  $A$  and  $T$ , i.e. for which (see (3), (4)) :

$$\int_1^{\rho} f(j) dj = T \quad (6)$$

$$\int_1^{\rho} jf(j) dj = A. \quad (7)$$

In other words, which  $A$  and  $T$  yield which type(s) of Lotkaian informetrics.

This problem will be solved in this paper. In the next section we will study the (limiting) case  $\rho = \infty$ . Here we will show that  $A$  and  $T$  ( $A > T$ ) determine one  $\alpha$  and  $D$  in (5) for which formulae will be given. The third section is then devoted to the (more intricate) solution where  $\rho < \infty$  is an extra parameter. Here we will show that any  $\alpha \leq 2$  yields a solution to the above problem and, if  $\alpha > 2$ , we show that there is a solution to the above problem if and only if

$$\frac{A}{T} = \mu < \frac{\alpha - 1}{\alpha - 2}. \quad (8)$$

This gives a complete characterization of Lotkaian informetrics, its coverage and also its limitations.

## **II. The limiting case** $\rho = \infty$ .

**Theorem II.1** : Given  $A, T \in \mathbb{R}^+$ ,  $A > T$  then the function  $f$  of (5) yields  $A$  and  $T$  ((6) and (7) for  $\rho = \infty$ ) if the following relations hold :

$$\alpha = \frac{2A - T}{A - T} \quad (9)$$

$$D = \frac{AT}{A - T} \quad (10)$$

**Proof** : The proof is easy and follows from the relations (cf. (6), (7))

$$\int_1^{\infty} \frac{D}{j^{\alpha}} dj = T \quad (11)$$

$$\int_1^{\infty} \frac{D}{j^{\alpha-1}} dj = A \quad (12)$$

Hence  $\alpha > 2$  (only in these cases can (12) be satisfied for  $A < \infty$ ) and we have

$$\frac{D}{\alpha - 1} = T \quad (13)$$

$$\frac{D}{\alpha - 2} = A \quad (14)$$

from which (9) and (10) easily follow (note that (9) implies that  $\alpha > 2$ ).

Denoting  $\mu = \frac{A}{T}$ , the average number of items per source, formulae (9), (10) can also be read as

$$\alpha = \frac{2\mu - 1}{\mu - 1} \quad (15)$$

$$D = \frac{A}{\mu - 1} \quad (16)$$

Note that  $\alpha$  only depends on  $\mu$  and not on  $A$  and  $T$  but  $D$  depends on both.

## **II.2 A result on Type/Token-Taken informetrics.**

In Egghe (2003) we describe the so-called Type/Token-Taken informetrics where one, in addition to sources and items, also describes the use that is made of the items. Here “Taken” stands for this use ; Type/Token refers (as in linguistics, see Herdan (1960)) to the source-item relationship.

The use of items is referring to the occurrence of items as they are encountered by information professionals or other information users. Examples are given in Egghe (2003) but the general idea behind the description of use is the following : sources that have many items will be encountered (say in a database) more often than sources with few items. This use leads to higher averages, i.e. average number of used-items per source, denoted  $\mu^*$ . In Egghe (2003) and with the notation of above, it is proved that

$$\mu^* = \frac{\int_1^{\infty} j^2 f(j) dj}{\int_1^{\infty} j f(j) dj} \quad (17)$$

$$\mu^* = \frac{1}{A} \int_1^{\infty} j^2 f(j) dj, \quad (18)$$

In Egghe (2003), other qualitative relations between  $\mu$  and  $\mu^*$  are given such as  $\mu^* \geq \mu$  and the fact that  $\mu^*$  is an increasing (implicite) function of  $\mu$ .

However, the following result is new, not appearing in Egghe (2003) where only the case  $\rho < \infty$  is described.

