NEW BRADFORDIAN LAWS EQUIVALENT WITH OLD LOTKA LAWS, EVOLVING
FROM A SOURCE-ITEM DUALITY ARGUMENT

L. EGGHE

LUC, Universitaire Campus, B-3610 Diepenbeek, Belgium (*)
UIA, Universiteitsplein 1, B-2610 Wilrijk, Belgium

Abstract

Based on the duality techniques in a previous paper (L. Egghe,
The duality of informetric systems with applications to the
empirical laws), we study general relationships between Bradfordian
and Lotka laws. This results in new Bradfordian laws which are

equivalent with the well-known Lotka laws $\psi(n) = \dfrac{B}{n^{\alpha}}$ ($\alpha > 1$). The

method also sheds some light on the question why $\alpha < 2$ is more
common than $\alpha > 2$. Also, the general law of Leimkuhler, as found
by Rousseau, is reproved and shown to be equivalent with the above
mentioned laws. Fitting methods are applied and give close results.

I. INTRODUCTION AND REVIEW OF KNOWN RESULTS

I.1. Bibliographies, duality

I.1.1. Bibliographies

We will use the formalism from [6]. A bibliography (in [6] we used the term
IPP for the same thing but we prefer here the more concrete term "bibliography")
is a triple (S,I,V) where S and I are sets (resp. of sources and items) and
where V is a relation from S into I. S and I can be countable (examples : all
practical bibliographies) or continuous intervals. In the latter case one
speaks about continuous bibliographies : a continuous bibliography is a tripel
(S,I,V) where S and I are intervals : S = [0,T], I = [0,A] and where V is a
strictly increasing differentiable function from S into I such that V(0) = 0
and V(T) = A. In this setting, $\forall r \in S$, $V(r) \in I$ is the cumulative number of
items in the sources s, for s $\in$ [T-r,T].

Continuous bibliographies are close models for large bibliographies and
certainly contain (in the sub-set sense) all the discrete ones. They also
give more insight in both the dual theory of bibliographies (see [6]) as well
as in Bradford's law (see again [6]). In the sequel we will only be dealing
with continuous bibliographies. We repeat the concept of duality in continuous
bibliographies (or IPP's, as they are called there), as this was introduced in
[6].

I.1.2. Duality - intuitively

It is clear that any bibliometric principle on bibliographies must be dealing
with sources and items. A probabilistic principle of this kind was formulated

---

(*) Permanent address.

by D. De Solla Price [3] in his so called "success-breeds-success" principle.

Duality in bibliometrics is another source-item principle which finds its origin in geometrh. Here one has a natural duality between lines and points in this way that any result on points vs. lines gives rise to a new result obtained from the former by interchanging the words "points" and "lines". *In fact the duality principle in geometry has given rise to a whole subject of geometry called projective geometry.*

I.1.3. Duality - mathematical

Let (S,I,V) be any continuous bibliography, where $S = [0,T]$ and $I = [0,A]$. The dual bibliography of (S,I,V) is the bibliography (cf. [6]) :

$$([0,A], [0,T], U) \tag{1}$$

where

$$U(i) = I - V^{-1}(A-i) \tag{2}$$

for every $i \in [0,A]$ (V^{-1} denotes the inverse function of V, which exists since V is strictly increasing). Let σ and ρ be the following functions :

$$\sigma_i = U'(i) \tag{3}$$

and

$$\rho_r = V'(r) \tag{4}$$

resp. for every $i \in [0,A]$ and for every $r \in [0,T]$ (' denotes the derivative), ρ_s expressed in function of i gives

$$\rho_i = V'(V^{-1}(i)) \tag{5}$$

In [6] the following easy lemma is proved (for completeness reasons we repeat the short proof).

Lemma :

$$\rho_i = \frac{1}{\sigma_{A-i}} \tag{6}$$

for every $i \in [0,A]$.

Proof :

For the used results from mathematical analysis we refer the reader to [1]. For every $i \in [0,A]$, using (2) one has

$$U'(i) = \frac{1}{V'(V^{-1}(A-i))}$$

$$= \frac{1}{\rho_{A-i}} \quad ,$$

by (5). Hence (3) gives

$$\sigma_i = \frac{1}{\rho_{A-i}}$$

or, what is the same,

$$\rho_i = \frac{1}{\sigma_{A-i}}$$

for every $i \in [0,A]$. □

This lemma, together with the definition of V, σ_i and ρ_i gives now the following result :

σ_i = the density function in the coordinate $i \in I$ of the sources per item $\qquad$ (7)

ρ_i = the density function in the coordinate A-i $\in I$ of the items per source. $\qquad$ (8)

Readers not interested in the above derivations can suffice by reading formulas (6), (7) and (8). They suffice to follow the rest of this paper.

