Study of rank- and size-frequency functions and their relations in a generalized Naranan framework

by

L. Egghe

Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium[1]
and
Universiteit Antwerpen (UA), IBW, Stadscampus, Venusstraat 35, B-2000 Antwerpen, Belgium
leo.egghe@uhasselt.be

ABSTRACT

The Naranan formalism supposes that the number of sources and the number of items in sources grows exponentially. Here we extend this formalism by assuming, very generally, that the number of sources grows according to a function $\varphi(t)$ and that the number of items in sources grows according to a function $\psi(t)$.We then prove formulae for the rank-frequency function $g(r)$ and the size-frequency function $f(j)$ in terms of the function $\varphi(t)$ and $\psi(t)$. As a special case we obtain Naranan's original result that $f(j)$ is the law of Lotka if φ and ψ are exponential functions.

---

[1] Permanent address

Key words and phrases: Naranan, rank-frequency function, size-frequency function, growth.

We also prove relations between the rank- and frequency functions of two systems where the second system is built on the same functions φ and ψ as the first system but in reverse order. Results if $\varphi = \psi$ follow from this as a consequence.

I. Introduction

General information production processes (IPPs) are systems where one has sources (e.g. journals, authors, …) and where these sources contain (or have or produce) items (e.g. articles produced by authors or published in journals). The classical way to describe such systems is by studying their rank- and size-frequency functions. They are defined as follows (cf. Egghe (2005a), Egghe and Waltmann (2011)) in a continuous framework. The size-frequency function $f(j)$ ($j>0$) is the density of sources with j items, i.e. for every $m,n>0$, $m<n$

$$\int_m^n f(j)dj \tag{1}$$

denotes the number of sources with a number of items between m and n . The corresponding rank-frequency function $g(r)$ is defined as

$$r=\int_j^\infty f(j')dj'=g^{-1}(j) \tag{2}$$

where $j=g(r)$, where g^{-1} denotes the inverse of the injective function g (injective because g^{-1} strictly decreases, by (2)). So $g(r)$ denotes the number of items in the source on rank r (where sources are arranged in decreasing order of their number of items). By (2), g follows from f but f also follows from g since (2) implies (supposing that g is differentiable)

$$f(j)=-\frac{1}{g'\left(g^{-1}(j)\right)} \tag{3}$$

General rank- and size frequency functions are studied in Egghe and Waltman (2011) where one proves the following results.

(i) f decreases if and only if g is convex

(ii) f is first increasing and then decreasing if and only if g has an S-shape: first convex and then concave.

See Egghe and Waltman (2011) and Egghe (2011a) (and references therein) for examples of both cases. In the same way one can prove characterizations of concave g or reversely S-shaped g (first concave and then convex) (see also Egghe (2011b).

A dynamic theory of IPPs (or, simply, systems) can be developed if we assume that the number of sources grow in time t and that the number of items in these sources also grow in time t. Naranan, in his historic paper in Nature (Naranan (1970)) studied this for the case of exponential growth of sources and of exponential growth of items in sources. The exact result is as in the next Theorem.

Theorem 1 (Naranan)

(i) Let the number of sources $\varphi(t)$ grow exponentially in time t:

$$\varphi(t) = c_1 a_1^t \tag{4}$$

$c_1 > 0$, $a_1 > 1$.

(ii) Let the number of items $\psi(t)$ in each source grow exponentially in time t and this function is the same in each source:

$$\psi(t) = c_2 a_2^t \tag{5}$$

$c_2 > 0$, $a_2 > 1$.

Then this system has a size-frequency function $f(j)$ which is a power law (i.e. Lotka's law, see Egghe (2005a))

$$f(j) = \frac{C}{j^\alpha} \tag{6}$$

$C > 0$, $\alpha > 1$, $j \geq 1$ and Lotka's exponent α equals

$$\alpha = 1 + \frac{\ln a_1}{\ln a_2} \tag{7}$$

Based on this, such systems can be considered as a self-similar fractal with fractal dimension $\alpha - 1$ (see Egghe (2005a,b)).

It is well-known that Lotka's law is equivalent with Zipf's law (as rank-frequency function):

$$g(r) = \frac{B}{r^\beta} \tag{8}$$

$B, \beta > 0$, $0 < r \leq T$ (the total number of sources) and where

$$\beta = \frac{1}{\alpha - 1} \tag{9}$$

The proof is based on relations (2), (3) and can be found in Egghe (2005a), Exercise II.2.2.6 or in Egghe and Rousseau (2006) where a proof is given in the Appendix.

