Malaysian Journal of Library & Information Science, Vol.5, no.1, July 2000: 37-51

APPLICATION OF STATISTICAL MODELS TO THE COLLABORATIVE PUBLICATIONS DATA IN THEORETICAL POPULATION GENETICS

B.M.Gupta [1]
C.R.Karisiddappa [2]

[1]National Institute of Science, Technology and Development Studies,
Dr K.S.Krishnan Marg, New Delhi 110 012, India
e-mail:bmgupta1@yahoo.com
[2]Department of Library and Information Science,
Karnatak University, Dharwad 580 003

ABSTRACT

Studies the growth and proportion of different types of co-authored publications in theoretical population genetics. Explores the applicability of appropriate statistical model to the decline in the proportion of single-authored publication with time. Studies the applicability of selected statistical models to the distribution of authorship in publications of theoretical population genetics with time.

Keywords: Collaboration; Publications; Statistical models; Theoretical population genetics.

INTRODUCTION

Collaboration in the area of scientific research is defined as: "a process of functional interdependence between scholars in their attempt to co-ordinate skills, tools and rewards" (Patel, 1972). As a direct consequence of collaboration, multi-authorship appears in research papers. The first collaborated paper was published in 1655 (Beaver and Rosen, 1978). Since then the proportion of total collaborated papers has been increasing with time. According to Beaver and Rosen (1978, 1979a, 1979b), collaboration was initiated in response to professionalisation and increased knowledge, based on their conclusions from the analysis of publication data from 17th century to 20th century. According to Pao (1992), professionalisation refers to a complex set of dynamic processes involving organisations and scientists, individually and collectively, with respect to both the scientific group and community at large. It also sets the criteria for the new entrants, and formalises rules of behaviour among members, and promotes interaction between the specific groups and outsiders. Some indicators of professional maturation of an individual or scientific field can take the form of initiating professional societies, specialised journals, innovative applications from research programmes, and institutionalisation of prizes and awards. The society, in turn, also recognises the contributions of an individual scholar or professional by taking his/her professional advice in the field of policy-making. Beaver and Rosen have maintained those collaborative functions as a social regulator within the profession, through their influence on recognition and subsequent access on resources, and thereby increases the visibility and productivity of authors.

It was during the 20th century that the professionalisation in science had its maximum impact on the members of the scientific community. Because of this factor, there had been an increasing trend in the growth of collaborative publications in all disciplines of science and technology (S&T). However, the increase in the number and proportion of collaborative publications and their growth rates have been observed to vary from one subject to another, one branch to another, within the same subject, and from one country to another. Changes in the growth rate of collaborative publications observed in subject disciplines are probably related to changes in the acknowledgement of teamwork, as an accepted practice in collaboration that simultaneously effect changes in the support of S&T.

Price (1965) studied the phenomenon of collaboration in chemistry publications during 1910-1960, as reflected in the increase in multi-authored publications in *Chemical Abstracts* database. In the 1960s, the proportion of two-authored publications had reached around 40 per cent of the total publication output, three-authored publications made up 15 per cent, and four-authored publications were around 5 per cent. The proportion of three-authored publications showed faster growth than that of two-authored publications and similarly the proportion of four-authored publications showed a more rapid growth than that of three-authored publications. In the field of astronomy and allied sciences, it was found that proportion of single-authored publications stood at 95 per cent in 1910, fell to 86 per cent in 1940, then to 70 per cent in the mid-fifties, and to less than 50 per cent in the late sixties (O'Connor, 1969). In physics the proportion of single-authored publications had fallen from 75 per cent in the 1920s to 39 per cent in the 1950s. Keenan and Atherton (1964) have documented collaboration trends in different sub-disciplines of physics and noticed that solid state physics and nuclear physics have almost identical proportion of single-authored publications, but the extent and size of multi-authored publications vary between them. Even though 14 per cent of the papers have 4 or more authors in solid state physics, the corresponding figure for nuclear physics is only 4 per cent. The extent of collaboration is relatively small in mathematics (Price, 1965), where approximately 94 per cent publications in 1940 were single-authored, and by 1960, this proportion remained at a relatively high level of 79 per cent. Apart from mathematics, the discipline of science with the lowest scores in collaborative publications during the twentieth century has been geology, where the proportion of single-authored publications had decreased from 88 per cent in 1940 to 78 per cent in 1960 (Yitzhaki and Ben-Tamar, 1990). In contrast, within the discipline of science, the largest number and proportion of collaborative publications is in biomedicine. In a study on selective biomedical publications of the United States, Clark (1964) observed that the proportion of single-authored publications dropped from about 30 per cent at the end of the Second World War to about 20 per cent in the early sixties. From the above presentation, it is evident that all science disciplines show an increasing trend in their number and proportion of multi-authored publications, the actual rates of their increase, however, vary considerably from one discipline to another.