I.2. Bradford's law for continuous bibliographies

Let $(S = [0,T], T = [0,A], V)$ be any continuous bibliography. We say that this bibliography satisfies Bradford's law if there exist constants C and K > 1 such that

$$\sigma_i = C.K^i \qquad (9)$$

for every $i \in [0,A]$ (cf. [6]). This definition is equivalent with the more classic one where one specifies the number of Bradford groups : for this, see [7]. The definition above has the advantage not to deal with (and hence being independent of) the number p of Bradford groups. This in turn implies that (9) is a fixed function for the bibliography (i.e. K is constant, contrary with the p-dependent versions where this so-called Bradford multiplicator k(p) is p-dependent). For this reason, the above formulation of Bradford's law is called the *group-free version of Bradford's law*. Remark that we need the fine structure of a continuous bibliography in order to define this concept. In [7] there is proved an explicit formula for K, in absolute known bibliography constants as well as in function of k(p) for every p.

As is well-known (cf. [4]), Bradford's law is equivalent with *Lotka's law*

$$\psi(j) = \frac{B}{j^2} \qquad (10)$$

(for every $j \geq 1$). Here $\psi(j)$ denotes the number of sources with j items and B is a constant. We refer also to [4] and [7] for the fact that the above laws are also equivalent with the so-called law of Leimkuhler (sometimes also called the Bradford-Zipf law) : if $R = U^{-1}$, then

$$R(r) = a \ln (1 + br) \qquad (11)$$

for every $r \in [0,T]$, where a and b are constants.

I.3. Outline of this article

This paper deals with the question : can the above mentioned duality technique be used to determine generalized Leimkuhler or Bradford laws, starting from the common general law of Lotka :

$$\psi(j) = \frac{B}{j^\alpha} \qquad (12)$$

for every $j \in [0,A]$, where $\alpha > 1$ and B is a constant.

It is well-known that a lot of bibliographies, satisfy Lotka's law for a certain $\alpha > 1$. Let us just mention [14] (and other work of Pao and others) for many pertinent examples. In [18], Rousseau shows that, if we have (12), then we have the following generalized law of Leimkuhler (in our notation) :

$$R(r) = \frac{B}{2 - \alpha} [y_m^{2-\alpha} + \frac{\alpha - 1}{B})^{\frac{2-\alpha}{1-\alpha}}] \tag{13}$$

where B and α are as in (8) and where y_m is the number of items in the most important source.

In this paper, we will reprove formula (13) and we will add to these two formulas the Bradford equivalent. Furthermore, we will show that we have equivalence of (12), (13) and the newly established generalized law of Bradford.

That this theory conforms well with discrete data has been shown (by theoretical example) in [18], and further, more practical investigations will follow at the end of this article. They will show that formula (13) (and hence our approach with continuous bibliographies) fits very well the data of practical bibliographies.

Incidentally we will also find an explanation why α in formula (12) cannot be far above 2. Stated in other terms, a Groos droop ($\alpha < 2$) in the Leimkuhler curves is more common than the opposite effect ($\alpha > 2$), as is also found to be true in practice (see e.g. [4] and [2]).

II. BASIC EQUATIONS

Let us take a general continuous bibliography ($[0,T]$, $[0,A]$, V) and denote by $\psi(j)$ the number of sources with j items. Then the following equations follow immediately from (7) and (8) :

Item-relationship :

$$\int_1^{\rho_i} \psi(j) \, jdj = i, \text{ for every } i \in [0,A] . \tag{14}$$

Source-relationship :

$$\int_0^i \sigma_{A-i'} \, di' = \int_1^{\rho_i} \psi(j) \, dj, \text{ for every } i \in [0,A] . \tag{15}$$

Equation (15) is a difficult-to-handle integral equation, while equation (14) is easy. Luckily, we have the following result.

Theorem II.2 : Given equation (6) then equations (14) and (15) are equivalent.