Based on this equivalence a second, shorter, proof of Naranan's theorem was given in Egghe (2010) by showing that from Naranan's assumptions it follows that (8), hence (6) is valid and where also (7) could be reproved.

In Egghe (2011b), a variant of Naranan's theorem is studied by replacing the exponential functions (4) and (5) by power functions. There we show that the rank-frequency function g can have the shapes: convex, concave, S-shaped (first convex, then concave), or reverse S-shaped (first concave, then convex). Note that in Naranan's case the function g can only be convex as follows from (8).

In this paper we replace the exponential functions (4) and (5) by general functions $\varphi(t)$ and $\psi(t)$, which are differentiable. In the next section we will give explicite formulae for the rank-frequency function g and the size-frequency function f in terms of general growth functions $\varphi(t)$ and $\psi(t)$. For all results we will give examples by using Naranan's exponential functions.

In the third section we will consider two systems: one as above with growth functions $\varphi(t)$ and $\psi(t)$ and a second one with $\varphi(t)$ and $\psi(t)$ interchanged. We prove relations between the rank- and size-frequency functions of both systems and we also prove relations between these functions if $\varphi(t)=\psi(t)$ as a special case.

The paper ends with a conclusion section where we also give suggestions for further research.

Theory of the generalized Naranan framework

The Naranan framework (4), (5) in Theorem 1 is generalized as follows.

Generalized Naranan framework

(i) Let the number of sources grow according to the differentiable injective function $\varphi(t)$.

(ii) Let the number of items in each source grow according to the differentiable infective function $\psi(t)$ which is the same for every source.

Theorem 2

For all t, the rank-frequency function $g(r)$ of this system is given by

$$g(r) = \psi\left(t - \varphi^{-1}(r)\right) \tag{10}$$

where φ^{-1} denotes the inverse of the injective function φ.

Proof

Let t be fixed and $\theta \in [0,1]$. By definition of the generalized Naranan framework we have that sources that are "born" at θt, have at time t, a period equal to $t - \theta t$ to "grow" items. Hence, ranking sources in decreasing order of their number of items, we have, by definition of the rank-frequency function $g(r)$:

$$r = \varphi(\theta t) \tag{11}$$

and

$$g(r) = g\left(\varphi(\theta t)\right) = \psi(t - \theta t) \tag{12}$$

Formulae (11) implies

$$\theta = \frac{1}{t}\varphi^{-1}(r) \tag{13}$$

, so (13) in (12) yields

$$g(r) = \psi\left(t - \varphi^{-1}(r)\right)$$

, proving (10). □

We have the following Lemma, in preparation of the calculation of the size-frequency function $f(j)$ (cf. formula (3)).

Lemma 3

$$g'(r) = -\frac{\psi'\left(t-\varphi^{-1}(r)\right)}{\varphi'\left(\varphi^{-1}(r)\right)} \tag{14}$$

and

$$r = g^{-1}(j) = \varphi\left(t-\psi^{-1}(j)\right) \tag{15}$$

Proof

Equation (14) follows from formula (10) by applying the well-known rules of differentiation. Also by formula (10),

$$j = g(r) = \psi\left(t-\varphi^{-1}(r)\right)$$

from which $r = g^{-1}(j)$ as in (15) readily follows. □

Theorem 4

For all t, the size-frequency function $f(j)$ of this system is given by

$$f(j) = \frac{\varphi'\left(t-\psi^{-1}(j)\right)}{\psi'\left(\psi^{-1}(j)\right)} \tag{16}$$

Proof

By formula (3):

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

$$= \frac{\varphi'\left(\varphi^{-1}\left(g^{-1}(j)\right)\right)}{\psi'\left(t-\varphi^{-1}\left(g^{-1}(j)\right)\right)}$$

(by (14))

$$= \frac{\varphi'\left(\varphi^{-1}\left(\varphi\left(t-\psi^{-1}(j)\right)\right)\right)}{\psi'\left(t-\varphi^{-1}\left(\varphi\left(t-\psi^{-1}(j)\right)\right)\right)}$$

by (15). Hence

$$f(j) = \frac{\varphi'\left(t - \psi^{-1}(j)\right)}{\psi'\left(\psi^{-1}(j)\right)}$$

So (16) is proved. □

Example

For the functions (4) and (5) of Naranan, we have

$$r = c_1 a_1^{\theta t}$$

so

$$\theta = \frac{\ln\left(\frac{r}{c_1}\right)}{t \ln a_1} \qquad (17)$$

So (12) gives, by (17)

$$g(r) = \frac{c_2 a_2^t}{a_2^{\frac{\ln\left(\frac{r}{c_1}\right)}{\ln a_1}}}$$

$$g(r) = \frac{B}{r^\beta} \qquad (18)$$

where

$$\beta = \frac{\ln a_2}{\ln a_1} \qquad (19)$$

and

$$B = c_2 a_2^t \left(c_1^{\frac{\ln a_2}{\ln a_1}} \right) \qquad (20)$$

after some elementary calculation.

