Mathematical characterizations of the Wu- and Hirsch-indices using two types of minimal increments

Leo Egghe

leo.egghe@uhasselt.be

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

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

Abstract

For a general increasing function $f(n)$ $(n = 1, 2, 3, ...)$ we can define the most general version of the Hirsch-index being the highest rank n such that all papers on ranks $1, ..., n$ each have at least $f(n)$ citations. The minimum configuration to have this value of n is n papers each having $f(n)$ citations, hence we have $nf(n)$ citations in total. To increase the value n by one we hence need (minimally) $(n+1)f(n+1)$ citations, an increment of $I_1(n) = (n+1)f(n+1) - nf(n)$ citations. Define the increment of second order as $I_2(n) = I_1(n+1) - I_1(n)$. We characterize the general Wu-index by requiring specific values of $I_1(n)$ and $I_2(n)$, hence also characterizing the Hirsch-index.

Conference Topic

Scientometrics Indicators (Topic 1)

Introduction

The most general Hirsch-type index can be defined by using a general increasing function $f(n)$ $(n=1,2,3,...)$. The definition is as follows. Let us have a set of papers where the i^th^ paper has c_i citations (i.e. received c_i citations). We assume that papers are arranged in decreasing order of received citations (i.e. $c_i \geq c_j$ if and only if $i \leq j$). The most general Hirsch-type index can be defined as the highest rank n such that all papers on ranks $1,...,n$ have at least $f(n)$ citations. Well-known examples are $f(n)=n$ for the classical Hirsch-index (h-index), Hirsch (2005), $f(n)=an$ $(a>0)$ for the general Wu-index (Egghe (2011) and Wu (2010) for $a=10$), $f(n)=n^a$ $(a>0)$ for the general Kosmulski-index (Egghe (2011) and Kosmulski (2006) for $a=2$). Note that the general Wu- and Kosmulski-indices reduce to the h-index for $a=1$.

It is important, at least from a theoretical point of view, to know for these h-type indices, how (e.g.) an author can increase his/her h-type index value from *n* to $n+1$ (for any *n = 1,2,…*). In other words, it is important to know what effort is required from an author to increase his/her h-type index by one.

In general $c_i \geq f(n)$ for $i=1,...,n$ but in many cases we will have $c_i > f(n)$. However the minimum situation to have an index equal to n is to have n papers with exactly $f(n)$ citations each and where the other papers have zero

citations. In this case we have a total of $nf(n)$ citations. To have the minimal situation for an index equal to $n+1$, we need $n+1$ papers with exactly $f(n+1)$ citations each and where the other papers have zero citations. Now we have a total of $(n+1)f(n+1)$ citations. We define the general increment of order 1 as, for every n:

$$I_1(n)=(n+1)f(n+1)-nf(n) \tag{1}$$

The general increment of order 2 is defined as

$$I_2(n)=I_1(n+1)-I_1(n) \tag{2}$$

which is equal to, by (1)

$$I_2(n)=(n+2)f(n+2)-2(n+1)f(n+1)+nf(n) \tag{3}$$

Examples:

1. For the general Wu-index ($f(n)=an$) we have

$$I_1(n) = a(2n+1) \tag{4}$$

$$I_2(n) = 2a \tag{5}$$

for all n, as is readily seen.

This gives for the h-index:

$$I_1(n) = 2n+1 \tag{6}$$

$$I_2(n) = 2 \tag{7}$$

for all n.

2. For the general Kosmulski-index ($f(n) = n^a$) we have

$$I_1(n) = (n+1)^{a+1} - n^{a+1} \tag{8}$$

$$I_2(n) = (n+2)^{a+1} - 2(n+1)^{a+1} + n^{a+1} \tag{9}$$

for all n.

3. For the threshold index (obtained for $f(n) = C$, a constant) (called the "highly cited publications indicator" in Waltman and van Eck (2012)) we have

$$I_1(n) = C \tag{10}$$

$$I_2(n) = 0 \tag{11}$$

for all n.

