RESEARCH ARTICLE

h-Type indices, partial sums and the majorization order

Leo Egghe[1] **and Ronald Rousseau**[2,3]

[1]University of Hasselt, Belgium
[2]University of Antwerp, Faculty of Social Sciences, B-2020 Antwerp, Belgium
[3]KU Leuven, MSI, Facultair Onderzoekscentrum ECOOM, Naamsestraat 61, Leuven B-3000, Belgium

an open access journal

Citation: Egghe, L., & Rousseau, R. (2020). *h*-Type indices, partial sums and the majorization order. *Quantitative Science Studies*, *1*(1), 320–330. https://doi.org/10.1162/qss_a_00005

DOI:
https://doi.org/10.1162/qss_a_00005

Received: 22 January 2019
Accepted: 27 July 2019

Corresponding Author:
Ronald Rousseau
ronald.rousseau@uantwerpen.be

Handling Editor:
Vincent Larivière

The MIT Press

Keywords: partial sums of an array, *h*-index, *g*-index, *R*-index, Gini index, Lorenz curve

ABSTRACT

We study the array of partial sums, P_X, of a given array X in terms of its *h*-type indices. Concretely, we show that $h(P_X)$ can be described in terms of the Lorenz curve of the array X and obtain a relation between the sum of the components of P_X and the Gini index of X. Moreover, we obtain sharp lower and upper bounds for *h*-type indices of P_X.

1. INTRODUCTION

h-type indices such as the *h*-index itself and the *g*-index have interesting mathematical properties as shown, for example, in Egghe and Rousseau (2019a), although they are only probably approximately correct (PAC) in research evaluation exercises (Bouyssou & Marchant, 2011; Waltman & van Eck, 2012; Rousseau, 2016). In this investigation we continue our theoretical investigation of the mechanism leading to *h*-type indices. Concretely, we study properties related to *h*-type indices of the array of partial sums of a given array X. We recall that these partial sums form the basis of the Lorenz curve and the related Gini index. As a consequence, we also obtain relations with the Gini index and the Lorenz curve of the original array X. We will further derive sharp lower and upper bounds for *h*-type indices of P_X.

In the following section we recall the definitions we will use in this investigation.

2. DEFINITIONS

Let $(\mathbf{R}^+)^N$ be the set of all arrays of length N with nonnegative real values. An array $X = (x_r)_{r=1,2,\ldots,N}$ in $(\mathbf{R}^+)^N$ is said to be decreasing if, for all $r = 1, 2, \ldots, N$, $x_r \geq x_{r+1}$. The set $(\mathbf{R}^+)^N$ has a natural partial order defined by $X \leq Y$ if, for all $r = 1, 2, \ldots, N$, $x_r \leq y_r$. Equality between X and Y only occurs if $x_r = y_r$ for all r. We denote the set of all decreasing arrays in $(\mathbf{R}^+)^N$ with at least one component larger than or equal to 1 by Φ^N.

Next we recall the definition of some *h*-type indices for arrays in $(\mathbf{R}^+)^N$.

2.1. The *h*-index (Hirsch, 2005)

Let $X = (x_r)_{r=1,2,\ldots,N} \in \Phi^N$. 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 array X are strictly smaller than 1, then $h(X) = 0$. Such arrays are not considered further on because we will work with arrays in Φ^N. If x_N, the last element in X, is larger than or equal to N, then $h(X) = N$.

Recall that an h-index (and similarly for the other h-type indices defined further on) can only be defined for decreasing arrays. Moreover, for $r \le h(X)$, $x_r \ge r$; conversely, if for all $r \le n$, $x_r \ge n$, then $n \le h(X)$. Further, if $r > h(X)$, then $x_r < h(X) + 1$.

2.2. The *g*-index (Egghe, 2006a, b)

Let $X = (x_r)_{r=1,2,\ldots,N} \in \Phi^N$. The g-index 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. If the sum of all coordinates of X is strictly larger than N^2, then we extend the array X with coordinates equal to zero, making it into an array in Φ^M, $M > N$, until it is possible to apply the definition.

2.3. The *R*-index (Jin et al., 2007)

Let $X = (x_r)_{r=1,2,\ldots,N} \in \Phi^N$. 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 if their properties are for our purposes the same, all concrete examples will be given for R^2.

2.4. Kosmulski's $h^{(2)}$ Index (Kosmulski, 2006)

Let $X = (x_r)_{r=1,2,\ldots,N} \in \Phi^N$. The $h^{(2)}$ or Kosmulski index of X, denoted $h^{(2)}(X)$, is the largest natural number $h^{(2)}$ such that the first $h^{(2)}$ coordinates have each at least a value $(h^{(2)})^2$.

2.5. The Majorization Order (Hardy et al., 1934)

Let $X, Y \in \Phi^N$, where N is any finite number in $\boldsymbol{N} = \{1, 2, 3, \ldots\}$. The array X is majorized by Y, or X is smaller than or equal to Y in the majorization order, denoted as $X \le Y$ if for all $i = 1, \ldots, N$

$$\begin{cases} \displaystyle\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i \text{ and} \\ \displaystyle\sum_{j=1}^{i} x_j \le \sum_{j=1}^{i} y_j; \forall \mathrm{i} = 1, \ldots, \mathrm{N}. \end{cases}$$

We note that this definition is also valid for arrays in which all values are between zero (included) and 1 (not included).