Compared to S&T, the situation is different in the social sciences. In sociology and psychology, the proportion of single-authored publications decreased from 98 per cent and 84 per cent in the 1920s to 72 per cent and 55 per cent in 1950s, respectively. In most other disciplines of social sciences (political science, anthropology, etc.), however, very little variation occurred in the proportion of single-authored publications. Similarly, nearly all arts disciplines continued to accumulate single-authored papers throughout the first half of the twentieth century (1974).

It is noticed that the proportion of single-authored publications is decreasing slowly with time in all disciplines, and sometimes, this can be as low as that for the biomedical publications. Thus, it becomes natural to ask as to what extent the decrease in proportion of single-authored papers will be and will this proportion become nonexistent after some time? The decline in proportion of single-authored papers in different disciplines can, in fact, be postulated through a simple model (O'Connor, 1969). The model should represent a decrease in proportion of single-authored publications as a function of time, by some kind of a quasi-logistic model (S-shaped curve), so that there is a preliminary period when the decline is slow, followed by a period of rapid decrease, and then a final period when the decreasing curve flattens again. He also stated that different disciplines follow this curve at varying speeds.

Efforts have also been made in the field to find appropriate probability distributions, which might describe the distribution of the number of authors per paper in different disciplines. Yitzhaki and Ben-Tamar (1990) published a comprehensive review indicating how the average number of authors per paper has changed with time in different disciplines.

Just as the number of papers published in journals by individual scholars can be considered as an indicator of their productivity, co-authorship in publications can be considered as an indicator of collaboration in a discipline. As this variable can only take discrete values, it is reasonable to assume that a discrete probability distribution would describe the distribution of number of authors per publication. Some scholars have studied this problem and their findings are briefly summarised here.

Price and Beaver (1966), using memos circulated among members of an invisible college, inferred a Poisson model for distribution of authorship in publications. Haitun (1982) used the Price and Beaver data, and classified the distribution of authorship using stationary scientometric distributions. One of the important reviews in this area was reported by Ajiferuke (1991), who studied 15 statistical probability distributions, namely Zipf, Mandelbrot, Geometric, Shifted Poisson, Shifted