Proof :

For the used results from mathematical analysis (here and in the sequel) we refer the reader to [1].

a. (14) is equivalent with the system

$$\begin{cases} \psi(\rho_i) \; \rho_i \rho_i' = 1 \\ \rho_0 = 1 \end{cases} \tag{16}$$

Indeed, (14) $\Rightarrow$ (16) follows by derivation and from

$$\int_1^{\rho_0} \psi(j) \; dj = 0 \quad (\text{i.e. } \rho_0 = 1) \text{ and } (16) \Rightarrow (14) \text{ is shown as follows :}$$

integrating (16) yields

$$\int_0^i \psi(\rho_i) \; \rho_{i'} \cdot \rho_{i'}' \cdot dj = i + C$$

Hence

$$\int_{\rho_0=1}^{\rho_i} \psi(j) \; jdj = i + C$$

From this it follows that C = 0 (take i = 0).
Finally we have

$$\int_1^{\rho_i} \psi(j) \; jdj = i \; ,$$

being (14).

b. (15) is equivalent with the system

$$\begin{cases} \sigma_{A-i} = \psi(\rho_i) \; \rho_i' \\ \rho_0 = 1 \end{cases} \tag{17}$$

This is shown in the same way as above.

c. Given (6), it is trivial to see that (16)$\Leftrightarrow$(17). Hence also (14)$\leftrightarrow$(15). $\square$

In view of the previous results we can skip equation (15) and work with the system

$$\begin{cases} \rho_i = \dfrac{1}{\sigma_{A-i}} \; , \text{ for every } i \in [0,A] \tag{6} \\[2ex] \int_1^{\rho_i} \psi(j) \; jdj = i, \text{ for every } i \in [0,A] \tag{14} \end{cases}$$

These equations are basic for this article and will yield most of our results.

III. EXCLUSION OF CERTAIN LOTKA LAWS ψ, GIVEN A BIBLIOGRAPHY

This paragraph does not deal with the problem of fitting practical data with a certain Lotka law. It will however shed some light on why certain Lotka laws are not very much encountered in practice.

We suppose, in this section, that ψ (the general Lotka law) is continuous and defined on the interval $[1,\infty[$. This does not mean that we have sources with an unlimited number of items. We just suppose the existence of the continuous function, being an extension of the used one. The function ψ is then, in practice, restricted to the interval $[1,y_m[$, where y_m is the number of items in the most productive source. We also suppose $\psi > 0$ on $[1,\infty[$. All the functions (12) satisfy these two requirements.

We have the following result.

Theorem III.1 : If ψ is a continuous positive function defined on the interval
$[1,\infty[$, and if σ and ρ are defined as in section I, then the
following assertions are equivalent :

(i) The function ρ_i exists.

(ii) The function σ_i exists.

(iii) $A > \int_1^\infty \psi(j) \, jdj$ (18)

Proof : (i) $\longleftrightarrow$ (ii)

Follows readily from (6).

(iii) $\to$ (i)

If $\int_1^x \psi(j) \, jdj \leq A$ for every $x \in [1,\infty[$, then

$$\int_1^\infty \psi(j) \, jdj = \lim_x \int_1^x \psi(j) \, jdj \leq A ,$$

a contradiction to (iii). Hence, there is a $x_0 \in [1,\infty[$ such that

$$A < \int_1^{x_0} \psi(j) \, jdj$$ (19)

Let $i \in [0,A]$ be fixed but arbitrary. Hence

$$\int_1^1 \psi(j) \, jdj = 0 \leq i \leq A < \int_1^{x_0} \psi(j) \, jdj$$ (20)

Furthermore, the function

$$x \to \int_1^x \psi(j) \, jdj$$

is continuous. So we have the existence of $x_i \in [0,A]$ such that

$$i = \int\limits_{1}^{x_i} \psi(j) \, dj$$

(cf. [1]). It then suffices to define

$$\rho_i = x_i$$

for every $i \in [0,A]$.

(i) → (iii)

Since ρ_i exists, for every $i \in [0,A]$, we have by (14) :

$$A = \int\limits_{1}^{\rho_A} \psi(j) \, dj \tag{21}$$

But, since $\psi > 0$ on $[1,\infty[$ and since ψ is continuous, we have that

$$\int\limits_{\rho_A}^{\infty} \psi(j) \, j \, dj > 0 \tag{22}$$

(suppose $\int\limits_{\rho_A}^{\infty} \psi(j) \, j \, dj = 0$. Then the function $j \to \psi(j) \, j$ is 0 almost everywhere on $[\rho_A,\infty[$, in the Lebesgue-sense. But ψ is continuous. Hence $j \to \psi(j) \, j$ is identically 0 on $[\rho_A,\infty[$. Hence $\psi = 0$ on $[\rho_A,\infty[$, a contradiction (see [1] for the used results)). (This argument is not needed further on in the proof). Hence (21) and (22) imply

$$A < \int\limits_{1}^{\infty} \psi(j) \, j \, dj \; . \quad \square$$

This result has an interesting consequence.

Corollary III.2 : Suppose that (S,I,V) is an arbitrary bibliography.