This is (8), hence, by (9) and (19) we have (6) and (7), hence Naranan's result, which can also be calculated from (16): calculating φ', ψ' and ψ^{-1} (from (4) and (5)) yields

$$f(j) = \frac{C}{j^{1+\frac{\ln a_1}{\ln a_2}}}$$

which is (6) and (7). Here

$$C = \frac{c_1 a_1^t \ln a_1}{\ln a_2} c_2^{\frac{\ln a_1}{\ln a_2}} \qquad (21)$$

for later use.

These are the basic models with which we can work further on in the sequel. In the next section we will compare this generalized Naranan framework with the one where the growth functions $\varphi(t)$ and $\psi(t)$ are interchanged.

Comparison of two systems

Let us first have a generalized Naranan system as described in the previous section: number of sources grow according to $\varphi(t)$ and number of items in sources grow according to $\psi(t)$. Let us have a second system where the functions $\varphi(t)$ and $\psi(t)$ are interchanged. Then we have the following Theorem.

Theorem 5

If we denote by f_i , g_i ($i = 1, 2$) the size-frequency functions, respectively the rank-frequency functions of system 1 and 2, with variables $x = j_1 = r_2$ and $y = r_1 = j_2$, then we have

$$g_2(x) = g_1^{-1}(x) \quad (22)$$

$$g_2^{-1}(y) = g_1(y) \quad (23)$$

$$g'_2(x) = -f_1(x) \quad (24)$$

$$g'_1(y) = -f_2(y) \quad (25)$$

Proof

Formula (23) is equivalent with (22) and they follow from (10) and (15). Formulae (24) and (25) follow from (14) and (16). That $x = j_1 = r_2$ and $y = r_1 = j_2$ follows from the interchanging of φ and ψ: by definition of the generalized Naranan frameworks:

$$r_1 = \varphi(t) = j_2 =: y$$

and

$$j_1 = \psi(t) = r_2 =: x \qquad \square$$

Example

In case we have functions (4) and (5), we have, for the first system

$$g_1(r_1) = \frac{B_1}{r_1^{\beta_1}} \tag{26}$$

with B_1 as in (20) and β_1 as in (19) and

$$f_1(j_1) = \frac{C_1}{j_1^{\alpha_1}} \tag{27}$$

with C_1 as in (21) and α_1 as in (7). For the second system we have

$$g_2(r_2) = \frac{B_2}{r_2^{\beta_2}} \tag{28}$$

with B_2 as in (20) but with the indices reversed:

$$B_2 = c_1 a_1^t \left(c_2^{\frac{\ln a_1}{\ln a_2}} \right) \tag{29}$$

and β_2 as in (19) but with the indices reversed:

$$\beta_2 = \frac{\ln a_1}{\ln a_2} \tag{30}$$

and

$$f_2(j_2) = \frac{C_2}{j_2^{\alpha_2}} \tag{31}$$

with C_2 as in (21) but with the indices reversed:

$$C_2 = \frac{c_2 a_2^t \ln a_2}{\ln a_1} c_1^{\frac{\ln a_2}{\ln a_1}} \tag{32}$$

and α_2 as in (7) but with the indices reversed:

$$\alpha_2 = 1 + \frac{\ln a_2}{\ln a_1} \tag{33}$$

It is now an elementary calculation to prove formulae (22) – (25) in this example, based on the results (26) – (33).

The case $\varphi = \psi$

Let us now suppose that the growth of the number of sources is equal to the growth of the number of items in the sources (but $\varphi = \psi$ remains a general injective diffentiable function). Then both systems in Theorem 5 are the same. Hence we have the result as in the next Theorem.

Theorem 6

In a system in the generalized Naranan framework, where both growth functions are equal, we have that

$$f(x) = -g'(x) \tag{34}$$

and

$$g(x) = g^{-1}(x) \tag{35}$$

Proof

This follows readily by Theorem 5 since $g = g_1 = g_2$ and $f = f_1 = f_2$. Note that here $x =: j = r$ since $\varphi = \psi$. □

Example

In the case of (4) and (5) we have that $\varphi = \psi$ implies $a_1 = a_2$. By (7), we have that $\alpha = 2$. Hence, by (6) and (8)

$$f(j) = \frac{C}{j^2} \tag{36}$$

and

$$g(r) = \frac{B}{r} \tag{37}$$

It is now readily verified that $f(x) = -g'(x)$ and $g^{-1}(x) = g(x)$ ($x = j = r$).

