

Mathematical derivation of the scale-dependence of the h-index and other h-type indices

Peer-reviewed author version

EGGHE, Leo (2011) Mathematical derivation of the scale-dependence of the h-index and other h-type indices. In: SCIENTOMETRICS, 87(2). p. 287-292.

DOI: 10.1007/s11192-010-0327-0

Handle: <http://hdl.handle.net/1942/11522>

# Mathematical derivation of the scale-dependence of the h-index and other h-type indices

by

L. Egghe

Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek,  
Belgium (\*)

and

Universiteit Antwerpen (UA), Stadscampus, Venusstraat 35, B-2000 Antwerpen

[leo.egghe@uhasselt.be](mailto:leo.egghe@uhasselt.be)

---

## **ABSTRACT**

We present a mathematical derivation of the scale-dependence of the h-index. This formula can be used in two cases: one where the units are scale-dependent and one where the units are not scale-dependent. Examples are given.

---

(\*) Permanent address

Key words and phrases: h-index, Hirsch- index, scale, h-type index

Acknowledgement: The author is grateful to two anonymous referees who gave many comments to improve the paper.

## Introduction

The Hirsh-index or  $h$ -index is defined for a decreasing sequence of positive numbers. In the original article Hirsch (2005) one has a decreasing sequence of citations to articles (e.g. of a researcher). As such, the units are clear: a unit is a citation. So, a scale variation (say multiplication of these units by a number  $\gamma > 0$ ) does not apply in practise although theoretically one can wonder what the new  $h$ -index will be if we multiply all these numbers by a fixed number  $\gamma > 0$ . Similar variants have been discussed in van Eck and Waltman (2008) and in Wu (2010) (see also Egghe (2011)).

However, this problem is not theoretical anymore if we want to apply the  $h$ -index to other units such as time or money (and many other features). Let us give an example. It would be nice to apply the  $h$ -index in econometrics, e.g. to measure poverty or wealth of a group of persons (cf. the discussion in Egghe (2010a), p.80). So in this case one has a decreasing sequence of numbers, representing amounts of money. Again, one can multiply all these amounts by a number  $\gamma > 0$ , meaning that all persons become richer (in case  $\gamma > 1$ ). Again one can wonder what the new  $h$ -index would be. But in this example one can interpret scale variability in a totally different way. Multiplying all these amounts by a number  $\gamma > 0$  can simply mean that we change the currency (e.g. from \$ to € or vice-versa). In this case we would prefer the  $h$ -index to be the same but this will not be the case. So, a formula for the  $h$ -index in function of  $\gamma > 0$  would help us to compare decreasing sequences of amounts of money expressed in different currencies. A concrete example is given by one of the referees. There are about 8 million millionaires in the US. Hence the  $h$ -index will be significantly smaller than 8 million. But expressed in Hong Kong dollars (1 US \$ = 8 HK \$) then there are about 8 million people in the US with more than 8 million Hong Kong dollars each, so the  $h$ -index in that currency is 8 million.

The same remarks can be made for time: do we express time in months or in years (cf. the discussion in Egghe, Bornmann and Guns (2010), see also Bornmann and Daniel (2010))? Also here there are two interpretations. If we keep the same time-units, multiplication by a number  $\gamma > 0$  means longer ( $\gamma > 1$ ) or shorter ( $0 < \gamma < 1$ ) time-periods. But if we do not keep

the same time-units, multiplication by a number  $\gamma > 0$  means e.g. a change of time-units (e.g. from years to months ( $\gamma = 12$  here)).

All these examples show that, for the  $h$ -index to be useable in connection with variables for which scales can be changed, we need a formula for the  $h$ -index in function of  $\gamma > 0$ : or we have the new value of  $h$ , or we have a scaling measure to compare two situations with different units (e.g. \$ vs. €).

This will be done in this paper using the Lotkaian framework. To be able to do this we will repeat a few well-known results from Lotkaian informetrics (Egghe (2005)). Consider an IPP (Information Production Process) consisting of sources and items (Egghe (2005)). This framework includes a wide variety of source-item applications such as author-article relations, journal-article relations or article-citation relations. Let  $f(j)$  be the size-frequency function of this system:  $f(j)$  is the number of sources with  $j$  items. If sources are ranked in decreasing order of their number of items, then  $g(r)$  is the rank-frequency function of the system:  $g(r)$  is the number of items in the source on rank  $r$ . In the sequel we will use a continuous approach since this is mathematically easier (Egghe (2005)). Let

$$f(j) = \frac{C}{j^\alpha} \quad (1)$$

$C > 0$ ,  $\alpha > 1$ ,  $j \geq 1$  be Lotka's law ( $f(j)$  is the density of sources with item density  $j$  - sources can, e.g., be articles and items can be citations to these articles as is the case in the classical Hirsch-index). The applicability of (1) to the article-citation framework is discussed in Egghe (2005), Chapter I, p.90 where the literature reports that the Lotkaian exponent  $\alpha$  varies between 1.3 and 3.

