International Journal of Information Science and Management
Vol. 16, No. 1, 2018, 15-26

On the Distribution of the H-index

Simon Brown
Deviot Institute, Deviot, Tasmania 7275,
Australia and College of Public Health, Medical and Veterinary Sciences,
James Cook University, Townsville, Queensland 4811, Australia
Simon.Brown@deviotinstitute.org

Abstract

The h-index has been considered in terms of both unrestricted integer partitions and fuzzy integrals, and the expected value of h for a given number of citations has been estimated. However, the distribution of h as a function of both the number of citations and the number of papers has not been considered explicitly. Using Durfee squares determined from restricted integer partitions, it is shown that for a small number of papers the expected value of h estimated from the unrestricted partitions of the number of citations is unreliable. Despite this, it is confirmed that the distribution of his asymptotically normal. This means that h-indices should be considered in the context of the number of publications unless that number is large.

Keywords: Distribution, H-Index, Integer Partition, Uncertainty

Introduction

The productivity of researchers and the impact of the work they do is a preoccupation of universities, research funding agencies and sometimes even researchers themselves. Various metrics have been used to measure these including journal impact factors, citation counts and publication rates. At present, however, the most popular of these metrics is the h-index (Hirsch, 2005). Hirsch's definition of the index is that h = m if m of a researcher's p papers have at least m citations each and each of the other papers has no more than m citations. As a guide, Hirsch (2005) suggested that a 'successful' scientist would have h = 20 after 20 years of work, whereas outstanding and 'truly unique' individuals would have h = 40 and h = 60, respectively, after 20 years of work. Subsequent work has shown that this is too great a generalisation, if only because h is highly discipline-specific and depends on circumstance, the comprehensiveness of the literature databases used to calculate the index and many other factors (Ruch & Ball, 2010; Vinkler, 2007). For example, very eminent mathematicians often have h < 10 and some Nobel laureates also have very small h-indices (Yong, 2014). The inevitable inference is an individual's h-index should be considered in the context of these factors and of the distribution of h for a given number of papers and citations appropriate to the individual researcher.

e distribution of h can be determined by considering that it is based on a distribution of c citations among p papers which might vary subject only to the constraint that c is constant. This is an example of partitioning the integer c into no more than p parts, for example, if there are c = 10 citations of p = 3 papers there are 14 possible partitions, including 8+1+1=4+3+3= 5+5=10 (the complete list is given in Table 1). In general, if the number of citations of paper i is xi, then, for $x_1 \geq x_2 \geq \ldots \geq x_p$,

$$x_1 + x_2 + \ldots + x_j + \ldots + x_p = c, \quad (1)$$

where c is the total number of citations of an author's papers and it is likely that some papers are not cited (xi = 0, i > j, where j is the number of papers that are cited). As the number of solutions of (1) increases very rapidly with c and p (Brown, 2009), it is useful to consider that most researchers have some uncited papers so that only the number of cited papers (pc < p) needs to be considered (Ruch & Ball, 2010). In either case, the distribution of c among p can be considered as an integer partition of c with a restricted number of parts (Andrews, 1976), because the number of citations (c) is usually much greater than the number of papers (p). The number of partitions of c with no more than p parts is given by the coefficient of xc of the series expansion of $\left(\prod_{k=1}^{p}\left(1-x^k\right)\right)^{-1}$. For example,

$$\frac{1}{\prod_{k=1}^{3}\left(1-x^k\right)} = 1 + x + 2x^2 + 3x^3 + 4x^4 + 5x^5 + 7x^6 + 8x^7 + 10x^8 + 12x^9 + 14x^{10} + \ldots \quad (2)$$

so as the coefficient of x10 is 14, there are 14 possible partitions of c = 10 among p = 3 papers (Table 1), whereas there are 42 unrestricted partitions of 10.

The link between h and integer partitions is even deeper, as becomes apparent from the geometric representation of a partition known as a Ferrers graph (Figure 1). This is constructed by tallying the number of citations of each paper, so if paper 4 has been cited three times, there would be three dots vertically in a line below 4, and so on for each of the other papers. This generates the Ferrers graph of the partition of the number of citations (c) into a number of parts corresponding to the number of cited papers pc that is no more than p, the number of papers (Figure 1). The grey region in Figure 1, enclosing the largest square that can be formed from the dots making up the Ferrers graph for this partition, is known as the Durfee square and every partition has one (even if some of them are only 1 × 1 squares). In this case the Durfee square has 3 dots on each side and for this particular combination of papers and citations h = 3. It is always the case that h is the number of dots along one side of the Durfee square in the Ferrers graph representing the citations of an author's papers (Anderson, Hankin & Killworth, 2008). By analogy with (2), the total number of Durfee squares of side d is given by the coefficient of the series expansion of $x^{d^2}\left(\prod_{k=1}^{d}\left(1-x^k\right)^2\right)^{-1}$. For example, for d = 3,