3. AN INEQUALITY RELATED TO THE *g*-INDEX AND THE MAJORIZATION ORDER

Theorem 1. $\forall\, N \in \mathbf{N}$: $X, Y \in \boldsymbol{\Phi}^N$: $X -< Y \Rightarrow g(X) \le g(Y)$ (*)

Proof. Although this theorem is implied in Egghe (2009, p. 487) we present here two short proofs.

First proof: For each $i \le g(X)$, $\sum_{j=1}^{i} y_j \ge \sum_{j=1}^{i} x_j \ge i^2$. This implies that $g(Y) \ge g(X)$.

Second proof: For each $i > g(Y)$, $\sum_{j=1}^{i} x_j \le \sum_{j=1}^{i} y_j < i^2$. Also this inequality implies that $g(X) \le g(Y)$.

Comments

A. This theorem proves that g is an order-preserving mapping from $(\mathbf{R}^+)^N$ with the majorization order to the positive real numbers with their natural order.
B. The converse of inequality (*) does not hold. Consider for instance $X = (5, 2, 2)$ and $Y = (4, 4, 1)$. Then $g(X) = g(Y) = 3$ but neither $X \le Y$ nor $Y \le X$ holds.

C. The inequality (*) can be strict. Indeed, take $X = (2, 1, 1)$ and $Y = (2, 2, 0)$. Then $X \leq Y$, but $g(X) = 1$ and $g(Y) = 2$.
D. Yet, inequality (*) cannot be strict for $N = 2$. Indeed, consider $X = (x_1, x_2)$ and $Y = (y_1, y_2)$, with $X \leq Y$. Then $1 \leq x_1 \leq y_1$ (hence $g(X), g(Y) \geq 1$) and $x_1 + x_2 = y_1 + y_2$. As $N = 2$, this sum completely determines the value of the g-index. Hence this value must be equal for X and Y. We note that even here there is no upper bound to the value of $g(X) = g(Y)$.
E. If it were allowed that $x_1 < 1$ then the previous Comment D is not valid. Indeed, take $X = (½, ½)$ and $Y = (1, 0)$ then $X \leq Y$, and $g(X) = 0$ and $g(Y) = 1$.
F. Inequality (*) does not hold for the h- or the R^2-index. Consider $X = (3, 3, 3)$ and $Y = (5, 2, 2)$. Then $X \leq Y$ but $h(X) = 3 > h(Y) = 2$. Moreover, $R^2(X) = 9 > R^2(Y) = 7$.

For small N we even have the opposite relation for the h-index. This is shown in the next proposition.

Proposition 1.

(a) For $X, Y \in \Phi^2$ and $X \leq Y$, $h(X) \geq h(Y)$
(b) For $X, Y \in \Phi^3$, $X \leq Y$ and if the components of X and Y are strictly positive natural numbers, then $h(X) \geq h(Y)$.

Proof.

(a). $N = 2$, $X \leq Y$ then $x_1 \leq y_1$ and $x_1 + x_2 = y_1 + y_2$. Hence $x_2 \geq y_2$. If now $h(X) = 1$, then $1 \leq y_1$ and $2 > x_2 \geq y_2$. This implies that $h(Y) = 1 = h(X)$. The case $h(X) = 2$ is trivial: As $N = 2$, it follows that $h(Y) \leq 2 = h(X)$.
(b). $N = 3$ and $X \leq Y$, then $x_1 \leq y_1$; $x_1 + x_2 \leq y_1 + y_2$ and $x_1 + x_2 + x_3 = y_1 + y_2 + y_3$. This already implies that $x_3 \geq y_3$. We now consider three cases: $h(Y) = 3$, $h(Y) = 2$ and $h(Y) = 1$.

Assume first that $h(Y) = 3$. Then $y_3 \geq 3$. Hence we see that $x_1 \geq x_2 \geq x_3 \geq y_3 \geq 3$, from which we derive that $h(X) = 3 = h(Y)$.

Assume next that $h(Y) = 2$. Then $y_3 < 3$, hence $y_3 = 2$ or $y_3 = 1$, and $y_1 \geq y_2 \geq 2$. We first consider the case $y_3 = 1$. We know already that x_3 is at least equal to 1. So, if x_3 is equal to 1, then $x_1 + x_2 = y_1 + y_2$. As $x_1 \leq y_1$ the previous equality implies that $y_2 \leq x_2$. Now $h(Y) = 2$ leads to $2 \leq y_2 \leq x_2$, or $h(X) \geq 2 = h(Y)$. Still with $y_3 = 1$ we now consider the case that $x_3 > 1$. Then $x_1 \geq x_2 \geq x_3 \geq 2$, leading to $h(X) \geq 2 = h(Y)$.

Next we consider the case $y_3 = 2$. Then $x_1 \geq x_2 \geq x_3 \geq y_3 = 2$, which implies that $h(X) \geq 2 = h(Y)$.

Finally, as components are assumed to be strictly positive natural numbers, $h(X)$ and $h(Y)$ are at least equal to 1. Hence $h(Y) = 1$ implies $h(X) \geq h(Y)$.

