Research Paper

Equalities between h-type Indices and Definitions of Rational h-type Indicators

Leo Egghe[1], Yves Fassin[2], Ronald Rousseau[3,4†]

[1]University of Hasselt, Belgium
[2]Department of Marketing, Innovation and Organisation, Faculty of Economics and Business Administration, Ghent University, Tweekerkenstraat, 2, 9000 Gent, Belgium
[3]University of Antwerp, Faculty of Social Sciences, Middelheimlaan 1, 2020, Antwerpen, Belgium
[4]KU Leuven, Facultair Onderzoekscentrum ECOOM, Naamsestraat 61, 3000 Leuven, Belgium

Citation: Leo Egghe, Yves Fassin, Ronald Rousseau (2019). Equalities between h-type indices and definitions of rational h-type indicators.
Vol. 4 No. 1, 2019
pp 22–31
DOI: 10.2478/jdis-2019-0002
Received: Nov. 24, 2018
Revised: Dec. 8, 2018
Accepted: Dec. 11, 2018

Abstract

Purpose: To show for which publication-citation arrays h-type indices are equal and to reconsider rational h-type indices. Results for these research questions fill some gaps in existing basic knowledge about h-type indices.

Design/methodology/approach: The results and introduction of new indicators are based on well-known definitions.

Findings: The research purpose has been reached: answers to the first questions are obtained and new indicators are defined.

Research limitations: h-type indices do not meet the Bouyssou-Marchant independence requirement.

Practical implications: On the one hand, more insight has been obtained for well-known indices such as the h- and the g-index and on the other hand, simple extensions of existing indicators have been added to the bibliometric toolbox. Relative rational h-type indices are more useful for individuals than the existing absolute ones.

Originality/value: Answers to basic questions such as "when are the values of two h-type indices equal" are provided. A new rational h-index is introduced.

Keywords h-index; g-index; Rational h-type indices; Relative rational h-index; Lotkaian framework

† Corresponding author: Ronald Rousseau (E-mail: ronald.rousseau@kuleuven.be).

1 Introduction

Definition: The classical h-index

Consider a set S of publications, ranked decreasingly according to the number of citations each of these publications has received. Publications with the same number of citations are given different rankings. Then the h-index of set S is h if the first h publications received each at least h citations, while the publication ranked h+1 received strictly less than h+1 citations. Stated otherwise: the h-index of set S is the largest natural number h such that the first h publications received at least h citations (Hirsch, 2005).

When applied to the publication list of a researcher the previous definition favors more prolific, e.g. older, scientists above those with less publications, e.g. younger ones. For this reason one may use a publication window in calculating an h-index. Also the citation window can be adapted to make a difference between short-term and long – term influence. As databases differ in content an h-index may also differ according to the used database. Besides these adaptations of the original definition it is also possible to calculate h-indices for other types of citations, e.g. of patents and for fractionally counted items.

Definition: the g-index

Additional citations to publications among the first h play no role at all. For this reason another indicator has been introduced. This is the g-index, proposed by Egghe (2006a). It is defined as follows: articles are ranked in decreasing order of received citations (as for the h-index). Then the g-index of this set of articles is defined as the highest rank g such that these g articles together received at least g^2 citations. If necessary, fictitious articles with zero citations are added to the publication list.

Definition: Kosmulski's index and its generalizations

Another variation on the h-index was introduced by Kosmulski (2006). He proposed the $h^{(2)}$-index as follows. Again one ranks the set of articles for which one wants to determine the $h^{(2)}$-index in decreasing order of received citations. Now this set (authors, journals, etc.) has an $h^{(2)}$-index equal to h_2 if $r = h_2$ is the highest rank such that the first h_2 articles each received at least $(h_2)^2$ citations.

As a next step, colleagues observed that one may define in a similar manner an $h^{(k)}$-index (k = 1, 2, 3, ….). This has been done e.g. in (Deineko & Woeginger, 2009), who proposed an axiomatic characterization of an even more general family of indices and in (Egghe, 2011), who studied this index in a Lotkaian framework.

Concretely the $h^{(3)}$ index is defined as follows. Consider a list of articles ranked decreasingly according to the number of citations each of these articles has received.

