STEPHEN BULICK

Book Use as a Bradford-Zipf Phenomenon

The Bradford distribution is introduced and explained; its ambiguity, arising from two different conceptions, and relationship to the Zipf distribution are discussed. The general situation that might be expected to yield such a distribution is described and shown to fit the interaction between book users and books available for use in a library. Circulation data from a major research library are then shown to follow the verbal rather than the graphical formulation of Bradford's law. Some possible implications are discussed.

IN HIS FAMOUS WORK *Documentation,* S. C. Bradford expressed an idea that has come to be called "Bradford's Law of Scattering" or simply "Bradford's Law." Bradford discovered a mathematical expression for the distribution of articles on a given subject in the scientific literature:[1]

If scientific journals are arranged in order of decreasing productivity of articles on a given subject, they 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 numbers of periodicals in the nucleus and succeeding zones will be as $1:n:n^2$

In this statement "n" is a constant particular to the subject involved.

He also developed a graphic formulation of the law.[2] Unfortunately, as Vickery and later Wilkinson noticed, the two formulations are different.[3,4] Because some researchers have used one formulation and some have used the other, confusion has

developed. Whichever formulation is chosen, the idea to be preserved is that, given some coherent population of items, a small number will be very productive (in terms of selection, citation, or whatever), a greater number will be less productive, still a greater number will be even less productive, and so on until the population of items has been exhausted.

There are, thus, "zones" of fairly constant total productivity but containing greater and greater numbers of individual items in each zone, i.e., the average productivity per item decreases in each succeeding zone. Furthermore, there is a direct relation between the cumulative productivity and the cumulative total of items. Though Bradford apparently felt that the zones were based on a theoretically derived nucleus, Vickery shows that Bradford's law holds for any number of zones of equal productivity.[5]

RELATION OF BRADFORD AND ZIPF LAWS

While originally proposed to describe the distribution of journal articles in the scientific literature and frequently used by librarians as an aid in preparing bibliographies and in maintaining journal coverage of various subject areas, Bradford's law seems applicable to other situations as well. Brookes points out that the Bradford distribution

closely resembles a Zipf distribution with certain restrictions.[6]

A strictly linear Zipf distribution arises from a situation in which items are chosen from a restricted population of possible items. Gradually some items emerge, on a "success breeds success" basis, as most popular and continue to be chosen at a greater rate. The total population of "used" items increases as well, since some items continue to be chosen for the first time. With no restrictions as to the availability of the popular items, the distribution of "uses" vs. "items used" becomes linear on a semilog scale. With such restrictions the graph curves at the bottom but rises to linearity. The Bradford distribution can thus be superimposed on the Zipf distribution and, in fact, becomes a special case of the Zipf distribution.

Leimkuhler has also noticed the close relationship between the Bradford and Zipf distributions.[7] Since the "success breeds success" mechanism and the restriction on most popular items are common to many situations, one would expect the Bradford-Zipf distribution to apply in other areas. Brookes describes the general situation in which one might expect to encounter such a distribution:

The Bradford-Zipf distribution can be expected to arise when selection is made of items, characterized by some common element, which are all equally open to selection for an equal period and subject to the "success-breeds-success" mechanism, but when the selection of a most popular group is also, but to a weaker extent, subject to restriction. It is thus a general law of concentration over an unrestricted range of items on which is superimposed a weaker law of dispersion over a restricted range of the most frequently selected items.[8]

Our notion is that this description fits the pattern of book use by library patrons and that Bradford's law should, therefore, describe that situation as well.

Two Formulations of Bradford-Zipf

Wilkinson performs a valuable service in working out both forms of Bradford's law and then applying each to the same set of data. Her conclusion is that the graphical formulation yields better results.[9] We shall borrow her expressions to show that the verbal formulation better describes library circulation data.

The underlying assumption of both forms of Bradford is the same: If items are ranked by productivity from high to low, there is a relationship between cumulative (and therefore relative) productivity and the logarithm of cumulative items and thus the rank. The verbal formulation, used by Vickery[10] and Leimkuhler,[11] is expressed as:

$R(n) = j \log (n/t + 1)$

where: $R(n)$ = cumulative productivity of the first n items

and j and t are constants.