Comments

A. Proposition 1(b) is not valid if some of the components are zero. This is illustrated as follows. Let $X = (2, 1, 1)$ and $Y = (2, 2, 0)$. Then $X \leq Y$, but $h(X) = 1$ and $h(Y) = 2$.
B. Proposition 1(b) is also not valid if some of the components are not natural numbers. Indeed, let $X = (2, 1.5, 1.5)$ and $Y = (2, 2, 1)$. Then $X \leq Y$, but $h(X) = 1$ and $h(Y) = 2$.
C. Propositions 1(a) and 1(b) are not valid for R^2.

(a) $N = 2$. Consider $X = (1, 1)$ and $Y = (2, 0)$. Then $X \leq Y$, $h(X) = h(Y) = 1$; $R^2(X) = 1$ and $R^2(Y) = 2$.

(b) $N = 3$. Consider $X = (2, 2, 2)$ and $Y = (3, 2, 1)$. Then $X \le Y$, $h(X) = 2 = h(Y)$ and $R^2(X) = 4 < R^2(Y) = 5$.

D. Proposition 1 with $N \ge 4$ is not valid for the *h*-index *h*.

Consider $X = (6, 5, 2, 2)$ and $Y = (6, 5, 3, 1)$. Then $X \le Y$, $h(X) = 2 < h(Y) = 3$. We further remark that $R^2(X) = 11 < R^2(Y) = 14$.

4. INTRODUCING THE ARRAY OF PARTIAL SUMS

Now we come to the main part of this article. First we introduce some notation. Let $X = (x_r)_{r=1,2,\ldots,N} \in \Phi^N$ and consider the partial sums:

$$\sum\nolimits_{j=1}^{i} x_j, i = 1, \ldots, N.$$

Ranking these partial sums again in decreasing order leads to the array P_X. The *i*th component of P_X, denoted as y_i, is equal to $\sum_{j=1}^{N-i+1} x_j$. An example: Let $X = (4, 3, 2, 1)$. Then $P_X = (10, 9, 7, 4)$ and $h(P_X) = 4$.

Remarks

1. $(P_X)_1 = A = \sum_{j=1}^{N} x_j$; $(P_X)_N = x_1$.
2. If X ends with p zeros, then P_X starts with $p + 1$ As.
3. Clearly $X \le P_X$, as $(P_X)_N = x_1$. Hence $h(X) \le h(P_X)$, $g(X) \le g(P_X)$ and $R(X) \le R(P_X)$. (Egghe & Rousseau, 2019a; Proposition 2).

If X denotes the number of received citations of an author's publications, then the indicator value $h(P_X)$ shows how many of the less cited publications can be removed so that the total number of the remaining items of X is higher than the rank of this total in the array P_X. This is another way of describing the impact of the most cited publications. Contrary to the case of $h(X)$, $h(P_X)$ may increase if a publication in X's *h*-core, not necessarily the most cited one, increases its number of citations. We provide an example: let $X = (4, 2, 0, 0, 0, 0, 0, 0)$. Then $h(X) = 2$; $P_X = (6, 6, 6, 6, 6, 6, 6, 4)$ and $h(P_X) = 6$. Consider now $X' = (4, 3, 0, 0, 0, 0, 0, 0)$. Then $h(X') = 2$; $P_{X'} = (7, 7, 7, 7, 7, 7, 7, 4)$ and $h(P_{X'}) = 7$.

5. A RELATION WITH THE GINI INDEX

We recall (Rousseau et al., 2018, formula (4.19)) that the Gini concentration index of a decreasing array X of nonnegative real numbers, $(x_j)_{j=1,\ldots,N}$ is obtained as

$$G(X) = \frac{1}{N}\left(N + 1 - \frac{2}{A}\sum\nolimits_{j=1}^{N} j\, x_j\right), \tag{1}$$

where $A = \sum_{j=1}^{N} x_j$. Consider now P_X. The sum of all components of P_X, denoted as $S(X)$, is

$$\sum_{j=1}^{N} x_j + \sum_{j=1}^{N-1} x_j + \ldots + \sum_{j=1}^{1} x_j = \sum_{j=1}^{N} (N - j + 1) x_j = (N + 1)A - \sum_{j=1}^{N} j x_j.$$

From this result we obtain a relation between $G(X)$ and $S(X)$:

$$G(X) = \frac{1}{N}\left(N+1-\frac{2}{A}((N+1)A-S(X))\right) = \frac{1}{N}\left(N+1-2(N+1)+\frac{2S(X)}{A}\right)$$
$$= \frac{1}{NA}(2S(X)-A(N+1)). \quad (2)$$

Conversely

$$S(X) = \frac{A}{2}(N(1+G(X))+1). \quad (3)$$

An example: If $X = (a, a, a, a)$, $a > 0$, then $G(X) = 0$ (by definition), $N = 4$, $A = 4a$ and $S(X) = 10a$. Now we check formula 3 and find that, indeed, $10a = \frac{4a}{2}(4 + 1)$.

