Infinite sequences and their h-type indices

Leo Egghe¹ and Ronald Rousseau^{2,3}

¹University of Hasselt, Belgium

leo.egghe@uhasselt.be

²University of Antwerp, Faculty of Social Sciences, B-2020 Antwerp, Belgium; ronald.rousseau@uantwerpen.be

and

³KU Leuven, Facultair Onderzoekscentrum ECOOM, Naamsestraat 61, Leuven B-3000, Belgium; ronald.rousseau@kuleuven.be

Abstract

Starting from the notion of h-type indices for infinite sequences we investigate if these indices satisfy natural inequalities related to the arithmetic, the geometric and the harmonic mean. If f denotes an h-type index, such as the h- or the g-index, then we investigate inequalities such as $\min(f(X), f(Y)) \leq f((X + Y)/2) \leq \max(f(X), f(Y))$. We further investigate if: $f(\min(X, Y)) = \min(f(X), f(Y))$ and if $f(\max(X, Y)) = \max(f(X), f(Y))$. It is shown that the h-index satisfies all the equalities and inequalities we investigate but the g-index does not always, while it is always possible to find a counterexample involving the R-index. This shows that the h-index enjoys a number of interesting mathematical properties as an operator in the partially ordered positive cone $(\mathbf{R}^+)^{\infty}$ of all infinite sequences with non-negative real values.

In a second part we consider decreasing vectors X and Y with components at most at distance d . Denoting by D the constant sequence $(d, d, d, \dots)$ and by $Y-D$ the vector $(\max(y_r - d, 0))_r$, we prove that under certain natural conditions, the double inequality $h(Y-D) \leq h(X) \leq h(Y+D)$ holds.

Keywords: inequalities; h-index; generalized h-index; g-index; R-index

1. Introduction

In this contribution we return to the investigation of mathematical properties of basic h- and g-indices (and a few generalizations such as Kosmulski's index) (Egghe, 2006a, 2006b; Hirsch, 2005; Kosmulski, 2006). We do this in a framework of infinite sequences of non-negative real values. Natural inequalities about h-type indices on the one hand, and the arithmetic, geometric and harmonic mean of infinite sequences, on the other are investigated. We further describe vectors situated in the neighborhood (defined in detail further on) of a vector with a given h, $h^{(2)}$, g or R index.

Although we realize that the h-index and other h-type indices are at best heuristics when applied in a research evaluation context, we still think that they are interesting operators on publication-citation arrays and their behavior in this context may lead to a deeper insight supporting their acceptance, say tolerating them, as useful heuristics (Bornmann & Marewski, 2018).

2. Definitions

Let $(\mathbf{R}^+)^{\infty}$ be the positive cone of all infinite sequences with non-negative real values. Addition and multiplication by a non-negative scalar are performed component wise. Similarly, multiplication of two such vectors is performed component by component. Finally, if $X = (x_r)_{r=1,2,\dots}$ is a vector in $(\mathbf{R}^+)^{\infty}$ then X^s , where s is any non-negative real number, is defined as $X^s = ((x_r)^s)_{r=1,2,\dots}$. Any finite sequence with non-negative values can be considered as an element in $(\mathbf{R}^+)^{\infty}$ by adding infinitely many zeros. The positive cone $(\mathbf{R}^+)^{\infty}$ has a partial order defined by $X \leq Y$ if for all $r = 1, 2, \dots$ $x_r \leq y_r$. Equality only occurs if for all r $x_r = y_r$. A vector $X = (x_r)_{r=1,2,\dots}$ in $(\mathbf{R}^+)^{\infty}$ is said to be decreasing if for all $r = 1, 2, \dots$ $x_r \geq x_{r+1}$.

Each set S of publications and corresponding citations can be represented by a vector in $(\mathbf{R}^+)^{\infty}$, denoted as X_S , and vice-versa: a vector X in $(\mathbf{R}^+)^{\infty}$ can be considered to be a representation of a, possibly infinite, abstract set S_X , of publications and citations. This abstract set is referred to as the underlying set of vector X . If the set U is the union of two sets S and T then $X_U \geq X_S$ and $X_U \geq X_T$.

Next we formulate the definition of some h-type indices for infinite sequences, i.e., for vectors in $(\mathbf{R}^+)^{\infty}$.

2.1 Definition: The h-index (Hirsch, 2005)

Let $X = (x_r)_{r=1,2,\dots}$ be a decreasing vector in $(\mathbf{R}^+)^{\infty}$. The h-index of X , denoted $h(X)$, is the largest natural number such that the first h

coordinates have each at least a value h . If all components of a decreasing vector X are strictly smaller than 1, then $h(X) = 0$. We will further consider only vectors X with at least one component larger than or equal to 1.

Note that an h -index, and similarly for the other h -type indices defined further, can only be defined for decreasing vectors. Moreover, we have the following characterizations of the h -index: let A and B be natural numbers and let $X = (x_r)_{r=1,2,\dots}$ be a decreasing vector in $(\mathbf{R}^+)^{\infty}$ then (for $r \leq A$: $x_r \geq A$) implies that $h(X) \geq A$; similarly: (for $r > B$: $x_r < r$) implies $h(X) \leq B$.

