The impact factor rank-order distribution revisited

L. Egghe

Universiteit Hasselt (Uhasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium[(*)]

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

leo.egghe@uhasselt.be

There is some controversy around the impact factor (IF) rank-order distribution. The controversy is around the shape of the IF-rank-order distribution. In some articles (Lancho-Barrantes et al. (2010), Guerrero-Bote et al. (2007)) one claims examples of convexly decreasing IF-rank-order distributions, while in other contributions the S-shape (first convex, then concave) is advocated (Mansilla et al. (2007), Martinez-Mekler et al. (2009), Campanario (2009), Egghe (2009)). Egghe and Waltman (2011) show both shapes.

The difference between the two shapes can be explained through the size-frequency distribution of impact factors. To be as clear as possible we will repeat here the definitions of rank-order distributions and size-frequency distributions. They can be given in a general context of sources and items (Egghe (2005)) but since, in this note we are only interested in the journal-IF relation, we will, mainly, use this terminology.

Let us have a set of journals (e.g. in a certain field), each having a value for the impact factor IF. It does not matter here what type of IF is used (i.e. two year IF, five-year IF, or any other variant). We define f to be the size-frequency distribution where for every positive real number n, $f(n)$ is the density of sources with n items, i.e. for every two positive numbers m,n, $m<n$,

$$\int_m^n f(n')dn' \qquad (1)$$

denotes the number of sources with a number of items between m and n. In our case this means the number of journals with IF between m and n.

The corresponding rank-frequency distribution g is defined as

$$r=\int_n^\infty f(n')dn'=g^{-1}(n) \qquad (2)$$

where $n = g(r)$ and where g^{-1} denotes the inverse function of the function g. This defines $g(r)$ as the IF of the journal on rank r. Equation (2) defines g given f but it also determines f given g since (2) is equivalent with

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

(see Egghe (2005) or Egghe and Waltman (2011)). Note that, by (2), g is always strictly decreasing but that is not the case for f. That is the key point we want to make here. We invoke the key results of Egghe and Waltman (2011).

1. f is decreasing if and only if g is convex.

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

In Egghe and Waltman (2011) examples of both shapes (for f as well as g in the case of IF distributions) are given so that, in practise, g can be convex but it can also be S-shaped. In fact, the clearest example of an S-shape for g is given in Fig. 2 in Egghe and Waltman (2011), third example: mathematics. It is clear from these examples that, the larger the increasing part of f, the more explicite is the S-shape of g.

Lancho-Barrantes et al. (2010) claim that there are no S-shaped examples of g but their own example (not surprisingly in a part of mathematics: algebra and number theory) shows an S-shape (see their Fig. 1 where the squared dots clearly decrease in a concave way at the right side of the graph). This is ignored in Lancho-Barrantes et al. (2010) since they fit the curve by a convexly decreasing logarithmic function. This is in contradiction with the above theory and with the practical example itself.

In Egghe (2009) we presented an argument for the S-shape of g assuming that the size-frequency function f resembles a normal (Gaussian) distribution. This was criticized in Waltman and van Eck (2009) and I respect this criticism since one can give arguments against this model (altough the most explicite examples of functions f that are first increasing and then decreasing resemble the Gaussian distribution very well). It is also a weakness of the paper Egghe (2009) since it gives only arguments for the S-shape of g and not for the convex shape of g, which, clearly also exists. So, here I leave this argument as it is but I claim that S-shapes for g are indeed existing in practise and that they can be explained (by purely mathematical arguments, independent from Gaussian distributions) by statement (2) above from Egghe and Waltman (2011) and that statement (1) above from Egghe and Waltman (2011) also covers the convex shape of g.

The convex fit of the S-shaped g in Lancho-Barrantes et al. (2010) is not correct and should be replaced by an S-shaped fit. One suggestion for this (but there might be other candidates) is given in Mansilla et al. (2007) (and repeated in Egghe and Waltman (2011)): fit g by the model

$$g(r) = K \frac{(N+1-r)^b}{r^a} \tag{4}$$

($a > 0, b \geq 0, K > 0$ are parameters, N is the total number of journals) for which we proved in Egghe and Waltman (2011) that

(i) g is strictly decreasing

(ii) g has an S-shape, first convex and then concave, if $0 < b < 1$

(iii) g is convex if either $b = 0$ or $b \geq 1$.

REFERENCES

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

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

L. Egghe (2009). Mathematical derivation of the impact factor distribution. Journal of Informetrics 3(4), 290-295.

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.

V. Guerrero-Bote, F. Zapico-Alonso, M.E. Espinosa-Calvo, R. Gómez-Crisóstomo and F. De Moya-Anegón (2007). Import-export of knowledge between scientific subject categories: The iceberg hypothesis. Scientometrics 71(3), 423-441.

B.S. Lancho-Barrantes, V.P. Guerrero-Bote and F. Moya-Anegón (2010). The iceberg hypothesis revisited. Scientometrics 85(2), 443-461.

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 on the arts and sciences. PLoS ONE 4(3), e4791.

L. Waltman and N.J. van Eck (2009). Some comments on Egghe's derivation of the impact factor distribution. Journal of Informetrics 3(4), 363-366.

2010019not