Let

$$g(r) = \frac{B}{r^\beta} \quad (2)$$

$B > 0$ ,  $\beta > 0$ ,  $r \in ]0, T]$  be Zipf's law ( $g(r)$  is the item density in the source on rank density  $r$  and where  $T$  denotes the total number of sources). Then the following Theorem is valid.

**Theorem 1:** The following assertions are equivalent

$$(i) f(j) = \frac{C}{j^\alpha}, C > 0, \alpha > 1, j \geq 1$$

$$(ii) g(r) = \frac{B}{r^\beta}, B > 0, \beta > 0, r \in ]0, T]$$

Moreover, the relations between the parameters are

$$B = \left( \frac{C}{\alpha - 1} \right)^{\frac{1}{\alpha - 1}} = T^{\frac{1}{\alpha - 1}} \quad (3)$$

$$\beta = \frac{1}{\alpha - 1} \quad (4)$$

The result can be found in Egghe (2005), Exercise II.2.26 or Egghe and Rousseau (2006) where also a proof is given in the Appendix.

Further in Egghe and Rousseau (2006), the following Theorem is proved.

**Theorem 2:** If we have Lotka's law with  $\alpha > 1$ . Then the  $h$ -index of this system is given by

$$h = T^{\frac{1}{\alpha}} \quad (5)$$

With this machinery we are able to present and prove a formula of the  $h$ -index where all items are multiplied by  $a > 0$  in function of the old  $h$ -index and, of course, in function of  $a$ .

## **The $h$ -index for scaled data**

Let us have a system as in Theorem 1 and hence the  $h$ -index  $h$  satisfies (5). Let us multiply all item values by a fixed number  $\gamma > 0$ . Denote by  $h(\gamma)$  the  $h$ -index of this system. Then we have the following result.

Theorem 3:

$$h(\gamma) = \gamma^{\frac{\alpha-1}{\alpha}} T^{\frac{1}{\alpha}} \quad (6)$$

$$h(\gamma) = \gamma^{\frac{\alpha-1}{\alpha}} h \quad (7)$$

Proof: (7) follows from (6) by (5). Since  $g(r)$  denotes the item densities on rank  $r$  in the original system, multiplying all item values by a number  $\gamma > 0$  means that, in the new system, we have the following law of Zipf:

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

The defining relation for the  $h$ -index  $h(\gamma)$  is:  $r = h(\gamma)$  if and only if

$$\frac{\gamma B}{r^\beta} = r$$

, so

$$h(\gamma) = r = (\gamma B)^{\frac{1}{\beta+1}} \quad (8)$$

But

$$\gamma B = \gamma T^{\frac{1}{\alpha-1}} \quad (9)$$

by (3) and

$$\frac{1}{\beta+1} = \frac{\alpha-1}{\alpha}$$

by (4). Hence

$$h(\gamma) = \gamma^{\frac{1}{\beta+1}} T^{\frac{1}{\alpha-1} \frac{1}{\beta+1}}$$

$$h(\gamma) = \gamma^{\frac{\alpha-1}{\alpha}} T^{\frac{1}{\alpha}} \quad (10)$$

again by (4).  $\square$

Corollary 4: If  $\alpha = 2$  (the “classical” Lotka exponent), then we have that

$$h(\gamma) = \sqrt{\gamma} h \quad (11)$$

for all  $\gamma > 0$ .

The general shape of the relations (7) is a concavely increasing function of  $\gamma$ , for all  $\gamma > 1$ .

Hence,  $h(\gamma)$  increases in  $\gamma$  but more slowly than multiplication with  $\gamma$  (which is the multiplier of the item values). Formula (7) was already proved implicitly in Egghe (2008) but there a longer intricate proof was given.

Example: Let us take Lotka’s law with  $\alpha = 2$ . Hence, by Theorem 1, we have Zipf’s law with  $\beta = 1$ :

$$g(r) = \frac{B}{r} \quad (12)$$

Take  $B = 100$ . Hence we have a decreasing sequence as in Table 1 (only the needed ranks are presented and the data are rounded-off to the nearest entire number)

Table 1. Data derived from (12)

$r$	$g(r)$
1	100
2	50
3	33
4	25
5	20
6	17
7	14
8	13
9	11
10	10
11	9
12	8
13	8
14	7
15	7

Hence  $h = 10$  in this case.

