Basic independence axioms for the publication-citation system

Ronald Rousseau [a,b,c] **and Fred Y Ye** [d,e]

*[a]KHBO (Association KU Leuven), Faculty of Engineering Technology, Zeedijk 101, B-8400 Oostende, Belgium
[b]KU Leuven, Dept. Mathematics, Celestijnenlaan 200B, B-3000 Leuven (Heverlee), Belgium
[c]Universiteit Antwerpen, IBW, Venusstraat 35, B-2000 Antwerpen, Belgium
[d]Dept. Information Resources Management, Zhejiang University, Hangzhou 310027, China
[e]School of Information Management, Nanjing University, Nanjing 210093, China*

ABSTRACT

Following the idea of the Bouyssou-Marchant independence axiom, a basic independence axiom for an indicator is introduced. It is shown that the average number of citations nor the h-index satisfy this basic axiom. Among the indicators considered here only the number of publications and the number of citations satisfy the requirements of this axiom. Weaker forms of the basic independence axiom are considered. The number of citations of the most-cited article satisfies these weaker forms. When studying the h-index the notion of an h-critical publication is introduced.

Keywords: Independence; consistency; publications; citations; indicators; h-index.

INTRODUCTION

In recent years it has come to the attention of informetricians that basic indicators used for research evaluation do not always have the properties one might expect. A well-known example is the discussion between ratios of averages and averages of ratios (or in other words, between the original Leiden crown indicator and the Karolinska indicator) (Lundberg, 2007; Opthof & Leydesdorff, 2010; Larivière & Gingras, 2011). Even the well-known impact factor may yield some surprises (Rousseau & Leydesdorff, 2011).

In order to solve these problems and to derive meaningful rankings based on indicators Bouyssou and Marchant, as well as Waltman and van Eck initiated a series of studies aiming at clarifying the properties of indicators, see (Marchant, 2009; Bouyssou & Marchant, 2010, 2011a, b; Waltman & van Eck, 2009, 2012). We will not provide all details but just focus on the independence axiom proposed by Bouyssou and Marchant (2011b).

***Corresponding author.**

E-mail: ronald.rousseau@khbo.be
yye@zju.edu.cn

DOI: 10.5530/jscires.2012.1.6

This independence axiom for an indicator f (any!) can be formulated as follows. Let $f(S) \leq f(T)$, where f(S), resp. f(T), denotes the value of the indicator f derived from sets of articles written by scientist S, resp. scientist T. If one adds now to both sets an article with n (a natural number) citations, then the axiom requires that the relation $f(S) \leq f(T)$ must still hold. Note that the above-mentioned sets of articles can be restricted by a specific publication window, and if the indicator f involves citations, then too this indicator can be restricted by a specific citation window; see (Liang & Rousseau, 2009).

We refer to this requirement as the B-M independence axiom. Bouyssou and Marchant (2011b) show that the h-index, the average number of citations per publication and even the median number of citations do not satisfy the B-M independence axiom. So-called scoring rules (see further for a definition) do.

Clearly the B-M independence axiom is a reasonable requirement. Yet, this requirement is not how authors' article sets grow. The aim of this article is to introduce independence axioms that are, in a sense, more basic (more natural) than Bouyssou and Marchant's. We investigate which indicators satisfy the new independence axioms.

Notation. If scientist S has published three articles that are, respectively, cited two times, once and four times then we denote this in the following ranked order: S = [4, 2, 1].

Basic Steps and Basic Independence Axioms

We first define two types of basic steps in a scientist's career:

Definition: two basic steps

Step (P): A basic publication step (P) occurs if a scientist publishes a new article, with no citations.

Step (C): A basic citation step (C) occurs if a scientist receives one citation to an already existing publication.

One may say that at any moment in a scientist's career nothing happens in terms of publications and citations, or one of these basic steps occurs.

The basic independence axiom for indicator f

If f(S) ≤ f(T) and the same type of basic step occurs to these two scientists then still f(S) ≤ f(T).

Note that this axiom is not a weak version of the B-M independence axiom. In the B-M independence axiom one always adds a new publication. When considering basic steps one may add a publication, but then this publication must have no citations. It is however also possible to add a citation and no new publication, which is not considered in the B-M axiom.