2.2 Definition: The $h^{(k)}$ -index (Deneko & Woeginger, 2009; Egghe, 2011; Fassin & Rousseau, 2018; Kosmulski, 2006).

Let $X = (x_r)_{r=1,2,\dots}$ be a decreasing vector in $(\mathbf{R}^+)^{\infty}$. The $h^{(k)}$ -index of X , denoted $h^{(k)}(X)$, is the largest natural number such that the first h coordinates have each at least a value equal to h^k .

Also here we have the characterizations: let A and B be natural numbers and let $X = (x_r)_{r=1,2,\dots}$ be a decreasing vector in $(\mathbf{R}^+)^{\infty}$ then (for $r \leq A$: $x_r \geq A^k$) implies that $h^{(k)}(X) \geq A$; similarly: (for $r > B$: $x_r < r^k$) implies $h^{(k)}(X) \leq B$.

2.3 Definition: The g -index (Egghe, 2006a, 2006b)

Let $X = (x_r)_{r=1,2,\dots}$ be a decreasing vector in $(\mathbf{R}^+)^{\infty}$. The g -index of X , denoted $g(X)$, is defined as the highest natural number g such that the sum of the first g coordinates is at least equal to g^2 .

2.4 Definition: The R -index (Jin et al., 2007)

Let $X = (x_r)_{r=1,2,\dots}$ be a decreasing vector in $(\mathbf{R}^+)^{\infty}$. The R -index of X is defined as the square root of the sum of all coordinates up to and including the one with index $h(X)$. Omitting the square root yields the R^2 -index. As it is easier to work with R^2 than with R while their properties are for our purposes the same all concrete examples will be given for R^2 .

2.5 Elementary propositions

Proposition 1

Every decreasing vector in $(\mathbf{R}^+)^{\infty}$ has an h -index.

We note that every decreasing vector is bounded. Assume now that $X = (x_r)_r$ has no h -index: then for every r : $x_r > r$. This contradicts the fact that X is bounded.

Similarly we see that every decreasing vector in $(\mathbf{R}^+)^{\infty}$ has an $h^{(k)}$, g and R -index.

Proposition 2

All h-type indices defined above are order-preserving. i.e., if $X \leq Y$ then $h(X) \leq h(Y)$ with similar inequalities for the other h-type indices. This follows immediately from their definition.

An indicator IND is independent in the sense of Bouyssou and Marchant (2001) if the following holds: suppose scientist S_1 is at least as good, according to the IND-value, as scientist S_2 . Suppose now that both of them publish one or more additional papers with the same number of citations. Then these two scientists should compare in the same way as previously, i.e., the IND-value of scientist S_1 is still equal to or larger than the IND-value of scientist S_2 . Bouyssou and Marchant (2011) as well as Waltman and van Eck (2012) have shown that the h-index is not independent. Yet, the h-index has the following property.

Proposition 3

If $X \leq Y$, and hence $h(X) \leq h(Y)$, with underlying sets S and U , and if one adds the same publications and citations to each of them, leading to underlying sets T and V then clearly $X_T \leq X_V$ and hence $h(X_T) \leq h(X_V)$.

3. Requirements for informetric indicators

3.1 Definitions

Let X, Y be decreasing vectors in $(\mathbf{R}^+)^{\infty}$. Then we form the following new vectors:

(i) For each $\theta \in [0, 1]$ we form the weighted arithmetic mean of X and Y :

$$\theta X + (1 - \theta) Y.$$

(ii) For each $s, t \in [0, 1]$, $s+t = 1$, we form the generalized geometric mean:

$$X^s Y^t.$$

3.2 Natural requirements for h-type indices

Let f be a candidate informetric indicator. Then we formulate two “natural” requirements, denoted as (WA) and (GG). Here WA stands for weighted average and GG stands for generalized geometric mean. We note that we do not expect all informetric indicators to meet these requirements, but, in our opinion, indicators related to measuring ‘impact’ should.

Property (WA). For each $\theta \in [0, 1]$:

$$\min(f(X), f(Y)) \leq f(\theta X + (1 - \theta) Y) \leq \max(f(X), f(Y))$$

Property (GG). For each $s, t \in [0,1]$, $s + t = 1$

$$\min(f(X), f(Y)) \leq f(X^s Y^t) \leq \max(f(X), f(Y))$$

In the next theorem we provide answers to the question if $h^{(k)}$ (for all k , including $k = 1$), g and R satisfy properties (WA) and (GG).

3.3 Relations between the h -, the g - and the R -index on the one hand and properties (WA) and (GG) on the other

Theorem 1

- a) The h -index and all $h^{(k)}$ indices satisfy properties (WA) and (GG).
- b) The g -index satisfies property (WA) but not property (GG). Yet, it satisfies the second inequality of property (GG).
- c) The R -index satisfies not a single inequality included in properties (WA) or (GG).

Proof.

For each index and each property we have to prove two inequalities, one on the left-hand side, involving a minimum, and one on the right-hand side, involving a maximum.

- a) Property (WA): we have to prove that for each $\theta \in [0,1]$:

$$\min(h(X), h(Y)) \leq h(\theta X + (1 - \theta) Y) \leq \max(h(X), h(Y))$$

First step: proof of $\min(h(X), h(Y)) \leq h(\theta X + (1 - \theta) Y)$.

