Relations between the continuous and the discrete Lotka power function

Peer-reviewed author version

EGGHE, Leo (2005) Relations between the continuous and the discrete Lotka power function. In: Journal of the American Society for Information Science and Technology, 56(7). p. 664-668.

DOI: 10.1002/asi.20157
Handle: http://hdl.handle.net/1942/737

Relations between the continuous and the discrete Lotka power function

by

L. Egghe

Limburgs Universitair Centrum (LUC), Universitaire Campus, B-3590 Diepenbeek, Belgium[1]
and
Universiteit Antwerpen (UA), Campus Drie Eiken, Universiteitsplein 1, B-2610 Wilrijk, Belgium

leo.egghe@luc.ac.be

ABSTRACT

The discrete Lotka power function describes the number of sources (e.g. authors) with $n = 1, 2, 3, \ldots$ items (e.g. publications). As in econometrics, informetrics theory requires functions of a continuous variable j, replacing the discrete variable n. Now j represents item densities instead of number of items. The continuous Lotka power function describes the density of sources with item density j. The discrete Lotka function is the one that one obtains from data, obtained empirically; the continuous Lotka function is the one needed when one wants to apply Lotkaian informetrics, i.e. to determine properties that can be derived from the (continuous) model.

It is, hence, important to know the relations between the two models. We show that the exponents of the discrete Lotka function (if not too high, i.e. within limits encountered in

[1] Permanent address: Limburgs Universitair Centrum, Universitaire Campus, B-3590 Diepenbeek, Belgium.
Acknowledgement: The author is indebted to Prof. Dr. R. Rousseau for interesting discussions on the topic of this paper.
Key words and phrases: Discrete Lotka function, continuous Lotka function, power function.

practise) and of the continuous Lotka function are approximately the same. This is important to know in applying theoretical results (from the continuous model), derived from pratical data.

I. Introduction

Lotka's law in its historical formulation in Lotka (1926) is formulated as follows: in a fixed group of authors (or in a bibliography), the number $f(n)$ of authors with $n(n = 1, 2, 3...)$ publications is given by the power law:

$$f(n) = \frac{K}{n^a} \quad (1)$$

where K and a are positive constants. The most classical value for the exponent is $a = 2$ but even Lotka himself pointed out in Lotka (1926) that other values of a might occur, but usually $a > 1$, making (1) a (fastly) decreasing function.

Lotka's law is found to be valid in many applications in or even outside the informetrics field (see e.g. Wilson (1999), Egghe (2004) and many references in both works). In general we can talk about sources (generalizing authors above) having (or producing) items (generalizing publications above) and in this framework, $f(n)$ in (1) denotes the number of sources with n Î ¥ items.

As in other "-metrics" theories (such as econometrics), instead of (1), one uses functions of a continuous variable j, replacing the discrete variable n above. This is done for mathematical reasons: models and their properties can better be understood when one can use the formalism of calculus (i.e. mathematical analysis). This is so because it is easier to evaluate derivatives and integrals than discrete differences or sums.

Hence, in the theory of Lotkaian informetrics (see e.g. Egghe (2004a)) one uses the continuous variant of (1), denoted by φ.

The continuous Lotka function φ is also given by a power function but of a continuous variable $j \geq 1$:

$$\varphi(j) = \frac{C}{j^{\alpha}} \qquad (2)$$

where again C and α are positive constants. In words: $\varphi(j)$ is the density of sources with item density $j \geq 1$.

As said above: knowing the continuous Lotka function (2) is important since (2) is the basis for many derived results, that are not possible to prove with (1), using discrete sums. More fundamentally, without a continuous model one has no relations with other informetric (even econometric or linguistical) distributions such as the ones of Pareto and Zipf.

The problem with the continuous model (2) above is that its parameters C and α (the most important one) cannot be determined directly from a concrete set of data (e.g. a bibliography). Such a set of data (obviously being discrete) gives information about the discrete function f in (1).

From the above it is clear that the following problem is worth studying.

