J Data Sci. Info. Citation Studies., 2023; 2(1):21-24.
https://www.jcitation.org

Research Article

# An Interpretation of *h, e* and *R* indices from the Viewpoint of Riemann Zeta Function

**Bidyarthi Dutta**

Department of Library and Information Science, Vidyasagar University, Midnapore, West Bengal, India.

**ABSTRACT**

The *h*-index is an indicator, which measures both the productivity and the effect of citation of the publications for an individual scientist or scholar, journal or institution. Besides, *R*-Index and *e*-Index are also significant as *h*-index ignores the information of excess citation beyond *h*-core zone. The *h*-index alone conveys partial citation information, which becomes complete only when it is countered with e Index and *R*-Index. The *h, e* and *R*-indices are defined as the positive square roots of *h*-core citation, net excess citation and total citation respectively. The *h*-Index holds only integral values and is a discrete variable, whereas *e* and *R*-indices may have any positive value, which are continuous variables. This paper shows that the ratio of total to excess citation, i.e. $\left(\frac{R}{e}\right)^2$ may be expressed as an infinite series of *h*-core to total citation, i.e., $\left(\frac{h}{e}\right)^2$. Also, the ratio of total citation to net excess citation is explained in the light of Riemann Zeta Function $(\zeta(r))$. It is found that only few Zeta Functions (from $\zeta(2)$ to $\zeta(6)$) may interpret the said ratio.

**Keywords:** *h*-index, *e*-Index, *R*-Index, h-core Citation, Excess Citation, Total Citation, Riemann Zeta Function.

**Correspondence:**
**Bidyarthi Dutta**
Department of Library and Information Science, Vidyasagar University, Midnapore- 721 102, West Bengal, India.
E-mail: bidyarthi.bhaswati@gmail.com

**Received:** 17-02-2023;
**Revised:** 28-02-2023;
**Accepted:** 02-03-2023.

## INTRODUCTION

The *h*-index basically is an author-level indicator, which measures both the productivity and the effect of citation of the publications, initially used for an individual scientist or scholar. The *h*-index correlates with obvious success indicators such as winning the Nobel Prize, being accepted for research fellowships and holding positions at top universities. The index is based on the set of the scientist's most cited papers and the number of citations that they have received in other publications. The index has more recently been applied to the productivity and impact of a scholarly journal as well as a group of scientists, such as a department or university or country. The index was suggested in 2005 by Jorge E. Hirsch, a physicist at UC San Diego, as a tool for determining theoretical physicists' relative quality and is sometimes called the Hirsch index or Hirsch number. The *h*-index is defined as the maximum value of *h* such that the given author/journal has published at least *h* papers that have each been cited at least *h* times. The index is designed to improve upon simpler measures such as the total number of citations or publications. The index works best when comparing scholars working in the same field, since citation conventions differ widely among different subject domains.

**DOI:** 10.5530/jcitation.2.1.3



### Literature Review and Research Gap

The *h*-index was introduced by Hirsch (2007) and used by major bibliographic or citation databases today. It is an author, institution or journal-level metrics. Followed by *h*-index, other associated metrics were subsequently introduced, e.g., e-index (Zhang, 2009, Jin, *et.al*. 2007), R-index (Zhang, 2009), a-index (Zhang, 2009, Jin, *et.al*. 2007) and g-index (Egghe, 2006), *et al*. Despite the effectiveness and simplicity, there are some demerits of h index, for instance, the loss of citation information and low resolution. There are many papers describing advantages, disadvantages and possible areas of applications of *h*-index as discussed by Dutta, (2020). Wu (2010) and Kosmulski (2009) introduced two *h*-type indices, viz. W-index and Kosmulski index, which were generalised by Egghe (2011) in Lotkaian framework. Fassin & Rousseau (2018) introduced h (3) index, another *h*-type index. Waltman and Van (2012) discussed various inconsistencies of *h*-index. Also, the *h*-index has a relatively narrow range. For example, in any field, scientists having an *h*-index larger than 100 (which means at least 10,000 citations) are rare. Therefore, due to low resolution, it is quite common for a group of scientists to have an identical *h*-index. This lacuna was resolved by the e-index, which is a real number to complement the *h*-index for the ignored excess citations. Mathematically speaking, the square root of the number of *h*-core citations, *h*-excess citations and total citations are defined as *h*-index, *e*-index and *R*-index, respectively.