Suppose we take

$$\psi(j) = \frac{B}{j^{\alpha}} \tag{12}$$

for $j \in [1,\infty[$. Then this function can never fit our bibliography (as a law of Lotka), if

$$\alpha \geq \frac{B}{A} + 2 \tag{23}$$

Here A is the total number of items.

Proof :

A function ψ as above does not fit our bibliography if the function $i \to \rho_i$ cannot be constructed from it (through (14)). This implies, according to theorem III.1 that

$$\int\limits_{1}^{\infty} \psi(j) \, j \, dj \leq A$$

For the function ψ as above, this yields easily

$$\alpha \geq \frac{B}{A} + 2 . \quad \square$$

The practical value of formula (23) will be investigated in another paper, but it is already clear from the above that high values of α are not expected very much. Stated in terms of Leimkuhler curves (cf. [4], [18]), this gives an explanation of the fact that more often one finds a Groos droop ($\alpha < 2$) than the opposite effect ($\alpha > 2$). In any case we also have the following, surprising result :

Corollary III.3 : If the Lotka function

$$\psi(j) = \frac{B}{j^{\alpha}}$$

($j \in [1,\infty[$) fits a bibliography, then

$$\alpha < 3 \quad .$$

Proof :

From corollary III.2 we readily have

$$\alpha < \frac{B}{A} + 2$$

But $B = \psi(1)$ and A is the total number of items. So, certainly

$$B < A .$$

Hence

$$\alpha < 3 . \quad \square$$

In some practical fitting procedures (such as the one of Pao [14]) one sometimes encounters $\alpha > 3$. In view of our theory above, this can only be so if we lower A, i.e. if we do not use the whole bibliography. This is indeed what is done in the Pao fitting procedure.

IV. THE NEW LAW OF BRADFORD THAT IS IMPLIED BY LOTKA'S LAW $\psi(j) = \frac{B}{j^{\alpha}}$

        After the preparation above it is now easy to find the general law of Bradford that is implied by the laws of Lotka :

$$\psi(j) = \frac{B}{j^{\alpha}}$$

for every $j \in [1,\infty[$. In fact, in the next section we will show that this law of Bradford is equivalent with (12).

Theorem IV.1 : Let (S,I,V) be a bibliography that satisfies Lotka's law (12) for a certain $\alpha \neq 2$. Let A be the total number of items. Then

$$\rho_i = (\frac{i(2-\alpha)}{B} + 1)^{\frac{1}{2-\alpha}} \tag{24}$$

and

$$\sigma_i = ((\frac{A(2-\alpha)}{B} + 1) - i \frac{2-\alpha}{B})^{-\frac{1}{2-\alpha}} \tag{25}$$

for every $i \in [0,A]$. Hence, the general law of Bradford $(\alpha \neq 2)$ is of the form

$$\sigma_i = (A_1 + iA_2)^{A_3}$$ (26)

(instead of (9) which is the function if $\alpha = 2$).

Proof :

(26) follows from (25) and (25) follows from (24), using (6). Hence all we have to do is to prove equation (24). From (14), we find

$$\int_1^{\rho_i} \frac{B}{j^{\alpha-1}} \, dj - i = 0$$

Hence

$$\frac{B}{2-\alpha} (\rho_i^{2-\alpha} - 1) - i = 0$$

This implies

$$\rho_i = (\frac{i(2-\alpha)}{B} + 1)^{\frac{1}{2-\alpha}}$$ (27)

if

$$\frac{i(2-\alpha)}{B} + 1 > 0$$ (28)

for every $i \in [0,A]$. But since our bibliography satisfies (12), we have, by corollary III.2 that

$$\alpha < \frac{B}{A} + 2$$

Hence , if $\alpha > 2$

$$\frac{A(2-\alpha)}{B} + 1 > 0$$ (29)

Now there are two cases :

(i) $\alpha < 2$. Then

$$\frac{i(2-\alpha)}{B} + 1 > 0$$ (30)

always, for every $i \in [0,A]$.

(ii) $\alpha > 2$. Then

$$\frac{A(2-\alpha)}{B} + 1 = \min_{i \in [0,A]} (\frac{i(2-\alpha)}{B} + 1)$$ (31)

Hence, (29) and (31) imply

$$\frac{i(2-\alpha)}{B} + 1 > 0$$

for every $i \in [0,A]$.

In conclusion : (28) is always satisfied and hence also (27). □

Remark : From formula (25) it follows that $\lim_{\alpha \to 2} \sigma_i$ is an exponential function. Hence our theory for $\alpha \neq 2$ gives the classical Bradford situation (when $\alpha = 2$)

as a limiting case (cf. (9)).