For all $r \leq h(X)$ we know that $x_r \geq h(X)$; similarly: for all $r \leq h(Y)$ we know that $y_r \geq h(Y)$. Hence: for all $r \leq \min(h(X), h(Y))$: $\theta x_r + (1 - \theta) y_r \geq \theta h(X) + (1 - \theta) h(Y) \geq \theta \min(h(X), h(Y)) + (1 - \theta) \min(h(X), h(Y)) = \min(h(X), h(Y))$. Hence $h(\theta X + (1 - \theta) Y) \geq \min(h(X), h(Y))$.

Second step: proof of $h(\theta X + (1 - \theta) Y) \leq \max(h(X), h(Y))$.

For all $r > h(X)$ we know that $x_r < r$; similarly: for all $r > h(Y)$ we know that $y_r < r$. Hence: for all $r > \max(h(X), h(Y))$: $\theta x_r + (1 - \theta) y_r = (\theta X + (1 - \theta) Y)_r < r$. Hence $h(\theta X + (1 - \theta) Y) \leq \max(h(X), h(Y))$.

- a) Property (GG): we have to show that for each $s, t \in [0,1]$, $s + t = 1$

$$\min(h(X), h(Y)) \leq h(X^s Y^t) \leq \max(h(X), h(Y))$$

First step : proof of the inequality: $\min(h(X), h(Y)) \leq h(X^s Y^t)$.

For all $r \leq \min(h(X), h(Y))$ we know that $x_r \geq r$ and $y_r \geq r$. Hence $(x_r)^s \cdot (y_r)^t \geq r^{(s+t)} = r$, leading to the conclusion that $h(X^s Y^t) \geq \min(h(X), h(Y))$.

Second step: proof of the inequality: $h(X^s Y^t) \leq \max(h(X), h(Y))$.

We know that for all $r > \max(h(X), h(Y))$: $(x_r)^s \cdot (y_r)^t < r^{(s+t)} = r$.

Similar proofs holds for all other $h^{(k)}$. Details can be obtained from the authors.

b) Property (WA): we have to prove that for each $\theta \in [0, 1]$:

$$\min(g(X), g(Y)) \leq g(\theta X + (1-\theta) Y) \leq \max(g(X), g(Y))$$

First step: proof of the inequality: $\min(g(X), g(Y)) \leq g(\theta X + (1-\theta) Y)$.

We know that for all $j \leq g(X)$: $\sum_{r=1}^j x_r \geq j^2$ and similarly, for all $j \leq g(Y)$: $\sum_{r=1}^j y_r \geq j^2$. From these two inequalities we see that for all $j \leq \min(g(X), g(Y))$: $\sum_{r=1}^j (\theta x_r + (1-\theta)y_r) = \theta(\sum_{r=1}^j x_r) + (1-\theta)(\sum_{r=1}^j y_r) \geq j^2$. From this we conclude that $g(\theta X + (1-\theta)Y) \geq \min(g(X), g(Y))$.

Second step: proof of the inequality: $g(\theta X + (1-\theta) Y) \leq \max(g(X), g(Y))$.

We know that for all $j > g(X)$: $\sum_{r=1}^j x_r < j^2$ and similarly, for all $j > g(Y)$: $\sum_{r=1}^j y_r < j^2$. From these two inequalities we derive that for all $j > \max(g(X), g(Y))$: $\sum_{r=1}^j (\theta x_r + (1-\theta)y_r) = \theta(\sum_{r=1}^j x_r) + (1-\theta)(\sum_{r=1}^j y_r) < j^2$. Hence $g(\theta X + (1-\theta)Y) \leq \max(g(X), g(Y))$.

b) Property (GG): we have to prove that for each $s, t \in [0, 1]$, $s + t = 1$

$$\min(g(X), g(Y)) \leq g(X^s Y^t) \leq \max(g(X), g(Y)).$$

We begin with the proof of $g(X^s Y^t) \leq \max(g(X), g(Y))$ as this inequality holds, while the other one does not.

From Hölder's inequality we derive that for $j > \max(g(X), g(Y))$:

$$\sum_{r=1}^j x_r^s \cdot y_r^t \leq \left(\sum_{r=1}^j x_r \right)^s \cdot \left(\sum_{r=1}^j y_r \right)^t < (j^2)^s \cdot (j^2)^t = j^2$$

Consequently: $g(X^s Y^t) \leq \max(g(X), g(Y))$.

Next we provide a counterexample, showing that the inequality $\min(g(X), g(Y)) \leq g(X^s Y^t)$ does not always hold.

Take $s=t = \frac{1}{2}$ (the geometric mean), $X=(3,1)$, $Y=(2,2)$. Then $g(X)=2$ and $g(Y)=2$; $\min(g(X),g(Y)) = 2$. Now $\sqrt{XY} = (\sqrt{6}, \sqrt{2})$; $\sqrt{6} > 1$, but $\sqrt{6} + \sqrt{2} \approx 3.86 < 4$. Hence $g(X^s.Y^t) = 1 < \min(g(X),g(Y))$, in this case for $s=t=1/2$.