Before we turn our attention to the basic independence axiom we study the outcome of a basic step on some indicators, namely the total number of publications (PUB), the total number of citations, counted as whole numbers (CIT), the h-index (h) (Hirsch, 2005), the average number of citations per publication (AVG) (this includes the journal impact factor), the median number of citations (MED), the number of citations received by the most-cited article (TOP), and the number of citations received by the top 3 most-cited articles (TOP3). If a scientist has less than three publications one counts the number of citations received by the articles he/she did publish. Of course one can also consider TOP5, TOP10 and so on (these will not be discussed as results are essentially the same as for TOP3).

There is, however, an exceptional case we have to deal with first. If a scientist has no publications then PUB = 0, CIT = 0 and all other indicators are undefined. If this scientist now publishes one article (step (P)) then PUB = 1, while CIT, h, AVG, MED, TOP and TOP3 are all equal to zero. For such a scientist step (C) cannot occur. From now on we always assume that a scientist has at least one publication. Now we discuss the influence of the two basic steps on scientists who have at least one publication.

1) A basic publication step (P)

a) The total number of publications (PUB). Its influence is that PUB increases by one.
b) The total number of citations (CIT). Its influence is that CIT stays the same.
c) The h-index. Adding one uncited publication never changes the h-index.
d) The average number of citations per publication. If the average was AVG = CIT/PUB, then this average will decrease and become CIT/(PUB + 1), unless CIT = 0, in which case AVG stays zero.
e) The median number of citations (MED).

   We recall that the median of a finite sequence (x_1, x_2, …, x_n), ranked in decreasing order, is defined as:

$$MED = \begin{cases} x_{(n+1)/2}, \text{if n is odd} \\ \frac{1}{2}\left(x_{n/2} + x_{(x+2)/2}\right), \text{if n is even} \end{cases} \quad (1)$$

   As step (P) does not add any citation, the corresponding article can be placed at the end of the row. Hence, if n was odd, the length of the sequence is now even and MED becomes $\frac{x_{(n+1)/2} + x_{(n+3)/2}}{2}$. If n was even, the length of the sequence is now odd and MED Becomes $x_{\left(\frac{n}{2}\right)+1}$.
f) The highest number of received citations (TOP). This indicator never changes by a basic step (P).
g) The number of citations received by the three most-cited articles (TOP3). This indicator never changes by a basic step (P).

2) A basic citation step (C)

a) The total number of publications (PUB) stays the same.
b) The total number of citations (CIT) increases by one.
c) The h-index. The h-index may stay the same, but it is possible that h increases by 1.
   Consider S = [2, 1, 0] with h-index equal to 1. If the second article receives one more citation then the h-index becomes 2. If, however, the first or the last article receives one more citation then the h-index stays equal to 1.
d) The average number of citations per publication (AVG). This number was CIT/PUB and becomes (CIT +1)/PUB.

e) The median number of citations (MED).
An article that receives one more citation may have no influence on the median, may increase the median by 0.5, or may increase the median by 1. We provide examples. If S = [2, 1, 0], with MED(S) = 1, and if the first article receives one more citation then the median stays the same. If the second article receives one more citation then the median becomes 2 (an increase by 1). If T = [2, 1, 1, 0], with MED(T) = 1, and if the second article receives one more citation then MED(T) becomes 1.5, an increase by 0.5.
f) The highest number of received citations (TOP). This indicator stays the same, unless it happens to be the most-cited article that receives the extra citation. In that case this indicator increases by one.
g) TOP3. Again this indicator stays the same unless one of the three most-cited articles receives the extra citation. In that case TOP3 increases by one.

A Discussion of the Basic Independence Axiom and the Indicators PUB, CIT, AVG, MED, TOP and TOP3

1) The total number of publications (PUB)

If PUB(S) = PUB(T) and step (P) occurs for S as well as for T, then trivially PUB(S) is still equal to PUB(T). Similarly, if PUB(S) < PUB(T), then this inequality stays the same after a (P)-step.

Step (C) has no influence on the number of publications, hence this requirement is always satisfied for PUB.

2) The total number of citations (CIT)

If CIT(S) = CIT(T) or CIT(S) < CIT(T) and step (P) occurs then, trivially, the respective total number of citations stay the same and hence their equality or inequality.