**Theorem II.2** :  $\mu^* = \infty$  if  $\alpha \leq 3$ . If  $\alpha > 3$ ,

$$\mu^* = \frac{T}{2T - A} = \frac{1}{2 - \mu} \quad (19)$$

**Proof** : We have that

$$\int_1^{\infty} j^2 f(j) dj = D \int_1^{\infty} j^{2-\alpha} dj$$

$$\begin{cases} = \frac{D}{3-\alpha} j^{3-\alpha} \Big|_1^{\infty} & (\text{if } \alpha \neq 3) \\ = D \ln j \Big|_1^{\infty} & (\text{if } \alpha = 3) \end{cases}$$

So

$$\int_1^{\infty} j^2 f(j) dj = \infty$$

if  $\alpha \leq 3$  and that

$$\int_1^{\infty} j^2 f(j) dj = \frac{D}{\alpha-3} \quad (20)$$

if  $\alpha > 3$ .

Hence  $\mu^* = \infty$  if  $\alpha \leq 3$ . If  $\alpha > 3$  we have, combining (9), (10), (18) and (20) :

$$\mu^* = \frac{1}{A} \frac{D}{\alpha-3} = \frac{T}{2T-A} = \frac{1}{2-\mu}$$

since  $\mu = \frac{A}{T}$ .

**Note** : We mentioned already that  $\mu^* \geq \mu$ , a fact that is proven in Egghe (2003). Note also that, in (19), the fact that  $\alpha > 3$  implies, using (9), that  $\mu < 2$ . Of course, since  $A > T$ , also  $\mu > 1$ . So formula (19) only applies in the region  $1 < \mu < 2$ . Fig. 1 depicts this functional relation between  $\mu^*$  and  $\mu$ . Note that the graph in (1,1) is tangent to the straight line  $y = x$

which visualizes the fact that  $\mu^* > \mu$  for  $\mu > 1$ . Formula (19) expresses a new direct functional relation between  $\mu^*$  and  $\mu$ .

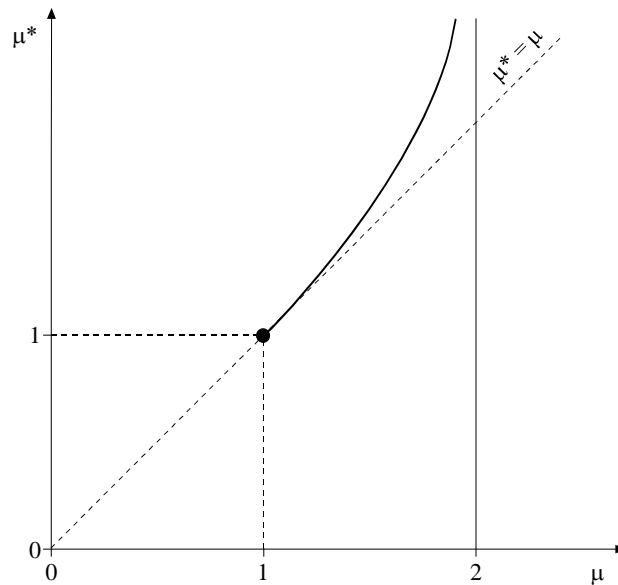


Fig. 1 : The function  $\mu^*(\mu)$ .

We will now illustrate the construction of the “right” function  $f$  (as described in Theorem II.1) by giving some examples.

### II.3 Examples.

1.  $A = 1,000$ ,  $T = 500$ . Hence  $\mu = 2$  and, by (15), (16),  $\alpha = 3$ ,  $D = 1,000$ . We have the function

$$f(j) = \frac{1,000}{j^3}$$

2.  $A = 10,000$ ,  $T = 1,000$ . Hence  $\mu = 10$  and  $\alpha = \frac{19}{9} = 2.11$ ,  $D = \frac{10,000}{9} = 1,111.11$

hence



$$f(j) = \frac{1,111.11}{j^{2.11}}$$

3.  $A = 100,000$ ,  $T = 2,000$ . Hence  $\mu = 50$  and  $\alpha = \frac{99}{49} = 2.02$ ,

$$D = \frac{100,000}{49} = 2,040.8 \text{ hence}$$

$$f(j) = \frac{2,040.8}{j^{2.02}}.$$

4.  $A = 1,000$ ,  $T = 750$ . Here  $\mu = 1.333$  and  $\alpha = 5$ ,  $D = 3,000$ , hence

$$f(j) = \frac{3,000}{j^5}$$

Note that here, the Type/Token-Taken average  $\mu^*$  equals (by (19)) :

$$\mu^* = 1.5$$

(note that  $\mu^* > \mu$ , in agreement with Fig.1).

