Interdisciplinary Description of Complex Systems 11(2), 227-237, 2013

DISTRIBUTION OF CITATIONS IN ONE VOLUME OF A JOURNAL

Antonio Protić[1], Biserka Runje[2] and Josip Stepanić[2, *]

[1]c/o Faculty of Mechanical Engineering and Naval Architecture – University of Zagreb
Zagreb, Croatia

[2]Faculty of Mechanical Engineering and Naval Architecture – University of Zagreb
Zagreb, Croatia

DOI: 10.7906/indecs.11.2.5
Regular article

Received: 2 April 2013.
Accepted: 14 April 2013.

ABSTRACT

Citations to published scientific articles are regularly collected and processed, bringing about the impact factor and a large number of other bibliometric indicators. We interpret the set of citations collected during fixed period as a characteristic statistical distribution of citations, argue about its properties and conjecture what statistical measures represent reliably such distributions. In that way we try to contribute to determining precisely the scope and level of suitability of impact factor if accompanied with a small set of additional indicators, all derived solely from the distribution function.

KEY WORDS

citation, bibliometric indicator, impact factor, distribution function, sensitivity

CLASSIFICATION

APA: 4050
JEL: D80, D81
PACS: 89.75.Kd

*Corresponding author, η: josip.j.stepanic@fsb.hr; +385 1 6168592;
FSB, I. Lučića 1, HR – 10 000 Zagreb, Croatia

INTRODUCTION

References and citations are two of necessary elements of scientific publications. Total number and other derivatives of citations of a scientific text are regularly used in diverse characterisations related to scientific publishing. Historical example is introduction of impact factor, a quantity determined annually for a given journal [1]. During times, use of impact factor both diffused and broadened. Nowadays, because of diffusion, the impact factor is important for careers of individual authors of scientific publications. Because of broadening, the use of impact factor is accompanied with the use of other bibliometric indicators, attributable more or less to individual authors. These processes were facilitated by the development of technologies for semi-automatic archiving and processing of scientific publications. In practice, the technologies enabled interested parties, such as librarians, scientists, governmental and non-governmental funding bodies, administrators of data-bases and search engines, etc. to develop their own indicators.

Moreover, a considerable number of studies of quantitative aspects of citation distributions have been undertaken, focusing onto different, still not completely clear aspects of citation distributions [2]. As a regular characteristic of a citation distribution they confirmed its skewness. The tail of citation distribution function is modelled as power-law [3-6], yet without consensus about its details. Eom and Fortunato [3] compare a lot of approaches to citation distributions and dynamics and extract shifted power-law as their underlying functional form. Golosovsky and Solomon [4] exhaustively treat citations to articles in physics journals in recent decades and show that heavy tail of the distribution is well described by power-law function, but that for extreme tail it is nevertheless improper, i.e. a runway behaviour is observed. In order to cover the skewness, or weightednedd, Wagner and Leydesdorf [5] discuss the integrated impact indicator. Leydesdorf [6], stressing that one is not allowed to compare impact factors even across neighbouring fields and subfields, discusses further aspects of intuitively simple indicator, aligned with the state-of-the-art and allowable for statistical testing.

Listed and similar studies were conducted for a rather large set of citations, spanning several decades and including many journals. In that sense they are representative of a given discipline, and are large enough to extract subtle effects, thus underlying distributions can be modelled in stringent details. In case of a smaller number of citations analysed, it was shown that indicators, introduced on the basis of large set of citations, show unpredicted effects [7]. In addition to citations analyses, indicators focused onto differently defined collection of articles have been developed. Rodríguez-Navarro [8] introduced x-index, a simple and precise indicator for high research performance related to countries and institutions, not the journals.

Radicchi and Castellano [9] concentrated on the problem that absolute values of some indicators systematically differ among various disciplines. In order to suppress that effect, they applied reverse engineering approach to study the citation patterns of millions of articles. As a result they derived transformations base on power-law function which suppress the disproportionate citation counts among scientific domains. Their result was further deepened with additional analyses conducted by Waltman, van Eck and van Raan [10] who concluded that although many fields indeed seem to have fairly similar citation distributions, there are quite some exceptions as well.