Problem I.1 : Determine relations between the discrete Lotka function f and the continuous Lotka function φ. Especially determine relations between the exponents a and α.

This problem will be studied in the next section. The relation between f and φ will be clarified using a third function: the discretized version of the continuous function φ, to be introduced in the sequel.

II. Comparing f and φ : introduction of the discretized version of φ

It is clear that, if $\alpha > 1$

$$T = \int_1^{\infty} \varphi(j)dj = \int_1^{\infty} \frac{C}{j^{\alpha}} dj = \frac{C}{\alpha - 1} \tag{3}$$

and, if $\alpha > 2$:

$$A = \int_1^{\infty} j\varphi(j)dj = \int_1^{\infty} \frac{C}{j^{\alpha - 1}} dj = \frac{C}{\alpha - 2} \tag{4}$$

Formulae (3) and (4) imply $\left(\mu = \frac{A}{T}\right)$

$$\alpha = \frac{2A - T}{A - T} = \frac{2\mu - 1}{\mu - 1} \tag{5}$$

$$C = \frac{AT}{A - T} = \frac{A}{\mu - 1} \tag{6}$$

In the same way, for the discrete Lotka function f, we have

$$T = \sum_{n=1}^{\infty} f(n) = \sum_{n=1}^{\infty} \frac{K}{n^{\alpha}} . \tag{7}$$

Hence

$$K = \frac{T}{\zeta(\alpha)} \tag{8}$$

Also since

$$A = \sum_{n=1}^{\infty} nf(n) = \sum_{n=1}^{\infty} \frac{K}{n^{a-1}} \tag{9}$$

we have

$$\mu = \frac{A}{T} = \frac{\zeta(a-1)}{\zeta(a)} \tag{10}$$

for $a > 2$, where $\zeta(a) = \sum_{n=1}^{\infty} \frac{1}{n^a}$ denotes Riemann's Zeta function. So the knowledge of T (the total number of sources) and of A (the total number of items) determine (within the range α, $a > 2$) α and C (the two parameters of the continuous Lotka function) via (5) and (6) and a and K via (8) and (10). The latter determination goes as follows: since we know A and T, we know $\mu = \frac{A}{T}$, the average number of items per source. Then we can determine a numerically if we have a table of a versus $\frac{\zeta(a-1)}{\zeta(a)}$ (which is provided in this article – see Appendix – from $a = 2.11$ on (in 0.01 increments)). Then, since T and a are known, K follows from (8), using a table of a versus $\frac{1}{\zeta(a)}$ which can be found in Egghe and Rousseau (1990) and Egghe (2004a).

Note that $K = f(1) \neq C = \varphi(1)$. In fact even more is true: if $\alpha > 2$, then

$$K = f(1) < T < C = \varphi(1), \tag{11}$$

showing that f and φ are different functions.

The first inequality is obvious since $f(1)$ is the number of sources with one item and since T is the total number of sources. The second inequality trivially follows from (3). That $\varphi(1) > T$

is not counter-intuitive since every j (hence also $j = 1$) denotes an item density (compare with a Gaussian density which has a total area of 1 but which can have a peak (at 0) as high as we wish).

The above yields f and φ, given A and T, so theoretically the relation between f and φ is established. However, e.g. due to the occurrence of the cumbersome function ζ in (8) and (10), we will continue our search for more practical relations between the functions f and φ. This is obtained by discretization of the function φ, which is defined now.

Definition II.1 : Let φ be as in (2). The discretized version of φ, denoted $I(\varphi)$, is defined on $\mathbb{N}$ as follows: for every $n = 1,2,3,...$

$$I(\varphi)(n) = \int_n^{n+1} \varphi(j)dj = \int_n^{n+1} \frac{C}{j^\alpha} dj \tag{12}$$

For $\alpha \neq 1$ we hence have

$$I(\varphi)(n) = \frac{C}{\alpha - 1}\left(\frac{1}{n^{\alpha-1}} - \frac{1}{(n+1)^{\alpha-1}}\right) \tag{13}$$