$$\frac{x^{3^2}}{\prod_{k=1}^{3}\left(1-x^k\right)^2} = x^9 + 2x^{10} + 5x^{11} + \ldots \quad (3)$$

so as the coefficient of x10 is 2, there are only 2 possible unrestricted partitions of c = 10 that have h = 3, although only one of these is among the restricted partitions listed in Table 1 (the 'missing' partition is 3 + 3 + 3 + 1 for which p ≥ 4 and so it is excluded from Table 1).

The average dimension of the Durfee square of the partitions of c is proportional to $\sqrt{c}$ (Canfield, Corteel & Savage, 1998), just as h is proportional to the square root of the number of citations (Hirsch, 2005; Ruch & Ball, 2010). Relying on the simulations of Canfield et al. (1998) using unrestricted partitions, Yong (2014) estimated that the expected value of h as

$$\langle h \rangle \approx \frac{\sqrt{6}\ln(2)}{\pi}\sqrt{c} \approx 0.54\sqrt{c} \quad , \tag{4}$$

where ≈ denotes approximately equality, and constructed the 95% confidence intervals of h values for selected values of c. He gave no consideration to the significance of the number of papers (p). Equation (4) is similar to Hirsch's (2005) own expression, which can be written as $h = \sqrt{c/a}$, where a is a constant between 3 and 5. In (4) the constant (0.54) corresponds to a value of 3.43 for the constant a.

The h-index can also be seen as an example of a fuzzy integral known as a Sugeno integral (Torra & Narukawa, 2008), which can be written in terms of the number of citations of paper i (xi) as

$$h = \max_i \min\{x_i, i\} \quad , \tag{5}$$

where the xi are sorted in descending order as for (1). Apart from this elegant expression, this view of the index has the advantage that it implies that h must have the properties of all Sugeno integrals (Arenas-Díaz & Ramírez-Lamus, 2013; de Campos & Bolaños, 1992; Torra & Narukawa, 2006). Confidence intervals for h have also been constructed based on this perspective of the index (Gagolewski, 2015), but this requires knowledge of the underlying distribution which, even given the necessary analytical expression (Marichal, 2006), is not straightforward to determine (Grabisch & Raufaste, 2008).

The partition approach is more simple in this respect. However, Yong's (2014) approach is based on the treatment of unrestricted partitions of Canfield et al. (1998) rather than on a restricted number of parts which is more appropriate. In the former, the number of publications (p) is treated as though it could equal the number of citations (c, so there would be 42 partitions of c = 10 citations), but in the latter p is limited (so there are 14 partitions of c = 10 citations of p = 3 papers in Table 1). However, no consideration has been given to the distribution of h for different values of c and p, despite the conclusion of Glänzel (2006) that h is proportional to p1/(α+1) where α > 1, so authors and the agencies with which they interact have no idea how their value of the h-index relates to the likely range of values. Here a simple description of the distributions for small c is provided and an expression for the 95% confidence interval of h for the mode of h for given values of c and p is constructed.

Methods

In determining the possible distributions of c citations among p papers, the partitions of c into no more than p parts were listed and for each partition, h was calculated using the size of the Durfee square. For values of c > 100, the stated number of random partitions was sampled and h was calculated in the same way. All calculations were carried out in R (Ihaka & Gentleman, 1996) using the rpartitions package (Hankin, 2006) to list all the (restricted) partitions and random partitions were generated using the rpartitions package (Locey & McGlinn, 2012).

Results

The number of unrestricted partitions rises very rapidly with the number of citations (c), but if the number of papers (p) is small the increase is very much less (Figure 2). For example, for c = 100 citations there are 51, 46262, 6292069, 97132873 and 190569292 partitions for $p \le 2$, $p \le 5$, $p \le 10$, $p \le 20$ and $p \le 100$ papers, respectively (Figure 2). If, as is often the case, p is much smaller than c, then the use of unrestricted partitions ($p \le c$) to estimate the distribution of h (Yong, 2014) is inappropriate and this discrepancy grows as c/p increases (Figure 2).