Stated references focus onto clarification of the scope and meaning of bibliometric indicators within the context of the traditional publishing, i.e. scientific journals. However, the digital age has brought about new forms for disseminating information, such as are web-based forms. Lozano, Larivière and Gingras [11] analyse different web-publishing forms and reach the conclusion that in the contemporary, digital age impact factor is less connected to articles' citations. Before the digital age, the citation rate of any given article and its journal's impact factor mutually reinforced each other. According to them, since 1990 and the advent of the digital age, the strength of the relation between impact factors and article citations has been decreasing [11]. In other words, since 1990, the proportion of highly cited papers coming from highly cited journals has been decreasing, and accordingly, the proportion of highly cited papers not coming from highly cited journals has also been increasing [11]. They project that such a trend will continue, and "should this pattern continue, it might bring forth the end to the use of the impact factor as a way to evaluate the quality of journals, papers and researchers and have interesting implications for the future of scientific literature" [11]. Similarly, Evans, Hopkins and Kaube analyse different publishing forms [12] and propose new bibliometric indicator that counts both citations and references. They apply such an indicator to set of publications from several, qualitatively different, well defined institutions and in addition to an internet archive. Attempts to resolve problems with improper use of existing bibliometric indices motivated Frittelli and Peri [13] to formulate scientific research measures, which originate from the more recent developments in the theory of risk measures. In particular, they are based on the Coherent Risk Measures.

While interdisciplinarity suppressed the use of impact factor as universal measures, along with other proposals to overcome it, e.g. [6, 9], Silva et al. [14] confirmed quantitatively that science fields are becoming increasingly interdisciplinary, with the degree of interdisciplinarity (for which they exploit entropy) correlating strongly with the in-strength of journals and with the impact factor. Albarrán, Ortuño and Ruiz-Castillo [15] analyse low-impact and high-impact measures as distinct measures to be applied for comparing the citation distributions of research units working in the same homogeneous field. They suggest using two real valued indicators to describe the shape of any distribution: a high-impact and a low-impact measure defined over the set of articles with citations above or below the critical citation level. The key to this methodology is the identification of a citation distribution with an income distribution [15].

There are two general characteristics of diverse bibliometric indicators, thus quantitative measures used for bibliometric characterisation of some entity, no matter whether that be an article, journal, individual author or an institution. First, bibliometric indicators are well defined and straightforwardly determined. Secondly, scientific works of higher quality by prescribed criteria, are to be considered within scientific community receive better funding. As a consequence, in many cases nowadays, quality of scientific work is related to quantitative indicators. The last sentence is generally treated as an oximoron, and a lot of efforts and researches have been conducted in order to either making smaller the gap between the quality and quantity in a context of scientific publications, or to clearly express the finite difference between them.

However, despite the fact that a lot of bibliometric indicators have been introduced after impact factor (as is stated, we are currently experiencing an explosion of research metrics [4]), yet we consider that the very journals and particular volumes are not covered appropriately.

We aim to contribute to broadening the characterisation of scientific impact of a given journal. Thus we concentrate on the set of citations to scientific articles published within a given volume of scientific journal. Our conjectures are that such a set of citations: (a) is proper basis for defining more comprehensive bibliometric indicators of the journal's scientific impact, (b) should be considered statistically as a representation of a proper distribution function.

In order to test these conjectures, we ask what could be additional bibliometric indicators which: (i) are straighforwardly determined from the set of collected citations for articles within one volume of a given scientific journal, (ii) are statistically consistent, which (iii) contribute to more detailed characterisation of scientific impact of a given journal, and which (iv) minutely depend on the large-scale changes in scientific publishing. Before proceeding let us emphasise that such indicators are to be used solely for characterisation of scientific journals, along with impact factor already determined for that purpose. The underlying reasoning is that by defining clear, unique, reliable set of bibliometric indicators of a given scientific journal one broadens understanding of what part of quality of a scientific work can be expressed using quantitative indicators.

In this article we describe initial set of such measure, put it into a proper context, and argue how they fulfil the listed points (i) and (ii). Along with the average, the measures are standard deviation, and corresponding sensitivities, of a distribution posed for a set of citations to scientific articles published in one volume of a given journal. For clarity, further in the text we assume without explicit mentioning that set of citations and citation distribution function refer to scientific articles published in one volume of a given scientific journal.