This is the first time that Bradford's law for the general Lotka law (12) is proved. In [4] we tried to put up a qualitative model for Bradford's law in case of formule (12). Although not perfect we predicted that (with our notation σ_i for the Bradford function) :

if $\alpha < 2$: $\dfrac{\sigma_i'}{\sigma_i}$ must increase with i

if $\alpha > 2$: $\dfrac{\sigma_i'}{\sigma_i}$ must decrease with i

and (as is well-known and trivial to prove) : if $\alpha = 2$ (hence in case we have the classical Bradford law (9)) :

$\dfrac{\sigma_i'}{\sigma_i}$ is a constant (independent of i) .

The theorem above confirms these qualitative predictions of [4].

Corollary IV.2 : In case our bibliography satisfies (12) then the corresponding law of Bradford σ_i satisfies

$\dfrac{\sigma_i'}{\sigma_i}$ increases with i if $\alpha < 2$

$\dfrac{\sigma_i'}{\sigma_i}$ decreases with i if $\alpha > 2$

$\dfrac{\sigma_i'}{\sigma_i}$ is constant if $\alpha = 2$.

Proof :

Suppose that $\alpha \neq 2$. Formula (26) yields

$$\sigma_i' = A_2 A_3 (A_1 + i A_2)^{A_3 - 1}$$

Hence

$$\frac{\sigma_i'}{\sigma_i} = \frac{A_2 A_3}{A_1 + i A_2}$$

Now, substituting the values of A_1, A_2 and A_3, in terms of A, B and α (using (25)) yields :

$$\frac{\sigma_i'}{\sigma_i} = \frac{1}{A(2-\alpha) + B - i(2-\alpha)} \tag{32}$$

This is an increasing function if $\alpha < 2$ and a decreasing one if $\alpha > 2$. If $\alpha = 2$, the result is well-known : Formula (9) yields

$$\frac{\sigma_i'}{\sigma_i} = \ln K , \qquad \text{a constant.} \quad \square$$

V. THE LAW OF LEIMKUHLER R THAT EVOLVES FROM $\psi(j) = \dfrac{B}{j^\alpha}$ AND PROOF OF THE
EQUIVALENCE OF THE FUNCTIONS ψ, σ AND R

This law of Leimkuhler has already been shown by R. Rousseau in [18]. With our
method we can give a second proof of it. Also, having the generalized Bradford
law at our disposition, we will also be able to show equivalences between the
laws ψ, σ and R.

Theorem V.1 : (Rousseau [18]) : Suppose we have a bibliography (S,I,V) that
satisfies Lotka's law

$$\psi(j) = \frac{B}{j^\alpha} \tag{12}$$

Then

$$R(r) = \frac{B}{2-\alpha} \left[y_m^{2-\alpha} - (y_m^{1-\alpha} + \tfrac{\alpha-1}{B} r)^{\frac{2-\alpha}{1-\alpha}} \right] \tag{13}$$

where y_m is the number of items in the most productive source.

Proof :

In our formalism, we clearly have :

$$r = \int_0^i \sigma_{i'} \, di' \tag{33}$$

and

$$R(r) = i \tag{34}$$

But, using theorem IV.1, formula (26) :

$$\int_0^i \sigma_{i'} \, di' = \frac{(A_1 + iA_2)^{1+A_3} - A_1^{1+A_3}}{A_2(1 + A_3)}$$

So, using (33) and (34) above and recalculating, we find

$$R(r) = \frac{1}{A_2} \left[(A_1^{1+A_3} + A_2(1 + A_3)r)^{\frac{1}{1+A_3}} - A_1 \right] \tag{35}$$

Hence, using the meaning of A_1, A_2 and A_3 in terms of A, B and α (formula (25)),
we get, after reworking :

$$R(r) = \frac{B}{2-\alpha} \left[\left(\frac{A(2-\alpha)}{B} + 1 \right) - \left(\left(\frac{A(2-\alpha)}{B} + 1 \right)^{\frac{1-\alpha}{2-\alpha}} - \frac{1-\alpha}{B} r \right)^{\frac{2-\alpha}{1-\alpha}} \right] \tag{36}$$

By the very definition of the dual Bradford function ρ_i we have that $y_m = \rho_A$
(A being the total number of items and hence also the end of the iteminterval
[0,A]). Using theorem IV.1 again (now formula (24)), we find

$$\rho_A = \left(\frac{A(2-\alpha)}{B} + 1 \right)^{\frac{1}{2-\alpha}} \tag{37}$$