Generalised Poisson, Logarithmic, Borel-Tanner, Shifted Yule, Shifted Generalised Warring, Shifted Inverse Gaussian-Poisson (IGP), Shifted Generalised Inverse Gaussian-Poisson (GIGP), Shifted Binomial, Shifted Beta-Binomial, Shifted Negative Binomial, and Shifted Generalised Negative Binomial and studied their applicability in 94 data sets on distribution of authorship, representing different disciplines from S&T, social sciences, and humanities. His results showed that Shifted Inverse Gaussian-Poisson distribution is the most appropriate statistical probability distribution, which best describes the distribution of authorship in publications. This statistical probability distribution is, however, difficult to understand by an average scholar, because it contains three parameters, besides a so-called modified Bessel function of the second kind. As a result, Rousseau (1994) tried selected one-parameter distributions, such as Lotka's, Geometric, and Zero-Truncated-Poisson distributions to the number of authors per publication in library and information science literature. He found that geometric and Truncated Poisson distributions most adequately describe the observed distribution of authorship in publications. Rousseau in collaboration with Gupta (1998), subsequently explored the Lotka's, Geometric, Truncated Poisson, Truncated Binomial, and Truncated Negative Binomial distributions in the distribution of authorship during 1976-80 in theoretical population genetics literature, derived from 11 core journals in the discipline. The results obtained by Rousseau in a study mentioned earlier is again corroborated and reinforced by these authors in this study which states that where single-authored publications dominate, Geometric and Truncated Poisson distribution most adequately describe the distribution of authorship. They have also noted that it is not clear whether the Truncated Binomial is an acceptable model for such observed distribution of authorship. Earlier scholars, to some extent, corroborated some of the results obtained by Rousseau and his colleagues, especially in the applicability of geometric distribution. For example, Goffman and Warren (1980) and Worthern (1978) found that geometric distribution adequately describes the distribution of authorship in schistosomiasis and drug publications, respectively.

OBJECTIVES

The main objectives of this study are:

(a) To study the growth in the number and proportion of collaborative publications in the field of theoretical population genetics research with time;

(b) To study the applicability of a suitable mathematical model in the decline of the proportion of single-authored publications with time; and

(c) To study the applicability of selected probability distributions to the distribution of authorship with time for publications in the field of theoretical genetics.

DATABASE AND METHODOLOGY

A bibliography entitled "*Bibliography of Theoretical Population Genetics"* (Rousseau, 1992*)* had been used as the main source of data for the present study. It covers different types of publications in the discipline of theoretical population genetics from 1881 to 1980. The present study covers publication from 1901 to 1980. Since the average number of publications per year covered in this bibliography is small (especially those before 1950), the publication data are cumulated in 5-year blocks. For studying the growth in the number and proportion of multi-authored publications in the field of theoretical population genetics, the entire data from the bibliography was used. This consists of 7,877 publications, of which 2700 are multi-authored. To explore the applicability of selected probability distributions to the number of authors per publication, publication data from 1941 to 1980 have been used because, prior to 1940, there were only a few collaborative publications.

DISCUSSION AND RESULTS

Overall Collaboration Profile in Theoretical Population Genetics

(a) Proportion of Collaborated Publications

The growth in the number of total publications and collaborative publications considered in 5-year period blocks have shown a consistent increasing trend with time since 1941-45 (Table 1). However, the proportion of collaborative publications in total publications has shown a consistent growth with minor fluctuations in certain period blocks.

Table 1: Total Publications and Collaborative Publications Between 1896-1980

Period blocks	**Number of publications**	**Number of collaborative publications**	
	Total	**Total**	**Percentage**
1896-1900	8	1	12.50
1901-1905	16	-	-
1906-1910	20	-	-
1911-1915	35	2	05.71
1916-1920	42	2	04.76
1921-1925	63	4	06.34
1926-1930	101	7	06.93
1931-1935	145	10	06.89
1936-1940	137	16	11.67
1941-1945	78	17	21.79
1946-1950	144	29	20.13
1951-1955	289	48	16.60
1956-1960	458	99	21.62
1961-1965	772	186	24.09
1966-1970	1426	420	29.45
1971-1975	1918	698	36.39
1976-1980	2225	862	38.74
Total	**7877**	**2401**	

Quantification of Publications by Number of Authors

In the total publication sample, 5,476 (69.51%) appeared as single-authored publications, 1,816 (23.05%) as two-authored publications, 443 (5.62%) as three-authored publications, 108 (1.37%) as four-authored publications, 22 (0.28%) as five-authored publications, and the rest 12 (0.15%) as more than five-authored publications (Table 2).