Similarly, a step (C) increases the total number of citations by one and hence the equality or inequality between the total number of citations of scientists S and T stays the same.

3) The average number of citations

If AVG(S) = AVG(T) and step (P) occurs then it is possible that AVG(S) > AVG(T). Indeed, take AVG(S) = 4/6 and AVG(T) = 2/3 (where numbers refer to actual data, before simplification of the fraction) then after step (P), the averages are AVG(S) = 4/7 and AVG(T) = 2/4, contradicting the basic independence axiom. Even if AVG(S) < AVG(T) and step (P) occurs then it is possible that AVG(S) > AVG(T). Indeed if AVG(S) =10/16 and AVG(T) = 2/3 then AVG(S) < AVG(T). However, after a basic step (P) we have: AVG(S) = 10/17 and AVG(T) = 2/4 hence AVG(S) > AVG(T), again contradicting the basic independence axiom.

If AVG(S) = AVG(T) and step (C) occurs then it is possible that AVG(S) > AVG(T). Indeed, take AVG(S) = 4/6 and AVG(T) = 2/3 then after step (C), the averages are AVG(S) = 5/6 and AVG(T) = 3/3, contradicting the basic independence axiom. Finally, even if AVG(S) < AVG(T) and step (C) occurs then it is possible that AVG(S) > AVG(T). Indeed if AVG(S) = 2/3 and AVG(T) = 8/11 then AVG(S) < AVG(T). However, after a basic step (C) we have: AVG(S) = 3/3 and AVG(T) = 9/11 hence AVG(S) > AVG(T), again contradicting the basic independence axiom.

4) The median number of citations

If MED(S) = MED(T) and step (P) occurs then it is possible that MED(S) > MED(T). Indeed, take S = [2, 2] and T = [4, 3, 1, 1] then MED(S) = MED(T) = 2. After step (P), MED(S) = 2 and MED(T)=1, contradicting the basic independence axiom. Even if MED(S) < MED(T) and step (P) occurs then it is possible that MED(S) > MED(T). Indeed if S = [2, 2] and T = [9, 9, 1, 1] then MED(S) = 2 < MED(T) = 5. However, after a basic step (P) we have MED(S) = 2 > MED(T) = 1.

If MED(S) = MED(T) and step (C) occurs then it is possible that MED(S) > MED(T). Indeed, take S = [2, 2] and T = [4, 3, 1, 1] then MED(S) = MED(T) = 2. If the most-cited article receives one more citation then MED(S) = 2.5, while MED(T) = 2.Finally, even if MED(S) < MED(T) and step (C) occurs then it is possible that MED(S) > MED(T). Indeed, if S = [4, 2, 1] with MED(S) = 2 and T = [4, 3, 2, 1] with MED(T) = 2.5 (hence MED(S) < MED(T)) and one adds a citation to the article with 2 citations (for S) and for the article with 4 citations (for T), then MED(S) = 3 > MED(T) = 2.5.

5) The number of citations received by the most-cited article

If TOP(S) = TOP(T) and step (P) occurs then TOP(S) is still equal to TOP(T). The same is trivially true if TOP(S) < TOP(T).

If TOP(S) = TOP(T) and step (C) occurs then it is possible that TOP(S) > TOP(T), namely if the most-cited article of T receives one more citation, while this is not the case for the most-cited article of S. If TOP(S) < TOP(T) then it is only possible that TOP(S) = TOP(T), namely if TOP(S) was equal to TOP(T)-1, and the most-cited article of S receives the extra citation, while this is not the case for the most-cited article of T. In any case, if TOP(S) < TOP(T) there is no violation of the basic independence axiom.

6) The sum of citations received by the three most-cited articles (TOP3)

If TOP3(S) = TOP3(T) and step (P) occurs then TOP3(S) is still equal to TOP3(T). Similarly step (P) never changes the relation TOP3(S) < TOP3(T).

If TOP3(S) = TOP3(T) and step (C) occurs then it is possible that TOP3(S) > TOP3(T), namely if one of the three most-cited article of T receives one more citation, while this is not the case for any of the three most-cited article of S. If TOP3(S) < TOP3(T) then step (C) leads either to either to TOP3(S) = TOP3(T), or the inequality TOP3(S) < TOP3(T) stays true. In either case there is no violation of the basic independence axiom.