It is clear that, in exact mathematical terms, $I(\varphi)$ is not a power function, hence $I(\varphi)$ is never equal to f. But both functions represent discrete size-frequency functions of the same informetric system. So f and $I(\varphi)$ should have the same "shape". Note that

$$I(\varphi)(n) = -\frac{C}{\alpha - 1}\left(\frac{1}{(n+1)^{\alpha-1}} - \frac{1}{n^{\alpha-1}}\right)$$

$$I(\varphi)(n) \approx -\frac{C}{\alpha - 1}\frac{d}{dn}\left(\frac{1}{n^{\alpha-1}}\right)$$

$$I(\varphi)(n) \approx \frac{C}{n^{\alpha}} \qquad (14)$$

So $I(\varphi)$ and φ have the same shape (heuristic argument). For $I(\varphi)$ and f to have the same shape we hence require $\frac{C}{n^{\alpha}}$ and $\frac{K}{n^{a}}$ to have the same shape. This implies

$$a \approx \alpha \qquad (15)$$

for values of a and α not too high (if these values are high then, even if they are very different,

$$\frac{C}{n^{\alpha}} \approx \frac{K}{n^{a}} \approx 0$$

for $n \geq 2$).

Whether (15) is true for values of a and α not too high can be proved in an exact (numerical) way as we will do in the sequel. Note first that it follows from the exhaustive list of examples in Egghe (2004a) (Chapter I) that most exponents are below 4 (and most commonly around 2). For these values we can indeed prove (15) as follows. Use (5) and (10) (and hence also the table in the Appendix) to obtain the following table of α- and a-values in function of $\mu = \frac{A}{T}$ (note that (5) and (10) imply that both a and α only depend on μ and not on the two values of T and A themselves!).

Table 1. Comparison of values of α and a for several

values of $\mu = \frac{A}{T}$

μ	α	a
1.5	4.00	2.81
2	3.00	2.48
2.5	2.67	2.34
3	2.50	2.27
3.5	2.40	2.22
4	2.33	2.18
4.5	2.29	2.16
5	2.25	2.14
5.5	2.22	2.13
6	2.20	2.12

We see that α and a are indeed comparable and closer to each other the closer they are to 2, the most common value. We also note from (4), (5) and (10) that $\alpha > 2$ if and only if $a > 2$ (hence also $\alpha \leq 2$ if and only if $a \leq 2$), even if a and α are very different (for high values). This is an important conclusion because many, from Lotka's function φ derived properties (e.g. shape of the cumulative first-citation distribution, existence of the Groos droop, concentration and fractal properties,… - see Egghe (2004a)), are common for different αs above 2 and for different αs below 2 but are different for values of α below and above 2.

Indeed, data show a Groos droop (see Groos (1967)) if and only if $\alpha < 2$ (see Egghe (1985, 2004a) or Egghe and Rousseau (1990)). The cumulative first-citation distribution is S-shaped if and only if $\alpha > 2$ and is concave if and only if $\alpha \leq 2$ (see Egghe (2000)). In all these results, the value $\alpha = 2$ is a turning point.

So, from the above, the practical calculation of the exponent a (based on the data) yields an estimate of α which determines $\varphi(j) = \frac{C}{j^{\alpha}}$.

Note also that (3) and (13) imply that

$$I(\varphi)(1) = T\left(1 - \frac{1}{2^{\alpha-1}}\right) < T \qquad (16)$$

, a property that is shared with the discrete function f (and not with the continuous function φ itself). In fact it is easy to see that $I(\varphi)(n) < T$ for every $n \in \mathbb{N}$. Inequality (16) can even be improved by the following proposition.