Analysing the percentage contribution of individual types of co-authored publications, a systematic increase is observed in all the categories of co-authored publications, although with different growth rates as described below:

(a) The proportion of two-authored publications has increased from 5.71 per cent during 1911-15 to 26.38 per cent during 1976-80, with an average of 14.7 per cent for the entire period.

(b) The proportion of three-authored publications has increased from 1.98 per cent during 1926-30 to 9.12 per cent during 1976-80; with an average of 0.94 for the entire period.

(c) The proportion of four-authored publications has increased from 0.06 per cent during 1951-55 to 2.24 per cent during 1976-80, with an average of 0.94 for the entire period.

(d) The proportion of five-authored publications has increased from 0.12 per cent during 1961-65 to 0.58 per cent during 1976-80, with an average of 0.27 for the entire period.

Table 2: Publications Classified by Number of Authors Between 1896-1980

Period blocks	Number of Publications	Number of publications classified by the number of authors				
	Total	One author	Two authors	Three authors	Four authors	More than 4 authors
1896-1900	8	7	-	1	-	
1901-1905	16	16	-	-	-	
1906-1910	20	20	-	-	-	
1911-1915	35	33	2	-	-	
1916-1920	42	40	2	-	-	
1921-1925	63	59	4	-	-	
1926-1930	101	94	5	2	-	
1931-1935	145	135	9	1	-	
1936-1940	137	121	15	1	-	
1941-1945	78	61	16	1	-	
1946-1950	144	115	28	1	-	
1951-1955	289	241	35	11	2	-
1956-1960	458	359	76	18	5	-
1961-1965	772	586	156	25	4	1
1966-1970	1426	1006	330	70	15	5
1971-1975	1918	1220	551	109	32	6
1976-1980	2225	1363	587	203	50	22
Total	**7877**	**5476**	**1816**	**443**	**108**	**34**

To obtain a picture of the relative growth of single-authored and multi-authored publications, data on co-authored publications was analysed in terms of the relative frequency of publications by number of authors for different period blocks from 1931-35 to 1976-80 (Table 3). The relative frequency of individual types of co-authored publications by number of authors is calculated (the total number of two-authored publications is 1816, and the relative frequency for 1976-80 is then calculated as 587/1816=0.3232). The analysis indicates that the growth in the frequency of all types of co-authored publications by the number of authors increased in different proportion, as we move from single-authored to four-authored publications. In single-authored publications the relative frequency has increased from 0.0246 during 1931-35 to 0.2489 during 1976-80, while in different types of multi-authored publications, it has increased from: 0.0049 during 1931-35 to 0.3232 during 1976-80 in two-authored publications; 0.0248 during 1951-55 to 0.458 during 1976-80 in three-authored publications; 0.0185 during 1951-55 to 0.4629 during 1976-80 in four-authored publications; and 0.0454 during 1961-65 to 0.5909 during 1976-80 in five-authored publications.

Table 3: Relative Frequency of Publications by Number of Authors Between 1931-50

Period blocks	**Relative frequency of publications by number of authors**				
	One author	**Two authors**	**Three authors**	**Four authors**	**Five authors**
1931-35	0.0246	0.0049			
1936-40	0.0221	0.0082			
1941-45	0.0111	0.0088			
1946-50	0.0210	0.0154			
1951-55	0.0440	0.0193	0.0248	0.0185	
1956-60	0.0655	0.0418	0.0406	0.0463	
1961-65	0.1070	0.0859	0.0564	0.0370	0.0454
1966-70	0.1837	0.1817	0.1580	0.1389	0.1818
1971-75	0.2228	0.3034	0.2460	0.2964	0.1818
1976-80	0.2489	0.3232	0.4582	0.4629	0.5909

Modelling the Contribution of Single-Authored Publications

Studying the percentage share of single-authored publications, it is observed that it has decreased from 94.28 per cent during 1911-1915 to 61.43 per cent during 1976-1980. The decreasing trend is observed in all period blocks, except during 1941-1945 and 1946-1950, which might be due to World War II. The decline in percentage share of single-authored publications in theoretical population genetics can be visualised in three stages: (a) Period of slow decline during1911-15 to 1931-35; (b) Period of rapid decline during 1936-40 to 1966-70; and (c) Period of comparative slower decline, than the second stage and seems to be stabilising.