Because we would like to have that TOP and TOP3 satisfy our basic independence axiom (or at least some version of it) we consider weaker versions.

Two Weaker Axioms

For the reason mentioned above, it might be useful to weaken the axiom related to (C). We propose two weaker forms:

Axiom WCR: a weak (C) rank form. If f(S) ≤ f(T) and either a new, uncited article, is added to S and T, or a citation is given to two articles on the same rank (where for this axiom, publications with the same number of citations are considered to have the same rank), then still f(S) ≤ f(T).

Axiom WCS: a weak (C) size form. If f(S) ≤ f(T) and either a new, uncited article, is added to S and T, or a citation is given to two articles with the same number of citations, then still f(S) ≤ f(T).

Clearly if an indicator f satisfies the basic independence axiom then it also satisfied the weaker forms, but the opposite is not true.

The indicators TOP and TOP3 do not satisfy the (C)-part of the basic independence axiom but TOP does satisfy the two weaker axioms. Indeed, TOP3 does not satisfy WCR but trivially satisfies WCS. The following example shows that TOP3 does not satisfy WCR. Let S = [6, 5, 1, 0] and T = [5, 4, 3, 3] then TOP3(S) = TOP3(T). Adding one citation to the 4th ranked article yields: TOP3(S) = 12 and TOP3(T) = 13, or TOP3(T) > TOP3(S), contradicting the WCR requirement. We also note that if S and T collaborated on the article that received one extra citation then we are automatically in the case that the size form of the weak axiom applies.

As the requirement related to the (P)-part has not changed every indicator that fails because of the (P) part of the basic independence axiom also fails the weaker axioms. This is the case for AVG and MED.

A Discussion of the Basic Independence Axiom and the h-index

In order to discuss this aspect we introduce the notion of an h-critical publication.

Definition: an h-critical publication

A publication is an h-critical publication if it is such that by receiving one more citation the h-index increases.

Of course, this increase in h-index is automatically by one. Hence, an h-critical publication always has h citations. In reality one might expect more h-critical publications for junior scientists than for senior ones, as in general young scientists may have many articles with similarly low numbers of citations. Senior scientists will probably have less often h-critical publications in their publication set.

Proposition

An actor's publication list (with h-index h) has h-critical publications if and only if the following two requirements are satisfied:

(1) There do not exist articles in the h-core with h citations;
(2) there exists an article in the h-tail with h citations.

Proof. Indeed, we first note that h-critical publications never belong to the h-core. If the article ranked h+1 has h citations and if it happens to be the article that receives an extra citation, then it enters the h-core. Yet, if the article ranked h has only h citations then it drops from the h-core and the h-index stays the same. Only if the article ranked h has at least h+1 citations the h-index will increase by one. Further, if the article ranked h+1 has strictly less than h citations it can never enter the h-core and the h-index can never increase to h+1.

Note

The largest possible set of h-critical publications, given PUB, consists of (PUB – h) articles. This happens if all articles in the h-tail have h citations.

When a step (P) occurs then this has no influence on the set of h-critical publications, unless this set contains articles with zero citations. This is the case if S has only uncited articles, for instance if S = [0 0 0]. Anyway, step (P) never leads to a case that the h-index violates the weak independence axiom.

If, however, step (C) occurs then this may or may not influence the set of h-critical publications, and hence lead to a violation of the basic independence axiom. An example: S = [3, 3, 2] and T = [3, 3, 2], hence h(S) = h(T) = 2. The critical sets of S and T are the same, namely the third ranked article. If now S receives an extra citation for its critical article, while T receives an extra citation for its most-cited article then h(S) ≤ h(T) (in this case: 2 ≤ 2) leads to h(S) = 3 > h(T) = 2, which is a contradiction of the basic independence axiom.

We consider now the weak forms and their relation with the h-index.

On the one hand, the h-index does not satisfy WCR. Indeed, consider S = [3, 3, 2] and T = [2, 2, 0]. Then h(S) = h(T) = 2, but adding one citation to the publication on rank 3 leads to h(S) = 3 > h(T) = 2.