6. A GEOMETRIC INTERPRETATION OF $H(P_X)$ IN TERMS OF THE LORENZ CURVE L_X

For the decreasing array X of nonnegative real numbers, $(x_j)_{j=1,\ldots,N}$ and for $a_j = \frac{x_j}{\sum_{k=1}^{N} x_k} = \frac{x_j}{N\bar{x}}$, the Lorenz curve of X, denoted as L_X, connects points with coordinates $\left(s = \frac{i}{N}, \sum_{j=1}^{i} a_j\right)$. The average of array X is denoted as $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$.

Now $h(P_X)$ is equal to the largest natural number i such that $y_i = \sum_{j=1}^{N-i+1} x_j \geq i$, which is also equal to $(N + 1)$ minus the smallest natural number i such that $y_{N-i+1} = \sum_{j=1}^{i} x_j \geq N - i + 1$. Dividing by the sum of all elements in X this yields

$$\sum_{j=1}^{i} a_j \geq \frac{N-i+1}{N * \bar{x}} = \frac{1}{\bar{x}}\left(1-s+\frac{1}{N}\right).$$

Consequently, $h(P_X)$ is equal to N (1 − the smallest s such that $L_X(s) \geq \frac{1}{\bar{x}}\left(1-s+\frac{1}{N}\right)$) + 1.

An illustration: If $X = (3, 2, 1, 0)$, $N = 4$, $\bar{x} = \frac{3}{2}$, $P_X = (6, 6, 5, 3)$ and $h(P_X) = 3$. Now $L_X\left(\frac{1}{4}\right) = \frac{3}{6} < \frac{2}{3}\left(1-\frac{1}{4}+\frac{1}{4}\right) = \frac{2}{3}$, but $L_X\left(\frac{2}{4}\right) = \frac{5}{6} \geq \frac{2}{3}\left(1-\frac{2}{4}+\frac{1}{4}\right) = \frac{2}{4}$. Hence, the smallest s is equal to 2/4 and $h(P_X) = 4(1 - 2/4) + 1 = 2 + 1 = 3$.

In Egghe and Rousseau (2019b) we studied $h(P_X)$ and its relation with the Lorenz curve in a continuous context. This led to a new geometric interpretation of the h-index.

7. BOUNDS ON h-TYPE INDICES

In the next sections we derive bounds for h-type indices of P_X. This is of importance for the following reason. A function relates an input to a unique output. In this way the standard h-index is a function which maps an array to a natural number. Yet it is not an explicit function, such as the function that maps the real number x to $x^2 + 4x + 7$ or the function which maps a finite array to its sum. Finding an h-index needs a procedure and hence it is not possible to study properties in an analytical way (e.g., using integrals). The bounds obtained in this article are explicit functions which can be studied using analytical methods.

We denote by $\lfloor a \rfloor$, the floor function of a (i.e., the largest integer smaller than or equal to a). We note that $a \geq \lfloor a \rfloor > a - 1$. Using the notation just introduced we come to the following interesting theorem.

Theorem 2. Let $X = (x_r)_{r=1,2,\ldots,N} \in \Phi^N$, then

$$\min(N, A) = \min(N, N\bar{x}) \geq h(P_X) \geq \left\lfloor (N+1)\,\frac{\bar{x}}{\bar{x}+1} \right\rfloor, \tag{4}$$

Before proving Theorem 2 we make three remarks:

1. The first inequality, namely $\min(N, A) \geq h(P_X)$ is easy to see because, on the one hand, an h-index can never be larger than the length of the array and on the other $1 \leq h(P_X) \leq \sum_{j=1}^{N-h(P_X)+1} x_j \leq A$.
2. $h(P_X) = N$ if and only if $x_1 \geq N$.
3. $h(P_X) = 0$ can never occur in our context. Indeed, this may only occur if all components are strictly smaller than 1, which is excluded. Yet, in Egghe and Rousseau (2019c) we showed that formula 4 is also correct in cases for which $h(P_X) = 0$.

Proof of Theorem 2. We only have to show the second inequality. By definition we know that $h(P_X)$ is equal to the largest index i such that $y_i = \sum_{j=1}^{N-i+1} x_j \geq i$. We know that $y_i = \sum_{j=1}^{N-i+1} x_j = (N-i+1)$.(the average of $(x_1, x_2, \ldots, x_{N-i+1})) \geq (N-i+1).\,\bar{x}$ (as the array X is ranked in decreasing order).

Now, if $(N - i + 1).\,\bar{x} \geq i$ then certainly $y_i \geq i$. Solving this inequality for i leads to

$$i \leq (N+1)\,\frac{\bar{x}}{\bar{x}+1}.$$

As the index i is a natural number it follows that $h(P_X) \geq \left\lfloor (N+1)\frac{\bar{x}}{\bar{x}+1} \right\rfloor$. This proves this theorem.

In order to make these bounds more concrete we provide a table (Table 1) for some values of N and $\bar{x}$ (or A), showing how sharp these bounds often are. Largest differences occur when the average number of items is one.

The next theorem shows that the second inequality in Theorem 2 becomes an equality for the array $\bar{X} = (\bar{x}, \bar{x}, \ldots, \bar{x}) \in (\mathbf{R}^+)^N$.