The graphical formulation, used by Brookes,[12] is expressed as:

$R(n) = k \log n/s$

where: $R(n)$ = cumulative productivity of first n items

and k and s are constants.

Furthermore, $j \neq k$ and $t \neq s$.

Application to Library Circulation Data

We now develop each of these expressions for library circulation data. Our items are books and monographs; our productivity is measured in "uses" or "transactions" of those items. All logarithms are to base e. Table 1 shows virtually complete external circulation data for calendar year 1972 from the Hillman Library at the University of Pittsburgh, a major research library and Association of Research Libraries member. (It should be noted that the circulation data exclude nonpatron transactions for binding, storage, mending, etc.) Items have been ranked according to frequency of use; columns are as follows:

A represents the number of times circulated.
B represents the number of individual items having the frequency in column A, e.g., seven items circulated fifteen or more times.
C cumulates column B, e.g., forty-six items circulated thirteen times or more.
D is column A times column B, i.e., the number of transactions (or circulations) accounted for by the items in column B, except for row 1, of course.
E cumulates column D, e.g., items that circulated 13 times or more accounted for 629 total transactions.

To ease the computational burden (and drop out a constant) we recast each of the

two expressions to get $F(x)$, the proportion of total productivity contained in the most productive fraction, x, of items.
Verbal:

$F(n/N) = R(n)/R(N)$, where N = total items count
$F(n/N) = j \log (n/t + 1) / j \log (N/t + 1)$
    $x = n/N$
$F(x) = \log (xN/t + 1) / \log (N/t + 1)$
    Let $V = N/t$, then
$F(x) = \log (Vx + 1) / \log (V + 1)$

Graphical:

$F(n/N) = (k \log n/s) / (k \log N/s)$
    $x = n/N$
    $F(x) = (\log xN/s) / \log N/s$
Let $G = N/s$, then
    $F(x) = \log Gx / \log G$

Constants for distributions of this type (gamma related) present a problem, since there is no strictly scientific way of determining what they should be. Both Wilkinson and Leimkuhler offer explanations.[13,14] Wilkinson's method is used here although Leimkuhler's is similar. To derive the constants t and s, using data in table 1:

Let p items contribute cumulative S transactions
    q items contribute cumulative $2S$ transactions
From table 1:
    $p = 25,525$ $S = 105,359$
    $q = 107,152$ $2S = 210,717$
$t = p^2/ (q - 2p)$
$t = 651,525,625 / (107,152 - 51,550)$
    $= 11,613$

$V = N/t$
    $= 9.227$
$s = p^2/ q$
    $= 651,525,625 / 107,152$
    $= 6,080$
$G = N/s$
    $= 17.624$

To obtain the actual productivity, in transactions, for any most productive fraction x of items, it is necessary to multiply 107,152, the total items population count, by the fraction, then look at table 1, interpolating as necessary, for the actual number of transactions produced by xN items, e.g.:

Let $x = 0.10$, then $xN = 10,715$.

From column C of table 1 we find that 10,715 lies between 7,294 and 13,603.

7,294 items produce cumulative 44,522 transactions.

We must account for the next 3,421 items (10,715—7,294), which will be drawn from the group used 4 times each, that is, the next 3,421 items yield $4 \times 3,421 = 13,684$ transactions.

$44,592 + 13,684 = 58,276$ transactions produced by 10,715 items.

Interpolation is necessary because each rank contains many ties.

Applying each version of Bradford to the data in table 1 and interpolating as necessary to get the actual transaction counts yields the results shown in table 2.

