

A rationale for the Hirsch-index rank-order distribution and a comparison with the impact factor rank-order distribution

Peer-reviewed author version

EGGHE, Leo (2009) A rationale for the Hirsch-index rank-order distribution and a comparison with the impact factor rank-order distribution. In: JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, 60(10). p. 2142-2144.

DOI: 10.1002/asi.21121

Handle: <http://hdl.handle.net/1942/9683>

A rationale for the Hirsch-index rank-order distribution and a comparison with the impact factor rank-order distribution

by

L. Egghe

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

leo.egghe@uhasselt.be

ABSTRACT

We present a rationale for the Hirsch-index rank-order distribution and prove that it is a power law (hence a straight line in the log-log scale). This is confirmed by experimental data of Pyykkö and by data produced in this paper on 206 mathematics journals. This distribution is of a completely different nature than the impact factor (IF) rank-order distribution which (as was proved in a previous paper) is S-shaped. This is also confirmed by our example.

Only in the log-log scale of the h-index distribution we notice a concave deviation of the straight line for higher ranks. This phenomenon is discussed.

Key words and phrases: power law, h-index distribution, Hirsch, impact factor distribution.

Acknowledgement: The author is grateful to M. Goovaerts for providing the experimental data.

I. Introduction

If one ranks a set of journals (e.g. in a field) in decreasing order of their impact factors (IF) then we showed in Egghe (2008) that this rank-order distribution $r \otimes IF(r)$ as well as the function $r \otimes \ln(IF(r))$ is S-shaped: convex for smaller r and concave for larger r . We could even show in Egghe (2008) that the concave part of the function $r \otimes \ln(IF(r))$ is larger than the one of the function $r \otimes IF(r)$. The shape of $r \otimes \ln(IF(r))$ was experimentally confirmed in Mansilla, Köppen, Cocho and Miramontes (2007).

Since we want to compare these theoretical and experimental results with analogous results on the Hirsch-index distribution for the same set of journals (which is the main topic of the present paper), we produced a new data set: 206 mathematics journals of which we traced citation data in the Web of Science (WoS). Fig. 1 gives the $r \otimes IF(r)$ function (distribution) and Fig. 2 gives the $r \otimes \ln(IF(r))$ function. We clearly see the S-shape in both graphs and confirmation is also given that the concave part in Fig. 2 is “larger” than in Fig. 1.