<u>Proposition II.2</u> : If $\alpha = a > 2$, then

$$I(\varphi)(1) < K = f(1) < T < C = \varphi(1) \qquad (17)$$

<u>Proof</u> : We only have to show that (using (11)) $I(\varphi)(1) < K$. Using (16) and (8) we hence have to show (using that $a = \alpha$) that

$$1 - \frac{1}{2^{\alpha-1}} < \frac{1}{\zeta(\alpha)}$$

or

$$\left(1 - \frac{1}{2^{\alpha-1}}\right)\sum_{k=1}^{\infty} \frac{1}{k^{\alpha}} < 1$$

Hence we must prove that

$$1 + \frac{1}{2^{\alpha}} + \frac{1}{3^{\alpha}} + \ldots - \frac{2}{2^{\alpha}} - \frac{2}{4^{\alpha}} - \frac{2}{6^{\alpha}} < 1$$

or

$$1 - \frac{1}{2^{\alpha}} + \frac{1}{3^{\alpha}} - \frac{1}{4^{\alpha}} + \frac{1}{5^{\alpha}} - \frac{1}{6^{\alpha}} \ldots < 1 \qquad (18)$$

But the left-hand side of (18) equals

$$1-\left(\frac{1}{2^{\alpha}}-\frac{1}{3^{\alpha}}\right)\left(\frac{1}{4^{\alpha}}-\frac{1}{5^{\alpha}}\right)\ldots$$

which is strictly inferior to 1 since all numbers between brackets are positive. ~

It is easy to see that, if $\alpha = a$ is large enough, we have that $I(\varphi)(n) > f(n)$. Indeed, this relation is satisfied if (use (3), (8) and (13))

$$\left(n-(n+1)\left(\frac{n}{n+1}\right)^{\alpha}\right)\zeta(\alpha) > 1$$

which is always true for $n = 2,3,\ldots$ and α large enough (since $\lim_{\alpha\to\infty}\zeta(\alpha) = 1$). Of course, by definition of $I(\varphi)$, we always have that $I(\varphi)(n) < \varphi(n)$ for every $n \in \mathbb{N}$.

We also have the following proposition, showing the closeness of f and $I(\varphi)$ for large α.

Proposition II.3 : If $\alpha = a$, then

$$\lim_{\alpha\to\infty}\left[f(n)-I(\varphi)(n)\right] = 0 \qquad (19)$$

for every $n \in \mathbb{N}$.

Proof : If $\alpha = a$, then, using (3), (8) and (13)

$$\lim_{\alpha\to\infty}\left[f(n)-I(\varphi)(n)\right] = \lim_{\alpha\to\infty}\left[\frac{T}{n^{\alpha}\zeta(\alpha)} - T\left(\frac{1}{n^{\alpha-1}}-\frac{1}{(n+1)^{\alpha-1}}\right)\right] \qquad (20)$$

since in the above limit we can assume that $\alpha > 2$. If $n = 1$ then (20) equals

$$T\left(\frac{1}{\lim_{\alpha\to\infty}\left(1+\frac{1}{2^{\alpha}}+\frac{1}{3^{\alpha}}+\ldots\right)}-\lim_{\alpha\to\infty}\left(1-\frac{1}{2^{\alpha-1}}\right)\right)=0$$

while, if $n \neq 1$, the same value 0 is trivially obtained. ~

Table 2 illustrates the above proposition for $n = 1$. We use the table of $\zeta^{-1}(\alpha)$ versus α that e.g. can be found in Egghe and Rousseau (1990). Note that the values $\zeta(2)$ and $\zeta(4)$ are explicitely known since

$$\zeta(2)=\sum_{n=1}^{\infty}\frac{1}{n^2}=\frac{\pi^2}{6} \tag{21}$$

and

$$\zeta(4)=\sum_{n=1}^{\infty}\frac{1}{n^4}=\frac{\pi^4}{90} \tag{22}$$

as e.g. can be found in Gradshteyn and Ryzhik (1965) or in Abramowitz and Stegun (1972).

Table 2. Comparison of $I(\varphi)(1)$ and K (both divided by T)

	$1-\frac{1}{2^{\alpha-1}}$	$\zeta^{-1}(\alpha)$
$\alpha = 1.5$	0.2929	0.3828
$\alpha = 2$	0.5	$\frac{6}{\pi^2} = 0.6079$
$\alpha = 2.5$	0.6464	0.7454
$\alpha = 3$	0.75	0.8319
$\alpha = 3.5$	0.8232	0.8875
$\alpha = 4$	0.875	$\frac{90}{\pi^4} = 0.9239$

Fig. 1 illustrates, qualitatively, the relation between f, φ and $I(\varphi)$.