Such a decline in percentage share of single-authored publications appears to have some resemblance with quasi-logistic model suggested by O'Connor (1969). In order to test whether this decline in percentage contribution of single-authored publications follows a logistic or some other kinds of growth model, the applicability of exponential, logistic, Gompertz, and power growth models was explored in the publication data. The power and logistic models were found to be the best amongst the four growth models studied, as seen from the parameter values obtained and fit statistics derived (Table 4).

Table 4: Fit Statistics Derived from the Application of Selected Growth Models to the Decline in Percentage Contribution of Single-authored Contributions

Types of growth model	**Fit statistics**	
	R^2	**F**
Exponential model	0.876	2865.56
Logistic model	0.914	2532.03
Gompertz model	0.876	1751.32
Power model	0.922	2782.81

For the application of logistic model, the following equation was used:

$$f(x) = 1/(k + ab^t),$$

where $k, a > 0$; $0 < b < 1$; $t \geq 0$; and $K = 1/k$

When this model was applied to the decline in percentage share of single-authored publications, the parameter values obtained are indicated below and fit statistics derived are shown in Table 4.

$k = 0.010 \pm 0.001$; $a = 0.007 \pm 0.0001$; and $b = 0.851 \pm 0.052$.

For the application of the power model, the following equation was used:

$$f(x) = \alpha + \beta t^{\gamma}$$

where $\alpha, \beta > 0$

When this model was applied to the decline in percentage share of single-authored publications, the parameter values obtained are indicated below and fit statistics derived are shown in Table 4.

$\alpha = 44.608 \pm 18.284$; $\beta = 16.012 \pm 15.896$; and $\gamma = 0.436 \pm 0.250$

It is concluded that on modelling the data, using selected growth models on the increasing number of publications having a decline in percentage of single-authored publications, power and logistic models have shown the positive fits.

However, these fits are not very good. Such results are partially in agreement with those suggested by 0'Connor (1969).

Indices of Collaboration

Some mathematical measures have been proposed by scholars in the past to study the extent and size of co-authorship, as reflected in publications. These measures are Degree of Collaboration (DC) first suggested by Subramanyam (1983), which takes the proportion of co-authored publications in total publications, Collaboration Index (CI), which takes the mean number of authors per publication, and Collaboration Coefficient (CC) first suggested by Ajiferuke, Burrell, and Tague (1988) that takes the proportional mean of the sum of publications and number of authors, and set the values between 0 and 1. In order to study the extent of collaboration in different period blocks in theoretical population genetics, the values of DC, CI, and CC were computed for publications and the results obtained are presented in Table 5.

As can be seen from Table 5, the mean number of authors per publication (as reflected in the value of CI) increased from 1.04 during 1916-20 to 1.56 during 1976-80. The growth in the proportion of collaborated publications and the proportional mean of the sum of the publications with each number of authors is clearly reflected in the decreasing value of DC and the increasing value of CC. The trend in the computed values of CI, DC, and CC of different period blocks is almost consistent, reflecting the growing collaboration and pointing towards increasing professionalisation in theoretical population genetics with time.