In all these examples, of course,  $j \in [1, +\infty)$ . Note that, always,  $\alpha > 2$ , as required.

### **III. The general case** $\rho < \infty$

#### **III.1 The existence theorem for f**

We start with the main result of this paper : the existence theorem for the function  $f$ , given  $A$ ,  $T$ ,  $A > T$ .

**Theorem III.1** : Let  $A, T \in \mathbb{R}^+$ ,  $A > T$  be given. Let  $\alpha > 1$ .

(i) If  $\alpha \leq 2$ , then  $\exists \rho > 1$  and  $D > 0$  such that

$$\begin{aligned} f : [1, \rho] &\rightarrow \mathbb{R}^+ \\ j \rightarrow f(j) &= \frac{D}{j^\alpha} \end{aligned} \quad (21)$$

yields  $A$  and  $T$  (i.e. (6) and (7)).

(ii) If  $\alpha > 2$ , then the conclusion of (i) is valid if and only if

$$\frac{A}{T} = \mu < \frac{\alpha - 1}{\alpha - 2}. \quad (22)$$

**Proof** : (i) and (ii) will be proved together. The proof takes two parts.

1.  $f$  as above, yielding  $A$  and  $T$ , exists if and only if

$$\frac{A}{T} = \mu < \frac{\int_1^\infty jf(j)dj}{\int_1^\infty f(j)dj}. \quad (23)$$

Proof :

Necessary condition.

If such an  $f$  exists, we need to show that (using (6) and (7))

$$\frac{\int_1^\rho jf(j)dj}{\int_1^\rho f(j)dj} < \frac{\int_1^\infty jf(j)dj}{\int_1^\infty f(j)dj}$$

which follows from the Lemma in the Appendix.

Sufficient condition.

Consider the function (for  $x > 1$ )

$$\varphi : x \rightarrow \frac{\int_1^x jf(j) dj}{\int_1^x f(j) dj} \quad (24)$$

Since each integral is continuous in  $x$ , cf. Apostol (1957), the same is true for  $\varphi$ . Hence  $\varphi$  attains all values between its minimum and maximum (Apostol (1957)), being

$$\lim_{x \rightarrow 1} \varphi(x) = \lim_{x \rightarrow 1} \frac{xf(x)}{f(x)} = 1$$

, using de l'Hôpital's rule (Apostol (1957)), and

$$\lim_{x \rightarrow \infty} \varphi(x) = \frac{\int_1^{\infty} jf(j) dj}{\int_1^{\infty} f(j) dj}$$

since  $\varphi$  is increasing (by the Lemma in the Appendix). Hence, since we have (23), there is an  $x = \rho \in ]1, \infty[$  such that

$$\mu = \frac{\int_1^{\rho} jf(j) dj}{\int_1^{\rho} f(j) dj}. \quad (25)$$

Note that  $\rho$  is a function of  $\mu$ . Therefore we will denote it by  $\rho(\mu)$ . Since  $\mu = \frac{A}{T}$  we have,

up to a constant, that

$$A = \int_1^{\rho(\mu)} jf(j) dj \quad (26)$$

$$T = \int_1^{\rho(\mu)} f(j) dj, \quad (27)$$

which determines D. Hence, there is an

$$\begin{aligned} f : [1, \rho(\mu)] &\rightarrow \mathbb{R}^+ \\ j &\rightarrow f(j) = \frac{D}{j^\alpha} \end{aligned}$$

that yields A and T, by (26) and (27).