Section two focuses onto a citation distribution function and presents both the methodology for its construction and its general characteristics. Section three contains results of determining citation distribution function for several journals. Section four discusses obtained results and section five presents conclusions along with projections of further work.

STATISTICAL DISTRIBUTION FUNCTION

CONSTRUCTION

The starting point is a set of citations to the group of articles which were published within a given volume of a chosen scientific journal. Citations included in the set generally occur in diverse publications during generally unspecified time interval. The elements of that set should be precisely defined in order to obtain the uniquely constructed set of citations for further analyses. Before proceeding, let us emphasise that we ask in this article what could be additional bibliometric indicators. Regarding that, we implicitly assume that the very definition and overall usefulness of such indicators do not crucially or substantially depend on the very volume of the citations considered. In that sense, in this article the chosen journal would be any of journals included in the Web of Science. Given volume of that journal is any volume published at least two years ago. Set of citations includes all citations as listed in the Web of Science database, collected during time interval from the year of publishing of the given volume to present time. Such a determination invokes systematic difference between e.g. the impact factor and here introduced indicator. However, we assume that the difference is of the similar amount in the most of cases. Naturally, the more stringent the assumed use of the indicator the more prescribed the set of citations.

Let A be the number of articles in the group. We enumerate them using index i = 1, ..., K by some non-specified criteria, e.g. ascending date when manuscripts were received. Number of citations for i-th article is n_i and the total number of citations for the given volume is $N = n_1 + ... + n_K$. Relative frequency of citation to the i-th article is $f_i = n_i/N$. The set

$$\{f_1, f_2, \dots, f_K\}, \tag{1}$$

is a distribution of citations (DC) to the group of articles which were published within a given volume of a chosen scientific journal.

GENERAL CHARACTERISTICS

Modelled citation probability distribution function (CPDF) is the following

$$p(x;a,b) = b^{1-1/a} \cdot \frac{\sin(\pi/a)}{(\pi/a)} \cdot \frac{1}{b+x^a}\,, \tag{2}$$

where x is the continuous variable, a generalisation of the number of citations, while a and b are two parameters to be determined from fitting (2) to (1) using the nonlinear least square method [16]. Parameter a is responsible for large citation tail of (2). In that region of x axis the function acquires the power-law form

$$p(x{>>}b;\, a,\, b) \sim x^{-a}, \tag{3}$$

while in the opposite region, $x << b$ function (2) tends to a constant value. The form (3) is aligned with a number of results based on analyses of a significant quantity of citations. Since N as defined here is considered to be much smaller than the number of citations utilised in other contexts, we expect that here discrimination of different, yet qualitatively similar forms, is not possible. In that sense, form (3) approximates sufficiently well expected tails of DC. In the region of small number of citations, $x << b$, similarly, we expect that the shape of DC could not discriminate among many functionally different, yet qualitatively similar functions which is why we choose a form that in a simple way grasps non-divergent behaviour of low-number of citations limit. Before proceeding, in (2) we assume that x is continuous, real, non-negative variable which integer values have the meaning of possible number of citations. It is interesting to note that the form (2) resembles modified Pareto distribution, originating in the economic context thus represents another contribution of other disciplines to formulating of bibliometric indicators [14-16]. To summarise, form (2) has two parameters which are separately related to the two observed, mutually different parts of the DC. We assume that in case of rather large N the DC tends to a definite CPDF.

First characteristic of the CPDF is its first moment, the expected value. It is closely related to the impact factor, the differences being stringently prescribed number of years during which the citations to a set of articles are collected as well as the set of literature sources within which citations are collected. Nevertheless, we assume that listed differences are not of substantial but technical character. Thus we consider the expected value of the CPDF to resemble the impact factor. For CPDF (2), the expected value exists for $a > 2$ and equals

$$\overline{x}(a,b) = \frac{b^{1/a}}{2\cos(\pi/a)}\,. \tag{4}$$

However, expected value is assumed on the basis of average value determined for DC. But, in statistics, average value is reasonable representative of a distribution if that distribution is a normal distribution. Since prevalently the DCs obtained are skewed distributions, average values do not reliable represent them. Instead, medians and related statistical quantities (deciles, quantiles, ...) are reliable representations. In order to relate as much as possible to impact factor, we proceed with (4). Its sensitivity κ_x is two-component function defined as follows:

$$\kappa_\mathrm{x}(a,b) = \left(\left(\frac{\partial \overline{x}}{\partial b}\right)_{a,b}, \left(\frac{\partial \overline{x}}{\partial a}\right)_{a,b} \right). \tag{5}$$