Table 5: Indices of Collaboration Obtained from Publications in Different Period Blocks during 1916-1980

Period blocks	Collaboration measures		
	CI	DC	CC
1916-20	1.04	0.9524	0.0239
1921-25	1.06	0.9365	0.0318
1926-30	1.09	0.9307	0.0380
1931-35	1.07	0.9311	0.0357
1936-40	1.12	0.8832	0.0597
1941-45	1.21	0.7821	0.1112
1946-50	1.21	0.7986	0.1019
1951-55	1.22	0.8339	0.0912
1956-60	1.22	0.7843	0.1174
1961-65	1.28	0.7591	0.1278
1966-70	1.37	0.7055	0.1592
1971-75	1.46	0.6361	0.1968
1976-80	1.56	0.6128	0.2178

Applicability of Selected Probability Distributions to the Distribution of Authorship in Publications

In this section the applicability of selected statistical probability distributions is explored for their goodness-of-fit in publication data on distribution of authorship in different period blocks in theoretical population genetics from 1941 to 1980. As a result, there were eight data sets under consideration pertaining to period blocks: 1941-45, 1946-50, 1951-55,.. ....,1976-80, as reported in Table 6. The applicability of Lotka's Law and two other statistical probability distributions, namely Geometric and Truncated Poisson, were explored in the above eight data sets.

Tables 6 and 7 report the basic statistics and parameter values obtained from the application of selected probability distributions in all the eight data files on distribution of authorship in different period blocks. Table 8 gives the Chi-square values and the degree of freedom of the fitted distributions of these eight data sets. Critical table values of the Chi-square test at 5% level are: 3.84 (df= 1), 5.99 (df= 2), 7.81 (df= 3), 9.49 (df= 4), etc.

Table 6: Basic Statistics Derived from Distribution of Authorship Data Between 1941-1980

Period blocks	**Basic statistics**	
	Mean	**Variance**
1941-45	1.2308	0.2032
1946-50	1.2083	0.1788
1951-55	1.2180	0.2881
1956-60	1.2773	0.3445
1961-65	1.2876	0.3163
1966-70	1.3759	0.4436
1971-75	1.4651	0.5126
1976-80	1.5627	0.7782

Table 7: Parameter Values Obtained from Application of Selected Probability Distributions to the Distribution of Authorship in Publications Between 1941-1980

Period blocks	**Parameter values**		
	α (Lotka)	**P (Geom.)**	**θ (Poison)**
1941-45	3.05	0.8125	0.4302
1946-50	3.14	0.8276	0.3916
1951-55	3.19	0.8210	0.4074
1956-60	2.95	0.7829	0.5098
1961-65	2.89	0.7726	0.5286
1966-70	2.68	0.7268	0.6756
1971-75	2.50	0.6825	0.8186
1976-80	2.39	0.6399	0.9704

These results indicate that Lotka's distribution is not appropriate for studying the distribution of authorship in publications. In this case, out of eight publication data sets, a positive fit was obtained only in one case. Geometric and Truncated Poisson distributions gave positive fits (indicated in bold in Table 8) in large number of the data sets (four out of eight cases through Geometric distribution and six out of eight cases through Truncated Poisson distribution). It is also observed that an adequate fits were obtained from these two distributions only at initial stages of the development of theoretical population genetics. This suggests that when the percentage of single-authored publications tends to be large, these distributions show better positive fits. These results match and are in agreement with those obtained by Rousseau (1994) in library and information science publications, and Gupta, Kumar, and Rousseau (1998) in theoretical population genetics publications.

Table 8: Chi-square Values and Degree of Freedom Obtained from Fitted Distributions to the Distribution of Authorship in Publications Between 1941-1980

	Statistical models and distributions					
Period blocks	**Lotka's inverse power law**		**Geometric distribution**		**Truncated Poisson distribution**	
	Chi-square value	Df	Chi-square value	Df	Chi-square value	df
1941-45	9.33	1	**2.20**	1	**0.177**	2
1946-50	16.96	1	4.67	1	**2.491**	2
1951-55	**4.49**	2	**3.19**	2	9.171	3
1956-60	16.98	3	**0.60**	2	**7.662**	3
1961-65	71.54	4	**5.99**	3	**0.239**	4
1966-70	173.63	5	11.79	4	**5.364**	5
1971-75	431.12	6	65.18	4	**0.130**	5
1976-80	420.52	9	26.19	5	29.28	6

