AN ANALYSIS OF BRADFORD MULTIPLIERS AND A MODEL TO EXPLAIN LAW OF SCATTERING

I. K. RAVICHANDRA RAO

Documentation Research and Training Centre Indian Statistical Institute
8th Mile, Mysore Road, Bangalore 560 059 (India)

(Received August 30, 1997)

In his book on "Documentation", Bradford derived the law of scattering, based on algebric explanation with the supposition that $n_1 = n_2 = n$. n_1 and n_2 are computed based on average no. of articles per journals in the first three zones. An analysis of a small sample of 12 data sets, using t-test suggests that it is unlikely that $n_1 = n_2$. Further an attempt has been made to identify a suitable model to explain the law of scattering; among the various models tried, log-normal fits much better than many models including the log-linear model.

Introduction

The topic in bibliometrics that has received a great deal of attention is the problem related to the scattering of articles. The tabulation of the distribution of the number of references on a specific subject area among journals is the traditional way of summarizing scattering of articles. In most of the bibliographies, covering a short period on a particular scientific subject, it may be observed that on a given subject:

i) Most of the journals contribute only one article each; the other articles in the journals obviously are not relevant to the said subject;
ii) a few journals contribute on an average 5 to 10 articles each;
iii) very few journals, comparing to the first two groups, contribute a large number of articles.

This was first observed by *Bradford* (1934). Bradford in his analysis on the data obtained from 169 journals in applied geophysics (1929–31) and from 102 journals in lubrication (1931–37), he noticed that when partial sum of number of references are plotted against the natural logarithm of the partial sum of the number of journals or periodicals, an almost a straight line graph results. i.e.,

$$F(x) = a + b \cdot \ln x$$

$F(x)$ is the references contained in the first x most productive journals and a & b are parameters of the subject; the curve is often known as the Bradford curve. Based on the graph, he further stated the following statement which is now called as "Bradford's Law of Scattering":

"If scientific journals are arranged in order of decreasing productivity of articles on a given subject, these may be divided into a nucleus of periodicals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus, when the number of periodicals in the nucleus and succeeding zones will be as $1 : n : n^2 : \ldots$" (*Bradford*, 1934, 1938).

Vickery (1948) observed the differences between graphical and verbal interpretation of Bradford's Law of Scattering and argued that "if the collection of journals are divided into arbitrary number of groups each containing r references and if S_{kr} is the cumulated number of journals in the most productive k groups ($k = 1, 2, \ldots$) then

$$S_{kr} = s\,(n^k - 1) \text{ where } s = S_r/(n-1)$$

He also further pointed out that the theoretical graph was a smooth curve and that Bradford's interpretation only predict the upper straight line portion of the curve.

Bradford multipliers

Later in 1948, by defining $n_i = r_i/r_{i+1}$, where r_i is the average number of articles per journal in the ith group/zone, when journals are arranged in decreasing productivity and with an assumption that $n_1 = n_2$, *Bradford* (1948) argued that the ratio of zone size will be as $1 : n : n^2 : \ldots n$ is known as Bradford multiplier.

How far this assumption is correct? An attempt has been made in this paper to test the hypothesis that $n_1 = n_2$; by defining $n_{1i} - n_{2i} = d_i$, for $i = 1, 2, \ldots$ a null hypothesis has been formulated as H_0: $\bar{D} = 0$. For twelve sets of data, collected by different authors, semi-log curves were drawn as explained by Bradford. Until a point from where a straight line begins is considered as nucleus zone for each of the data sets and as explained by Bradford, the next two zones are identified. Thus m_1, m_2 and m_3 (the number of journals in the nucleus and in the next two succeeding zones) and the corresponding values of r_1, r_2 and r_3 (average numbers of articles per journal in three zones). As explained by Bradford, $m_1 r_1 = m_2 r_2 = m_3 r_3$. The values of m_1, m_2, m_3, r_1, r_2, r_3, n_1, n_2 and d (the difference between n_1 and n_2 for each of the twelve data sets are given in Table 1. t-test is carried out to test H_0: $\bar{D} = 0$ vs H_1: $\bar{D} < 0$. t-statistic is computed, under the assumption that H_0 is true. It is given by:

$$t = \frac{\bar{d} - \bar{D}}{\sigma_d / \sqrt{n}} = \frac{\sqrt{n}\,\bar{d}}{\sigma_d}$$