Second characteristic of a distribution is its variance, calculated as a combination of distribution's second and first moments. Standard deviation of CPDF (2) exists for $a > 3$ and equals

$$s(a,b) = b^{2/a}\left[\frac{\sin(\pi/a)}{\sin(3\pi/a)} - \frac{1}{4\cos^2(\pi/a)}\right]. \tag{6}$$

while corresponding sensitivity κ_s equals

$$\kappa_s(a,b) = \left(\left(\frac{\partial s}{\partial b}\right)_{a,b}, \left(\frac{\partial s}{\partial a}\right)_{a,b}\right). \tag{7}$$

Sensitivities are quantities which measure to what amount will CPDF (2) change, for given values of a and b, in case of a minute, formally unit, change in the DC (1). Figure 1 shows CPDF introduced in (1), while Figure 2 shows expected value (4) and variance (6). Graphs of different components of the two sensitivities have the form similar to that of the expected value and variance, especially in the region close to the lower boundary of the range of a for which they are defined. In that sense, for $a > 3$, components of sensitivity (7) of standard deviation have more pronounced dependence on the small changes in a and are more suitable for discriminating seemingly similar, yet quantitatively different CPDFs.

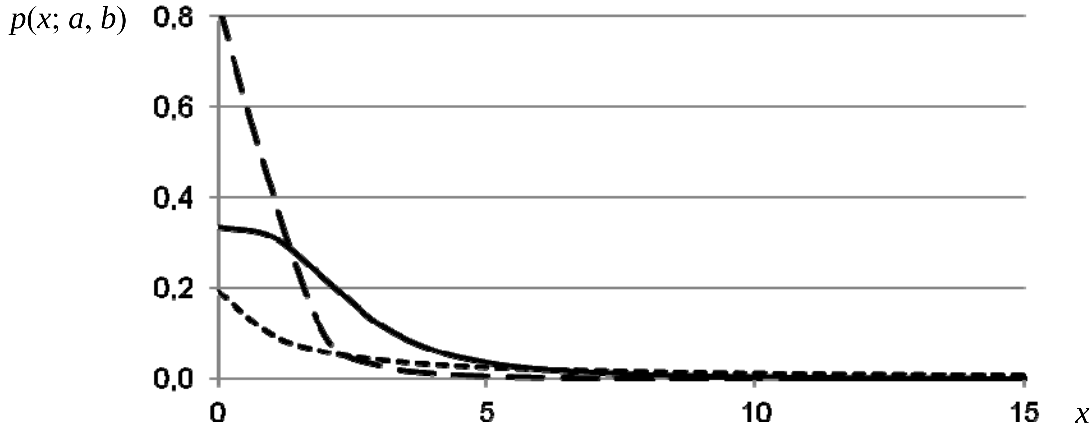

Figure 1. Graph of model CPDF (1) for a = 3 and b = 15 (solid line), a = 3 and b = 1 (dashed line), a = 1,2 and b = 1 (dotted line).