For any combination of p papers and c citations a range of values of the h-index is possible. As is confirmed by a complete listing of all the restricted partitions (as for Figure 3), the lower bound is always h = 1, corresponding to the unlikely case that only one paper of p is cited. For unrestricted partitions the correspondence between the Durfee square and h implies that $h \le \lfloor \sqrt{c} \rfloor$, where $\lfloor x \rfloor$ is the largest integer $\le$ x, and for restricted partitions, $h \le \min\left(p, \lfloor \sqrt{c} \rfloor\right)$. Taking the examples of Figure 2, if c = 100 and $p \le 2$, only 2 partitions ((100, 0) and (99, 1)) have h = 1 and for the remaining 49 partitions, h is equal to 2, so the mean h is $\langle h \rangle = \lfloor 1.96 \rfloor = 1$. For p = 10 and p = 20, $\langle h \rangle = \lfloor 5.6 \rfloor = 5$ and $\langle h \rangle = \lfloor 5.7 \rfloor = 5$, respectively, as would be estimated using (4) for c = 100, and in either case $h \in (1, 10)$ (Figure 3). From this, it is reasonable to infer that $\langle h \rangle$ is over-estimated by (4) when p is much smaller than c and c is relatively small, although the discrepancy declines as p increases. To extend this to larger c (Figure 4), for which it is impractical to list all of the partitions, 20 blocks of 5000 random restricted partitions were generated using the rpartitions package (Locey & McGlinn, 2012) and h was determined from the dimension of the Durfee square of each of these 100000 partitions. Using this approach, (4) provides a reasonable estimate of $\langle h \rangle$ for c at least as large as 1250 and it makes little difference whether p = 20, 50 or 100 (Figure 4). As is the case for smaller c (Figure 3), 95% of h indices lie within a very narrow band (reaching approximately ± 4 by c =1000) around $\langle h \rangle$, although the potential range of values is very much greater (Figure 4). Based on these simulations, the standard deviation of h depends only on c (the correlation between the standard deviation of h and c is r = 0.985, F = 817.8, P < 0.001, n = 27), so the empirical approximation of the 95% confidence band using (4) is

$$\left(0.54\sqrt{c} \pm 1.96\left(0.57 + 0.045\sqrt{c}\right)\right), \qquad (6)$$

where $c \ge 7$ for the lower bound. The contribution of p = 20, 50, 100 is not statistically significant (P = 0.538), although this is not the case for smaller p (Figure 3).

The distribution of the h-index for c = 100 citations is unimodal, but, for p = 6, 8, 10 or 20 papers, it is not normal (p < 0.001) based on the Anderson-Darling normality test (Figure 5). This is shown by the disagreement between the actual values and the curves obtained using the normal probability distribution function and the mean and standard deviation of the actual values (Figure 5). The difference is not due to the truncation of the h-index to $1 \le h \le \min\left(p, \lfloor \sqrt{c} \rfloor\right)$ because a similar discrepancy is apparent when the truncated normal probability distribution function is used. Consistent with the simulations in Figure 3, the discrepancy declines as p increases from 6 (Figure 5A) to 20 (Figure 5D). However, for small p the effect of truncation is significant, because $h \leq p$ which makes the distribution asymmetric (Table 1 and Figure 5A).

For small c, the distribution of h is not normal, but as c increases (for moderate p = 100) from 200 (Figure 6A) to 1000 (Figure 6 D) a normal distribution provides an increasingly good approximation (based on the Anderson-Darling normality test). The truncation of h has little effect because the variance is small and so the probability of observing values of h close to either boundary is correspondingly small. As can be seen in Figure 4, in the simulations, the variance of h is less than 4 for all c and p.

The approximate 95% confidence interval (6) provides a useful means of estimating the significance of an h index. Three examples are considered. First, Ruch and Ball (2010) report data for ‘neurologists’ and ‘particle physicists’, each of whom had published more than 50 papers, so the effect of p is likely to be limited (Figure 4). These data prompt the inference that the greater citation and publication rates of the physicists yield larger h-indices. However, considered in the context of (4) and (6), the indices of the two disciplines are within 2 standard deviations of $\langle h \rangle$ with the exception of three physicists who had smaller h-indices than this (Figure 7). Second, Anderson et al. (2008) reported data for six ‘randomly chosen fellows’ of the Royal Society (each of whom had p > 50 by 2006) and Yong (2014) reported data for Fields medallists and Abel prize recipients (although he gives no indication of p) (Figure 8A). Again only three of these individuals had h-indices outside the bounds defined by (6), although one of these had c = 40094 and h = 35 for which (4) yields $\langle h \rangle \approx 108$. Third, Yong (2014) reported data for 119 mathematicians who are members of the United States National Academy of Sciences (Figure 8B). Of these, only 14 had h-indices just outside the lower bound given by (6).