Articles with the same number of citations are given different rankings. Then the $h^{(3)}$-index of this set S is h_3 if the first h_3 articles received each at least $(h_3)^3$ citations, while the article ranked h_3+1 received strictly less than $(h_3+1)^3$ citations. Stated otherwise: the $h^{(3)}$-index of a set S is the largest natural number h_3 such that the first h_3 publications each received at least $(h_3)^3$ citations (Fassin & Rousseau, 2018).

In this contribution we represent the units of attention (authors, journals, research groups, etc.) as a finite array such as A = (10, 7, 7, 2, 0). This symbol shows that author A has five publications with respective (ranked) citations equal to 10, 7, 7, 2 and 0. Clearly author A has an h-index equal to 3 and a g-index equal to 5. The $h^{(2)}$-index is equal to 2 and the $h^{(3)}$-index is equal to 1. The number of items with a non-zero number of citations is called the length of the array. This array has length 4. For simplicity we will always assume that values in array A are natural numbers (including the value zero). In this contribution we restrict our attention to the g, h, $h^{(2)}$ and the $h^{(3)}$ index, to which we refer as h-type indices.

2 When do we have equality?

It follows from their definitions that always $g \geq h \geq h^{(2)} \geq h^{(3)}$. In this section we tackle the question: for which arrays are two different h-type indices equal?

A. When is $h = h^{(2)}$?

We recall the two conditions: a set of articles has h-index h if the first h articles received at least h citations and the article ranked h+1 received strictly less than h+1 citations. Similarly: a set of articles has $h^{(2)}$-index h_2 if the first h_2 articles received at least $(h_2)^2$ citations each and the article ranked h_2+1 received strictly less than $(h_2+1)^2$ citations. If the two conditions must be satisfied at the same time, then the first h articles must have received at least h^2 citations each and the article ranked h+1 must have strictly less than h+1 citations. Obviously, $h = h^{(2)}$ can only occur for an array of length at least equal to h.

The following array A = (100, 30, 9, 3) is an example for which $h = h^{(2)} = 3$.

The least number of citations for the case $h = h^{(2)} = 3$, occurs for the array (9, 9, 9, 0). We added a non-essential zero at the end to make it clear that the length of this array is three. Generally, the least number of citations for the case $h = h^{(2)}$ is $\left(\underbrace{h^2, h^2, \ldots, h^2}_{h\ times}, 0\right)$. Of course, there is no upper limit to the corresponding number of citations.

B. When is $h = h^{(3)}$ or equivalently, when is $h = h^{(2)} = h^{(3)}$?

As, by definition $h^{(2)}$ is always situated between h and $h^{(3)}$, it suffices to solve the problem: when is $h = h^{(3)}$?

Again we recall the two conditions: a set of articles has h-index h if the first h articles received at least h citations and the article ranked h+1 received strictly less than h+1 citations. Similarly: a set of articles has $h^{(3)}$-index h_3 if the first h_3 (here equal to h) articles received at least $(h_3)^3$ citations each and the article ranked h_3+1 received strictly less than $(h_3+1)^3$ citations. If the two conditions must be satisfied at the same time, then the first h articles must have received at least h^3 citations each and the article ranked h+1 must have strictly less than h+1 citations.

The following array A = (100, 30, 27, 3) is an example for which $h = h^{(3)} = 3$.

The least number of citations for the case $h = h^{(3)} = 3$, occurs for the array (27, 27, 27, 0). Generally, the least number of citations for the case $h = h^{(3)}$ is $\left(\underbrace{h^3, h^3, \ldots, h^3}_{h\ times}, 0\right)$. Of course, even when $h = h^{(2)} = h^{(3)}$ there is no upper limit to the corresponding number of citations.

C. When is $h^{(2)} = h^{(3)}$?

Recall that a set of articles has $h^{(2)}$-index h_2 if the first h_2 articles received at least $(h_2)^2$ citations each and the article ranked h_2+1 received strictly less than $(h_2+1)^2$ citations; and a set of articles has $h^{(3)}$-index h_3 if the first h_3 articles received at least $(h_3)^3$ citations each and the article ranked h_3+1 received strictly less than $(h_3+1)^3$ citations. If the two conditions must be satisfied at the same time then the first h_2 articles must have received at least $(h_2)^3$ citations each and the article ranked h_2+1 must have strictly less than $(h_2+1)^2$ citations.