Where $\bar{d}$ is the mean of the 12 values of the differences between n_1 and n_2. For the sample of 12 data sets, $\bar{d}$ and σ_d are −1.4392 and 3.57382 respectively; the computed value of t is −1.3983. For $\alpha = 0.10$ and for 11 degrees of freedom $t_\alpha = 1.363$. Since $|t| > t_\alpha$, H_0 may be rejected; it indicates that H_1 may be accepted; i.e., $n_1 < n_2$ and thus indicating that they are unlikely to be equal.

If n is sufficiently large one can perhaps reject H_0: $\bar{D} = 0$ even at $\alpha = 0.05$ level.

This study thus suggests that Bradford's assumption that $n_1 = n_2$ is unlikely to be correct and log-linear model unlikely to explain the law of scattering.

Table 1
Values of Bradford multiples (n_1 and n_2)

No.	Source of data	m_1	m_2	m_3	r_1	r_2	r_3	n_1	n_2	d_i
1.	Bradford (1934) (Geography)	6	29	87	60.67	32.59	4.18	4.8189	3.0120	1.8069
2.	Bradford (1934) (Lubrication)	6	23	75	18.33	4.74	1.47	3.8671	3.2245	0.6426
3.	Kendall (ORSA) (1960)	7	85	604	99.14	8.17	1.15	12.1346	7.1043	5.0303
4.	Geffman & Warren (Most cell) (1969)	16	47	188	42.375	14.43	3.61	2.9366	3.9972	1.0606
5.	Saeb (1986) (Stat Methods)	8	32	309	42.75	10.72	1.1	3.9879	9.7455	−5.7576
6.	Depew (1986) (Lib. Sc.)	5	21	86	39.20	9.19	2.28	4.2655	4.3007	0.2546
7.	Cook (1989) (Popular Music)	36	73	119	33.17	16.37	10.08	2.0263	1.6240	0.4023
8.	Lawani (1972) (Tropi Agri.)	13	25	46	139.62	72.8	39.78	1.9179	1.6301	0.0878
9.	Lipatov (1986) (Polymer)	9	42	527	99.11	21.19	1.7	4.6772	12.4647	−7.7875
10.	Warren & Nevin (Schisto somicia) (1967)	25	165	1759	135.0	20.47	1.92	6.5950	10.6614	−4.0664
11.	Pope (1975) (Inf. Sc.)	11	116	1626	245.0	23.75	1.65	9.5220	15.5939	−6.0719
12.	IKR Rao (economics) (1990)	33	78	240	35.82	15.12	4.94	2.3690	3.0607	−0.6917
	$\bar{d}$ = 1.43592			σ_d = 3.573828			Computed value of t = −1.3983			

Lognormal model

Since 1948, many have worked in this area and suggested different models to explain law of scattering. *Simon* (1955) proposed a Beta model under the following two assumptions:

i) There is a constant probability α that the k^{th} paper be published in a new journal that has not published in the first $(k-1)$ papers.
ii) The probability the k^{th} paper is published in a journal that has published i papers is propositional to $i*f(i, k-1)$; i.e. to the total number of papers of all journals that have published exactly i papers. The β model is as follows:

$$j(r) = \frac{N}{r^{(1+\rho)}}$$

N is the total number of periodicals containing at least one paper on the subject, ρ is the distance between origin and the point at which straight line meets at x-axis, and $j(r)$ is the distribution of the number of journals j with exactly r papers.