Discussion

As $1 \leq h \leq \min\left(p, \left\lfloor \sqrt{c} \right\rfloor\right)$, it is unsurprising (i) that the distribution of h depends on both c and p and, in at least some circumstances, (ii) that $\langle h \rangle$ also depends on p as well as c. The approximation of $\langle h \rangle$ by (4) is reasonable for large numbers of citations (Figure 4), but for smaller c it is not necessarily appropriate (Figure 3). Moreover, for small numbers of publications (4) appears to over-estimate the average h (Figure 3). Similarly, for small c the distribution of h is not normal (Figure 5), especially for small p (Figure 5A), but as c increases the distribution tends to become more normal (Figure 6), consistent with least some of the simulations of Grabisch and Raufaste (2008) and with the demonstration of Canfield et al. (1998) that the distribution of the dimension of the Durfee square (which is equivalent to h) of unrestricted partitions is asymptotically normal.

The work described here allows three interesting inferences to be drawn. These relate to (i) the likelihood of achieving each of the milestone h-indices specified by Hirsch (2005), (ii) the differences between disciplines in h and (iii) the usefulness of h-indices in identifying outstanding individuals.

First, Hirsch's (2005) milestone h-indices, that is h = 20, 40, 60, correspond to very high citation rates that may be unachievable in some disciplines. For it to be even theoretically possible to reach h = 20, 40, 60, the correspondence of h with the dimension of the Durfee square means that it is necessary that c = 400, 1600, 3600, respectively. As there are more than 6.617 × 1015 partitions of c = 400 into no more than 20 parts only one of which corresponds to h = 20, the probability of observing this is less than 2 × 10-16 and the chances of observing h = 40 and h = 60 for c = 1600 and c = 3600, respectively, are even smaller. For about 2.5% of cases h = 20, 40, 60 for c ≤ 904, 3834, 8792 and for about 50% of cases this will be so for c ≤ 1372, 5487, 12350 (Figure 4).

Second, the h-indices of 'neurologists' and 'particle physicists' are not inconsistent with (4) because the data lie within the bounds of the 95% confidence interval (Figure 7). This prompts the inference that one factor underlying the difference between these two groups is their publication practice. For example, the number of their publications differ (averaging 93 and 303 for 'neurologists' and 'physicists', respectively), as do their number of citations (360 compared with 4323). However, while the correlation between p and c is weak in these two cases (r = 0.34 and r = 0.63 for the numbers of publications and citations of 'neurologists' and 'physicists', respectively, and r = 0.91 for the numbers of publications and citations of the 'neurologists' and 'physicists' combined, consistent with the greater range of values), a large number of publications in a discipline tends to be associated with a greater number of citations (Larivière & Costas, 2015).

Third, the h-indices of outstanding scientists and mathematicians are not conspicuously different from those of their colleagues with a similar number of citations (Figure 8). Moreover, the h-index of such individuals can be lower than would be expected from (4) and (6), for example, one of the Fellows of the Royal Society cited by Anderson et al. (2008) had c = 40094 from which (4) would yield $\langle h \rangle$ ≈ 108 rather than h = 35 as observed (indicated by the arrow in Figure 8A). What is conspicuous, however, is that none of these outstanding individuals has an h-index greater than the upper bound given by (6) (Figure 8), as is also true of the data for their colleagues in other disciplines (Figure 7). One of the weaknesses of this consideration of the h-index is that the comparisons are based on data from individuals at different career stages, which affects the time available for the identification of important papers and the accumulation of citations. Nevertheless, it is apparent that the h-index is not an obviously useful tool for distinguishing between Hirsch's (2005) categories of 'successful', 'outstanding' and 'truly unique'. Indeed, of the individuals represented in Figures 7 and 8, only a handful had h ≥ 40 and none had h ≥ 60.

Conclusions

The h-index is widely used to measure the productivity of researchers and the impact of the work they do, but, especially for those with small numbers of publications, this is not reasonable (Figure 3). This statistical effect is likely to be exacerbated by the restricted range of journals represented in databases such as Web of Science, Scopus and Google Scholar, and by the algorithms used to match citations with individuals and their publications. Together, these factors lead to underestimates of the actual h-index for all researchers when such databases are used. Conversely, for those with a small number of publications, the expected value of h given by (4) is an overestimate (Figure 3). This means that h-indices should be considered as an underestimate and should be considered in the context of the number of publications especially if that number is not large.