We close this section by the following Theorem.

Theorem 7

Under the conditions of Theorem 6 we have

$$f(x)\, f\left(g^{-1}(x)\right) = 1 \tag{38}$$

Proof

This follows readily by (34) and (3). □

Example

In the case of (4) and (5) we have (36) and (37). These formulae yield (noting that $g^{-1}(x) = g(x)$ here) formulae (38).

Some further comparisons of two systems

It is clear from (10) and (16) that multiplying ψ with a constant yields a rank-frequency function being the first one multiplied by this constant. Also: multiplying φ with a constant yields a size-frequency function being the first one multiplied by this constant. Taking a power of the function ψ yields a rank-frequency function being the first one to this power. The latter is not true for the size-frequency function when we take a power of the function φ.

Conclusions and suggestions for further research

The Naranan formalism describes the growth of the number of sources and the growth of the number of items in these sources by exponential functions. In this paper these exponential functions are replaced by general functions $\varphi(t)$ and $\psi(t)$. Explicite formulae for the rank-frequency function $g(r)$ and size-frequency function $f(j)$ are presented.

These systems are compared with similar systems but where the growth functions $\varphi(t)$ and $\psi(t)$ are interchanged. Relations between the rank- and size-frequency function of both systems are proved. As a consequence, results are obtained in case both growth functions are the same: $\varphi = \psi$.

Further research on such “generalized Naranan” systems is needed, both in the very general case $\varphi(t)$ and $\psi(t)$ and in the case of special functions φ and ψ. Which functions can be taken should follow from practical experiments on growth properties of sources and items.

Can Naranan’s framework be further generalized to the situation where the number of items in sources do not grow in the same way in different sources? Are there other ways to describe the dynamics of IPPs apart from already existing models of IPPs dependent on time t (see e.g. Egghe (2005a), Chapter 1) ?

Can the general rank-frequency function

$$g(r) = K\frac{(N+1-r)^b}{r^a} \tag{39}$$

be generated from a generalized Naranan framework? This function appears in Mansilla et al. (2009) and Campanario (2009) and is capable to describe convexly decreasing rank-frequency functions ($b \geq 1$ and $b = 0$, corresponding to Zipf's law) as well as S-shaped rank-frequency functions ($0 < b < 1$) – see Egghe and Waltman (2011).

More generally, can "any" given rank- (or size-) frequency function be generated from a generalized Naranan framework ?

As suggested by one of the referees it is a nice challenge to investigate how this general framework can be applied in ecological or biological studies.

References

J.M. Campanario (2009). Distribution of changes in impact factors over time. Scientometrics 84(1) 35-42.

L. Egghe (2005a). Power laws in the Information Production Process: Lotkaian Informetrics. Elsevier, Oxford, UK.

L. Egghe (2005b). The power of power laws and the interpretation of Lotkaian informetric systems as self-similar fractals. Journal of the American Society for Information Science and Technology 56(7), 669-675.

L. Egghe (2010). A new short proof of Naranan's theorem, explaining Lotka's law and Zipf's law. Journal of the American Society for Information Science and Technology 61(12), 2581-2583.

L. Egghe (2011a). The impact factor rank-order distribution revisited. Scientometrics 87(3), 683-685.

L. Egghe (2011b). Study of the rank- and size-frequency functions in case of power law growth of sources and items and proof of Heaps's law. Preprint.

L. Egghe and R. Rousseau (2006). An informetric model for the Hirsch-index. Scientometrics 69(1), 121-129.

L. Egghe and L. Waltman (2011). Relations between the shape of a size-frequency distribution and the shape of a rank-frequency distribution. Information Processing and Management, to appear.

R. Mansilla, E. Köppen, G. Cocho and P. Miramontes (2007). On the behavior of journal impact factor rank-order distribution. Journal of Informetrics 1(2), 155-160.

G. Martinez-Mekler, R. Alvarez Martinez, M. Beltrán del Rio, R. Mansilla, P. Miramontes and G. Cocho (2009). Universality of rank-order distributions in the arts and sciences. PLos ONE 4(3), e4791.

S. Naranan (1970). Bradford’s law of bibliography of science: an interpretation. Nature 227, 631-632.