CONCLUSIONS

The proportion and extent of collaborative publications have shown a systematic increase with time along with the growth of total number of publications in the discipline of theoretical population genetics. Of the total publications, 65.72% appeared as single-authored publications, 23.05% as two-authored publications, and 5.61% as three-authored publications. The single-authored publications decreased from 94.28% during 1911-15 to 61.43% during 1976-80 and multi-authored publications increased from 5.71% during 1911-15 to 26.38% during 1976-80 for two-authored publications, 1.98% during 1926-30 to 9.12% during 1976-80 for three-authored publications, and 0.06% during 1951-55 to 2.24% during 1976-80 for four authored publications, respectively. Although the decline in the proportion of single-authored publications appeared to have some resemblance with quasi-logistic model, in actual practice, Power and Logistic growth models showed positive fits in the data.

When different simple probabilistic distributions were explored for their goodness-of-fit in the publication data on the number of authors per publication, it was observed that Geometric and Truncated Poisson distributions adequately described the distribution of the number of authors per publication. These results are in agreement with those obtained by Rousseau in library and information science and Gupta, Kumar, and Rousseau in theoretical population genetics.

REFERENCES

Ajiferuke, I., Q. Burell and J. Tague. 1988. Collaboration coefficient: A single measure of the degree of collaboration in research. *Scientometrics*, Vol.14: 421-433.

Ajiferuke, Isola. 1991. A probabilistic model for the distribution of authorship. *Journal of the American Society for Information Science*, Vol.42: 279-289.

Beaver, D. De B. and R. Rosen. 1979. Studies in scientific collaboration 2. Scientific co-authorship, research productivity, and visibility in the French scientific elite. *Scientometrics*, Vol.1: 133-449.

Beaver, D. De B and R. Rosen. 1978. Studies in scientific collaboration 1. The professional origin of scientific co-authorship. *Scientometrics*, Vol.1: 65-84.

Beaver, D. De B. and R. Rosen. 1979. Studies in scientific collaboration 3. Professionalisation and the natural history of co-authorship. *Scientometrics*, Vol.1: 231-245.

Clark, B. 1964. Multiple authorship trends in scientific papers. *Science*, Vol.143: 820-824.

Felsenstein, J. *Bibliography of Theoretical Population Genetics*. Dowden, Hutchinson & Ross, 1981.

Goffman, W. and Warren, K.S. S*cientific Information Systems and the Principles of Selectivity*. New York; Praeger Publisher, 1980.

Gupta, B.M., S. Kumar, S. and R. Rousseau. 1998. Applicability of selected probability distributions to the number of authors per article in theoretical population genetics. *Scientometrics*, Vol. 42: 325-333.

Haitun, S. D. 1982.Stationary scientometric distributions. Part 2. Non-Gaussian nature of scientific activities. *Scientometrics*, Vol.4: 89-104.

Keenan, S. and P. Atherton. 1964. *Journal Literature of Physics.* New York; American Institute of Physics.

Meadows, A.J. 1974. *Communication in Science.* London : Butterworth: p.199.

O'Connor, Jean G. 1969. *DRTC Seminar,* Vol.7: 463.

Pao, M.L. 1992. Global and local collaboration. A study of scientific collaboration. *Information Processing and Management*, Vol.33: 99-109.

Patel, Narsi. 1972. Collaboration in the professional growth of American sociology. *Social Science Information*, Vol.12: 77-92.

Price, D. J. De. S and D. De B. Beaver 1966. Collaboration in an invisible college. *American Psychology*, Vol.21: 1011-1018.

Price, D.J.De S. 1965. *Little Science, Big Science.* Columbia: Columbia University Press: p.86-91.

Rousseau, R. 1992. Breakdown of the robustness property of Lotka's law: The case of adjusted counts for multi-authorship attribution. *Journal of the American Society for Information Science,* Vol.43: 645-647.

Rousseau, R. 1994. The number of authors per article in library and information science can often be described by a simple probability distribution. *Journal of Documentation*, Vol.50: 134-141.