Kendall (1960) in his analysis of the bibliography on operations research arranged that the scattering of articles in journal is similar to that of income distribution. *Cole* (1958) suggested a semi-log model, as mentioned by Bradford to explain law of scattering.

Goffman and *Warren* (1969) in their study introduced a notion of sub-dividing the literature into a maximal number of years instead of any number of years. They observed adherence to Bradford's law. *Groos* (1967) observed a S-shape curve (with a droop, at the end of the curve) to explain law of scattering. *Leimkuhler* (1967) suggested a distribution function:

$$F(y) = \frac{\log(\beta y + 1)}{\log(1+\beta)}, \; 0 \le y \le 1, \; 0 \le F(y) \le 1, \; \beta > 0$$

to explain law of scattering. $F(y)$ is the relative total number of reference contained in the topmost y proportion of journals. *Brookes* (1968) also suggested two different models (log-linear model and non-linear model) to explain lower and upper parts of the curve.

Fairthorne (1969) and also *Asai* (1981) suggested a log model and *Naranan* (1971) suggested a power model. *Karmeshu* and others (1982) presented two models to explain the mechanism that could produce Bradford distributions. These models are called as subdivisions model and multiple factor model.

Burrell (1988) suggested Waring process to explain general features of Bradford's law. *Basu* (1992, 1995) suggested a model to explain distribution of articles in journals based on probabilistic considerations.

Some of these models are based on size-frequency approach and most of them are based on rank-frequency approach. Also some of them are only of theoretical models and not tested with real-life data.

To identify a suitable model to explain the law of scattering about 24 different models were fitted to the 12 different sets of data. The list of models are given in Table 2. A software package called MACE was used to fit the models, based on regression analysis.

The values of the parameters – best fitting model, the log-normal model and the log-linear model are given in Table 3.

Table 4 gives the values of R^2 for best fitting model, log-normal model and log-linear model.

For the data collected by Bradford (in Geophysics) and Kendall (Operations Research), the best fitting model is modified Hoerl function. Hoerl function fits fairly well to Cook's data. Reciprocal hyperbola function fits fairly well to Sach's and Pope's data. For rest of the seven data sets lognormal model fits very well. y gives the number of articles (cumulative) contained in the x most periodicals. However, even for the other five data sets (Bradford, Geophysics); Kendall, Cook, Sach & Pope), the lognormal model is the second best; For all the 12 data sets, the value of R^2 for log-normal is at least 0.995. This study thus indicates that the lognormal model is perhaps explains well the law of scattering.

Table 2
Equations fitted using this program

1.	Y=A+B*X	STR.LINE	2.	Y=B*X	LINE THRU ORG.
3.	Y=1/(A+B*X)	REC.STR LINE	4.	Y=A+B*X+C/X	LINE AND RECIP.
5.	Y=A+B/X	HYPERBOLA	6.	Y=X/(A*X+B)	RECIP HYPERBOLA
7.	Y=A+B/X+C/X*X	2ND ORD HYP	8.	Y=A+B*X+C*X*X	PARABOLA
9.	Y=A*X+B*X*X	PAR AT ORIGIN	10.	Y=A*X^B	POWER
11.	Y=A*B^X	MOD.POWER	12.	Y=B^(1/X)	ROOT
13.	Y=A*X^(B*X)	SUPER GEOMET	14.	Y=A*X^(B/X)	MOD GEOMETRIC
15.	Y=A*e^(B*X)	EXPONENTIAL	16.	Y=A*e^(B/X)	MOD EXPONENTIAL
17.	Y=A+B*lnX	LOGARITHMIC	18.	Y=1/(A+B*lnX)	RECIP LOG
19.	Y=A*B^X*X^C	HOERL FUNCTION	20.	Y=A*B^(1/X)*X^C	MOD HOERL
21	Y=A*e^(((X–B)^2)/C)	NORMAL	22.	Y=A*e^((lnX–B)^2/C)	LOG NORMAL
23.	Y=A*X^B*(1–X)^C	BETA	24.	Y=A*(X/B)^C*e^(X/B)	GAMA
25.	Y=1/(A*(X+B)^2+C)	CAUCHY			