Yet, on the other hand, the h-index does not satisfy WCS either. Indeed, if S = [3, 3, 2] and T = [2, 2, 2], then h(S) = h(T) = 2. Adding one more citation to a publication with 2 citations leads to h(S) = 3 > h(T) = 2. This again contradicts the WCS requirement.

Scoring Rules

Definition (Bouyssou & Marchant, 2011b)

A scoring rule is a function $s(S)=\sum_{n=0}^{\infty} pub(n)*u(n)$, where pub(n) is the number of publications, (co)-authored by scientist S, with n citations, and u(n) is a fixed value associated with a publication with n citations.

PUB and CIT satisfy our axiom for basic independence and are also scoring rules (for PUB u(n) = 1 while for CIT u(n) = n). A general scoring rule will however not satisfy our axiom for basic independence, because we only require an increase by one citation but do not specify how many citations this publication already had.

TOP satisfies the weak form (WCS) but is not a scoring rule. If u(0) = 0 or 1 then a scoring rule satisfies the (P)-part of the basic independence axiom. A scoring rule always satisfies (WCS).

CONCLUSION

In this article, we introduced the basic independence axiom for an indicator. This axiom is derived from the idea to consider only basic steps such as adding one (uncited) publication or adding one citation to an already existing publication. It is shown that AVG, MED, the h-index, TOP and TOP3 do not satisfy this requirement, while PUB and CIT do. In general scoring rules do not satisfy this axiom. Next two weaker axioms were considered. TOP satisfies these two weaker axioms, while TOP3 satisfied the rank-related one. The h-index, AVG and MED do not satisfy any of the weaker axioms.

While studying the h-index the notion of h-critical publications has been introduced and the set of h-critical publications is characterized.

This article has been written considering scientists, publications and citations, but can, of course, be applied to many other source-item sets.

ACKNOWLEDGEMENT

We acknowledge the National Natural Science Foundation of China (NSFC Grant No No.7101017006 and 71173187) for financial support. The authors thank Raf Guns for useful remarks. RR thanks his colleagues Fred Ye at Zhejiang University and Mu-Hsuan Huang and Dar-Zen Chen at the National University of Taiwan for their hospitality during his stay when this article has been written.

REFERENCES

Bouyssou D and Marchant T. (2010). Consistent bibliometric rankings of authors and of journals. *Journal of Informetrics*, 4(3), 365–78.

Bouyssou D, and Marchant T. (2011a). Bibliometric rankings of journals based on impact factors: An axiomatic approach. *Journal of Informetrics*, 5(1), 75–86.

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

Hirsch JE. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Science of the USA*, 102, 16569–72.

Larivière V and Gingras Y. (2011). Averages of ratios vs. ratios of averages: an empirical analysis of four levels of aggregation. *Journal of Informetrics*, 5(3), 392–9.

Liang LM and Rousseau R. (2009). A general approach to citation analysis and an *h*-Index based on the standard impact factor framework. In: *ISSI 2009* (B. Larsen & J. Leta, eds.). Sao Paulo: BIREME & Federal University of Rio de Janeiro, pp. 143–53.

Lundberg J. (2007). Lifting the crown – citation z-score. *Journal of Informetrics*, 1(2), 145–54.

Marchant T. (2009). An axiomatic characterization of the ranking based on the h-index and some other bibliometric rankings of authors. *Scientometrics*, 80(2), 325–42.

Opthof T and Leydesdorff L. (2010). Caveats for the journal and field normalizations of the CWTS ("Leiden") evaluations of research performance. *Journal of Informetrics*, 4(3), 423–30.

Rousseau R and Leydesdorff L. (2011). Simple arithmetic versus intuitive understanding: the case of the impact factor. *ISSI Newsletter*, 7(1), 10–14.

Waltman L and van Eck JN. (2009). A taxonomy of bibliometric performance indicators based on the property of consistency. In: *ISSI 2009* (B. Larsen & J. Leta, eds.). Sao Paulo: BIREME & Federal University of Rio de Janeiro, pp. 1002–3.

Waltman L and van Eck NJ. (2012). The inconsistency of the *h*-index. *Journal of the American Society for Information Science and Technology*, 63(2), 406–15.