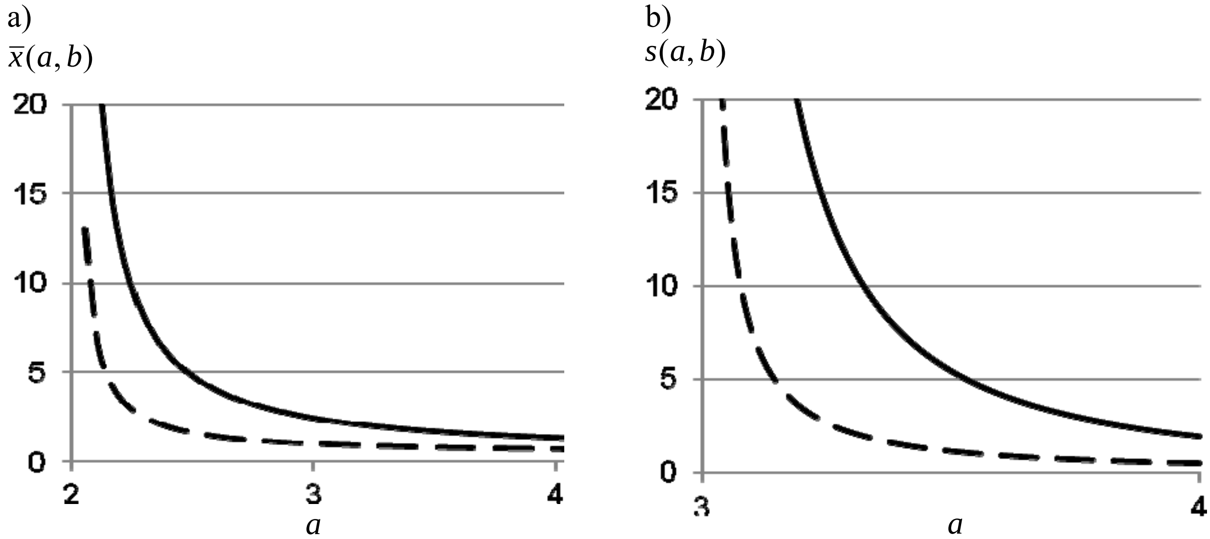

Figure 2. Graphs of model CPDF's a) average value, b) standard deviation as a function of a. Solid lines are for b = 15 and dashed for b = 1.

RESULTS

Citations were counted for three international scientific journals in the Web of Science database. We will refer to them as to the journals A, B and C, without listing explicitly their titles. All three journals belong to the same scientific discipline. Journal A is an international-level journal publishing scientific reviews, journal B is an international-level journal publishing prevalently regular scientific articles, while journal C is regional-level scientific journal publishing regular scientific articles. Impact factor of the journal A is the largest among these journals, and that of journal C the smallest. Impact factors of journals A and B are significantly larger than median for the underlying scientific discipline, while impact factor of the journal C is much smaller than that median.

All journals contain one volume per year, and the volume spans the complete calendar year. We considered volumes of the journal A published in the period from 2007 to 2011, volumes of the journal B published in 2010 and 2010, as well as the volume of the journal C published in 2010. Data were collected in November 2012 and included all citations that were recorded and available then in the Web of Science database. Number of articles and corresponding total number of citations are listed in Table 1. Medians, formally calculated average values and variances for the three journals are given in Table 2. Distributions of citations are shown as histograms in Figure 3.

Obtained histograms were fitted to (3), with the linear transformation in which a representative value of number of citations in a given bin (x) was linearly transformed to x' which coincides with the ordinal of that bin. There were no additional criteria for determining the width of the bins.

The software R-project Ver. 2.15.2 was utilised for fitting the expression $N_A \cdot p(x; a, b)$ to collected data within the nonlinear least square approach. Formally, in order to suppress all normalisations and include described linear transformation, data presented in Figure 3 were fitted to the form

$$\frac{k_1}{1+k_2 x^{k_3}}, \tag{8}$$

with the values obtained for coefficients $k_{1,2,3}$ as listed in Table 3. Graphs of functions (8) for a particular journal are in Figure 3 incorporated into histograms in order to make the comparison easier.

Table 1. Number of articles A published per considered journal and corresponding total number of citations N_A.

Journal	*A*	N_A
A	99	6067
B	159	2691
C	73	56

Table 2. Elementary statistical quantities for the DCs of three considered journals: medians, formally calculated average values and standard deviations.

Journal	Median	Average value	Standard deviation
A	31	61,3	76,9
B	12	16,9	1,3
C	0	0,8	15,5

Table 3. Values of coefficients in obtained for modelled CPDF (8).