Theorem 3. Let $X = (x_r)_{r=1,2,\ldots,N} \in \Phi^N$, then

$$\min(N, A) = \min(N, N\bar{x}) \geq h(P_{\bar{X}}) = \left\lfloor (N+1)\,\frac{\bar{x}}{\bar{x}+1} \right\rfloor, \tag{5}$$

Table 1. Some specific bounds for $h(P_X)$ according to Eq. 4

	$\bar{x}$					
N	0.1	0.5	1	2	5	10
10	$1 \geq h(P_X) \geq 1$	$5 \geq h(P_X) \geq 3$	$10 \geq h(P_X) \geq 5$	$10 \geq h(P_X) \geq 7$	$10 \geq h(P_X) \geq 9$	$10 \geq h(P_X) \geq 10$
30	$3 \geq h(P_X) \geq 2$	$15 \geq h(P_X) \geq 10$	$30 \geq h(P_X) \geq 15$	$30 \geq h(P_X) \geq 20$	$30 \geq h(P_X) \geq 25$	$30 \geq h(P_X) \geq 28$
100	$10 \geq h(P_X) \geq 9$	$50 \geq h(P_X) \geq 33$	$100 \geq h(P_X) \geq 50$	$100 \geq h(P_X) \geq 67$	$100 \geq h(P_X) \geq 84$	$100 \geq h(P_X) \geq 91$
200	$20 \geq h(P_X) \geq 18$	$100 \geq h(P_X) \geq 67$	$200 \geq h(P_X) \geq 100$	$200 \geq h(P_X) \geq 134$	$200 \geq h(P_X) \geq 167$	$200 \geq h(P_X) \geq 182$

Proof. We see that $P_{\bar{X}} = (N\bar{x}, (N-1)\bar{x}, \ldots, 2\bar{x}, \bar{x})$. Then $h(P_{\bar{X}})$ is the largest natural number i such that $(N - i + 1)\bar{x} \geq i$. We observe that then $h(P_{\bar{X}})$ is equal to the largest natural number i such that $i \leq (N+1)\frac{\bar{x}}{\bar{x}+1}$ and hence $h(P_{\bar{X}}) = \left\lfloor (N+1)\frac{\bar{x}}{\bar{x}+1} \right\rfloor$. This proves Theorem 3.

We next present some examples, illustrating different aspects of the previous results.

Example 1. Returning to the example introduced before, we have $X = (4, 3, 2, 1)$, with $\bar{x} = 2.5$ and $P_X = (10, 9, 7, 4)$. Now $N = h(P_X) = 4 > \left\lfloor (N+1)\frac{\bar{x}}{\bar{x}+1} \right\rfloor = \left\lfloor 5 \times \frac{2.5}{2.5+1} \right\rfloor = \lfloor 3.571 \rfloor = 3$. This illustrates that the second inequality in Theorem 2 can be strict. Continuing now with $\bar{X}$ we see that $N = 4 > h(P_{\bar{X}}) = h(10, 7.5, 5, 2.5) = 3 = \left\lfloor 5 \times \frac{2.5}{3.5} \right\rfloor$.

Example 2. Consider $X = (4, 2, 1, 1)$ with $\bar{x} = 2$ and $P_X = (8, 7, 6, 4)$. Now $N = h(P_X) = 4 > \left\lfloor (N+1)\frac{\bar{x}}{\bar{x}+1} \right\rfloor = \left\lfloor 5 \times \frac{2}{2+1} \right\rfloor = \left\lfloor \frac{10}{3} \right\rfloor = 3$. Continuing with $\bar{X}$ we see that $N = 4 > h(P_{\bar{X}}) = h(8, 6, 4, 2) = 3 = \left\lfloor 5 \times \frac{2}{3} \right\rfloor = \left\lfloor \frac{10}{3} \right\rfloor$. This example illustrates that the floor function is really needed, because 3 < 10/3.

Example 3. Let $X = (4, 0, 0, 0)$ with $\bar{x} = 1$ and $P_X = (4, 4, 4, 4)$. Now $N = h(P_X) = 4 > \left\lfloor 5 \times \frac{1}{1+1} \right\rfloor = \left\lfloor 5 \times \frac{1}{2} \right\rfloor = \lfloor 2.5 \rfloor = 2$. This is another example that the second inequality in Theorem 2 can be strict. Continuing with $\bar{X}$ we see that $N = A = 4 > h(P_{\bar{X}}) = h(4, 3, 2, 1) = 2 = \left\lfloor 5 \times \frac{1}{2} \right\rfloor = \lfloor 2.5 \rfloor$. This is not only another example that the floor function is really needed, but it also illustrates that the first inequality in Theorem 3, and hence also in Theorem 2, can be strict.

Example 4. In the previous examples $h(P_X) = N$. Next we present an example where $h(P_X) < N$. Let $X = (3, 2, 1, 0)$. Then $\bar{x} = 3/2$ and $P_X = (6, 6, 5, 3)$. Now $N = 4 > h(P_X) = 3 \geq \left\lfloor 5 \times \frac{3/2}{(3/2)+1} \right\rfloor = \left\lfloor 5 \times \frac{3}{5} \right\rfloor = \lfloor 3 \rfloor = 3$. Continuing with $\bar{X}$ we see that $N = 4 > h(P_{\bar{X}}) = h(6, 4.5, 3, 1.5) = 3 = \left\lfloor 5 \times \frac{3}{5} \right\rfloor = 3$.