Giving examples we show that the R^2 -index and hence the R -index, does not satisfy any of the two properties.

c) Property (WA). We consider the inequalities below and show that a parameter $\theta \in [0,1]$ exists such that the two inequalities

$$\min(R^2(X), R^2(Y)) \leq R^2(\theta X + (1-\theta) Y) \leq \max(R^2(X), R^2(Y))$$

do not hold.

First step. A counterexample for $\min(R^2(X), R^2(Y)) \leq R^2(\theta X + (1-\theta) Y)$.

Let $\theta = \frac{1}{2}$ (the arithmetic mean); take $X = (3,1)$ and $Y = (2,2)$. Then $h(X)=1$ and $h(Y) = 2$; $R^2(X)= 3$ and $R^2(Y)= 4$. Now $(X+Y)/2 = (5/2, 3/2)$ and hence $h((X+Y)/2) = 1$; $R^2((X+Y)/2) = 5/2 < \min(R^2(X), R^2(Y)) = 3$ (as $R^2(X)=3$ and $R^2(Y)=4$).

Second step. A counterexample for $R^2(\theta X + (1-\theta) Y) \leq \max(R^2(X), R^2(Y))$.

Again let $\theta = \frac{1}{2}$; take $X = (4,4,4,0)$ and $Y = (6,6,2,0)$. Then $h(X)=3$ and $h(Y) = 2$; $R^2(X) = 12$ and $R^2(Y)=12$. Now $(X+Y)/2 = (5,5,3,0)$ and hence $h((X+Y)/2) = 3$ and $R^2((X+Y)/2) = 13 > \max(R^2(X), R^2(Y)) = 12$.

c) Property (GG). We consider the inequalities below and show that there exist $s, t \in [0,1]$, $s + t = 1$, such that the inequalities

$$\min(R^2(X), R^2(Y)) \leq R^2(X^s Y^t) \leq \max(R^2(X), R^2(Y)).$$

do not hold.

First step. A counterexample for $\min(R^2(X), R^2(Y)) \leq R^2(X^s Y^t)$.

We take $X = (3,1)$ and $Y = (2,2)$, with $s=t=1/2$. Now $\sqrt{XY} = (\sqrt{6}, \sqrt{2})$; $h(\sqrt{XY}) = 1$ and $R^2(\sqrt{XY}) = \sqrt{6} \approx 2.45 < \min(R^2(X), R^2(Y)) = 3$.

Second step. A counterexample for $R^2(X^s Y^t) \leq \max(R^2(X), R^2(Y))$.

Again we take $s=t=1/2$. We take $X = (5,5,5,0)$ and $Y = (8,7,2, 0)$. Now, $h(X) = 3$; $h(Y) = 2$; $R^2(X) = 15$ and $R^2(Y)=15$.

$\sqrt{XY} = (\sqrt{40}, \sqrt{35}, \sqrt{10}, 0)$; $h(\sqrt{XY}) = 3$ and $R^2(\sqrt{XY}) = \sqrt{40} + \sqrt{35} + \sqrt{10} \approx 15.4 > \max(R^2(X), R^2(Y)) = 15$.

3.4 A property related to the harmonic mean

Although probably of less importance than the previous properties, for completeness sake we also consider a property related to the harmonic mean.

We recall that if a and b are non-negative numbers then

$$\min(a,b) \leq \text{hm}(a,b) \leq \text{geom}(a,b) \leq \text{arith}(a,b) \leq \max(a,b) \quad (*),$$

where hm stands for the harmonic mean, geom for the geometric mean and arith for the arithmetic mean.

The harmonic mean of two vectors X and Y in $(\mathbf{R}^+)^{\infty}$ is defined as the vector $\text{HM}(X,Y)$, with components $(\text{HM}(X,Y))_r = \text{hm}(x_r, y_r) = \frac{2x_r y_r}{x_r + y_r}$, if x_r or $y_r \neq 0$, and equal to 0 otherwise.

Now we investigate the property:

$$(\text{HM}) : \min(f(X), f(Y)) \leq f(\text{HM}(X,Y)) \leq \max(f(X), f(Y))$$

for $f = h, g$ and R . This leads to the following Theorem.

Theorem 2

a) $\min(h(X), h(Y)) \leq h(\text{HM}(X,Y)) \leq \max(h(X), h(Y))$

b) The g -index satisfies the relation $g(\text{HM}(X,Y)) \leq \max(g(X), g(Y))$, i.e., the right-hand side inequality in (HM), but does not satisfy the inequality on the left-hand side of (HM).

c) The R -index does not satisfy any of the two inequalities in (HM).

Proof

a) First step: proof of $\min(h(X), h(Y)) \leq h(\text{HM}(X,Y))$.

For all $r \leq h(X)$ we know that $x_r \geq h(X)$; similarly: for all $r \leq h(Y)$ we know that $y_r \geq h(Y)$. Hence: for all $r \leq \min(h(X), h(Y))$, $\min(x_r, y_r) \geq \min(h(X), h(Y))$. Hence by (*): $\text{hm}(x_r, y_r) \geq \min(h(X), h(Y))$. This proves that $\min(h(X), h(Y)) \leq h(\text{HM}(X,Y))$.