The following array A = (100, 30, 27, 15) is an example for which $h^{(2)} = h^{(3)} = 3$. Note that this array has an h-index equal to 4.

The least number of citations for the case $h^{(2)} = h^{(3)} = 3$, occurs for the array (27, 27, 27, 0). Generally, the least number of citations for the case $h^{(2)} = h^{(3)}$ is

$$\left(\underbrace{h^3, h^3, \ldots, h^3}_{h\ times}, 0\right).$$

D. When is h = g?

A set of articles has g-index h if the sum of the citations of the first h articles is at least h^2 and the sum of the first h+1 articles is strictly less than $(h+1)^2$. If $X = (x_1, x_2, \ldots x_j, \ldots)$ then we see that if g(X) = h, then $\sum_{i=1}^{h} x_i \geq h^2$. This inequality always holds if the h-index of X is equal to h. Now from $\sum_{i=1}^{h+1} x_i < (h+1)^2$ and the fact that $\sum_{i=1}^{h} x_i \geq h^2$, we see that if x_{h+1} = h then $h^2 \leq \sum_{i=1}^{h} x_i < h^2 + h + 1$.

For h = g = 3, $x_{h+1} = x_4 = 3$ and for the largest possible number of citations for x_1 we have: (6, 3, 3, 3) as an example. If $x_{h+1} = x_4 = 0$ we have (9, 3, 3, 0) again for the largest possible value of x_1. An example, still for g=h=3, of an intermediate case is (5, 4, 3, 2). In general, again trying to give the first item the largest possible value, we have for the largest possible integer value, namely h, for item h+1 an array of the form $\left(2h, \underbrace{h, h, \ldots, h}_{h-1\ times}, h\right)$. If item h+1 has value zero, then we have: $\left(3h, \underbrace{h, h, \ldots, h}_{h-1\ times}, 0\right)$.

E. When is g = h = $h^{(2)}$ = $h^{(3)}$?

If h = 3 then the condition h = $h^{(2)}$ = $h^{(3)}$ leads to an array of the form (27, 27, 27, 0), or with higher values. As 27+27+27+0 = 81 = 9^2 we observe that the g-index is at least 9. Hence the equality g = h = $h^{(2)}$ = $h^{(3)}$ is not possible for h=3, and certainly not for higher values.

If h = 2, then h = $h^{(2)}$ = $h^{(3)}$ leads to an array of the form (8, 8, 0), or with higher values. As 8+8+0=16=4^2 the g-index is at least 4. Hence, also for h=2 it is impossible to have equality.

Finally for h=1 it is easy to find examples for which g = $h^{(3)}$ such as (2, 1). As the sum of the first two citations must be at most equal to 3, this example is an extreme. Similarly (3, 0) is an extreme. Among publication-citation arrays of length one (3), (2) and (1) are the only three cases; among publication arrays of length two we have (2, 1) and (1, 1). From this we conclude that equality among the four indices can only occur for h=1 and, even then, occurs in just a few cases.

Note that there are no conditions on the tail so that there is no condition on the total number of citations. The array (2, 1, 1, 1,...) has h-index = g-index = $h^{(3)}$-index = 1, but there is no upper limit on the total number of received citations. If the number of articles, N, is given, then the upper limit for the number of received citations is N+1; the lower limit is 1.

3 Rational indices

Rational h-type indices can be used to make a distinction between cases with the same h-type value. The rational variant of the h-index, denoted as h_{rat}, was introduced by Ruane and Tol (2008) in the context of publications and citations. It is defined as follows.

