LIFETIME CITATION RATES AS A BASIS FOR COMPARISONS WITHIN A SCIENTIFIC FIELD

Nancy L. Geller, John S. de Cani and Robert E. Davies

University of Pennsylvania

0. Abstract

Under reasonable assumptions a formula is derived to project the number of citations likely to be received in the future by the usual type of scientific research paper. This projection, the paper's Lifetime Citation Rate, is based on the paper's past citations and the regularity of citation patterns. Hypothesis tests of the equality of expected Lifetime Citation Rates per paper and per authorship are also derived. Under appropriate precautions comparisons may then be made of the work of scientists in similar fields who have been publishing for different lengths of time. The method is applied to the members of a university department in which discrimination on the basis of sex was alleged.

1. Introduction

With several important precautions, the number of citations to a paper in major scientific journals can give a measure of the importance of the work. Lists of such citations since 1961 can be found in the volumes of Science Citation Index (1). Herein we report a method for comparing the work of scientists who have been publishing in the same field for different lengths of time. Under certain assumptions we project the number of citations a paper will receive in the forty years following its publication based on the regularity of citation patterns, exponential growth of the number of scientific papers, and the number of citations the paper has already received. We call this projection the "Lifetime Citation Rate" for the paper.

We also use statistical inference to compare the Lifetime Citation Rates of different authors. Comparisons are then made for members of a preclinical department of an eastern university where discrimination on the basis of sex was alleged.

2. Derivation of Lifetime Citation Rate Formula

We first assume:
(A1) Exponential growth of the number of scientific publications, where we put α as the power of 2 to indicate the doubling rate, and (A2) u_j, the expected number of citations of a paper of age j (selected at random), is a function of its age, not its date of publication.

Assumption (A1) is well-justified in the literature (c.f. (2), (3)). Assumption (A2) may be justified by the following: Consider all citations (recorded in S.C.I.) in a given year, say 1972. Each citation of 1972 refers to a paper written in 1972 or before. Thus we may calculate what percent of citations of 1972 refer to papers written in 1972, 1971, ... The age by years of papers cited in 1972 versus the proportion of papers of each age is shown by the step function in Figure 1. The smooth curve in Figure 1 corresponds to the "best fit" (by least squares) of the density function of the sum of two independent exponential random variables.

The general shape of the curve(s) in Figure 1 is maintained when we look at data for years other than 1972.

Let $P_j(t)$ be the probability a (randomly selected) paper cited in S.C.I. in year t is of age j. Under (A1) and (A2).

$$P_j(t) = \alpha^{t-j} u_j / \sum_{k=0}^{t} \alpha^{t-k} u_k .$$

If, further,

(A3) $\{u_k\}$ is bounded,

then as $t \to \infty$

$$P_j(t) \to \alpha^{-j} u_j / \sum_{k=0}^{\infty} \alpha^{-k} u_k = \pi_j . \quad (1)$$

We estimate π_j by $\hat{\pi}_j$, the proportion of papers of age j cited in any year, as given by the most recent citation decay curve, such as Figure 1.

Assume also
(A4) u_{ij}, the actual number of citations of paper i at age j, has a Poisson distribution with expected value proportional to u_j. From (1),

$$E(u_{ij}) = c_i \alpha^j \pi_j .$$

Thus the expected number of citations in its lifetime (truncated after 40 years) is:

$$c_i \sum_{j=0}^{39} \alpha^j \pi_j . \quad (2)$$

For a particular author, values of u_{ij} (since '61) are available from S.C.I.

and various estimates of α have been given in the literature (c.f. (2), (4)).

We estimate c_i by Maximum Likelihood Estimation:

$$\hat{c}_i = \Sigma u_{ij} / \Sigma \alpha^j \hat{\pi}_j \qquad (3)$$

where both the summations in (3) are on the index j, from max (0, 1962-t) to (1972-t) and t is the year of publication of the ith paper.

The total number of citations the ith paper will receive in its lifetime may thus be estimated using (2) and (3) by

$$\hat{c}_i \sum_{j=0}^{39} \alpha^j \hat{\pi}_j \qquad (4)$$

As an example, Professor B published a paper in 1964 which received 18 citations from the time it was published through 1972. Thus $\Sigma u_{ij} = 18$. We take $\alpha = \sqrt[8]{2}$ (c.f. (2)), and calculating the other factor, we get 2.649. Thus the estimate of the lifetime citations for this paper is 18 x 2.649 = 47.68.