References

Anderson, T. R., Hankin, R. K. S., & Killworth, P. D. (2008). Beyond the Durfee square: enhancing the h-index to score total publication output. *Scientometrics*, 76(3), 577-588. Doi: 10.1007/s11192-007-2071-2

Andrews, G. E. (1976). *The theory of partitions*. Reading: Addison-Wesley Publishing Company.

Arenas-Díaz, G., & Ramírez-Lamus, É. (2013). Medidas difusas e integrales difusas. *Universitas Scientiarum*, 18(1), 7-32. doi:10.11144/Javeriana.SC18-1.mdid

Brown, S. (2009). On the distribution of the largest part of small integers. *Electronic Journal of Applied Statistical Analysis*, 2(1), 1-10. doi:10.1285/i20705948v2n1p1

Canfield, E. R., Corteel, S., & Savage, C. D. (1998). Durfee polynomials. *Electronic Journal of Combinatorics*, 5, (32). Retrieved from http://www.combinatorics.org/ojs/index.php/eljc/article /view/v5i1r32.

de Campos, L. M., & Bolaños, M. J. (1992). Characterization and comparison of Sugeno and Choquet integrals. *Fuzzy Sets and Systems*, 52(1), 61-67. doi:10.1016/0165-0114(92) 90037-5

Gagolewski, M. (2015). Sugeno integral-based confidence intervals for the theoretical h-index. In P. Grzegorzewski, M. Gagolewski, O. Hryniewicz, O. & M. Á. Gil, (Eds.), Strengthening links between data analysis and soft computing, Springer, 233–240. doi:10.1007/978-3-319-10765-3

Glänzel, W. (2006). On the h-index – a mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, 67(2), 315-321. doi:10.1556/ Scient.67.2006.2.12

Grabisch, M., & Raufaste, E. (2008). An empirical study of statistical properties of Choquet and Sugeno integrals. *IEEE Transactions on Fuzzy Systems*, 16(4), 839-850. doi:10.1109/TFUZZ.2008.917295.

Hankin, R. K. S. (2006). Additive integer partitions in R. *Journal of Statistical Software*, 16, Code Snippet 1. doi:10.18637/jss.v016.c01

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the USA*, 102(46), 16569-16572. doi:10.1073/pnas.0507655102.

Ihaka, R., & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299-314. doi:10.2307/1390807

Larivière, V., & Costas, R. (2015). How many is too many? On the relationship between output and impact in research. In A. A. Salah, Y. Tonta, A. A. A. Salah, C. R. Sugimoto & U. Al (Eds.), *Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference*, Bogaziçi University Printhouse, 590-595.

Retrieved from http://www.issi2015.org/files/downloads/all-papers/0590.pdf

Locey, K., & McGlinn, D. (2012). Rpartitions: code for integer partitioning. *Weecology*, Logan. R package version 0.1. Retrieved from https://github.com/klocey/partitions

Marichal, J. L. (2006). Cumulative distribution functions and moments of lattice polynomials. *Statistics and Probability Letters*, 76(12), 1273–1279. doi:10.1016/j.spl. 2006.01.001

Ruch, S., & Ball, R. (2010). Various correlations between the H-Index and citation rate (CPP) in neuroscience and quantum physics: new findings. *International Journal of Information Science and Management*, 8(1), 1-19. Retrieved from http://ijism.ricest.ac.ir/ index.php/ ijism/article/view/21.

Torra, V., & Narukawa, Y. (2006). The interpretation of fuzzy integrals and their application to fuzzy systems. *International Journal of Approximate Reasoning*, 41(1), 43-58. doi:10.1016/j.ijar.2005.08.001

Torra, V., & Narukawa, Y. (2008). The h-index and the number of citations: two fuzzy integrals. *IEEE Transactions on Fuzzy Systems*, 16(3), 795-797. doi:10.1109/TFUZZ. 2007.896327

Vinkler, P. (2007). Eminence of scientists in the light of the h-index and other scientometric indicators. *Journal of Information Science*, 33(4), 481-491. doi:10.1177 /0165551506072165

Yong, A. (2014). Critique of Hirsch's citation index: a combinatorial Fermi problem. *Notices of the American Mathematical Society*, 61(9), 1040-1050. doi: 10.1090/noti1164

Table 1

The possible distributions of c = 10 citations among p = 3 papers and the corresponding h-indices.