The ratios of total citation to net excess citation (denoted by $\varepsilon'$) and *h*-core citation to total citation (denoted by $\delta$) are mathematically analysed in this paper. It is found that $\varepsilon'$ may be expressed as a converging infinite series of $\delta$. The ratio ($\varepsilon'$) is expressed here as Riemann Zeta Function and the occurrence of *h*-core citation is interpreted in the light of Zeta function.

## Mathematical Formulation

Let us consider an arbitrary item (a journal, an author or an institution), which started receiving citations since some particular point of time. The ratio of total citation and net excess citation may be expressed as a converging infinite series of the ratio of *h*-Index and *R*-Index. The ratio of total citation to net excess citation for any single cited item may be expressed as given in Equation (1). Similarly the ratio of *h*-core citation to total citation for any single cited item may be represented by Equation (2).

$$\frac{\text{Total Citation}}{\text{Net Excess Citation}} = \frac{R^2}{e^2} = \frac{R^2}{(R^2-h^2)} = \frac{R^2}{R^2\left(1-\frac{h^2}{R^2}\right)} = \left[1-\left(\frac{h}{R}\right)^2\right]^{-1} = [1-\delta]^{-1}, \quad \ldots(1)$$

$$\text{Now, } [1-\delta]^{-1} = 1+\delta+\delta^2+\delta^3+\cdots\ldots \quad \ldots(2)$$

$$\text{where } \delta = \left(\frac{h}{R}\right)^2, \text{ say; } \delta < 1$$

Here, $\delta$ can not be equal to one, otherwise which indicates equality between *h*-core citations and total citations. But this is possible only if excess citation becomes zero. But, zero excess citation may be theoretically correct only and very hardly observed in reality. Thus, for non-zero excess citation, the value of $\delta$ will lie between 0 and 1, i.e. $0 < \delta < 1$.

$$\text{As } \frac{e^2}{R^2} = \varepsilon^2, \quad \therefore \frac{R^2}{e^2} = \frac{1}{\varepsilon^2} = \varepsilon'(\text{say}) = [1-\delta^2]^{-1}; \text{ or}$$

$$\frac{\text{Total Citation}}{\text{Net Excess Citation}} = \varepsilon' = 1+\delta+\delta^2+\delta^3+\cdots\ldots \quad \ldots(3)$$

The ratio of total citation to net excess citation $\left(\frac{\text{Total Citation}}{\text{Net Excess Citation}}\right)$ may thus be expressed as a converging infinite series of the ratio of *h*-core citation to total citation $\left(\frac{\text{h}-\text{Core Citation}}{\text{Total Citation}}\right)$; or

$$\left(\frac{\text{Total Citation}}{\text{Net Excess Citation}}\right) = 1+\left(\frac{\text{h}-\text{Core Citation}}{\text{Total Citation}}\right)+\left(\frac{\text{h}-\text{Core Citation}}{\text{Total Citation}}\right)^2+\left(\frac{\text{h}-\text{Core Citation}}{\text{Total Citation}}\right)^3+\cdots\ldots$$

Now, $\delta < 1$ (Equation 2)) as therefore it may be expresses as, $\delta = \frac{1}{n^s}$, where $s > 1$ and $n = 2, 3, 4, \ldots\ldots.(4)$

Substituting equation (4) in equation (3), it is obtained:

$$\varepsilon' = 1+\frac{1}{n^s}+\frac{1}{n^{2s}}+\frac{1}{n^{3s}}\cdots\cdots\cdots \quad (5)$$

$$\text{Or, } \varepsilon' = 1+n^{-s}+n^{-2s}+n^{-3s}+\cdots\ldots$$

$$\varepsilon' = \sum_{n=1}^{\infty} n^{-ps} \ \ldots\ldots\ldots\ldots.(5A)$$

Where $s > 1$ and p= 1, 2, 3,…….

Calling ps = r, equation (5A) may be rewritten as:

$$\varepsilon' = \sum_{n=1}^{\infty} n^{-r}; r > 1 \ \text{(Since, r = ps; s > 1 and p = 1, 2, 3,....)} \quad \ldots(6)$$

Calling $\varepsilon'$ as $\varepsilon(r)$, as $\varepsilon'$ is a function of r, Equation (6) may be rewritten as:

$$\varepsilon(r) = \sum_{n=1}^{\infty} n^{-r}. r > 1 \quad \ldots\ldots(7),$$

The series denoted in Equation (7) represents Riemann Zeta Function.

$$\zeta(r) = \sum_{n=1}^{\infty} n^{-r}; r > 1 \ \ldots\ldots.(8)$$

Hence, the ratio of total citation to net excess citation for any single cited item may be represented as Riemann Zeta Function of $n^{-r}$, where n = 1, 2, 3.... upto $\infty$ and $r > 1$, as evident from the Equations (7) and (8).