2. We now show that, for f as in (21), we have that

$$\frac{\int_1^\infty jf(j) dj}{\int_1^\infty f(j) dj} = \infty \quad (28a)$$

if  $\alpha \leq 2$  and that

$$\frac{\int_1^\infty jf(j) dj}{\int_1^\infty f(j) dj} = \frac{\alpha - 1}{\alpha - 2} \quad (28b)$$

if  $\alpha > 2$ , which then proves (i) and (ii) at the same time.

Proof : it is easy to see that

$$\int_1^\infty f(j) dj = \frac{D}{\alpha - 1}$$

(for all  $\alpha > 1$ ) and

$$\int_1^\infty jf(j) dj = \frac{D}{\alpha - 2}$$

for all  $\alpha > 2$ . If  $\alpha \leq 2$  then

$$\int_1^{\infty} j^{\alpha} f(j) dj = \infty.$$

This proves part 2 and hence the theorem.

**Corollary III.2** : The function

$$\rho = \rho(\mu)$$

is strictly increasing in  $\mu$

**Proof** : This follows from the Lemma in the Appendix and (25).

**Note** : The number  $\rho$  that we found in the above theorem is the same as the one introduced in section I (by (6) and (7)). The duality theory described there defined  $\rho$  as a function of A or of T. The above theorem shows that  $\rho$  in fact is only dependent on  $\mu = \frac{A}{T}$ , which is a new result for the dual informetrics model.

Now that the existence theorem for f is proved we can present the method to calculate f (i.e. D and  $\rho$ , given any  $\alpha > 1$ ) that yields a given A and T,  $A > T$ . This is done in the next subsection.

### **III.2 Construction of f.**

By (26) and (27) we have

$$\int_1^{\rho(\mu)} \frac{D}{j^{\alpha}} dj = T.$$

Hence

$$\frac{D}{1-\alpha}(\rho(\mu)^{1-\alpha} - 1) = T$$

for all  $\alpha > 1$  and also

$$\int_1^{\rho(\mu)} \frac{D}{j^{\alpha-1}} dj = A.$$

Hence

$$\frac{D}{2-\alpha}(\rho(\mu)^{2-\alpha} - 1) = A \quad (30)$$

if  $\alpha \neq 2$ . If  $\alpha = 2$  we have

$$D \ln \rho(\mu) = A. \quad (31)$$

To simplify the notation we denote  $x = \rho(\mu)$ . (29) and (30) yield  $\left(\mu = \frac{A}{T}\right)$

$$\frac{1-\alpha}{2-\alpha} \frac{x^{2-\alpha} - 1}{x^{1-\alpha} - 1} = \mu$$

or

$$\frac{\alpha-1}{\mu(\alpha-2)} x^{2-\alpha} - x^{1-\alpha} + 1 - \frac{\alpha-1}{\mu(\alpha-2)} = 0 \quad (32)$$

if  $\alpha \neq 2$ . For  $\alpha = 2$  we have, using (29) and (31), the equation

$$\ln x + \frac{\mu}{x} - \mu = 0. \quad (33)$$

The equations (32) and (33) are the key equations that must be solved numerically (the same equations were already used in Egghe (2003)). Because of Theorem III.1 we have now that, under the conditions of this theorem, equations (32) and (33) always have a solution  $x > 1$ . Once  $x = \rho(\mu)$  is determined, then (29) or (30) determine  $D$ , hence  $f$ . We will illustrate this with some examples.

### III.3 Examples

Since now we can have solutions for every  $\alpha > 1$ , we will use this as an extra parameter.