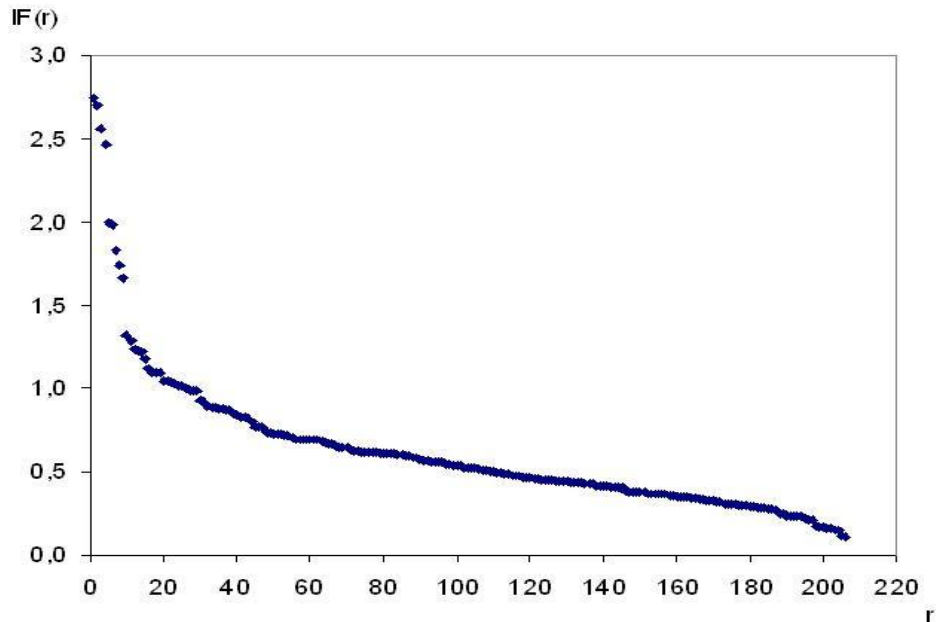


Fig. 1 $IF(r)$ in function of r

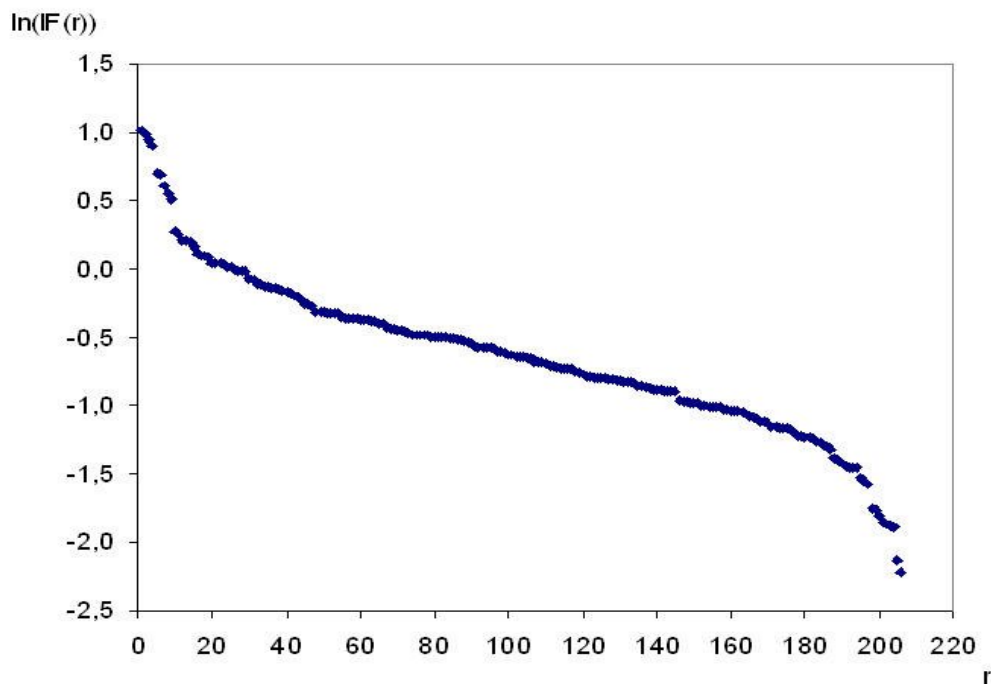


Fig. 2 $\ln(\text{IF}(r))$ in function of r

In the next section we will present a rationale for the Hirsch (h-)index rank-order distribution based on results from Egghe (2005), Chapter 2 (relation between a size-frequency function and a rank-frequency function) and from Egghe (2007) where a rationale is given that the size-frequency function of the h-index is a power law (law of Lotka).

In the third section we present the graph of $r \otimes h(r)$ which should be convexly decreasing (being a decreasing power law): this is confirmed. Since power laws in a log-log scale are straight lines we also give this graph in Section 3. The straight line is confirmed except for large ranks r . The same shapes were experimentally recovered in Pyykkö (2008).

In the fourth section we give conclusions and discuss the deviation from the straight line encountered in the previous section. We link this phenomenon with similar deviations in paper-citation graphs in Molinari and Molinari (2008), Radicchi, Fortunato and Castellano (2008), Lehmann, Jackson and Lautrup (2008) and van Raan (2001a,b).