Second step: proof of $h(\text{HM}(X,Y)) \leq \max(h(X), h(Y))$

For all $r > h(X)$ we know that $x_r < r$; similarly: for all $r > h(Y)$ we know that $y_r < r$. Hence: for all $r > \max(h(X), h(Y))$: $\max(x_r, y_r) < r$. Hence, again by (*): $\text{hm}(x_r, y_r) = (\text{HM}(X,Y))_r < r$. This proves that $h(\text{HM}(X,Y)) \leq \max(h(X), h(Y))$.

b) The g-index and its relation with the harmonic mean of vectors. We investigate if

$$\min(g(X), g(Y)) \leq g(\text{HM}(X, Y)) \leq \max(g(X), g(Y))$$

Similar to Property (GG) the inequality on the left-hand side does not hold, but that on the right-hand side does. We first prove the inequality which holds.

We know that for all $j > g(X)$: $\sum_{r=1}^j x_r < j^2$ and similarly, for all $j > g(Y)$: $\sum_{r=1}^j y_r < j^2$. If now $j > \max(g(X), g(Y))$, we obtain from a variant of the Cauchy-Schwarz-Bunyakovski inequality and (*) that:

$$\sum_{r=1}^j hm(x_r, y_r) \leq \sum_{r=1}^j geom(x_r, y_r) = \sum_{r=1}^j \sqrt{x_r y_r} \leq \sqrt{\sum_{r=1}^j x_r} \sqrt{\sum_{r=1}^j y_r} < j \cdot j = j^2$$

Consequently: $g(\text{HM}(X, Y)) \leq \max(g(X), g(Y))$.

Next we provide a counterexample, showing that the inequality $\min(g(X), g(Y)) \leq g(\text{HM}(X, Y))$ does not always hold.

Let $X = (9, 0, 0, 0, \dots)$ and $Y = (3, 3, 3, 3, \dots)$. Then $g(X) = g(Y) = 3$. Now $\text{HM}(X, Y) = (54/12, 0, 0, 0, \dots) = (4.5, 0, 0, 0, \dots)$. Then $g(\text{HM}(X, Y)) = 2 < 3 = \min(g(X), g(Y))$.

c) The R-index and its relation with the harmonic mean of vectors. We investigate if

$$\min(R(X), R(Y)) \leq R(\text{HM}(X, Y)) \leq \max(R(X), R(Y))$$

Also here R does not satisfy any of these inequalities.

First step. A counterexample for $\min(R(X), R(Y)) \leq R(\text{HM}(X, Y))$.

Reconsidering the previous counterexample we see that for $X = (9, 0, 0, \dots)$, $h(X) = 1$ and $R^2(X) = 9$; for $Y = (3, 3, 3, 3, \dots)$, $h(Y) = 3$ and $R^2(Y) = 9$. As $\text{HM}(X, Y) = (4.5, 0, 0, 0, \dots)$ we see that $h(\text{HM}(X, Y)) = 1$ and $R^2(\text{HM}(X, Y)) = 4.5 < 9 = \min(R^2(X), R^2(Y))$ providing a counterexample for the first inequality.

Second step. A counterexample for $R(\text{HM}(X, Y)) \leq \max(R(X), R(Y))$.

Let $X = (8, 8, 8, 8, 8, 8, 0, 0, \dots)$. Then $h(X) = 6$ and $R^2(X) = 48$. Let $Y = (9, 9, 9, 9, 9, 5, 5, \dots)$. Then $h(Y) = 5$ and $R^2(Y) = 45$. As $\text{HM}(X, Y) = (144/17, 144/17, 144/17, 144/17, 80/13, 0, 0, \dots)$, $h(\text{HM}(X, Y)) = 6$ and $R^2(\text{HM}(X, Y)) \approx 48.51 > \max(R^2(X), R^2(Y)) = 48$.

3.5 Min and max properties

Definition

If X and Y are decreasing vectors in $(\mathbf{R}^+)^{\infty}$ then we define the vectors $\min(X,Y)$ and $\max(X,Y)$ pointwise, i.e.,

$$(\min(X,Y))_r = \min(x_r, y_r) \text{ and } (\max(X,Y))_r = \max(x_r, y_r)$$

We immediately note that $\min(X,Y)$ and $\max(X,Y)$ are also decreasing vectors in $(\mathbf{R}^+)^{\infty}$.

In this section we investigate if $f(\min(X,Y)) = \min(f(X), f(Y))$ and if $f(\max(X,Y)) = \max(f(X), f(Y))$, again for $f = h, h^{(k)}, g$ and R .

We first note that $\min(X,Y) \leq X$ (and $\leq Y$), hence by Property 2, $f(\min(X,Y)) \leq \min(f(X), f(Y))$. Similarly X (and Y) $\leq \max(X,Y)$ and hence $\max(f(X), f(Y)) \leq f(\max(X,Y))$. In order to prove or disprove the announced equalities we only have to study the opposite inequality.

3.5.1 The h -index

Proposition 4: $\min(h(X), h(Y)) \leq h(\min(X,Y))$.

Proof.

Without restriction we may assume that $h(X) \leq h(Y)$. Hence we have to show that $h(X) \leq h(\min(X,Y))$. Now $y_{h(X)} \geq y_{h(Y)} \geq h(Y) \geq h(X)$ and $x_{h(X)} \geq h(X)$. Hence $\min(X,Y)_{h(X)} \geq h(X)$ which shows that $h(\min(X,Y)) \geq h(X) = \min(h(X), h(Y))$.