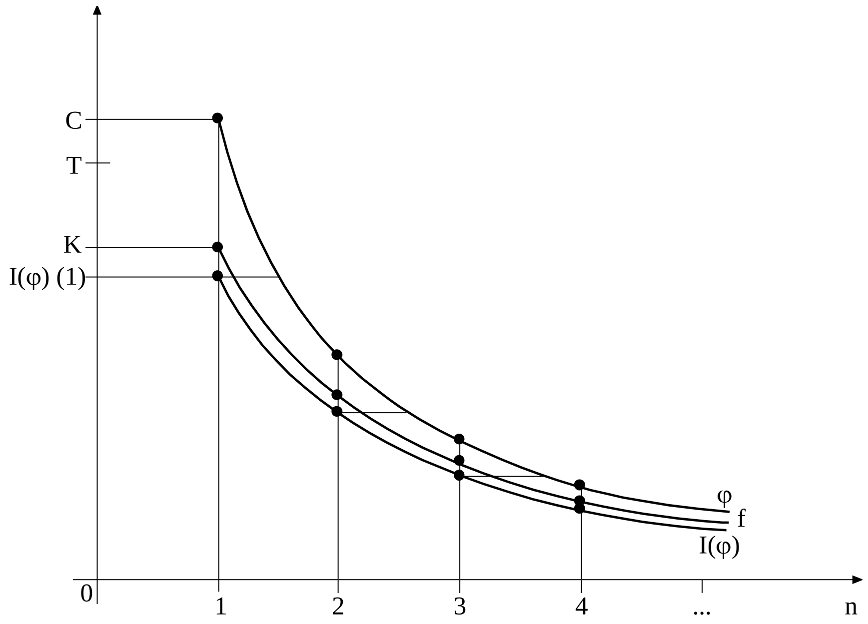

Fig. 1 Qualitative illustration of the relation between f, φ and $I(\varphi)$

Note that, from (17), the order of the function values in 1 is as depicted. For $n = 2, 3, \ldots, I(\varphi)$ can intersect f (if α is large enough) but remains below φ for every $n \in \mathbb{N}$. Note also, however, that $f(n) < \varphi(n)$ for all $n \in \mathbb{N}$. This is seen as follows. Formulae (3) and (8) imply that $f(n) < \varphi(n)$ is equivalent with (for $\alpha > 1$)

$$(\alpha - 1)\zeta(\alpha) > 1 \qquad (23)$$

Putting $\delta = \alpha - 1 > 0$, (23) is equivalent with

$$\sum_{n=1}^{\infty} \frac{\delta}{n^{\delta+1}} > 1 \qquad (24)$$

But, $\forall n \in \mathbb{N}$:

$$\frac{\delta}{n^{\delta+1}} > \int_n^{n+1} \frac{\delta}{j^{\delta+1}} dj$$

as is obvious since $j > n$ on $]n, n+1]$. So

$$\sum_{n=1}^{\infty} \frac{\delta}{n^{\delta+1}} > \sum_{n=1}^{\infty} \int_n^{n+1} \frac{\delta}{j^{\delta+1}} dj$$

$$= \int_1^{\infty} \frac{\delta}{j^{\delta+1}} dj = 1$$

since $\delta > 0$. This proves (24) and hence that

$$f(n) < \varphi(n) \tag{25}$$

for all $n \in \mathbb{N}$.

III. Conclusions

In this paper we discussed the discrete Lotka function f (formula (1)), obtained from practical data, and the continuous Lotka function φ (formula (2)), needed for applying results from Lotkaian informetrics theory (see Egghe (2004a)).

We showed that the discrete Lotka function f and the theoretical Lotka function φ, although they are different power laws, can be calculated from each other in an exact (but numerical) way. We provide the numeric “key” to calculate one from the other.