II. The rank-frequency function $h(r)$ of the h-index

We briefly repeat the argument that the size-frequency function $f(h)$ (the number of journals (or authors) with h-index h (see also Egghe (2007)) is a power law (law of Lotka). Suppose a journal or author has T articles. Supposing the article-citation relation to be Lotkaian, there exists an exponent $\alpha > 1$ such that

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

as proved in Egghe and Rousseau (2006). Then supposing that the number of articles of a journal or an author is also distributed according to Lotka's law (in terms of authors this was the original finding of Lotka himself ! – see Lotka (1926)). Then there exists another exponent $\alpha^* > 1$ such that the number of journals (or authors) with T articles is given by

$$f(T) = \frac{C}{T^{\alpha^*}} \quad (2)$$

where $C > 0$. Hence (1) and (2) imply that the number of journals or authors with h-index h is given by

$$f(h) = \frac{C}{h^{\alpha\alpha^*}} \quad (3)$$

again a power law (law of Lotka): the size-frequency function of the h-index is a law of Lotka.

This ends the first part of the argument. The second part is the derivation of the rank-frequency function $h(r)$ of the h-index based on the size-frequency function (3). This is a classical result, known for over 20 years – see e.g. Egghe (2005), Chapter 2: the rank r of the journal or author with h-index j is given by (by definition of f), denoting h^{-1} for the inverse function of $h(r)$:

$$r(j) = h^{-1}(j) = \int_j^{\infty} f(j') dj' \quad (4)$$

But, since $\alpha\alpha^* > 1$ we have

$$r = r(j) = \int_j^{\infty} f(j') dj' = \frac{C}{\alpha\alpha^* - 1} j^{1-\alpha\alpha^*} \quad (5)$$

So, from (4) and (5) we have

$$j = h(r) = \left(\frac{\alpha\alpha^* - 1}{C} r \right)^{\frac{1}{1-\alpha\alpha^*}} \quad (6)$$

$$h(r) = \frac{D}{r^\beta} \quad (7)$$

where

$$D = \left(\frac{C}{\alpha\alpha^* - 1} \right)^{\frac{1}{1-\alpha\alpha^*}} \quad (8)$$

is a constant and

$$\beta = \frac{1}{\alpha\alpha^* - 1} \quad (9)$$

Hence the rank-frequency distribution of the h-index is a power law, i.e. the so-called law of Zipf (cf. Egghe (2005)).

It is clear that (7) is a convexly decreasing function, hence different from the S-shaped distribution $IF(r)$ of the impact factor (see previous section).

We will, in the next section verify this on the same dataset (206 mathematics journals) on which we calculated $IF(r)$.

III. Experimental verification

Fig. 3 gives the graph of $h(r)$ for 206 mathematics journals with citation data from the WoS (the same data on which we calculated $IF(r)$). It is clear that we indeed obtained a convexly decreasing function $h(r)$. This graph is similar to the one obtained in Miller (2006) (50 journals in physics).