In the next section we will characterize the functions $f(n)$ for which (4) is valid. It turns out that we obtain a class of functions much wider than $f(n) = an$ and from this we will characterize the general Wu-index. From this we will also obtain a characterization of the h-index. The same will be done for the threshold index.

In the third section we will characterize the functions $f(n)$ for which (5) is valid. Again it turns out that we obtain a class of functions much wider than $f(n) = an$ and from this we will newly characterize the general Wu-index. From this we will also refind a characterization of the h-index, already proved in Egghe (2012).

The paper ends with a conclusions section and with suggestions for further research.

Characterization of functions $f(n)$ that satisfy $I_1(n) = a(2n+1)$ for all n and characterization of the Wu- and Hirsch-indices and analogue for the threshold index

So we put, for all n,

$$I_1(n) = (n+1)\,f(n+1) - nf(n) = (2n+1)a \tag{12}$$

Hence

$$f(n+1) = \frac{n}{n+1}\,f(n) + a\,\frac{2n+1}{n+1} \tag{13}$$

This shows that we can choose one free parameter: $f(1) > 0$. From (13) we now have

$$f(2) = \frac{1}{2}\,f(1) + a\,\frac{3}{2} \tag{14}$$

$$f(3) = \frac{1}{3}\,f(1) + \frac{8}{3}a \tag{15}$$

(now also using (14))

$$f(4) = \frac{1}{4}\,f(1) + \frac{15}{4}a \tag{16}$$

(now also using (15)).

From this mechanism we can formulate and prove the next Theorem.

Theorem 1:

$$I_1(n) = a(2n+1)$$

for all n if and only if

$$f(n) = \frac{1}{n}\,f(1) + \frac{n^2-1}{n}a \tag{17}$$

for all n.

Proof:

The proof is by complete induction. It is clear that (17) is valid for $n = 1$ and we proved (17) for $n = 2,3,4$. Now we suppose that (17) is true for n. For $n+1$ we have by (12) (hence (13))

$$f(n+1) = \frac{n}{n+1}\,f(n) + a\,\frac{2n+1}{n+1}$$

By (17) we have

$$f(n+1) = \frac{n}{n+1}\left[\frac{1}{n}\,f(1) + \frac{n^2-1}{n}a\right] + a\,\frac{2n+1}{n+1}$$

$$f(n+1)=\frac{1}{n+1}f(1)+\frac{a}{n+1}\left(n^2-1+2n+1\right)$$

$$f(n+1)=\frac{1}{n+1}f(1)+\frac{(n+1)^2-1}{n+1}a$$

which is (17) for $n+1$. Hence (17) is valid for all n.

Reversely, if we have (17), we have to show that (12) is valid. Indeed, for all n

$$I_1(n)=(n+1)f(n+1)-nf(n)$$

$$I_1(n)=(n+1)\left[\frac{1}{n+1}f(1)+\frac{(n+1)^2-1}{n+1}a\right]-n\left[\frac{1}{n}f(1)+\frac{n^2-1}{n}a\right]$$

$$I_1(n)=(2n+1)a$$

Hence (12) is valid for all n. □

Note that, for $a=1$, we have a characterization of the Hirsch-type increment $I_1(n)=2n+1$ (see (6)).

From Theorem 1 we can prove a characterization of the general Wu-index.

Theorem 2: We have equivalent of

(i) $I_1(n)=a(2n+1)$ for all n and $f(1)=a$

(ii) $f(n)=an$ for all n (i.e. we have the Wu-index)

Proof:

(i) => (ii)

By formula (17) in Theorem 1 we have for all n

$$f(n)=\frac{a}{n}+\frac{n^2-1}{n}a$$

$$f(n)=na$$

(ii) => (i)

It was already shown in the introduction that the Wu-index satisfies (12). □

Note that Theorem 2 for $a=1$ yields a characterization of the Hirsch-index.