Subramanyam, K. 1983. Bibliometric studies of research collaboration: A review. *Journal of Information Science*: 33-38.

Yitzhaki, Moshe and David Ben-Tamar. 1990. Multiple authorship in biochemistry and other fields: A case study of the *Journal of Biological Chemistry*. In. L. Egghe and R. Rousseau, eds. *Informetrics 89/90*. Amsterdam : Elsevier:373-389.

Appendix - 1

Observed and the Expected Number of Authors Per Paper Using Lotka's Law and other Statistical Distributions in Different Period Bocks

NAPP	OBNAP	Estimated Number of Authors		
		LIPL	GD	TPD
1	1363	1603.11	1423.80	1317.42
2	587	305.94	512.64	639.24
3	203	116.15	184.68	206.70
4	50	58.30	66.53	50.06
5	13	34.27	24.03	9.79
6	4	22.25	8.68	1.56
7	1	15.35	3.11	0.22
8	2	11.12	1.11	-
9	1	8.46	0.45	-
10	1	6.45	-	-

Observed and the Expected Number of Authors Per Paper Using Lotka's Law and other Statistical Distributions 1971-1975

NAPP	OBNAP	Estimated Number of Authors Per Paper		
		LIPL	GD	TPD
1	1220	1429.68	1309.20	1238.84
2	551	252.79	415.63	507.12
3	109	91.68	131.81	138.29
4	32	44.69	41.81	28.39
5	4	25.51	13.23	4.60
6	1	16.30	4.22	0.58
7	1	11.12	1.34	0.00

Observed and the Estimated Number of Authors Per Paper Using Lotka's Law and Other Statistical Distributions 1966-1970

NAPP	OBNAP	Estimated Number of Authors Per Paper		
		LIPL	GD	TPD
1	1006	1113.85	1036.40	998.20
2	330	173.83	283.20	337.11
3	70	58.61	77.29	75.86
4	15	27.09	21.10	12.83
5	4	14.97	5.70	1.71
6	1	9.13	1.57	0.14

Observed and the Estimated Number of Authors Per Paper Using Lotka's Law and Other Statistical Distributions 1961-1965

NAPP	OBNAP	Estimated Number of Authors Per Paper		
		LIPL	GD	TPD
1	586	630.03	599.61	585.87
2	156	85.00	133.94	154.86
3	25	26.33	29.88	27.25
4	4	11.50	6.72	3.63
5	1	6.02	1.47	0.39

Observed and the Estimated Number of Authors Per Paper Using Lotka's Law and Other Statistical Distributions 1956-1960

NAPP	OBNAP	Estimated Number of Authors Per Paper		
		LIPL	GD	TPD
1	359	377.80	358.57	351.15
2	76	48.87	77.86	89.49
3	18	14.79	16.90	15.21
4	5	6.32	3.66	1.92

Observed and the Estimated Number of Authors Per Paper Using Lotka's Law and Other Statistical Distributions 1951-1955

NAPP	OBNAP	Estimated Number of Authors Per Paper		
		LIPL	GD	TPD
1	241	247.36	237.27	234.12
2	35	27.11	42.45	47.69
3	11	7.43	7.60	6.47
4	2	2.98	1.36	0.66

Observed and the Estimated Number of Authors Per Paper Using Lotka's Law and Other Statistical Distributions 1951-1955

NAPP	OBNAP	Estimated Number of Authors Per Paper		
		LIPL	GD	TPD
1	115	122.40	119.17	117.75
2	28	13.88	20.55	22.95
3	1	3.89	3.54	2.98

Observed and the Estimated Number of Authors Per Paper Using Lotka's Law and Other Statistical Distributions 1945-1950

NAPP	OBNAP	Estimated Number of Authors Per Paper		
		LIPL	GD	TPD
1	61	65.41	63.38	62.42
2	16	7.90	11.88	13.42
3	1	2.29	2.23	1.93