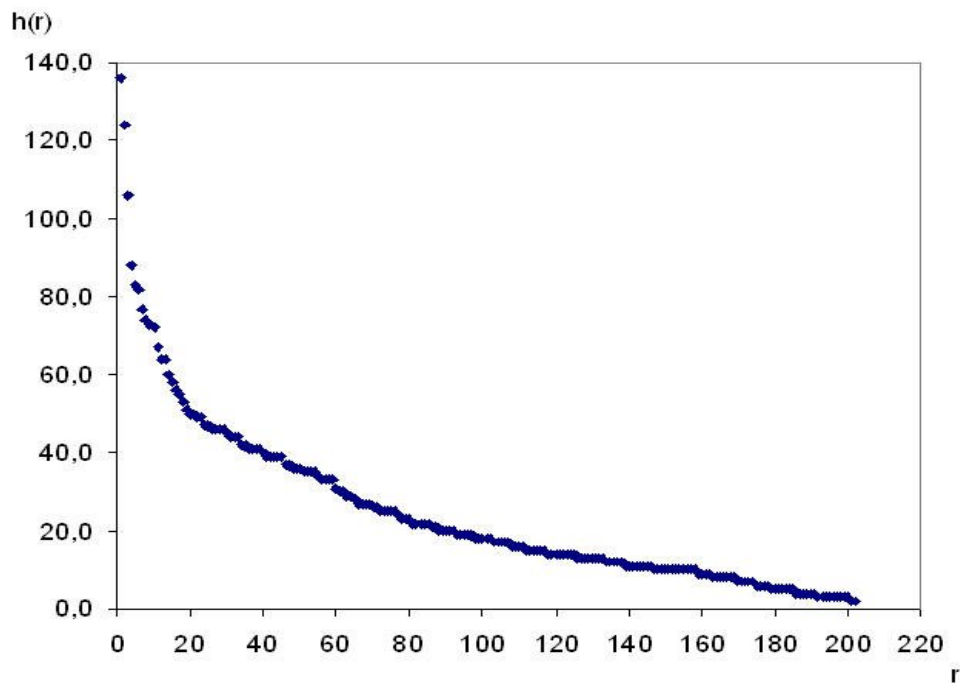


Fig. 3 $h(r)$ in function of r

Since (7) is a power law we have that the graph of $\ln r$ versus $\ln h(r)$ is linear. Graphs of power laws are, usually, presented in this so-called log-log format. Therefore we also produced the log-log version of Fig. 3 – see Fig. 4. The expected linear shape is true except for higher ranks.

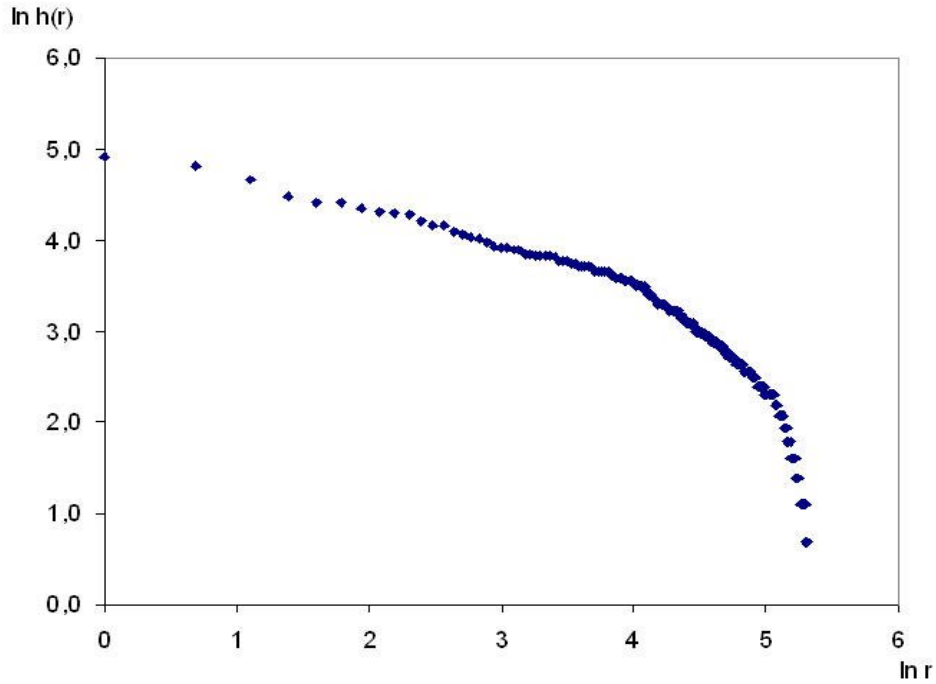


Fig. 4 $\ln h(r)$ in function of $\ln r$

We believe that Figs. 3 and 4 confirm, to a good extent the theoretical findings of the previous section. For a discussion of the concave part in Fig. 4 we refer the reader to the next section.

These experimental findings are in line with those of Pyykkö (Pyykkö (2008)): there one has more graphs where the log-log version of $h(r)$ is (almost perfectly) linear (hence the predicted power law $h(r)$) but we feel that many graphs are truncated for high ranks r .

IV. Discussion, conclusions and an open problem

In this paper we used a data set of 206 mathematics journals in the WoS. Both the rank-order distributions of the impact factor and of the Hirsch-index were constructed. The $IF(r)$ and $\ln(IF(r))$ -graphs showed an S-shape as predicted in Egghe (2008), with a “larger” concave part for $\ln(IF(r))$ than for $IF(r)$.