Note that $f(n)$ in (17) increases if $a\geq\frac{f(1)}{2}$:

$$f'(n) = \frac{n^2 a - f(1) + a}{n^2} \geq 0$$

if and only if

$$\left(n^2+1\right)a \geq f(1)$$

for all n. It suffices to require

$$2a \geq f(1)$$

or

$$a \geq \frac{f(1)}{2}$$

Now we will prove the analogue result for the threshold index. So let $f(n) = C > 0$ for all n (C: a constant). We showed in the introduction that $I_1(n) = C$ for all n. Let us characterize all functions $f(n)$ that satisfy this. So

$$I_1(n) = (n+1)\, f(n+1) - nf(n) = C \tag{19}$$

for all n. Hence

$$f(n+1) = \frac{n}{n+1} f(n) + \frac{C}{n+1} \tag{20}$$

Again we use the general parameter $f(1) > 0$. We have, by (20)

$$f(2) = \frac{1}{2} f(1) + \frac{C}{2} \tag{21}$$

$$f(3) = \frac{1}{3} f(1) + \frac{2C}{3} \tag{22}$$

(now also using (21))

$$f(4) = \frac{1}{4} f(1) + \frac{3C}{4} \tag{23}$$

(now also using (22)). Hence we can formulate and prove Theorem 3

Theorem 3: $I_1(n) = C$ for all n if and only if

$$f(n) = \frac{1}{n} f(1) + \frac{n-1}{n} C \tag{24}$$

for all n.
Proof:

The proof is by complete induction. We have already (24) for $n=1$ and proved (24) for $n=2,3,4$. Now we suppose (24) is valid for n. For $n+1$ we have by (20)

$$f(n+1)=\frac{n}{n+1}f(n)+\frac{C}{n+1}$$

$$f(n+1)=\frac{n}{n+1}\left[\frac{1}{n}f(1)+\frac{n-1}{n}C\right]+\frac{C}{n+1}$$

$$f(n+1)=\frac{1}{n+1}f(1)+C$$

which is (24) for $n+1$. So (24) is proved for all n.

Reversely, if we have (24) for all n, we have

$$I_1(n)=(n+1)f(n+1)-nf(n)$$

$$I_1(n)=(n+1)\left[\frac{1}{n+1}f(1)+\frac{n}{n+1}C\right]-n\left[\frac{1}{n}f(1)+\frac{n-1}{n}C\right]$$

$$I_1(n)=C$$

for all n. □

From Theorem 3 we can prove a characterization of the threshold index.

Theorem 4: We have equivalency of

(i) $I_1(n)=C$ for all n and $f(1)=C$

(ii) $f(n)=C$ for all n (hence the threshold index).

Proof:

(i) => (ii)
This is clear from (24), using that $f(1)=C$

(ii) => (i)
This was already proved in the introduction. □

Note that $f(n)$ in (24) increases if and only if $C\geq f(1)$. Indeed

$$f'(n) = \frac{C - f(1)}{n^2} \geq 0$$

if and only if $C \geq f(1)$.

Characterization of functions $f(n)$ that satisfy $I_2(n) = 2a$ for all n and characterization of the Wu- and Hirsch-indices and analogue for the threshold index

So we put, for all n

$$I_2(n) = (n+2)f(n+2) - 2(n+1)f(n+1) + nf(n) = 2a \tag{25}$$

Hence

$$f(n+2) = \frac{2(n+1)}{n+2}f(n+1) - \frac{n}{n+2}f(n) + \frac{2a}{n+2} \tag{26}$$

for all n. Hence we can choose two free parameters: we choose $f(1)$, $f(2)$. Since we only want to work with increasing functions $f(n)$ we suppose $f(2) \geq f(1)$. By (26) we have

$$f(3) = \frac{4}{3}f(2) - \frac{1}{3}f(1) + \frac{2a}{3} \tag{27}$$

$$f(4) = \frac{6}{4}f(2) - \frac{2}{4}f(1) + \frac{6a}{4} \tag{28}$$

(now also using (27))

$$f(5) = \frac{8}{5}f(2) - \frac{3}{5}f(1) + \frac{12}{5}a \tag{29}$$

(now also using (28)).