Number of citations (c)			h
paper 1	paper 2	paper 3	
10	0	0	1
9	1	0	1
8	2	0	2
8	1	1	1
7	3	0	2
7	2	1	2
6	4	0	2
6	3	1	2
6	2	2	2
5	5	0	2
5	4	1	2
5	3	2	2
4	4	2	2
4	3	3	3

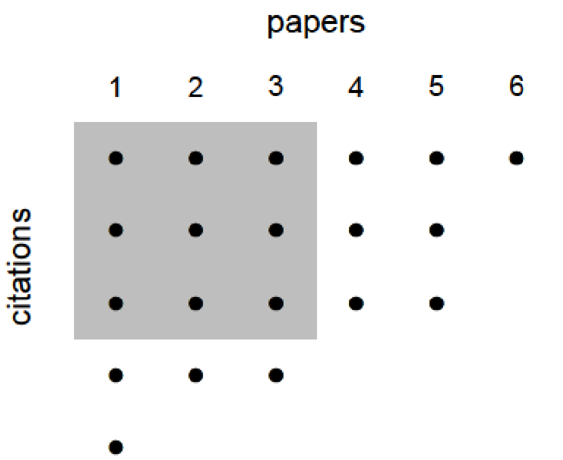

Figure 1. Example of a Ferrers graph corresponding to xi = (5, 4, 4, 3, 3, 1), representing c=5+4 +4 +3+3+1=20 citations of p=6 papers for which h=3. The grey region encloses the 3×3 Durfee square of this partition.

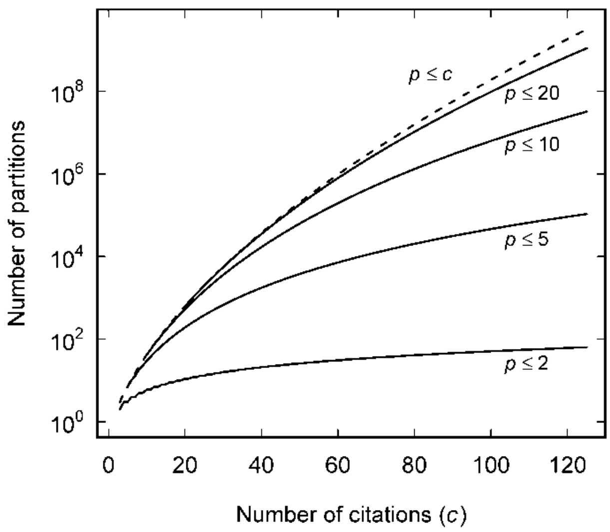

Figure 2. Number of partitions of c citations for p ≤ 2, 5, 10 or 20 papers and the limiting case of p≤c (dashed curve).

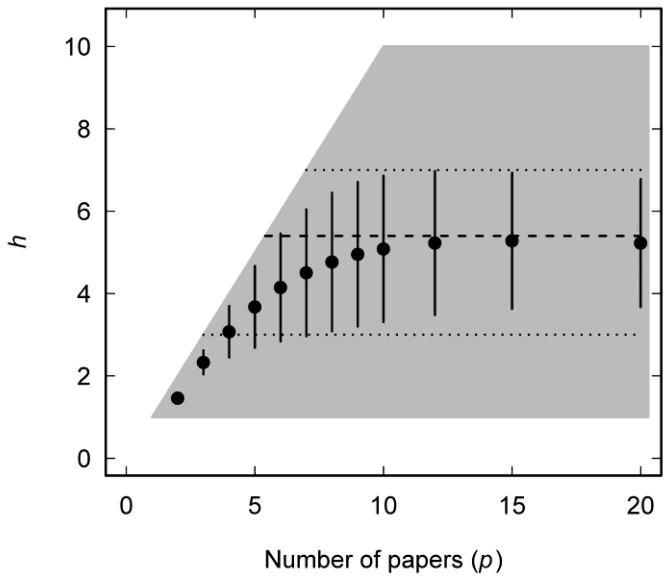

Figure 3. Comparison of the mean h for c = 100 citations of a range of p papers (●) with the value estimated using (4) (dashed line). The horizontal dotted lines represent Yong's (2014) 95% confidence interval [3, 7] for the unrestricted partitions of c = 100 citations. The error bars represent approximate 95% confidence intervals for <h> determined from the complete listings of the partitions. The grey region represents the observed range of h.