Take  $\gamma = 2$ . Then we have data as in Table 2.

Table 2. The  $g(r)$ -values in Table 1, multiplied by 2

$r$	$2g(r)$
1	200
2	100
3	66
4	50
5	40
6	34
7	28
8	26
9	22
10	20
11	18
12	16
13	16
14	14
15	14

Now  $h(2) = 14 \approx \sqrt{2} h$  since  $h = 10$ . So (11) (i.e. (7) for  $\alpha = 2$ ) is verified. In case the data in Table 2 are really increased data (i.e. when we keep the same item-units), then  $h(2)$  is the real  $h$ -index of this system. If Table 2 is only the result of a change of units, then data in the original units (as in Table 1) can (as far as the  $h$ -index is concerned) only be compared with data in the new units (as in Table 2) provided we apply the correction factor  $\sqrt{2}$ .

Remark: Theorem 3 also applies to many other  $h$ -type indices such as the  $g$ -index (Egghe (2006)), the R-index (Jin, Liang, Rousseau and Egghe (2007)), the (generalized) Wu-index (Egghe (2011)) and Randić's H-index (Egghe (2010b)). In these references and in the Lotkaian framework it is proved that they are all a fixed multiple of the  $h$ -index (if Lotka's  $\alpha$  is fixed): denote by  $M$  any such index, then

$$M = \varphi(\alpha)h \quad (13)$$

where  $\varphi$  is a function of  $\alpha$ .



Denoting by  $M(\gamma)$  any such index of the data multiplied by  $\gamma > 0$ , then (7) and (13) imply

$$\begin{aligned} M(\gamma) &= \varphi(\alpha)h(\gamma) \\ &= \varphi(\alpha)\gamma^{\frac{\alpha-1}{\alpha}}h \\ M(\gamma) &= \gamma^{\frac{\alpha-1}{\alpha}}M \end{aligned} \tag{14}$$

, hence showing that Theorem 3 is also valid for all these indices.

## **Discussion**

The mathematical dependence of the h-index on scaling has been proved (formula (7)). It could be possible to use this result to solve the problem of field-dependence of the h-index by standardizing the citations.

As remarked by both referees, the results are limited to the validity of Lotka's (or Zipf's) Law. Empirical citation distributions might follow only approximately such distributions. It can then be expected that the here obtained results are approximately true. In addition to this, sources with zero-items (e.g. uncited articles) are not included in the Lotkaian framework. It must however be remarked that such sources do not contribute to the h-index (by definition).

## References

- L. Bornmann and H.-D. Daniel (2010). The citation speed index: a useful bibliometric indicator to add to the h index. *Journal of Informetrics*, 4(3), 444-446.
- L. Egghe (2005). *Power Laws in the Information Production Process: Lotkaian Informetrics*. Elsevier, Oxford, UK.
- L. Egghe (2006). Theory and practise of the g-index. *Scientometrics* 69(1), 131-152.
- L. Egghe (2008). Examples of simple transformations of the h-index: qualitative and quantitative conclusions and consequences for other indices. *Journal of Informetrics* 2(2), 136-148.
- L. Egghe (2010a). The Hirsch index and related impact measures. *Annual Review of Information Science and Technology*, Volume 44 (B. Cronin, ed.), 65-114, Information Today, Inc., Medford, New Jersey, USA.
- L. Egghe (2010b). Mathematical results on the H-index and H-sequence of Randić, *Research Evaluation*, to appear.
- L. Egghe (2011). Characterizations of the generalized Wu- and Kosmulski-indices in Lotkaian systems, to appear.
- L. Egghe, L. Bornmann and R. Guns (2010). A proposal for a First-Citation-Speed-Index, to appear.
- L. Egghe and R. Rousseau (2006). An informetric model for the Hirsch-index. *Scientometrics* 69(1), 121-129.
- J. Hirsch (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America* 102(46), 16569-16572.
- B. Jin, L. Liang, R. Rousseau and L. Egghe (2007). The R- and AR- indices: Complementing the h-index. *Chinese Science Bulletin* 52(6), 855-863.
- N.J. van Eck and L. Waltman (2008). Generalizing the h- and g-indices. *Journal of Informetrics* 2(4), 263-271.
- Q. Wu (2010). The w-index: A measure to assess scientific impact by focussing on widely cited papers. *Journal of the American Society for Information Science and Technology* 61(3), 609-614.