Hence, combining (36) and (37), yields :

$$R(r) = \frac{B}{2-\alpha} \left[y_m^{2-\alpha} - (y_m^{1-\alpha} - \tfrac{1-\alpha}{B} r)^{\frac{2-\alpha}{1-\alpha}} \right] ,$$

being exactly the function as found by Rousseau in [18]. Here $r \in [0,T]$, T being the total number of sources. □

Remark : As shown in [18], the function R (formula (13)), shows a Groos droop for $\alpha < 2$ and has no Groos droop for $\alpha > 2$ (as predicted also in [4]). If there is a Groos droop, the inflection point is given by

$$r_d = \frac{B}{2-\alpha} (\frac{A(2-\alpha)}{B} + 1)^{\frac{1-\alpha}{2-\alpha}}$$

or, using (37) :

$$r_d = \frac{B}{2-\alpha} y_m^{1-\alpha} \tag{38}$$

(see [18], where also practical calculations are made).

So far we came across the following functions, for $\alpha \neq 2$:

(old) general Lotka :

$$\psi(j) = \frac{B}{j^\alpha}, \; j \geq 1 \tag{12}$$

(new) general Bradford :

$$\sigma_i = ((\frac{A(2-\alpha)}{B} + 1) - i \frac{2-\alpha}{B}))^{-\frac{1}{2-\alpha}} \tag{25}$$

or

$$\sigma_i = (y_m^{2-\alpha} - i \frac{2-\alpha}{B})^{-\frac{1}{2-\alpha}} \tag{39}$$

and

(new) general Leimkuhler :

$$R(r) = \frac{B}{2-\alpha} [y_m^{2-\alpha} - (y_m^{1-\alpha} - \frac{1-\alpha}{B} s)^{\frac{2-\alpha}{1-\alpha}}] \tag{13}$$

These three laws constitute a new closed circuit of equivalent laws :

Theorem V.2 : The general laws of Lotka, Bradford and Leimkuhler are equivalent.

Proof : Lotka → Bradford :

This is in fact theorem IV.1.

Bradford → Leimkuhler :

This is what is shown in theorem V.1 (see the proof).

Leimkuhler → Bradford :

Formula (13), which is given, is simply rewritten as :

$$R(r) = C [D - (E + Fr)^G] \tag{40}$$

where C, D, E, F and G are constants.
From formulas (33) and (34) we have

$$i = C [D - (E + F . \int_0^i \sigma_{i'} di')^G]$$ (41)

This yields :

$$\int_0^i \sigma_{i'} di' = \frac{(D - \frac{i}{C})^{\frac{1}{G}} - E}{F}$$

Hence

$$\sigma_i = - \frac{1}{GFC} (D - \frac{i}{C})^{\frac{1}{G} - 1}$$ (42)

which is of the form (39).

Bradford ⇒ Lotka :

We have given (25) (or (39)), hence also ρ_i, using (6). This ρ_i is of the form (see e.g. (24)) :

$$\rho_i = (1 + Hi)^I$$ (43)

where H and I are constants.
Using the universally valid equation (14) :

$$\int_1^{\rho_i} \psi(j) j dj = i$$ (14)

we also have, by derivation

$$\psi(\rho_i) \rho_i \rho_i' = 1$$ (44)

But, using (44) :

$$\rho_i' = IH (1 + Hi)^{I-1}$$

$$\rho_i' = \frac{IH}{1 + Hi} \rho_i$$ (45)

Using again (43) we see that

$$1 + Hi = \rho_i^{\frac{1}{I}}$$

We substitute this in equation (45) yielding

$$\rho_i' = IH \rho_i^{1 - \frac{1}{I}}$$ (46)

(44), (43) and (46) now yield

$$\psi(\rho_i) = \frac{1}{\rho_i \rho_i'}$$

$$= \frac{1}{IH \rho_i^{2 - \frac{1}{I}}}$$

Hence, we have the form

$$\psi(j) = \frac{B}{j^\alpha} \tag{12}$$

($\alpha \neq 2$), being Lotka's law. □

Remark : Both the (new) general Bradford law (25) or (39) and the (new) general Leimkuhler law (13) have 3 parameters, namely α, B and A (or y_m being ρ_A). In the (old) general Lotka law (12) we see only 2 parameters explicitely, namely α and B. But also here, A is present in an implicit way : j is restricted to the interval $[1,y_m] = [1,\rho_A]$.

VI. FITTING THE GENERALIZED LEIMKUHLER LAW AND EXAMPLES

In this last section it is our purpose to see whether (and how) the theoretically calculated general law of Leimkuhler

$$R(r) = \frac{B}{2-\alpha} [y_m^{2-\alpha} - (y_m^{1-\alpha} - \frac{1-\alpha}{B})^{\frac{2-\alpha}{1-\alpha}}] \tag{13}$$

is fitting practical (hence non-continuous or discrete) bibliographies.