Example 5. Finally, we present an example where $\min(N, A) = A < N$. Let $X = (2, 0, 0, 0)$. Then $\bar{x} = ½$ and $P_X = (2, 2, 2, 2)$. Now $A = 2 = \min(N, A) = h(P_X) = 2 \geq \left\lfloor 5 \times \frac{1/2}{(1/2)+1} \right\rfloor = \left\lfloor 5 \times \frac{1}{3} \right\rfloor = 1$. Continuing with $\bar{X}$ we see that $A = 2 = \min(N, A) > h(P_{\bar{X}}) = h(2, 3/2, 1, 1/2) = 1 = \left\lfloor 5 \times \frac{1}{3} \right\rfloor = 1$.

Corollaries

A. If $\bar{x} \geq N$, then $h(P_X) = N$.

Proof. As $\lim_{t \Rightarrow \infty} \frac{t}{t+1} = 1$, there exist a number t_0 such that for all $t > t_0$.

$$N \leq (N+1)\frac{t}{t+1} < N+1.$$

This double inequality clearly holds if we take $t_0 = N$. With $\bar{x}$ in the role of t we see that in these circumstances $\left\lfloor (N+1)\frac{\bar{x}}{\bar{x}+1} \right\rfloor = N$ and thus by Theorem 2, Corollary A is proved.

B. $\lim_{\bar{x} \Rightarrow \infty} h(P_X) = N$

This follows immediately from Corollary A.

Remark

When applied to publications, corollaries A and B show that for large $\bar{x}$ we only need those publications in X with the highest citations to determine $h(P_X)$. This is in accordance with the principle and meaning of an h-index.

8. PARTIAL SUMS AND THE *G*-INDEX

Using the same notations as before, we next prove the analogue of Theorem 2 for the g-index. We recall that the g-index has no upper limit.

Theorem 4. Let $X = (x_r)_{r=1,2,\ldots,N} \in \Phi^N$. Then

$$g(P_X) \geq \begin{cases} \left\lfloor (2N+1)\dfrac{\bar{x}}{\bar{x}+2} \right\rfloor & \text{if} \quad \displaystyle\sum_{j=1}^{N} y_j < N^2 \ \text{(6a)} \\ \left\lfloor \sqrt{N(N+1)\dfrac{\bar{x}}{2}} \right\rfloor & \text{if} \quad \displaystyle\sum_{j=1}^{N} y_j \geq N^2 \ \text{(6b)} \end{cases}$$

Proof.

$$g(P_X) = \begin{cases} \max\left\{ i \in \mathbf{N}; \sum_{j=1}^{i} y_j \geq i^2 \right\} & \text{if} \quad \sum_{j=1}^{N} y_j < N^2 \\ \max\left\{ i \in \mathbf{N}; \sum_{j=1}^{N} y_j \geq i^2 \right\} & \text{if} \quad \sum_{j=1}^{N} y_j \geq N^2 \end{cases}$$

Now, $\sum_{j=1}^{i} y_j = \sum_{j=1}^{i} \left(\sum_{k=1}^{N-j+1} x_k \right) = \sum_{s=1}^{N} x_s + \sum_{s=1}^{N-1} x_s + \ldots + \sum_{s=1}^{N-i+1} x_s \geq N\bar{x} + (N-1)\,\bar{x} + \ldots + (N-i+1)\,\bar{x}$ (because X is ordered decreasingly) $= \bar{x} \times \left(\frac{N(N+1)}{2} - \frac{(N-i)(N-i+1)}{2} \right) = \frac{\bar{x}}{2} i(2N-i+1)$.

Now we require that this expression is larger than or equal to i^2. This leads to:

$$i \leq \frac{(2N-i+1)\bar{x}}{2}.$$

Solving for i yields: $i \leq (2N+1)\frac{\bar{x}}{\bar{x}+2}$. Taking into account that $g(P_X)$ is an integer, we obtain that if $\sum_{j=1}^{N} y_j < N^2$ then $g(P_X) \geq \left\lfloor (2N+1)\frac{\bar{x}}{\bar{x}+2} \right\rfloor$.

If $\sum_{j=1}^{N} y_j \geq N^2$ then we have to study $\sum_{j=1}^{N} y_j \geq i^2$. In the same way as above we find that $\sum_{j=1}^{N} y_j \geq \bar{x}. \left(\frac{N(N+1)}{2} \right) \geq i^2$ (is all we need). Hence $i \leq \sqrt{\frac{\bar{x}}{2} N(N+1)}$ or $i_{\max} = \left\lfloor \sqrt{\frac{\bar{x}}{2} N(N+1)} \right\rfloor$, where $i_{\max}$ denotes the maximal value the index i can take here.

Consequently, if $\sum_{j=1}^{N} y_j \geq N^2$ then $g(P_X) \geq \left\lfloor \sqrt{\frac{\bar{x}}{2} N(N+1)} \right\rfloor$.

Similar to the theory for the h-index, the next theorem shows that inequality in Theorem 4 becomes an equality for the array $\bar{X}$.