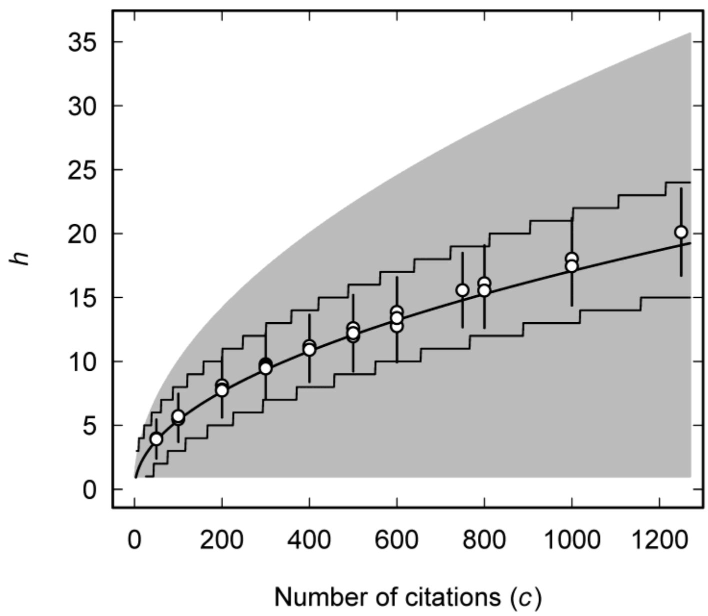

Figure 4. Comparison of h for a range of citations (c) of p = 20, 50 or 100 papers (○) with the value estimated using (4) (solid line). The stepped curves represent the estimated 95% confidence band (6) for the unrestricted partitions of c citations. The error bars represent approximate 95% confidence band for <h> determined from 100000 random restricted partitions. The grey region represents the possible range of h for each value of c.

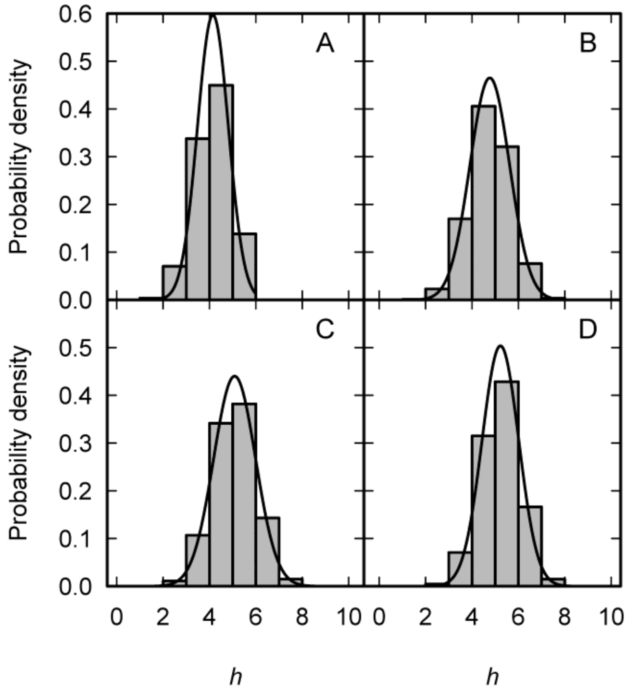

Figure 5. Distributions of h for p = 6 (A), 8 (B), 10 (C) and 20 (D), and c = 100. In each case the smooth curve is the normal probability density using the mean and standard deviation of h obtained from the dimensions of the Durfee squares of all of the restricted partitions.

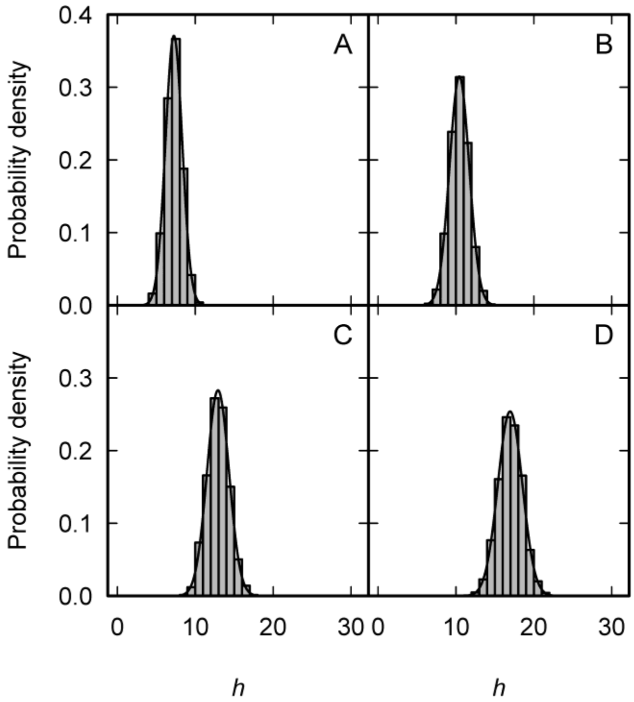

Figure 6. Distributions of h for c = 200 (A), 400 (B), 600 (C) and 1000 (D) citations of p = 100 papers. In each case the smooth curve is the normal probability density using the mean and standard deviation of h obtained from the dimensions of the Durfee squares of 5000 random restricted partitions.