Table 3
Values of parameters (a, b and c)

Sl. No.	Sources of data ( ) refer to references	Values of parameters for the best fitting model			Values of parameters for log-normal			Values of parameters for log-linear model	
		a	b	c	a	b	c	a	b
1.	Bradford (Geophysics) (1934)	242.9	0.5706	0.3125	1558.0	6.5737	−18.01	−99.63	222.9
2.	Kendall (ORSA) (1960)	479.6	0.4785	0.8335	1754.0	69.21	−25.13	162.5	264.9
3.	Cook (Music) (1939)	160.5	0.9495	0.6915	[illegible]	12.85	20.59	1143.0	905.2
4.	Lipatov (Polinsy) (1986)	-		-	2593.0	5.793	12.61	53.27	654.20
5.	Bradford (Lubrication) (1934)				4557	6.593	14.24	25.71	71.39
6.	Lawani (Agriculture) (1972)		–		16340.0	7.647	15.00	1670.0	2630.0
7.	Goffman & Warren (Mast Cell) (1969)		–	–	2349.0	6.766	−12.70	−375.1	417.7
8.	IKR Rao (Economics) (1990)	-		-	5671.0	8.662	17.19	911.4	689.8
9.	Depew (Lib. Sc.) (1986)		–		851.20	7.867	−26.93	28.25	112.40
10.	Warren & Newill (1967)		–		8795.0	7.185	−16.00	−1184.0	3475.0
11.	Sachs (Stat Methods) (1986)	0.001081	0.01503	–	813.5	4.843	−8.864	12.39	179.7
12.	Pope (Inf. Sc.) (1975)	0.001547	0.009658	–	6624.0	61.88	−12.06	−671.0	246.0

Table 4
Values of R^2

S no.	Sources of data (( ) refers to references)	Best Fitting Model and R^2 value	R^2 for log-normal Model $y = a.e^{\frac{(\log x - b)^2}{c}}$	R^2 for log-linear Model $y = a + b \log_e x$
1.	Bradford (Geophysics) (5)	$y = ab^{1/x}x^c$ (0.9976)	0.9947	0.9752
2.	Kendall (ORSA) (17)	-do- (0.9953)	0.9935	0.7213
3.	Cook (Music) (10)	$y = ab^{x}x^c$ (0.9970)	0.9953	0.8623
4.	Lipatov (Polinsy) (21)	$y = a.e^{\frac{(\log x - b)^2}{c}}$	0.9984	0.6149
5.	Bradford (Lubrication) (5)	-do-	0.9999	0.9525
6.	Lawani (Agriculture) (18)	-do-	0.9995	0.8047
7.	Goffman & Warren (Mast Cell) (14)	-do-	0.9992	0.5609
8.	IKR Rao (Economic) (24)	-do-	0.9992	0.5846
9.	Depew (Lib. Sc.) (12)	-do-	0.9987	0.5249
10.	Warren & Newill (28)	-do-	0.9979	0.5179
11.	Sachs (Stat Methods) (25)	$y = \frac{x}{a^{x} \cdot b}$ (0.995)	0.9942	0.6735
12.	Pope (Inf. Sc.) (23)	-do- (0.9996)	0.9972	0.4882

Conclusion

Bradford in 1948 in his book on *Documentation* argued that "we have no reason why n_1 and n_2 should differ and the simple supposition we can make is that they are equal". He thus assumed $n_1 = n_2 = n$. Based on a small sample of twelve data sets, it has been shown in this study that n_1 and n_2 are not likely equal. Thus it may be believed that Bradford multipliers vary from zone to zone.

Further Bradford suggested a log-linear model to explain the law of scattering. Again based on a small sample of 12 data sets, it has been argued that log-normal model fits much better than the log-linear model. The log-normal model will be much more accurate in predicting the total number of articles covered in a given number of core journals.