Proposition 5: $h(\max(X,Y)) \leq \max(h(X), h(Y))$.

Proof.

Again we assume that $h(X) \leq h(Y)$. Hence we have to show that $h(\max(X,Y)) \leq h(Y)$. Now $x_{h(Y)+1} \leq x_{h(X)+1} < h(X)+1 \leq h(Y)+1$ and $y_{h(Y)+1} < h(Y)+1$. This shows that $\max(x_{h(Y)+1}, y_{h(Y)+1}) < h(Y)+1$ and thus: $h(\max(X,Y)) \leq h(Y) = \max(h(X), h(Y))$.

Corollary 1. $h(\max(X,Y)) - h(\min(X,Y)) = |h(X) - h(Y)|$.

Proof. Assume, that $h(X) \geq h(Y)$, then $h(\max(X,Y)) = \max(h(X), h(Y)) = h(X)$ and $h(\min(X,Y)) = \min(h(X), h(Y)) = h(Y)$. Consequently, $h(\max(X,Y)) - h(\min(X,Y)) = h(X) - h(Y) = |h(X) - h(Y)|$. A similar argument yields the same result in the case that $h(X) \leq h(Y)$.

Remark 1. By property 2, $h(\min(X,Y)) \leq h(\max(X,Y))$. Now we provide an example where $\min(X,Y) < \max(X,Y)$ and yet there is equality for their h -indices. Let $X=(3,3,0)$, $Y=(5,2,0)$. Then $\min(X,Y) = (3,2,0)$, with $h(\min(X,Y)) = 2$; $\max(X,Y) = (5,3,0)$, with $h(\max(X,Y)) = 2$. Next we provide an example where again $\min(X,Y) < \max(X,Y)$, but $h(\min(X,Y)) < h(\max(X,Y))$. Let $X = (3,3,0)$ and $Y=(4,1,0)$. Then $\min(X,Y) = (3,1,0)$ with $h(\min(X,Y)) = 1$; $\max(X,Y) = (4,3,0)$ and $h(\max(X,Y)) = 2 > h(\min(X,Y))$.

Remark 2. As $h(\min(X,Y))=\min(h(X),h(Y))$ it is often (unless X or $Y = \min(X,Y)$), possible to find a vector strictly smaller than X and Y , namely $\min(X,Y)$ such that at least one of the h -indices of X or Y stays the same. Similarly we observe that from the equality $h(\max(X,Y)) = \max(h(X),h(Y))$ it is possible to find a vector strictly larger than X and Y , namely $\max(X,Y)$ such that at least one of the h -indices of X and Y stays the same (again, unless X or $Y = \max(X,Y)$).

It is clear that these properties also hold for all $h^{(k)}_S$.

3.5.2 The g -index.

We provide a counterexample for the two inequalities we have to study.

Let $X = (4,3,2)$, $Y= (11,2,2)$. Then $g(X) = 3$, $g(Y) = 3$. $\min(X,Y) = (4,2,2)$, with $g(\min(X,Y)) = 2 < \min(g(X), g(Y))$.

$\max(X,Y) = (11,3,2)$, with $g(\max(X,Y)) = 4 > \max(g(X),g(Y))$.

3.5.3 The R -index

Also here we can provide counterexamples.

Let $X = (4,3,2)$, $Y = (5,2,2)$. Then $h(X) = 2$, $R^2(X) = 7$; $h(Y) = 2$, $R^2(Y) = 7$. $\min(X,Y) = (4,2,2)$, with $h(\min(X,Y)) = 2$ and $R^2(\min(X,Y)) = 6 < \min(R^2(X), R^2(Y)) = 7$.

$\max(X,Y) = (5,3,2)$ with $h(\max(X,Y)) = 2$ and $R^2(\max(X,Y)) = 8 > \max(R^2(X), R^2(Y)) = 7$.

3.6 Min and max properties for the h -index of more than two vectors

Definition

If n is a natural number at least equal to one and if $X_1, \dots, X_n$ are decreasing vectors in $(\mathbf{R}^+)^{\infty}$ then the vectors $\min(X_1, \dots, X_n)$ and $\max(X_1, \dots, X_n)$ are defined pointwise, similar to the case $n=2$.

Clearly $\min(X_1, \dots, X_n)$ and $\max(X_1, \dots, X_n)$ are also decreasing vectors in $(\mathbf{R}^+)^{\infty}$.

We now prove the following theorem for every natural number n , at least equal to one.

Theorem 3

- A. $h(\min(X_1, \dots, X_n)) = \min(h(X_1), \dots, h(X_n))$
- B. $h(\max(X_1, \dots, X_n)) = \max(h(X_1), \dots, h(X_n))$.

The proof goes by complete induction.