Hence we can formulate and prove Theorem 5.

Theorem 5: $I_2(n) = 2a$ for all n if and only if

$$f(n) = \frac{1}{n}\left[2(n-1)f(2) - (n-2)f(1) + (n-1)(n-2)a\right] \tag{30}$$

for all n.

Proof:

The proof is by complete induction. We already proved (30) for $n = 3,4,5$ and is easy to see for $n = 1,2$. Now we suppose that (30) is valid for n and $n+1$. For $n+2$ we have , by (25)

$$f(n+2) = \frac{2(n+1)}{n+2}\left[\frac{2nf(2)-(n-1)f(1)+n(n-1)a}{n+1}\right] - \frac{n}{n+2}\left[\frac{2(n-1)f(2)-(n-2)f(1)+(n-1)(n-2)a}{n}\right] + \frac{2a}{n+2}$$

$$f(n+2) = \frac{1}{n+2}\left[2(n+1)f(2)-nf(1)+n(n+1)a\right] \tag{31}$$

after an elementary calculation. Now (31) is (30) for $n+2$.

Reversely, if (30) is valid for all n, it is an elementary calculation, using (25), that $I_2(n) = 2a$ for all n.

□

From Theorem 5 we can prove a characterization of the general Wu-index.

Theorem 6: We have equivalency of

(i) $I_2(n) = 2a$, for all n and $f(1) = a$ and $f(2) = 2a$

(ii) $f(n) = na$ for all n (hence we have the general Wu index).

Proof:

(i) => (ii)
It follows from (30) in Theorem 5 that, for $f(1) = a$, $f(2) = 2a$ that $f(n) = na$ for all n.

(ii) => (i)
We proved in the introduction that the Wu-index satisfies $I_2(n) = 2a$ for all n.

□

Note that, for $a = 1$, Theorem 6 is a characterization of the Hirsch-index, which appeared already in Egghe (2012).

Note: It is easy to see that $f(n)$ in (30) is an increasing function. This can be shown using (30) by calculating $f'(n)$ or by (26) using complete induction (and, in both cases, using that $f(1) \le f(2)$).

For the sake of completeness we also mention the following characterization of $I_2(n) = 0$ for all n and of the threshold index.

Theorem 7 (Egghe (2012)): $I_2(n)=0$ for all n if and only if

$$f(n)=\frac{2(n-1)f(2)-(n-2)f(1)}{n} \tag{32}$$

for all n.

Theorem 8 (Egghe (2012)): The following assertions are equivalent:

(i) $I_2(n)=0$ for all n, $f(1)=f(2)=C$ a positive constant.

(ii) $f(n)=C$ for all n, i.e. we have the threshold index.

Conclusions and suggestions for further research

In this paper we characterized functions for which $I_1(n)=(2n+1)a$ for all n. As a consequence we proved a characterization of the general Wu-index, hence also of the h-index.

We then characterized functions for which $I_2(n)=2a$ for all n. As a consequence we proved a new characterization of the general Wu-index, hence also of the h-index.

For the threshold index we executed the same exercise leading to characterizations of the threshold index.

We invite the reader to elaborate further studies on $I_1(n)$ and $I_2(n)$, hereby characterizing other known and new impact indices. We stress the importance of such studies, at least from a theoretical point of view. Characterizing indices which require a certain increment of citations in order to increase the index with one unit shows what effort is required from the author to reach this increase.

References

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

Egghe L. (2012). A mathematical characterization of the Hirsch-index by means of minimal increments. Preprint.

Hirsch J.E. (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.

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.

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

Wu Q. (2010). The w-index: A measure to assess scientific impact by focusing on widely cited papers. *Journal of the American Society for Information Science and Technology*, 61(3), 609-614.