Law of scattering is an area where much work has been done. However, till now, no one has come out with a single model which fits fairly well to the most of data sets. This study suggests that log-normal model fits fairly well to the most of the available data. Also, this study indicates that Bradford multipliers vary from zone to zone.

References

1. ASAI, I. A. (1981), General formulation of Bradford's distribution: The graph-orient approach, *Journal of American Society for Information Science*, 32(2), 113–119.
2. BASU, ARPANA (1992), Hierarchical distributions and Bradford's law, *Journal of American Society for Information Science*, 43(7), 494–500.
3. BASU, ARPANA (1995), Models of Bradford's law as solutions of differential equations, In: M. E. D. KOENING (Ed.), *Biannual Proc. of the 5th Conf. of the ISSI*.
4. BRADFORD, S. C. (1948), *Documentation*, London, Crosby Lockwood.
5. BRADFORD, S. C. (1934), Sources of information on specific subject, *Engineering*, Jan 26, 85–86.
6. BROOKES, B. C. (1968), The derivation and application of the Bradford-zipf distribution, *Journal of Documentation*, 24(4), 247–265.
7. BURRELL, Q. L. (1988), Modeling the Bradford phenomenon, *Journal of Documentation*, 44(1), 1–18.
8. COOK, KELVIN L. (1989), Law of scattering applied to popular music, *Journal of American Society for Information Science*, 40(4), 277–283.
9. COLE, P. F. (1958), Analysis of reference question records guide to information requirements of scientists, *Journal of Documentation*, 14, 197–207.
10. DEPEW, J. N., BASU, SANTI, (1986), A critical study of American literature: Implications for Collection Management, *Collection Management*, 8(2), 49–52.
11. FAIRTHORNE, R. A. (1969), Empirical hyperbolic distribution (Bradford–Zipf–Mandelbrot) for bibliometric description and prediction, *Journal of Documentation*, 25(3), 19–43.
12. GOFFMAN, W., WARREN, K. S. (1969), Dispersion of papers among journal based on a mathematical analysis of two diverse medical literature, *Nature*, 221 (March 29), 1205–1207.
13. GROSS, O. V. (1967), Bradford's law and the Keenan-Atherton data, *American Documentation*, 18(1), 46.

14. KARMESHU, N. C., LIND and CANO, V. (1984), Rationales for Bradford's law, *Scientometrics*, 6(4), 233–241.
15. KENDALL, M. G. (1960), The bibliography of operations research, *Operations Research Quarterly*, 11, 31–36.
16. LAWANI, S. M. (1972), Periodical literature of tropical and subtropical agriculture, *UNESCO Bulletin for Libraries*, 26(2), 88–93.
17. LEIMKUHLER, F. F. (1967), The Bradford distribution, *Journal of Documentation*, 23(3), 197–207.
18. LIPATOV, YU. S., DENISENKO, L. V. (1986), On the behavior of information laws in multi component polymer systems research, *Scientometrics*, 9(5–6), 197–208.
19. POPE, A. (1975), Bradford's law and periodical literature of information science, *Journal of American Society for Information Science*, 26(4), 207–213.
20. RAVICHANDRA RAO, I. K. (1990), Journal productivity in economics. In: L. EGGHE, R. ROUSSEAU (Eds), *Proceedings of Informetrics 89/90*, 251–280.
21. SACHS, S. L. (1986), *A Guide to Statistical Methods and to the Pertinent Literature, Literatur zur angewandten statistik*, Springer-Verlag.
22. SIMON, H. A. (1955), On a class of skew distribution function, *Biometrika*, 42, 425–440.
23. VICKERY, B. C. (1948), Bradford's law of scattering, *Journal of Documentation*, 28, 2, 127–130.
24. WARREN, K. S., NEWIN, V. A. (1967), *Schistosomiasis, a Bibliography of the World's Literature from 1852–1962*, Western Reserve univ. Press, Cleveland.