Theorem 5. Let $X = (x_r)_{r=1,2,\ldots,N} \in \Phi^N$, then

$$g(P_{\bar{x}}) = \begin{cases} \left\lfloor (2N+1)\dfrac{\bar{x}}{\bar{x}+2} \right\rfloor & \text{if} \quad \bar{x} < \dfrac{2N}{N+1} \text{ (7a)} \\ \left\lfloor \sqrt{N(N+1)\dfrac{\bar{x}}{2}} \right\rfloor & \text{if} \quad \bar{x} \geq \dfrac{2N}{N+1} \text{ (7b)} \end{cases}$$

Proof. Now: $\sum_{j=1}^{N} y_j = \sum_{j=1}^{N} \left(\sum_{k=1}^{N-j+1} \bar{x} \right) = \sum_{s=1}^{N} \bar{x} + \sum_{s=1}^{N-1} \bar{x} + \ldots + \sum_{s=1}^{1} \bar{x} = N\bar{x} + (N-1)\,\bar{x} + \ldots + 1.\ \bar{x} = \bar{x}.\ \left(\frac{N(N+1)}{2}\right)$.

Hence, $\sum_{j=1}^{N} (N-j+1)\bar{x} < N^2 \Leftrightarrow \bar{x} < \frac{2N}{N+1}$.

Similarly, $\sum_{j=1}^{N} (N-j+1)\bar{x} \geq N^2 \Leftrightarrow \bar{x} \geq \frac{2N}{N+1}$.

Comment. Also here we can make the remark that lower bounds for $g(P_X)$ and $g(P_{\bar{x}})$ depend only on N and $\bar{x}$.

Examples

Example 1. Take $X = (4, 4, 4, 4)$, $\bar{x} = 4$ and $P_X = (16, 12, 8, 4)$. Then $g(P_X) = 6$ (as $40 > 6^2$ and $40 < 7^2$). As this is a case where $40 > 4^2$ we have to check formula 6b. This formula states that $6 = g(P_X) \geq \left\lfloor \sqrt{\frac{4\times4\times5}{2}} \right\rfloor = \lfloor \sqrt{40} \rfloor = 6$. Thanks to the use of the floor function we obtain an equality.

Example 2. Take $X = (4, 3, 2, 1)$, $\bar{x} = 2.5$ and $P_X = (10, 9, 7, 4)$. Then $g(P_X) = 5$ (as $30 > 5^2$ and $30 < 6^2$). Also here we have to check formula 6b. We see that $5 = g(P_X) \geq \left\lfloor \sqrt{\frac{2.5\times4\times5}{2}} \right\rfloor = \lfloor \sqrt{25} \rfloor = 5$. This is an example where the floor function is not necessary.

Example 3. For $X = (4, 0, 0, 0)$, $\bar{x} = 1$ and $P_X = (4, 4, 4, 4)$. Here the sum, namely 16, is larger than or equal to $N^2 = 4^2$; hence we have to check formula 6b. This leads to $4 = g(P_X) \geq \left\lfloor \sqrt{12 \times \frac{1}{2}} \right\rfloor = \lfloor \sqrt{6} \rfloor = 2$. This is another case where we have strict inequality.

Example 4. For $X = (2, 0, 0, 0)$, $\bar{x} = 0.5$ and $P_X = (2, 2, 2, 2)$. Here the sum, namely $8 < 4^2$, hence we have to check formula (6a). This leads to $2 = g(P_X) \geq \left\lfloor \frac{9*0.5}{2.5} \right\rfloor = 1$. Also here we have strict inequality.

e) Finally we consider a case for which $N \neq 4$. Let $X = (5, 4, 3, 2, 1)$, $\bar{x} = 3$ and $P_X = (15, 14, 12, 9, 5)$. Here the sum namely $55 > 5^2$; hence we check formula 6b. We first note that $g(P_X) = 7$ ($55 > 7^2$ and $55 < 8^2$). Now $7 = g(P_X) \geq \left\lfloor \sqrt{\frac{3\times5\times6}{2}} \right\rfloor = \lfloor \sqrt{45} \rfloor = 6$. This is again a case with a strict inequality.

Corollary

$$\lim_{\bar{x} \Rightarrow \infty} g(P_X) = \infty$$

Proof. If $\bar{x}$ is large, then we have to consider formula 6b. Then the right-hand side of formula 6b becomes unlimited large and hence this also holds for $g(P_X)$. This result confirms the fact that the g-index has no upper limit.

9. PARTIAL SUMS, THE $R(R^2)$-INDEX AND KOSMULSKI'S INDEX $H^{(2)}(X)$

In the previous sections we studied the h-index and the g-index. As a final case we mention the R^2-index and Kosmulski's $h^{(2)}$-index. For proofs of the results we refer the reader to Egghe and Rousseau (2019c).

Theorem 6. Let $X = (x_r)_{r=1,2,\ldots,N} \in \Phi^N$. Then

$$R^2(P_X) > \frac{\bar{x} \times (N+1)}{2(\bar{x}+1)^2} \times \left(N \times \bar{x}^2 + 2N \times \bar{x} - \bar{x} - 2\right). \tag{8}$$

Theorem 7. If $(N + 1)\,\frac{\bar{x}}{\bar{x}+1}$ is a natural number and $X = (x_r)_{r=1,2,\ldots,N} \in \Phi^N$, then

$$R^2(P_{\bar{X}}) = \frac{(\bar{x})^2(N+1)(N\bar{x}+2N+1)}{2(\bar{x}+1)^2}. \tag{9}$$

Finally, we extend our results to the case of Kosmulski's index, denoted as $h^{(2)}$, referring to Egghe and Rousseau (2019c) for proofs.