Definition: Consider a researcher with h-index h. Let n be the smallest possible number of citations necessary to reach an h-index equal to h+1, then the rational h-index, denoted h_{rat}, is defined as:

$$h_{rat} = h + 1 - \frac{n}{2h+1} \tag{1}$$

We next explain this formula. If a researcher has h-index h, then one may ask about the minimum number of citations necessary to reach an h-index equal to h + 1. This number is denoted here as n. The next question is now: if you only know that this scientist's h-index is h what is then the largest number of citations that this researcher needs to reach an h-index equal to h+1. The answer is 2h+1, corresponding with the "worst case scenario" that there are h publications with h citations each and the publication at rank h + 1 has 0 citations. This explains the occurrence of the factor 2h+1 in the formula for the rational h-index (Rousseau et al., 2018). In a similar way a rational g-index was introduced in (Guns & Rousseau, 2009). Next we define the rational $h^{(2)}$ and $h^{(3)}$ indices.

Similar to the case of the h-index we note that the worst case for a set of articles with $h^{(2)}$ index equal to h_2 happens when the first h_2 articles received $(h_2)^2$ citations and the article ranked h_2+1 has no citations. Such an article needs h_2 times $(h_2+1)^2-(h_2)^2 = h_2(2h_2+1)$ extra citations plus $(h_2+1)^2$ new citations, leading to a total of $3(h_2)^2+3h_2+1$ citations. Consequently:

$$\left(h_2\right)_{rat} = h_2 + 1 - \frac{n_2}{3h_2^2 + 3h_2 + 1} \tag{2}$$

where n_2 is the minimum number of citations necessary to reach an $h^{(2)}$-index equal to $h^{(2)} + 1$.

Finally, for the $h^{(3)}$ index we note that the worst case for a set of articles with $h^{(3)}$ index equal to h_3 happens when the first h_3 articles received $(h_3)^3$ citations and the article ranked h_3+1 has no citations. Such an article needs h_3 times $(h_3+1)^3-(h_3)^3 = 3(h_3)^2 + 3(h_3) +1$ extra citations plus $(h_3+1)^3$ new citations, leading to a total of $4(h_3)^3 + 6(h_3)^2 + 4h_3 + 1$ citations. Consequently:

$$\left(h^{(3)}\right)_{rat} = h^{(3)} + 1 - \frac{n_3}{4h_3^3 + 6h_3^2 + 4h_3 + 1} \tag{3}$$

where n_3 is the minimum number of citations necessary to reach an $h^{(3)}$-index equal to $h^{(3)} + 1$.

An example. Array A = (100, 30, 27, 3) has a rational $h^{(3)}$-index of $3 + 1 - \frac{0+34+37+61}{4*27+6*9+4*3+1} = 4 - \frac{132}{175} \approx 3.246$.

4 The relative rational h-index

When researchers reach an h-index of h, it will rarely occur that they really need 2h+1 new citations to reach the value h+1. Usually some of these citations may already have been received. In the extreme case they will only need two new citations, namely when their publication-citation array is $\left(\underbrace{h+1, h+1, \ldots, h+1}_{h-1\, times}, h, h\right)$ or with more citations. If, at the moment a researcher reaches an h-index of h, they need m new citations to reach an h-index of h+1, then at a later moment their relative (or individual) rational h-index, denoted $h_{r,rat}$, is

$$h_{r,\,rat} = h + 1 - \frac{n}{m} \tag{4}$$

where n has the same meaning as before, namely: the minimum number of citations still necessary to reach an h-index equal to h + 1. As $m \leq 2h+1$, $h_{r,rat} \leq h_{rat}$. For an individual researcher this relative rational h-index is clearly more meaningful than the absolute one. An example: if $A_0 = (6, 5, 3, 1)$ when an h-index of 3 was reached and if this researcher's publication-citation array is now A = (9, 6, 4, 2), then their relative rational h-index is 4 – 2/4 = 3.5; the absolute one would be 4-2/7 ≈ 3.71. Similarly, one may define relative rational g, $h^{(2)}$ and $h^{(3)}$ indices and apply them not only to persons, but also to journals or other units of interest.