Journal	k_1	k_2	k_3
A	277,4	3,870	1,895
B	45,89	0,002	4,792
C	64,80	0,079	6,107

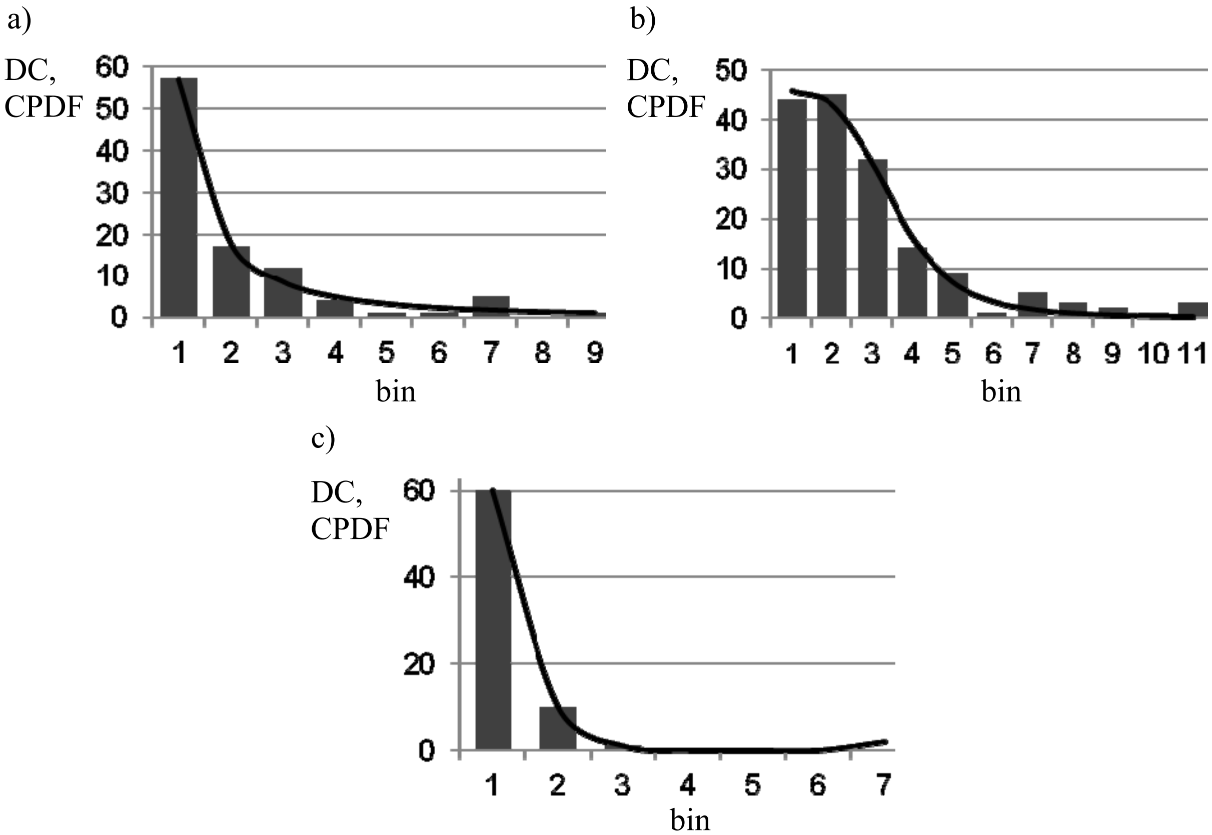

Figure 3. Graphs showing dependencies of DCs and CPDFs of citations for journal a) A, b) B and c) C. Grey bars refer to histogram of counted citations, while solid black lines refer to fitted values based on (2). Values on the axis abscissas are ordinals of the bins of histograms. These ordinals are linearly related to the corresponding average number of citations in a bin.

We utilised the nonparametric Kolmogorov–Smirnov test for testing the goodnes of fit. The stated hypothesis was that the data follows the specified distribution, and the alternative hypothesis was that the data does not follow a specified distribution. For each journal, the test statistic D is less than the critical value D_{critical}, therefore we do not have sufficient evidence to reject the stated hypothesis. It can be concluded that, with the level of significance being 0,05, the data follows the specified distribution.

Table 4. Data for the performed nonparametric Kolmogorov-Smirnov test for the significance level 0,05.