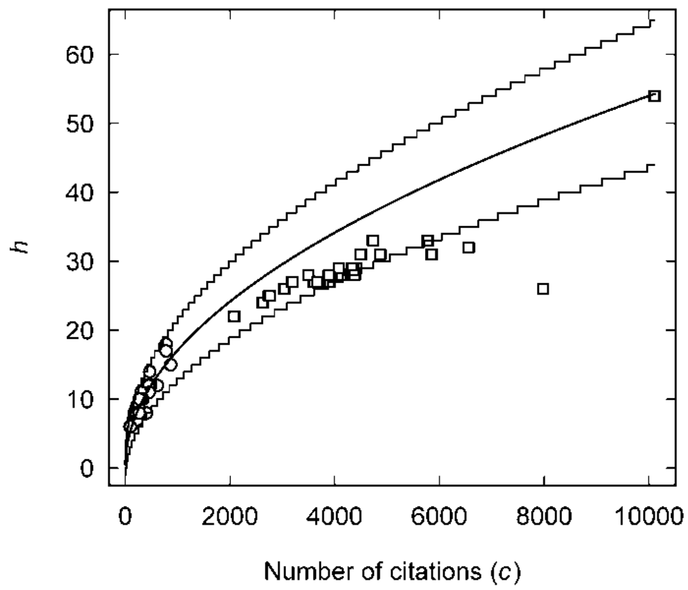

Figure 7. The data of Ruch and Ball (2010) for ‘neurologists’ (○) and ‘particle physicists’ (□). The solid line is the expected value (4) and the discontinuous curves represent the estimated 95% confidence bands (6).

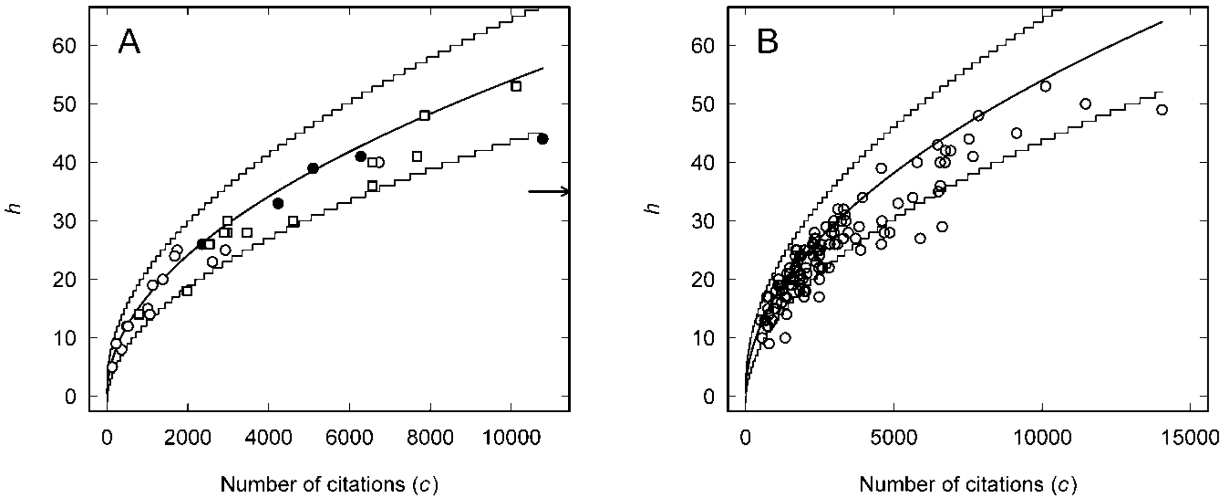

Figure 8. Comparison of h-indices of (A) members of the Royal Society (●), Fields medallists (○) and Abel Prize recipients (□) and (B) members of the United States National Academy of Sciences. The solid line is the expected value (4) and the discontinuous curves represent the estimated 95% confidence bands (6). In (A) the arrow indicates the h-index of of a Royal Society member for whom c = 40094. Data for the Royal Society members are taken from Anderson et al. (2008) and all other data are from Yong (2014).