5 Equality between h and g in a Lotkaian framework

In this section we use a continuous framework. This has no direct application in research evaluation, but it is part of a context in which researchers use a continuous version of h-type indices for modelling purposes (Egghe, 2005). We first recall the definition of the h- and the g-index in this framework. If f(r) is a given rank-frequency function (Zipf-type) then the h-index is the solution of the equality (in r):

$$f(r) = r \tag{5}$$

while the g-index is the solution, g, of the equality

$$\int_0^g f(r)\,dr = g^2 \tag{6}$$

We recall that always (in a continuous as well as a discrete framework) g ≥ h. It has been shown (Egghe & Rousseau, 2006) that in a Lotkaian framework

$$h = T^{1/\alpha} \quad \text{for } \alpha > 1 \tag{7}$$

where α is the exponent of the underlying Lotka (power) function and T is the total number of sources. Similarly, it has been shown (Egghe, 2006b) that

$$g = \left(\frac{\alpha-1}{\alpha-2}\right)^{(\alpha-1)/\alpha} T^{1/\alpha} \text{ for } \alpha > 2. \tag{8}$$

Now we prove that, for α > 2, and for fixed T, g-h is decreasing in α.

Proof. We consider the derivative of g-h with respect to α and prove that this derivative is always negative.

$$\begin{aligned}
\frac{d}{d\alpha}\left(g(\alpha)-h(\alpha)\right) &= \frac{d}{d\alpha}\left(\left(\left(\frac{\alpha-1}{\alpha-2}\right)^{(\alpha-1)/\alpha}-1\right)T^{1/\alpha}\right)\\
&= \left(\left(\frac{\alpha-1}{\alpha}\right)\left(\frac{\alpha-1}{\alpha-2}\right)^{\frac{\alpha-1}{\alpha}-1}\frac{\alpha-2-(\alpha-1)}{(\alpha-2)^2}+ln\left(\frac{\alpha-1}{\alpha-2}\right)\left(\frac{\alpha-1}{\alpha-2}\right)^{\frac{\alpha-1}{\alpha}}\frac{\alpha-(\alpha-1)}{\alpha^2}\right)T^{1/\alpha}\\
&\quad+\left(\left(\left(\frac{\alpha-1}{\alpha-2}\right)^{(\alpha-1)/\alpha}-1\right)\ln(T)T^{1/\alpha}\left(-\frac{1}{\alpha^2}\right)\right)\\
&= T^{1/\alpha}\left\{\left(\left(\left(\frac{\alpha-1}{\alpha-2}\right)^{(\alpha-1)/\alpha}-1\right)\left(-\frac{\ln(T)}{\alpha^2}\right)\right)+\left(\frac{\alpha-1}{\alpha-2}\right)^{\frac{\alpha-1}{\alpha}}\left(\left(\frac{-1}{\alpha(\alpha-2)}\right)+ln\left(\frac{\alpha-1}{\alpha-2}\right)\frac{1}{\alpha^2}\right)\right\}
\end{aligned} \tag{9}$$

The first factor is positive; the first term of the second factor is clearly negative, being a product of a positive and a negative factor. Now the second term of the second factor is a positive number multiplied by negative one (shown below), so that the derivative of g-h with respect to α is negative. This proves that, for fixed T, the difference between g and h decreases with α.

Now we consider the factor $\left(\frac{-1}{\alpha(\alpha-2)}\right)+ln\left(\frac{\alpha-1}{\alpha-2}\right)\frac{1}{\alpha^2}$ and show that it is negative. This holds if $ln\left(\frac{\alpha-1}{\alpha-2}\right)<\frac{\alpha}{\alpha-2}$. Taking exponentials we have to show that $\frac{\alpha-1}{\alpha-2}<e^{\alpha/(\alpha-2)}=1+\frac{\alpha}{\alpha-2}+\frac{1}{2}\left(\frac{\alpha}{\alpha-2}\right)^2+\ldots$. Now $e^{\alpha/(\alpha-2)}=1+\frac{\alpha}{\alpha-2}+\frac{1}{2}\left(\frac{\alpha}{\alpha-2}\right)^2+\ldots>1+\frac{\alpha}{\alpha-2}=\frac{2\alpha-2}{\alpha-2}$. As α > 2, this boils down to the inequality 2α-2 > α-1, which is clearly true. This proves the required inequality.