Of course, this is a double problem : what α and B should be used : α and B come from Lotka's law :

$$\psi(j) = \frac{B}{j^\alpha} \tag{12}$$

and hence we face the problem : find α such that both laws (12) and (13) fit our practical data. From this, B follows easily since

$$T = \sum_{j=1}^{y_m} \psi(j) = B \sum_{j=1}^{y_m} \frac{1}{j^\alpha}$$

and hence

$$B \approx \frac{T}{\zeta(\alpha)} \tag{47}$$

where $\zeta(\alpha)$ is the classical zèta-function. B can then be derived from a table of $\zeta(\alpha)^{-1}$ which is appearing for instance in [10], p.444.

On fitting Lotka's law (12) to practical data, there have been a lot of papers : [9], [10], [12], [13], [14], [19]. In them few methods to derive a good α have been constructed, some better than others. It is our feeling that this point needs more investigation : some of it will be done in a forthcoming paper.

In this paper we will suffice by investigating whether some α and B that yield a well fitting Lotka law (12) (to the practical data) will also yield a well fitting general Leimkuhler law. We will give three examples.

Example 1 :

The Pao data on computational musicology [11]. For this, the maximum
likelihood method of Nicholls [9] yields α = 2.2000 and B/T =0.6709 giving
a good fit of (12). The fit for (13) with this α and B is as follows :

r	R(r) observed	R(r) calculated (13)
1	40	35.8
2	74	65.2
3	95	90.4
4	111	112.5
5	125	132.2
6	138	150.0
7	151	166.3
8	164	181.3
9	176	195.3
10	188'	208.3
11	200	220.6
12	212	232.1
13	222	243.1
14	232	253.5
15	242	263.4
16	252	272.9
17	260	282.0
18	268	290.7
19	276	299.1
20	283	307.1
26	325	350.3
36	383	408.1
46	433	454.7
56	475	494.0
66	515	528.1
76	545	558.4
86	575	585.6
96	605	610.3
106	627	633.1
126	667	673.9
146	707	709.8
166	747	741.8
186	775	770.8
206	795	784.4
226	815	809.9
250	839	849.1
300	889	899.5
350	939	943.3
400	989	982.3
450	1039	1017.4
500	1089	1049.4

The fit is very good. Using the Kolmogorov-Smirnov test one has that the
maximal relative deviation is

$$D_{max} = 0.0412$$

while the critical value (at the 5 % level) is approximately $\frac{1.36}{\sqrt{500}}$ = 0.0608.
Hence, the model

$$R(r) = \frac{0.6709 \times 500}{-0.2} [40^{-0.2} - (40^{-1.2} + \frac{1.2}{0.6709 \times 500})^{0.1667}] \quad (48)$$

is accepted.

Example 2 :

The Murphy data [8] (see also [16] or [14]). For this, the least square method of Nicholls [9] yields $\alpha = 2.104$ and $B/T = 0.6424$. The fit of the corresponding Lotka-function (12) is very good (see [9]). With this α and B we have also a good fit for the generalized Leimkuhler function (13).

r	R(r) observed	R(r) calculated (13)
1	5	4.9
2	9	9.5
3	13	13.9
4	17	18.1
5	21	22.1
6	25	26.0
7	29	29.7
8	33	33.3
9	37	36.7
10	40	40.0
11	43	43.3
12	46	46.4
13	49	49.4
14	52	52.3
15	55	55.2
16	58	57.9
17	61	60.6
18	64	63.2
19	66	65.8
20	68	68.3
30	88	90.2
40	108	108.2
50	118	123.6
70	138	148.9
90	158	169.3
110	178	186.6
130	198	201.5
150	218	214.7
170	238	226.5

Here $D_{max} = 0.0665$ but the 5 % critical value is approximately $\frac{1.36}{\sqrt{170}} = 0.1043$.
Again we can accept our general Leimkuhler function :

$$R(r) = \frac{0.6424 \times 170}{-0.1047} [5^{-0.1047} - (5^{-1.1047} + \frac{1.1047}{0.6424 \times 170} r)^{0.0948}]$$

$$(49)$$

Example 3 :

The Radhakrishnan-Kernizan data [15], see also [14]. In this case the Nicholls least squares method yields α = 3.4880, B/T = 0.8864 and the maximum likelihood method gives α = 3.4000, B/T = 0.8782. Both methods give a fit to Lotka's law (12) (although not very splendid) but a very bad fit to R(r) ((13)). In a forthcoming paper, I develop a simple method different from the Nicholls methods that in these cases can yield a better fit. Our method yields α = 2.9907, B/T = 0.8306. We have a much better fit of the Lotka function (12) and furthermore, (13) is also a well-fitting function.