Now, for r = 1 and integral values of ‘n’, Equation (8) may be written as:

$$\zeta(1) = \sum_{n=1}^{\infty} n^{-1} = \sum_{n=1}^{\infty}\frac{1}{n} = 1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots\ldots \quad \ldots.(9)$$

The Equation (9) represents a harmonic series that is formed by summing up all positive unit fractions. The first ‘*n*’ terms of the harmonic series sum to approximately [ln(n) + γ], where(ln) indicates natural logarithm and $\gamma \approx 0.577$, which is the Euler–Mascheroni constant. As the number of terms tends to infinitely large value, the harmonic series does not have a finite limit and also tends to '$\infty$' infinity. It is thus a divergent series. For n=5, 10, 15…..100, the ζ(1) values are found as:

$$\zeta(1)=\sum_{n=1}^{5} n^{-1}=\ln(5)+0.577=2.18643$$
$$\zeta(1)=\sum_{n=1}^{10} n^{-1}=\ln(10)+0.577=2.87959$$
$$\zeta(1)=\sum_{n=1}^{15} n^{-1}=\ln(15)+0.577=3.28505 \qquad \text{…..(10)}$$
$$\zeta(1)=\sum_{n=1}^{100} n^{-1}=\ln(100)+0.577=5.18217$$
$$\zeta(1)=\sum_{n=1}^{1000} n^{-1}=\ln(1000)+0.577=7.48476$$

Hence, the values of ζ(1) are gradually increasing with the values of 'n', but it is very slowly increasing. However, for infinite number of 'n' values, $\zeta(1)\to\infty$.

Now, for r = 2 and integral values of 'n', Equation (8) may be written as:

$$\zeta(2)=\sum_{n=1}^{\infty} n^{-2}=\sum_{n=1}^{\infty}\frac{1}{n^2}=\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\cdots\ldots=\frac{\pi^2}{6}=1.64493 \qquad \text{…..(11)}$$

The ζ(2) is a converging series that yields the value of 1.64493.

Similarly,

$$\zeta(3)=\sum_{n=1}^{\infty} n^{-3}\approx\sum_{n=1}^{1000}\frac{1}{n^3}=\frac{1}{1^3}+\frac{1}{2^3}+\frac{1}{3^3}+\frac{1}{4^3}+\cdots\ldots+\frac{1}{1000^3}\ldots=1.20205 \qquad \text{…….(12)}$$

ζ(3) = 1.20205, for, n = 1000, but ζ(3) has no exact value like ζ(2) for $n\to\infty$. Hence, ζ(3) is approximated for n = 1000.

Now,

$$\zeta(4)=\sum_{n=1}^{\infty} n^{-4}=\sum_{n=1}^{\infty}\frac{1}{n^4}=\frac{1}{1^4}+\frac{1}{2^4}+\frac{1}{3^4}+\frac{1}{4^4}+\cdots\ldots=\frac{\pi^4}{90}=1.08232 \qquad \text{…..(13)}$$

The values of ζ(r) for r = 5 to 10 are listed below:

ζ(5)=1.03692,ζ(6)=1.01734,ζ(7)=1.00832,ζ(8)=1.00407,ζ(9)=1.00200, ζ(10)=1.00099, and so on ………………..(14)

An analysis of the Equations (11) to (14) instantly reveals that the values of ζ(r) gradually tend to one (1) with the increase in the values of 'r'. Now, the tendency of ζ(r) to 'one' implies ε′ tends to 1 and 'δ' tends to zero (Equation (3)), which in turn indicates that *h*-core citation tends to zero. But, it is not practically possible, i.e. the case of zero *h*-core citations is not practically possible. If any item receives just one citation, even then its *h*-index is reckoned as '1'. It may thus be logically inferred that 'δ' can not be equal to zero. Consequently, the values of ζ(r) for $r\geq 7$ gradually loses significance as ζ(7)=1.00832, i.e., the '0' occurs upto second decimal place for the first time so that from ζ(7) onwards the values may be approximated to '1'. The values of ζ(5) and ζ(6) are approximately 1.04 and 1.02 respectively, which are slightly greater than '1'.