1.  $\alpha = 1.5$ ,  $A = 15,000$ ,  $T = 10,000$ , hence  $\mu = 1.5$ . We have the equation (32) which takes the form

$$-\frac{x^{0.5}}{1.5} - x^{-0.5} + \frac{2.5}{1.5} = 0$$

which can be solved for  $x > 1$  ( $x = 1$  is always an (imported) solution !) using e.g. the MATHCAD 4.0 software. We find  $x = 2.251 = \rho(\mu)$ . From (30) (but (29) also yields this) we find  $D = 14,990$ . In conclusion, the Lotka function

$$f(j) = \frac{14,990}{j^{1.5}}$$

for  $j \in [1, 2.251]$  yields the given  $A$  and  $T$  (for  $\alpha = 1.5$ ).

2.  $\alpha = 2$ ,  $A = 50,000$ ,  $T = 20,000$ , hence  $\mu = 2.5$ . We have the equation (33) which takes the form

$$\ln x + \frac{2.5}{x} - 2.5 = 0$$

which is solved in MATHCAD 4.0 as  $x = 9.322 = \rho(\mu)$ . From (31) or (29) we now have  $D = 22,398$ . Hence the Lotka function

$$f(j) = \frac{22,398}{j^2}$$

For  $j \in [1, 9.322]$  yields the given A and T (for  $\alpha = 2$ ).

3.  $\alpha = 2.5$ ,  $A = 20,000$ ,  $T = 10,000$ , hence  $\mu = 2$ . Note that the conditions

$\mu = 2 < \frac{\alpha - 1}{\alpha - 2} = 3$  is satisfied so that there is a solution. We have now (32) :

$$\frac{3}{2}x^{-0.5} - x^{-1.5} - \frac{1}{2} = 0$$

yielding  $x = 7.464 = \rho(\mu)$ . From (30) or (29) we find  $D = 15,773$ . In conclusion, the Lotka function

$$f(j) = \frac{15,773}{j^{2.5}}$$

for  $j \in [1, 7.464]$  yields the given A and T (for  $\alpha = 2.5$ ).

4.  $\alpha = 3.5$ ,  $A = 20,000$ ,  $T = 10,000$ , hence  $\mu = 2$  (as above). Now

$$\mu = 2 > 1.6667 = \frac{\alpha - 1}{\alpha - 2}.$$

So, according to the theorem III.1 now there is no  $x = \rho(\mu)$  and function  $f = \frac{D}{j^{3.5}}$  that yields A and T. Equation (32) now has no solution in  $x > 1$ .



## Appendix

**Lemma** : For any integrable positive function  $f$  we have that  $1 < x < y \leq \infty$  implies

$$\int_1^y \int_1^x uf(u)f(v)dudv < \int_1^y \int_1^x vf(u)f(v)dudv \quad (\text{A1})$$

**Proof** :

$$\begin{aligned} & \int_1^y \int_1^x (v-u)f(u)f(v)dudv \\ &= \int_1^x \int_1^x (v-u)f(u)f(v)dudv + \int_x^y \int_1^x (v-u)f(u)f(v)dudv \\ &= \int_1^x \int_1^x vf(u)f(v)dudv - \int_1^x \int_1^x uf(u)f(v)dudv + \int_x^y \int_1^x (v-u)f(u)f(v)dudv \end{aligned}$$

Since the first two terms are equal and since in the third term  $u \in [1, x]$  and  $v \in [x, y]$ , the result follows from the fact that  $f$  is positive and that  $1 < x < y$ .

## **References**

- T.M. Apostol (1957). *Mathematical Analysis. A modern Approach to advanced Calculus.* Addison-Wesley, Reading, Massachusetts, 1957.
- 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 (2003). *Type/Token-Taken informetrics.* *Journal of the American Society for Information Science and Technology*, 54(7), 603-610, 2003.
- L. Egghe and R. Rousseau (1990). *Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science,* Elsevier, Amsterdam, 1990.
- G. Herdan (1960). *Type-Token Mathematics. A Textbook of mathematical Linguistics.* Mouton, 's Gravenhage, 1960.
- A.J. Lotka (1926). *The frequency distribution of scientific productivity.* *Journal of the Washington Academy of Sciences*, 16, 317-323, 1926.