r	R(r) observed	R(r) calculated (13)
1	7	6.4
2	13	12.0
3	18	17.0
4	22	21.6
5	26	25.8
6	30	29.8
7	34	33.5
8	38	37.1
9	41	40.4
10	44	43.7
11	47	46.8
12	50	49.8
13	53	52.7
14	56	55.5
15	59	58.2
20	69	70.8
30	89	92.3
40	109	110.7
50	129	127.0
100	180	191.5
200	280	283.5
300	380	354.2

We have D_{max} = 0.086 which is at about the 1 % level. Hence we obtain a 1 %-level fit here (contrary to the other methods). The law in question here is

$$R(r) = \frac{0.8378 \times 301}{-1.0442} [7^{-1.0442} - (7^{-2.0442} + \frac{2.0442}{0.8378 \times 301} r)^{0.5108}]$$

(50)

Remark : The simpler model of (13), namely (11), valid if Lotka's α = 2, contains two parameters. As is well-known this function is able to fit the core and middle part of the Leimkuhler observed cumulative distribution (cf. also [5] for extensive calculations on this respect). Formula (13), contains three parameters and, according to the above theory and examples, seems to be able to fit also the last part of the Leimkuhler observed cumulative distribution (Groos droop or the opposite effect, which is present whenever $\alpha \neq 2$).

REFERENCES

[1]   T.M. Apostol, Mathematical Analysis, second edition. Addison-Wesley
      Publ. Co., 1974.

[2]   Y.-S. Chen and F.F. Leimkuhler, Bradford's law : an index approach.
      Scientometrics 11 (3-4), p.183-198, 1987.

[3]   D. De Solla Price, A general theory of bibliometric and other cumulative
      advantage processes. J. Amer. Soc. Inf. Sci. 27, p.292-306, 1976.

[4]   L. Egghe, Consequences of Lotka's law for the law of Bradford.
      J. Documentation 41(3), p.173-189, 1985.

[5]   L. Egghe, Applications of the theory of Bradford's law to the
      calculation of Leimkuhler's law and to the completion of bibliographies.
      J. Amer. Soc. Inf. Sci., to appear, 1990.

[6]   L. Egghe, The duality of informetric systems with applications to the
      empirical laws. J. Inf. Sci., 16(1), p.17-27, 1990.

[7]   L. Egghe, A note on different Bradford multiplicators. J. Amer. Soc. Inf.
      Sci., to appear, 1990.

[8]   L.J. Murphy, Lotka's law in the Humanities? J. Amer. Soc. Inf. Sci. 24,
      p.461-462, 1973.

[9]   P.T. Nicholls, Empirical validation of Lotka's law. Inf. Proc. and
      Manag. 22(5), p.417-419, 1986.

[10] P.T. Nicholls, Estimation of Zipf parameters. J. Amer. Soc. Inf. Sci. 38,
      p.443-445, 1987.

[11] M.L. Pao, Bibliometrics and computational musicology. Coll. Manag. 3,
      p.97-109, 1979.

[12] M.L. Pao, Lotka's test. Coll. Manag. 4, p.111-124, 1982.

[13] M.L. Pao, Lotka's law : a testing procedure. Inf. Proc. and Manag. 21(4),
      p.305-320, 1985.

[14] M.L. Pao, An empirical examination of Lotka's law. J. Amer. Soc. Inf.
      Sci. 37(1), p.26-33, 1986.

[15] T. Radhakrishnan and R. Kernizan, Lotka's law and computer science
      literature. J. Amer. Soc. Inf. Sci. 30, p.51-54, 1979.

[16] I.K. Ravichandra Rao, The distribution of scientific productivity and
      social change. J. Amer. Soc. Inf. Sci. 31, p.111-121, 1980.

[17] R. Rousseau, Relations between continuous versions of bibliometric laws.
      J. Amer. Soc. Inf. Sci., 1990 (to appear).

[18] R. Rousseau, Lotka's law and its Leimkuhler representation.
      Lib. Sci. with a Slant to Documentation, 25, 150-178, 1988.

[19] J. Tague and P. Nicholls, The maximal value of Zipf size variable :
      sampling properties and relationship to other parameters. Inf. Proc.
      and Manag. 23(3), p.155-170, 1987.

(*) One can also consult the Ph. D. thesis of L. Egghe (with the same title),
    written under supervision of Prof. Dr. B.C. Brookes (City Univ. London,
    UK), 1989.