Theorem 8. Let $X = (x_r)_{r=1,2,\ldots,N} \in \Phi^N$. Then

$$\min\left(N, \sqrt{A}\right) \geq h^{(2)}(P_X) \geq \left\lfloor \frac{1}{2}\left(\sqrt{(\bar{x})^2 + 4(N+1)\bar{x}} - \bar{x}\right)\right\rfloor \tag{10}$$

Similarly to Theorem 3 we have

Theorem 9. Let $X = (x_r)_{r=1,2,\ldots,N} \in \Phi^N$. Then

$$\min\left(N, \sqrt{A}\right) \geq h^{(2)}(P_{\bar{X}}) = \left\lfloor \frac{1}{2}\left(\sqrt{(\bar{x})^2 + 4(N+1)\bar{x}} - \bar{x}\right)\right\rfloor \tag{11}$$

10. DISCUSSION AND CONCLUSION

In this article we studied arrays of partial sums, P_X, of a given array X in terms of their h-type indices. We showed that $h(P_X)$ can be described in terms of the Lorenz curve of the array X. Moreover, we obtained sharp lower and upper bounds for these h-type indices. We found bounds that only depend on N, the length of the array, and the average of array X, or equivalently, on the length of the array and the total sum of all items in the array.

As $h(P_X)$ is an h-index it is not surprising that it is not strictly independent in the sense of Bouyssou and Marchant (2011). This means that if $h(P_X) < h(P_Y)$ and if one adds to X and Y the same items (X becomes X', and Y becomes Y') then it is possible that $h(P_{X'}) > h(P_{Y'})$. An example: Let $X = (2, 0, 0, 0, 0)$ and $Y = (1, 1, 1, 1, 1)$. Then $P_X = (2, 2, 2, 2, 2)$ with $h(P_X) = 2$, and $P_Y = (5, 4, 3, 2, 1)$ with $h(P_Y) = 3$, hence $h(P_X) < h(P_Y)$. Adding 5 times 1 to each of them yields $X' = (2, 1, 1, 1, 1, 1, 0, 0, 0, 0)$, $P_{X'} = (7, 7, 7, 7, 7, 6, 5, 4, 3, 2)$ with $h(P_{X'}) = 6$, and $Y' = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$, $P_{Y'} = (10, 9, 8, 7, 6, 5, 4, 3, 2, 1)$ with $h(P_{Y'}) = 5$, hence $h(P_{X'}) > h(P_{Y'})$.

A reviewer asked if $h(P_X)$ can be described in terms of Vannucci's (2010) dominance dimension. It can: Using Vannucci's notation we see that $h(P_X) = \mathrm{dom}\left[\left(\mathrm{proj}_n^n\left(\sum_{j=1}^{N+1-n} x_j\right)\right)_n |(n^1)_n\right]$, where $X = (x_n)_{n=1,\ldots,N}$ is an array of length N.

Our investigation illustrated the rich mathematical structure hidden in the mechanism leading to h-type indices (see also Egghe & Rousseau, 2019d). In this article we considered the discrete case, requiring the floor function in order to get the correct results. In further research we intend to study the continuous case, where by definition no floor function will be needed. Then, bounds will be differentiable and integrable functions.

ACKNOWLEDGMENTS

The authors thank anonymous reviewers for useful suggestions to improve the presentation of this article.

AUTHOR CONTRIBUTIONS

Leo Egghe: conceptualization; formal analysis; investigation; methodology; writing—original draft; writing—review and editing. Ronald Rousseau: validation; writing—review and editing.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

No funding has been received.

DATA AVAILABILITY

Not applicable.

REFERENCES

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.

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. (2009). An econometric property of the *g*-index. *Information Processing & Management*, *45*(4), 484–489.

Egghe, L., & Rousseau, R. (2019a). Infinite sequences and their *h*-type indices. *Journal of Informetrics*, *13*(1), 291–298.

Egghe, L., & Rousseau, R. (2019b). A geometric relation between the *h*-index and the Lorenz curve. *Scientometrics*, *119*(2), 1281–1284.

Egghe, L., & Rousseau, R. (2019c). h-Type indices, partial sums and the majorization order: A first approach. E-LIS, http://hdl.handle.net/10760/34395

Egghe, L., & Rousseau, R. (2019d). Solution by step functions of a minimum problem in $L^2[0, T]$, using generalized *h*- and *g*-indices. *Journal of Informetrics*, *13*(3), 785–792.

Hardy, G. H., Littlewood, G. E., & Pólya, G. (1934). *Inequalities*. Cambridge: Cambridge 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.

Rousseau, R., Egghe, L., & Guns, R. (2018). *Becoming Metric-Wise. A Bibliometric Guide for Researchers*. Kidlington: Chandos-Elsevier.

Vannucci, S. (2010). Dominance dimension: A common parametric formulation for integer-valued scientific impact indices. *Scientometrics*, *84*(1), 43–48.

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.