The value of ζ(2) ≈1.62, which indicates the ratio of total citation to net excess citation is 1.62,

Or, $\delta+\delta^2+\delta^3+\cdots\ldots=0.62$, as evident from Equation (3). Now, δ represents the ratio of *h*-core citation to total citation, which is less than 1 (δ < 1) and the series ($\delta+\delta^2+\delta^3+\cdots\ldots=0.62$) is a converging infinite series. It is obvious in the summation of the above series that the major contribution of the sum (0.62) comes from δ followed by $\delta^2$, $\delta^3$ … and so on, i.e., higher orders of δ. Here, if around 60% contribution comes from δ, i.e., the 60% of total citation resides in the *h*-core zone. Thus ζ(2) ≈ 1.62 means *h*-core citation amounts approximately 60% of the total citation. Similarly, ζ(3) ≈1.20 means *h*-core citation amounts approximately 20% of the total citation. Also, ζ(4) ≈1.08, ζ(5) ≈1.04 and ζ(6) ≈1.02 mean *h*-core citation amounts approximately 8%, 4% and 2% of the total citations respectively. The higher order Riemann Zeta Function thus indicates gradually lower magnitude of *h*-core citation compared to total citation. The size of *h*-core zone or *h*-core square gradually diminishes with the higher order Riemann zeta Function, which almost vanishes at 8th order Zeta Function. This is the reason, the higher order (>7) Zeta Functions do not stand logically as *h*-core square cannot vanish totally at all.

## CONCLUSION

A new mathematical approach has been presented here involving Riemann Zeta function with two variables, viz. ε′ and δ. The two variables, i.e. ε′ and δ may be regarded as two different indicators and there is scope for further research with these two ratios. These two indicators will be tested for different cited items (authors, journals etc.) to observe the pattern. Also, the distribution of these two ratios for several cited items may be analysed to find out their normal pattern.

## ACKNOWLEDGEMENT

This work is executed under the research project entitled Design and development of comprehensive database and scientometric study of Indian research output in physics and space science since independence sponsored by Department of Science and Technology, Govt. of India under NSTMIS scheme, (Vide F. No. DST/NSTMIS/05/252/2017-18 dated 11/01/2018).

## REFERENCES

Dutta, B. (2020). Citation Swing Factor: An Indicator to Measure the Diffusion of Cited Items. Journal of Scientometric Research, 9(2), 214–218. https://doi.org/10.5530/jscires.9.2.26.

Egghe, L. (2006). An Improvement of the *h*-index. G-Index. ISSI Newsletter, 2(1), 8–9.

Egghe, L. (2006). Theory and Practise of the g-index. Scientometrics, 69(1), 131–152. https://doi.org/10.1007/s11192-006-0144-7

Egghe, L. (2011). Characterizations of the generalized Wu- and Kosmulski-indices in Lotkaian systems. Journal of Informetrics, 5(3), 439–445. https://doi.org/10.1016/j.joi.2011.03.006

Fassin, Y., & Rousseau, R. (2018). The h (3)–index for academic journals. Malaysian Journal of Library and amp. Information Sciences, 24(2), 41–53.

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. Proceedings of the National, 102(46), 16569–16572. arXiv:physics/0508025. https://doi.org/10.1073/pnas.0507655102

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. https://doi.org/10.1007/s11434-007-0145-9

Kosmulski, M. (2006). A new Hirsch-type index saves time and works equally well as the original index. ISSI Newsletter, 2(3), 4–6.

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. https://doi.org/10.1002/asi.21678

Wu, Q. (2010). The w-index: A Measure to Assess Scientific Impact by Focusing on Widely Cited Papers. Journal of the American Society for Information Science and Technology, 61(3), n/a–n/a. https://doi.org/10.1002/asi.21276

Zhang, C. T. (2009). The *e*-index, complementing the *h*-index for excess citations. PLOS ONE, 4(5), e5429. https://doi.org/10.1371/journal.pone.0005429

**Cite this article:** Dutta B. An interpretation of *h, e* and *R* indices from the viewpoint of Riemann Zeta Function. Journal of Data Science, Informetrics, and Citation Studies. 2023;2(1):21-4.