- 1) We observe that the theorem holds for $n = 1$ (trivial) and $n=2$ (shown above).
- 2) Assume now that the theorem holds for $n-1$. Then we note that, by construction, the vector $\min(X_1, \dots, X_n) = \min(\min(X_1, \dots, X_{n-1}), X_n)$. As we know the result for $n = 2$, we have $h(\min(X_1, \dots, X_n)) = \min(h(\min(X_1, \dots, X_{n-1})), h(X_n))$. By the induction hypothesis this leads to $h(\min(X_1, \dots, X_n)) = \min(\min(h(X_1), \dots, h(X_{n-1})), h(X_n)) = \min(h(X_1), \dots, h(X_n))$.
- 3) The same arguments hold with $\min$ replaced by $\max$, showing part B.

4. Arrays related to h-type indices

In this section we describe vectors situated in the neighborhood of a vector with a given h , $h^{(k)}$, g or R index.

For $d > 0$ we denote by D the constant sequence $(d, d, d, \dots)$.

If $Y \in (\mathbf{R}^+)^{\infty}$ then the vector $Y-D$ is defined as $(\max(y_r - d, 0))_{r=1,2,\dots}$

The next result describes the h -index of vectors related to a vector with h -index $h(X)$.

Theorem 4

Let $d > 0$ and a decreasing vector $X = (x_r)_{r=1,2,\dots}$ in $(\mathbf{R}^+)^{\infty}$ be given. Let further Y be a decreasing vector in $(\mathbf{R}^+)^{\infty}$ such that $\sup(|y_r - x_r|) \leq d$, then $h(Y-D) \leq h(X) \leq h(Y+D)$.

Proof.

We first note that the vectors $Y+D$ and $Y-D$ are decreasing, and hence have an h -index.

As $\sup(|y_r - x_r|) \leq d$ we derive that $\forall r = 1, 2, \dots: |y_r - x_r| \leq d$, hence: $y_r - d \leq x_r$. Now, $r > h(X)$ implies that $x_r < r$. Hence: for all $r > h(X)$: $y_r - d < r$ or $h(Y-D) \leq h(X)$.

Similarly, from the fact that $\forall r = 1, 2, \dots: |y_r - x_r| \leq d$ we obtain $y_r + d \geq x_r$. Applying this inequality for $r=1, \dots, h(X)$, yields: $y_r + d \geq x_r \geq h(X)$. Consequently: $h(Y+D) \geq h(X)$.

Corollary 2. Let $d > 0$ and a decreasing vector $X = (x_r)_{r=1,2,\dots}$ in $(\mathbf{R}^+)^{\infty}$ be given. Let further Y be a decreasing vector in $(\mathbf{R}^+)^{\infty}$ such that for some $p \geq 1$, $\|X-Y\|_p \leq d$, then $h(Y-D) \leq h(X) \leq h(Y+D)$.

Proof. Using Jensen's inequality: $\sup(|y_r - x_r|) = \|X-Y\|_{\infty} \leq \|X-Y\|_p$ (Hardy et al., 1934, p. 28-29) this result follows immediately from Theorem 4.

We observe that this corollary holds in particular for the Euclidean norm, $p = 2$.

Two examples for the case of the Euclidean norm. Let $X_1 = (10, 5, 3, 0, 0, \dots)$, $h(X_1) = 3$. Let $(Y_1) = (15, 10, 5, 0, 0, \dots)$, $h(Y_1) = 3$. $\|X_1 - Y_1\|_2 = \sqrt{5^2 + 5^2 + 2^2} = \sqrt{54} < 8$. Now we form the vectors $Y_1 - 8 = (7, 2, 0, 0, 0, \dots)$ and $Y_1 + 8 = (23, 18, 13, 8, 8, \dots)$. We see that $h(Y_1 - 8) = 2 < h(X) = 3 < h(Y_1 + 8) = 8$. This is an example where we have strict inequalities.

Next we consider $X_2 = (12, 6, 4, 0, 0, \dots)$ with $h(X_2) = 3$ and $Y_2 = (11, 7, 6, 0, 0, \dots)$ with $h(Y_2) = 3$. $\|X_2 - Y_2\|_2 = \sqrt{1^2 + 1^2 + 2^2} = \sqrt{6} < 3$. Forming the vectors $Y_2 - 3 = (8, 4, 3, 0, 0, \dots)$ and $Y_2 + 3 = (14, 11, 9, 3, 3, \dots)$. We see that $h(Y_2 - 3) = 3 = h(X) = h(Y_2 + 3)$. This is an example where we have equalities.

Proposition 6

Let $d > 0$ and a decreasing vector $X = (x_r)_{r=1,2,\dots}$ in $(\mathbf{R}^+)^{\infty}$ be given. Let further Y be a decreasing vector in $(\mathbf{R}^+)^{\infty}$ such that $\sup(|y_r - x_r|) \leq d$, then, for each $k = 1, 2, \dots$ $h^{(k)}(Y-D) \leq h^{(k)}(X) \leq h^{(k)}(Y+D)$.

The proof follows along the same lines as that of Theorem 4.

Theorem 5 provides an analogue of Theorem 4 for the g -index.

Theorem 5.

Let $d > 0$ and a decreasing vector $X = (x_r)_{r=1,2,\dots}$ in $(\mathbf{R}^+)^{\infty}$ be given. Let further Y be a decreasing vector in $(\mathbf{R}^+)^{\infty}$ such that $\sup(|y_r - x_r|) \leq d$, then $g(Y-D) \leq g(X) \leq g(Y+D)$.