The simplest and most important result is the fact that, for values of the exponents a and α not too high (which is usually true in practise) we have that $a \approx \alpha$, showing that Lotkaian informetrics theory can be applied using the empirical value of a.

References

M. Abramowitz and I.A. Stegun (1972). Handbook of mathematical functions, with formulas, graphs and mathematical tables. Dover, New York (USA), 1972).

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

L. Egghe (2000). A heuristic study of the first-citation distribution. Scientometrics 48(3), 345-359, 2000.

L. Egghe (2004a). Lotkaian Informetrics. Elsevier, Oxford, to appear, 2004.

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

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

I.S. Gradshteyn and I.M. Ryzhik (1965). Table of Integrals, Series and Products. Academic Press, New York, USA, 1965.

O.V. Groos (1967). Bradford's law and the Keenan-Atherton data. American Documentation 18, 46, 1967.

A.J. Lotka (1926). The frequency distribution of scientific productivity. Journal of the Washington Academy of Science 16(12), 317-324, 1926.

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

Appendix

Values of a versus $\frac{\zeta(a-1)}{\zeta(a)}$

a	ζ(a-1)/ζ(a)
2.11	6.235
2.12	5.773
2.13	5.382
2.14	5.049
2.15	4.763
2.16	4.510
2.17	4.286
2.18	4.088
2.19	3.911
2.20	3.752
2.21	3.607
2.22	3.477
2.23	3.356
2.24	3.248
2.25	3.147
2.26	3.054
2.27	2.969
2.28	2.888
2.29	2.815
2.30	2.745
2.31	2.681
2.32	2.619
2.33	2.563
2.34	2.509
2.35	2.459
2.36	2.411
2.37	2.366
2.38	2.324
2.39	2.283
2.40	2.245
2.41	2.208
2.42	2.174
2.43	2.141
2.44	2.109
2.45	2.079
2.46	2.051
2.47	2.023
2.48	1.997
2.49	1.972
2.50	1.947
2.51	1.924
2.52	1.902
2.53	1.880
2.54	1.860
2.55	1.840
2.56	1.821
2.57	1.802

a	ζ(a-1)/ζ(a)
2.58	1.785
2.59	1.767
2.60	1.751
2.61	1.735
2.62	1.719
2.63	1.705
2.64	1.690
2.65	1.676
2.66	1.662
2.67	1.649
2.68	1.636
2.69	1.624
2.70	1.612
2.71	1.601
2.72	1.589
2.73	1.578
2.74	1.568
2.75	1.557
2.76	1.547
2.77	1.537
2.78	1.528
2.79	1.519
2.80	1.509
2.81	1.500
2.82	1.492
2.83	1.484
2.84	1.475
2.85	1.467
2.86	1.460
2.87	1.452
2.88	1.445
2.89	1.438
2.90	1.431
2.91	1.424
2.92	1.417
2.93	1.410
2.94	1.404
2.95	1.398
2.96	1.392
2.97	1.386
2.98	1.380
2.99	1.374
3.00	1.368
3.01	1.363
3.02	1.358
3.03	1.352
3.04	1.347

a	ζ(a-1)/ζ(a)
3.05	1.342
3.06	1.337
3.07	1.332
3.08	1.328
3.09	1.323
3.10	1.318
3.11	1.314
3.12	1.309
3.13	1.305
3.14	1.301
3.15	1.297
3.16	1.293
3.17	1.289
3.18	1.285
3.19	1.281
3.20	1.278
3.21	1.274
3.22	1.270
3.23	1.267
3.24	1.263
3.25	1.260
3.26	1.256
3.27	1.253
3.28	1.250
3.29	1.247
3.30	1.244
3.31	1.240
3.32	1.237
3.33	1.234
3.34	1.232
3.35	1.229
3.36	1.226
3.37	1.223
3.38	1.220
3.39	1.217
3.40	1.215
3.41	1.212
3.42	1.210
3.43	1.207
3.44	1.205
3.45	1.202
3.46	1.200
3.47	1.197
3.48	1.195
3.49	1.193