Assuming independence of the number of citations a paper receives from one year to the next, we can also estimate the variance of a paper's Lifetime Citation Rate and hence the standard error. In the above example, our estimate plus or minus one standard error is 47.7 ± 11.2.

We now define an author's Lifetime Citation Rate per paper (LCR) as the sum of (4) over all the author's papers divided by the number of papers the author published.

In summary, to find the Lifetime Citation Rate for an author, we multiply the number of citations received by each paper by a factor which is a function of the paper's age. Then we add over all papers. It follows from (A4) that we are dealing with a weighted sum of Poisson variates.

In further work we derived an exact test for the equality of several weighted sums of Poisson variates. Here we approximate a weighted sum by a single weighted Poisson variate. The weight factor is dependent on the year in which the author first published and another assumption:

(A5) an author's remaining publications had equal probability of occuring in any of the years following his date of initial publication.

For the cases we considered the two estimates agreed quite well.

3. Hypothesis Test of Equality of Two LCR's

We now assume that each person's LCR is a Poisson variate v times a constant Θ. We wish to test the null hypothesis that the expected value of $v\Theta$ is the same for the two authors:

H_0: $E(v_1\Theta_1) = E(v_2\Theta_2)$ versus the alternative

H_1: $E(v_1\Theta_1) \neq E(v_2\Theta_2)$.

If two professors have LCR's equal in value,

v_1 should be close to $(v_1+v_2)\Theta_2/(\Theta_1+\Theta_2)$ and

v_2 should be close to $(v_1+v_2)\Theta_1/(\Theta_1+\Theta_2)$.

Thus we may use a χ^2 statistic with one degree of freedom to compare the actual values of v_1 and v_2 with their expected values under H_0:

	Actual	Expected
Dr. 1	v_1	$(v_1+v_2)\ \Theta_2/(\Theta_1+\Theta_2)$
Dr. 2	v_2	$(v_1+v_2)\ \Theta_1/(\Theta_1+\Theta_2)$

$$\chi^2 = ((v_1-(v_1+v_2)\Theta_2/(\Theta_1+\Theta_2))^2/(v_1+v_2)\Theta_2+ (v_2-(v_1+v_2)\Theta_1/(\Theta_1+\Theta_2))^2/(v_1+v_2)\Theta_1) (\Theta_1+\Theta_2).$$

The decision rule is: Reject H_0 if χ^2 is large.

4. Comparisons of LCR's for a Preclinical Department

Table 1 compares the LCR's of members of a preclinical department at an eastern university. Doctor A is a full professor and chairman of the department in question. Doctors B and C, both male, were promoted to associate professorships with tenure whereas Doctor D, a female, was refused promotion and tenure. Professors E,F,G are also full professors. Two calculations were made for Doctors A,B,C, and D: one based on all papers they wrote, attributing all their citations to them ("all") and one based on independent publications - those papers on which they were the author senior in rank ("ind"). Only the first calculation is done for Doctors E,F, and G. In further work we consider attributing an equal fraction of the citations of each paper to each author of that paper.

When measured by either version of the LCR, Doctor D's work is of higher equality than Doctor B's. When all papers are considered, Doctors A,C, and D have LCR's which are approximately equal. When only independent papers are considered, Doctors D and A have LCR's which are approximately equal, and Doctor C's LCR is lower. It thus appears that the quality of Doctor D's work is high compared to Doctors A,B, and C when measured by Lifetime Citation Rate.

The results of the hypothesis tests for Doctor D and the various members of her department are shown in Table 2. On the whole Doctor D's work is significantly better in quality than that of the two men who were promoted to associate professorships with tenure and fully comparable to the full professors in her department when measured by LCR.

5. Conclusions

Under reasonable assumptions we project the number of citations a scientific publication will receive in forty years based on the number of citations it already has received. We also make statistical comparisons between the projected citations per paper for different authors. Such information ought to be useful in promotion and tenure considerations since citations are a respected measure of quality of scientific work (c.f. (5), (6)). In applying our methods to a preclinical department of a medical school of an eastern university, we conclude that the quality of work of a woman who was not promoted nor given tenure was significantly better than the work of two men who were promoted and granted tenure. Moreover, by our measure, the woman's work was comparable in quality to full professors in her department.

References

(1) Science Citation Index 1972, Institute for Scientific Information, Inc., (ISI), Phila.,Pa., 19106.

(2) Price, D., Little Science, Big Science, Columbia University Press, 1963, p.9.