Journal	D	D_{critical}
A	0,030	0,388
B	0,064	0,452
C	0,027	0,381

DISCUSSION

Data collected represent rather small set, sufficient for presenting introduced idea, yet clearly too small to obtain precise amounts of bibliometric indicators for their utilisation in other areas. Yet, collected data are of rather large variability within a given discipline since they include review journal and scientific journal, international and regional journal, with rather large and with rather small average number of citations (collected within the given time span) etc. All stated brought about manifestly rather different histograms of DCs, shown in Fig. 3. The chosen model of CPDF as given by (3), previously encountered prevalently within the context of measuring income, with its free parameters adapts rather well to collected data.

While coefficient k_1 is combination of a and b (for normalised CPDF) as well as total number of citations N for non-normalised CPDF, coefficients k_2 and k_3 have simpler interpretation

$$b = 1/k_2,\ a = k_3. \tag{9}$$

In that sense, it is seen that among the analysed journal's citation sets, that of the journal A correspond to the set that has rather suppressed influence of the articles with low number of citations (i.e. has the smallest b). Correspondingly, that journal's citation set analysis resulted in the most pronounced influence of the articles with large number of citations (i.e. has the smallest a). For different choice of bins in obtaining histograms, it is expected that values in Table 3 would change, but without the need for changing the underlying functional forms, i.e. (8) or (2).

In deriving (4) and (6), the restrictions $a > 2$ and $a > 3$, respectively, were encountered. Set of journals B and C falls within the available region of a, while set of journal A violates both of imposed restrictions. That has consequences in further extracting and prescribing bibliometric indicators. Since such indicators will be calculated on a given DC, they will naturally be finite, yet one must additionally analyse what is the range within which calculated estimates can be reliably used if their theoretical limits diverge for some set of journals, i.e. DCs. Alternatively, instead of (2) one can use its finite version (c.f. equation (4) in [16]).

CONCLUSIONS AND PERSPECTIVE

This article emphasises importance to introduce reliable set of bibliometric indicators which are intrinsic to the journal in the sense that they take explicitly as few as possible indicators not belonging to a given journals, such as are quantities determined using data for other journals as well. In that way, the emphasis is in building the complete set of coherent intrinsic bibliometric indicators for a journal. Following that, the foundation for such an approach is formulated, illustrated and analysed, based formally on the Pareto-like functional form for modelled citation probability distribution function.

For the group of obtained distributions of citations it was shown that prescribed form (3) makes possible their reliable representation in the two-parameter set spanned by power-law exponent a that was previously encountered in analyses of large-citation tail of distribution functions, and by the parameter b representing influence of the small-citation end of the distribution function. Currently, there is a large number of parameters to be check and modified if needed, thus the first step in prospective research is to apply the function (3) (or its modification given by equation (4) in [16]) onto sets of distributions of citations for larger number of scientific journals.

REFERENCES

[1] Jokić, M. and Ball, R.: *Qualität und Quantität wissenschaftlicher Veröffentlichungen.* Forschungszentrum Jülich GmbH Zentralbibliothek, 2006,