We presented a theory proving that the rank-order distribution of the h-index has a different shape: $h(r)$ is a decreasing power law and hence is convexly decreasing. The log-log graph of $h(r)$ shows a linear decrease (as it should for a power law) except for large ranks r . We can give a partial explanation for this. The argument goes as follows. In the derivation of the power law (7) for $h(r)$ we made two assumptions.

- (i) the article-citation relation follows Lotka's law
- (ii) the journal (or author)-article relation also follows Lotka's law.

Now, we think that assumption (ii) is non-controversial since it reformulates the historical law of Lotka (Lotka (1926)). Although assumption (i) is intuitively correct we found deviations of this law in the literature. If assumption (i) should be true, then the article-citation relation in log-log scale should be linear. In van Raan (2001a,b) (Fig. 4), Radicchi, Fortunato and Castellano (2008) (Fig. 1), Brantle and Fallah (2007) (Fig. 3) and Lehmann, Jackson and Lautrup (2008) (Figs. 1,2) we see, however, that the article-citation relation (size-frequency function) in log-log scale is concavely decreasing: only the highly cited papers follow a straight line. This deviation of Lotka's law (and hence the (partial) violation of assumption (i)) is the reason for the deviation from formula (7).

It is very rare that size-frequency functions deviate from Lotka's law (see Egghe (2005), Chapter 1). An explanation of the van Raan curve (and hence also of the Radicchi et al. curves and the Lehmann et al. curves since they are similar) is given in van Raan (2001a,b), using Bessel functions.

A similar deviation, but now on the rank-frequency function can be seen in Molinari and Molinari (2008).

Rousseau (2009) communicated to me that the rank-order distribution of the h-index resembles well a lognormal distribution. A rationale for this is lacking.

References

- T.F. Brantle and M.H. Fallah (2007). Complex innovation networks, patent citations and power laws. PICMET 2007 Proceedings, 540-549.
- L. Egghe (2005). Power Laws in the Information Production Process: Lotkaian Informetrics. Elsevier, Oxford, UK.
- L. Egghe (2007). Distributions of the h-index and the g-index. Proceedings of the 11th International Conference of the international Society for Scientometrics and Informetrics, Madrid (Spain) (D. Torres-Salinas and H.F. Moed, eds.), 245-253, CSIC, Madrid (Spain).
- L. Egghe (2008). Mathematical derivation of the impact factor distribution. Journal of Informetrics, to appear.
- L. Egghe and R. Rousseau (2006). An informetric model for the Hirsch-index. Scientometrics 69(1), 121-129.
- S. Lehmann, A.D. Jackson and B.E. Lautrup (2008). A quantitative analysis of indicators of scientific performance. Scientometrics 76(2), 369-390.
- A.J. Lotka (1926). The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences 16(12), 317-324.
- 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.
- C.W. Miller (2006). Superiority of the h-index over the impact factor for physics. arXiv:physics/0608183v1.
- A. Molinari and J.-F. Molinari (2008). Mathematical aspects of a new criterion for ranking scientific institutions based on the h-index. Scientometrics 75(2), 339-356.
- P. Pyykkö (2008). Power-law distribution of individual Hirsch indices, the comparison of merits in different fields, and the relation to a Pareto distribution. Preprint.
- F. Radicchi, S. Fortunato and C. Castellano (2008). Universality of citation distributions: towards an objective measure of scientific impact. Proceedings of the National Academy of Sciences of the United States of America 105(45), 17268-17272.
- R. Rousseau (2009). Personal communication.

- A.F.J. van Raan (2001a). Competition amongst scientists for publication status: Toward a model of scientific publication and citation distributions. *Scientometrics* 51(1), 347-357.
- A.F.J. van Raan (2001b). Two-step competition process leads to quasi power-law income distributions. Application to scientific publication and citation distributions. *Physica A* 298, 530-536.