Proof. As in Theorem 4 we derive that $\forall r = 1, 2, \dots \quad y_r - d \leq x_r$. Hence: for $i=1, 2, \dots$

$$\sum_{r=1}^i (y_r - d) \leq \sum_{r=1}^i x_r$$

We know that for all $i > g(X)$: $\sum_{r=1}^i x_r < i^2$. Hence also $\sum_{r=1}^i (y_r - d) < i^2$ from which we conclude that $g(Y-D) \leq g$.

Similarly, we can derive that $\forall r = 1, 2, \dots: y_r + d \geq x_r$. Now, we have for all $i=1, \dots, g(X)$: $\sum_{r=1}^i (y_r + d) \geq \sum_{r=1}^i x_r \geq i^2$. Hence: $g(Y+D) \geq g(X)$.

Corollary 3. Let $d > 0$ and a decreasing vector $X = (x_r)_{r=1,2,\dots}$ in $(\mathbf{R}^+)^{\infty}$ be given. Let further Y be a decreasing vector in $(\mathbf{R}^+)^{\infty}$ such that for some $p \geq 1$, $\|X-Y\|_p \leq d$, then $g(Y-D) \leq g(X) \leq g(Y+D)$.

Information for the R - and the R^2 -index follows from Theorem 4.

Corollary 4. Let $d > 0$ and a decreasing vector $X = (x_r)_{r=1,2,\dots}$ in $(\mathbf{R}^+)^{\infty}$ be given. Let further Y be a decreasing vector in $(\mathbf{R}^+)^{\infty}$ such that $\sup(|y_r - x_r|) \leq d$, then $R^2(Y-D) \leq R^2(X) \leq R^2(Y+D)$.

Proof. By definition we have: $R^2(Y-D) = \sum_{r=1}^{h(Y-D)} (y_r - d) \leq \sum_{r=1}^{h(Y-D)} x_r \leq \sum_{r=1}^{h(X)} x_r$ (this inequality follows from Theorem 4) $= R^2(X)$.

Further, we have by definition that: $R^2(Y+D) = \sum_{r=1}^{h(Y+D)} (y_r + d) \geq \sum_{r=1}^{h(Y+D)} x_r \geq \sum_{r=1}^{h(X)} x_r$ (this inequality too follows from Theorem 4) $= R^2(X)$.

Corollary 5. Let $d > 0$ and a decreasing vector $X = (x_r)_{r=1,2,\dots}$ in $(\mathbf{R}^+)^{\infty}$ be given. Let further Y be a decreasing vector in $(\mathbf{R}^+)^{\infty}$ such that for some $p \geq 1$, $\|X - Y\|_p \leq d$, then $R^2(Y-D) \leq R^2(X) \leq R^2(Y+D)$.

5. Discussion

We recognize the fact that all h -type indicators do not always behave in a logical way, especially in an evaluation context; concretely they do not

meet the independence requirement (Bouyssou and Marchant, 2011; Waltman and van Eck, 2012). Like many other indicators they are only PAC (Probably Approximately Correct) (Rousseau, 2016) and as such should be considered as heuristics when applied in research evaluations.

Our investigations clarify why the h-index is not independent. Indeed, if two arrays are comparable for the partial order $\leq$ in the positive cone $(\mathbf{R}^+)^{\infty}$ then they can never lead to a counterexample for the independence requirement. Yet, real publication-citation arrays are not always comparable for this natural ordering. For this reason we may state that the absence of a total order in $(\mathbf{R}^+)^{\infty}$ is one of the reasons why the h-index is not independent.

Nevertheless, the h-index enjoys a number of interesting mathematical properties as an operator in the ordered positive cone $(\mathbf{R}^+)^{\infty}$ of all infinite sequences with non-negative real values. The g-index and the $R(R^2)$ -index do not, or only partially, have these mathematical properties. These properties deal with natural inequalities about h-type indices on the one hand, and the arithmetic, the geometric and the harmonic mean, on the other. Moreover, we proved the double inequality $h(Y-D) \leq h(X) \leq h(Y+D)$, where D is a constant vector and X and Y are decreasing vectors in $(\mathbf{R}^+)^{\infty}$.

Acknowledgment. The authors thank a reviewer for pointing out an error in an earlier version of the proof of Theorem 1.

References

- Bornmann, L., & Marewski, J.N. (2018). Heuristics as a conceptual lens for understanding and studying the usage of bibliometrics in research evaluation. <https://arxiv.org/abs/1807.05115>
- 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. (2006a). An improvement of the h-index: the g-index. *ISSI Newsletter*, 2(1), 8-9.

- Egghe, L. (2006b). Theory and practise 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.
- Fassin, Y., & Rousseau, R. (2018). The $h^{(3)}$ – index for academic journals. Preprint.
- Hardy, G.H., Littlewood, G.E., & Pólya, G. (1934). *Inequalities*. Cambridge, at the University Press.
- 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.
- Jin, B.H., Liang, L.M., Rousseau, R., & Egghe, L. (2007). The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin*, 52(6), 855-863.
- Kosmulski, M. (2006). A new Hirsch-type index saves time and works equally well as the original index. *ISSI Newsletter*, 2(3), 4-6.
- Rousseau, R. (2016). Citation data as 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.
- 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.