Now $\lim\limits_{\substack{\alpha\to\infty \\ T \text{ constant}}}(g-h)=\lim\limits_{\substack{\alpha\to\infty \\ T \text{ constant}}}\left(\left(\left(\frac{\alpha-1}{\alpha-2}\right)^{(\alpha-1)/\alpha}-1\right)T^{1/\alpha}\right)=0$. This shows that for fixed T, g tends to h. This is easy to understand: indeed, if the Lotka-coefficient α tends to infinity, the Zipf-coefficient β (the coefficient of the Zipf distribution equivalent of the Lotka distribution with coefficient α) tends to zero (recall that $\beta=\frac{1}{\alpha-1}$). Now a Zipf-coefficient equal to zero corresponds to a ranking in which all elements are equal, which means that g = h.

6 Discussion and conclusion

We derived conditions under which h-type indices, and in particular the h and the g-index, are equal. Next we introduced the rational $h^{(2)}$ and $h^{(3)}$-index. We moreover proposed a relative or individual rational h-index. Finally we studied the limiting behavior of the difference g-h in a continuous Lotkaian framework.

Although this article is explicitly meant to be a contribution in theoretical informetrics, it might have some practical use. This holds, in particular, for the introduction of the relative rational h-index.

We recognize that all h-type indicators do not always behave in a logical way (Bouyssou & Marchant, 2011; Waltman & van Eck, 2012). Like many other indicators the h, $h^{(2)}$, $h^{(3)}$ and g indicator are only PAC (Probably Approximately Correct) (Rousseau, 2016). However, this practical observation has no direct relation with the mathematical properties studied in this contribution. We further note that these h-type indices may play a role in heuristic approaches to support informed peer review (Bornmann et al., 2018).

Acknowledgement

The authors thank Raf Guns (University of Antwerp) and anonymous reviewers for useful remarks about this contribution.

Author Contributions

All the authors, Leo Egghe (leo.egghe@uhasselt.be), Yves Fassin (fassin@skynet.be, Yves.Fassin@Ugent.be), and Ronald Rousseau (ronald.rousseau@uantwerpen.be, ronald.rousseau@kuleuven.be) conceived and designed the analysis, contributed to the development of original idea, and worte the paper.

References

Bornmann, L., Hug, S., & Marewski, J.N. (2018). Bibliometrics-based heuristics: What is their definition and how can they be studied? arXiv:1810.13005

Bouyssou, D., & Marchant, T. (2011). Ranking scientists and departments in a consistent manner. Journal of the American Society for Information Science and Technology, 62(9), 1761–1769.

Deineko, V.G., & Woeginger, G.J. (2009). A new family of scientific impact measures: The generalized Kosmulski-indices. Scientometrics, 80(3), 821–828.

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

Egghe, L. (2006a). An improvement of the h-index: The g-index. ISSI Newsletter, 2(1), 8–9.

Egghe, L. (2006b). Theory and practice of the g-index. Scientometrics, 69(1), 131–152.

Egghe, L. (2011). Characterizations of the generalized Wu- and Kosmulski-indices in Lotkaian systems. Journal of Informetrics, 5(3), 439–445.

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

Fassin, Y., & Rousseau, R. (2018). The $h^{(3)}$ – index for academic journals. Preprint.

Guns, R., & Rousseau, R. (2009). Real and rational variants of the h-index and the g-index. Journal of Informetrics, 3(1), 64–71.

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences USA 102(46), 16569–16572.

Kosmulski, M. (2006). A new Hirsch-type index saves time and works equally well as the original h-index. ISSI Newsletter, 2(3), 4–6.

Rousseau, R. (2016). Citation data as a proxy for quality or scientific influence are at best PAC (Probably Approximately Correct). Journal of the Association for Information Science and Technology, 67(12), 3092–3094.

Rousseau, R., Egghe, L., & Guns, R. (2018). Becoming metric-wise. A bibliometric guide for researchers. Kidlington (UK): Chandos-Elsevier.

Ruane, F., & Tol, R.S.J. (2008). Rational (successive) h-indices: An application to economics in the Republic of Ireland, Scientometrics, 75(2), 395–405.

Waltman, L., & van Eck, N.J. (2012). The inconsistency of the h-index. Journal of the American Society for Information Science and Technology, 63(2), 406–415.