[2] Leydesdorff, L.; Bornmann, L.; Mutz, R. and Opthof, T.: *Turning the tables on citation analysis one more time: Principles for comparing sets of documents*.
Journal of the American Society for Information Science and Technology **62**(7), 1370-1381, 2011,
http://dx.doi.org/10.1002/asi.21534,
[3] Eom, Y.-H. and Fortunato, S.: *Characterizing and Modeling Citation Dynamics*.
PLoS ONE **6**(9), e24926, 2011,
http://dx.doi.org/10.1371/journal.pone.0024926,
[4] Golosovsky, M. and Solomon, S.: *Runaway Events Dominate the Heavy Tail of Citation Distributions*.
arXiv:1206.1999v1 [physics.soc-ph],
http://arxiv.org/abs/1206.1999v1, accessed 1st April 2013,
[5] Wagner, C.S. and Leydesdorff, L.: *An Integrated Impact Indicator (I3): A New Definition of "Impact" with Policy Relevance*.
arXiv:1205.1419v1 [cs.DL],
http://arxiv.org/abs/1205.1419v1, accessed 1st April 2013,
[6] Leydesdorff, L.: *Alternatives to the journal impact factor: I3 and the top-*10% *(or top-*25%*?) of the most-highly cited papers*.
Scientometrics **92**(2), 355-365, 2012,
http://dx.doi.org/10.1007/s11192-012-0660-6,
[7] Schreiber, M.: *Inconsistencies of Recently Proposed Citation Impact Indicators and how to Avoid Them*.
arXiv:1202.3861v2 [stat.AP],
http://arxiv.org/abs/1202.3861v2, accessed 1st April 2013,
[8] Rodríguez-Navarro, A.: *A Simple Index for the High-Citation Tail of Citation Distribution to Quantify Research Performance in Countries and Institutions*.
PLoS ONE **6**(5), e20510, 2011,
http://dx.doi.org/10.1371/journal.pone.0020510,
[9] Radicchi, F. and Castellano, C.: *A Reverse Engineering Approach to the Suppression of Citation Biases Reveals Universal Properties of Citation Distributions*.
PLoS ONE **7**(3), e33833, 2012,
http://dx.doi.org/:10.1371/journal.pone.0033833,
[10] Waltman, L.; van Eck, N.J. and van Raan, A.F.J.: *Universality of citation distributions revisited*.
arXiv:1105.2934v3 [cs.DL],
http://arxiv.org/abs/1105.2934v3 accessed 1st April 2013,
[11] Lozano, G.A.; Larivière, V. and Gingras, Y.: *The weakening relationship between the Impact Factor and papers' citations in the digital age*.
arXiv:1205.4328v1 [cs.DL],
http://arxiv.org/abs/1205.4328v1, accessed 1st April 2013,
[12] Evans, T.S.; Hopkins, N. and Kaube, B.S.: *Universality of Performance Indicators based on Citation and Reference Counts*.
arXiv:1110.3271v2 [physics.soc-ph],
http://arxiv.org/abs/1110.3271v2, accessed 1st April 2013,
[13] Frittelli, M. and Periy, I.: *From Risk Measures to Research Measures*.
arXiv:1205.1012v1 [q-fin.RM],
http://arxiv.org/abs/1205.1012v1 accessed 1st April 2013,
[14] Silva, F.N.; Rodrigues, F.A. Oliveira Jr, O.N. and da F. Costa, L.: *Quantifying the interdisciplinarity of scientific journals and fields*.
arXiv:1203.4807v1 [physics.soc-ph],
http://arxiv.org/abs/1203.4807v1, accessed 1st April 2013,

[15] Albarrán, P.; Ortuño, I. and Ruiz-Castillo, J.: *The Measurement of Low- and High-Impact in Citation Distributions: Technical Results*.
Working Paper 09-57 Economics Series (35), Universidad Carlos III de Madrid, 2009, http://www.eco.uc3m.es/temp/The%20measurement%20of%20low%20and%20high-impact.pdf,
[16] Stepanić, J.; Štefančić, H. and Zlatić, V.: *Social Free Energy of a Pareto-like Resource Distribution*.
Interdisciplinary Description of Complex Systems **4**(2), 136-143, 2006, http://indecs.eu/index.php?s=x&y=2006&p=136-143.

RASPODJELA CITATA U JEDNOM GODIŠTU ČASOPISA

A. Protić[1], B. Runje[2] i J. Stepanić[2]

[1]Sveučilište u Zagrebu – Fakultet strojarstva i brodogradnje
Zagreb, Hrvatska
[2]Sveučilište u Zagrebu – Fakultet strojarstva i brodogradnje
Zagreb, Hrvatska

SAŽETAK

Citiranja objavljenih znanstvenih radova redovito se pohranjuju i obrađuju, čime se dolazi do faktora utjecaja i velikog broja drugih bibliometrijskih indikatora. Citate prikupljene tijekom određenog razdoblja razmatramo kao karakterističnu statističku raspodjelu citata te diskutiramo o njenim svojstvima i postavljamo tvrdnje o tome koje statističke mjere pouzdano predstavljaju takve raspodjele. Time nastojimo doprinijeti preciznom određivanju opsega i razine primjenjivosti faktora utjecaja u slučaju kad je nadopunjen manjim brojem dodatnih indikatora, koji su svi određeni isključivo na temelju raspodjele citata.

KLJUČNE RIJEČI

citat, bibliometrijski indikator, faktor utjecaja, funkcija raspodjele, osjetljivost