(3) Anderla, J.G., The Growth of Science and Technical Information: A Challenge, N.S.F., Washington, D.C., 1974, p.6.

(4) Meadows, A.J., Communication in Science, Butterworth, London, 1974, Ch. 1.

(5) Cole, J.R. and Cole, S., Social Stratification in Science, University of Chicago Press, Chicago, 1973, Ch. 2.

(6) Menard, H.W., Science: Growth and Change, Harvard University Press, Cambridge, Mass., 1971, Ch. 2

TABLE 1

Comparisons of citation rates per paper
of Dr. D and other members
of her department

	Date of 1st Publication	Number of Papers	Number of Citations	Citations Per Paper	Av. Lifetime Citation Rate Per Paper (approx.)
A (all)	1956	38	589	15.5	51.4
(ind)	1960	29	486	16.8	59.5
B (all)	1963	13	67	5.2	21.8
(ind)	1969	6	12	2.0	21.2
C (all)	1963	11	134	12.2	50.9
(ind)	1967	3	8	2.7	15.4
D (all)	1962	16	218	13.6	53.5
(ind)	1962	15	211	14.1	55.0
E (all)	1939	123	988	8.0	30.9
F (all)	1955	46	766	16.7	55.6
G (all)	1936	158	2082	13.2	52.3

TABLE 2

Results of Hypothesis tests of Equality of
LCR's of Dr. D and others in her department
("all" papers and "independent" papers)

	D compared with:
A (Full professor and Chairman)	not significant
B (Associate Professor)	D better
C (Associate Professor	D better for independent papers not significant for all papers
E (Full Professor)	D better
F (Full Professor)	not significant
G (Full Professor)	not significant

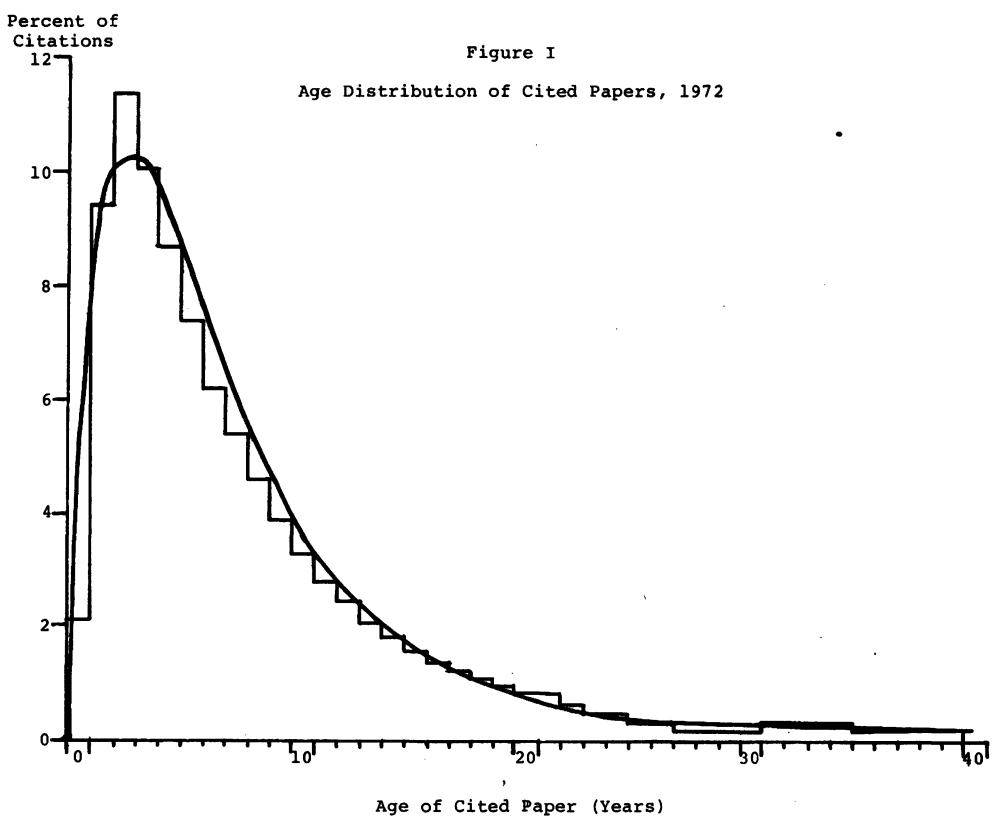

Figure I

Age Distribution of Cited Papers, 1972