TABLE 1

CIRCULATION DATA, 1972, HILLMAN LIBRARY, UNIVERSITY OF PITTSBURGH

A No. of times circulated	B No. of items circulated	C Cumulative no. of items circulated	D No. of transac- tions $(A \times B)$	E Cumulative no. of transactions
15+	7	7	110	110
14	12	19	168	278
13	27	46	351	629
12	37	83	444	1,073
11	62	145	682	1,755
10	120	265	1,200	2,955
9	281	546	2,529	5,484
8	503	1,049	4,024	9,508
7	965	2,014	6,755	16,263
6	1,859	3,873	11,154	27,417
5	3,421	7,294	17,105	44,522
4	6,309	13,603	25,236	69,758
3	11,756	25,359	35,268	105,026
2	23,898	49,257	47,796	152,822
1	57,895	107,152	57,895	210,717

TABLE 2
VALUES PREDICTED BY BRADFORD-ZIPF COMPARED WITH ACTUAL VALUES

Most productive fraction (x)	Items in that fraction (xN)	Predicted transactions verbal	Predicted transactions graphical	Actual transactions	Percent of error* verbal error	Percent of error* graphical error
0.01	1,072	7,999	. . . . . .†	9,669	17.27	. . . . .
0.02	2,143	15,349	. . . . . .	17,037	9.91	. . . . .
0.03	3,215	22,147	. . . . . .	23,469	5.63	. . . . .
0.04	4,286	28,471	. . . . . .	29,482	3.43	. . . . .
0.05	5,358	34,382	. . . . . .	34,842	1.32	. . . . .
0.06	6,429	39,931	4,101	40,197	0.66	89.80
0.07	7,501	45,160	15,422	45,350	0.42	65.99
0.08	8,572	50,103	25,229	49,634	0.94	49.17
0.09	9,644	54,791	33,880	53,922	1.61	37.17
0.10	10,715	59,253	41,617	58,276	1.68	28.59
0.20	21,430	94,773	92,526	93,239	1.65	0.76
0.30	32,146	120,228	122,301	118,600	1.37	3.12
0.40	42,861	140,080	143,428	140,030	0.04	2.43
0.50	53,576	156,354	159,816	157,141	0.50	1.70
0.60	64,291	170,147	173,205	167,855	1.37	3.19
0.70	75,006	182,115	184,527	178,571	1.98	3.34
0.80	85,722	192,685	194,334	189,287	1.80	2.67
0.90	96,437	202,150	202,984	200,002	1.07	1.49

Predicted values for verbal and graphical may contain 00.1% error due to rounding of logs. Values for xN and $F(x)$ are rounded to nearest whole item or transaction.

verbal: $F(x)=\ln(Vx+1)/\ln(V+1)$
$V=N/t = 9.227$
graphical: $F(x)=\ln Gx/\ln G$
$G=N/s = 17.624$

*% error = | actual - predicted | /actual
† when $xN < s$, $\ln Gx < 0$

CONCLUSIONS

As the data in table 2 show, the verbal formulation consistantly produces a lower error percentage than the graphical and is, in fact, a good overall predictor. Book use, then, is a Bradford-Zipf phenomenon.

Figure 1 shows a graph of cumulative transactions against cumulative items on a semi-log scale for the real data shown in table 1. This is a standard Bradford-Zipf graph of the type used by Brookes and Wilkinson. It clearly shows the curvilinearity described by Brookes. If we accept Brookes' explanation (though not his formulation) of the Bradford-Zipf phenomenon, then the curving part of the line represents demand interference, i.e., items represented on this part of the line would have been used even more if they had somehow been more available. The ideal would be a straight line

meaning that every item had lived up to its use potential.

This offers support for the strategy of shortening loan periods for frequently circulated items. Other potential uses for a function that describes the distribution of book use over a collection might include core collection determination or the derivation of a marginal utility function describing the effect on total use of adding to a collection.

REFERENCES

1. S. C. Bradford, *Documentation* (London: Crosby Lockwood & Son, 1953), p.154.
2. Ibid., p.153.
3. B. C. Vickery, "Bradford's Law of Scattering," *Journal of Documentation* 4:198–203 (1948).
4. E. A. Wilkinson, "The Ambiguity of Bradford's Law," *Journal of Documentation* 28:122–30 (1972).
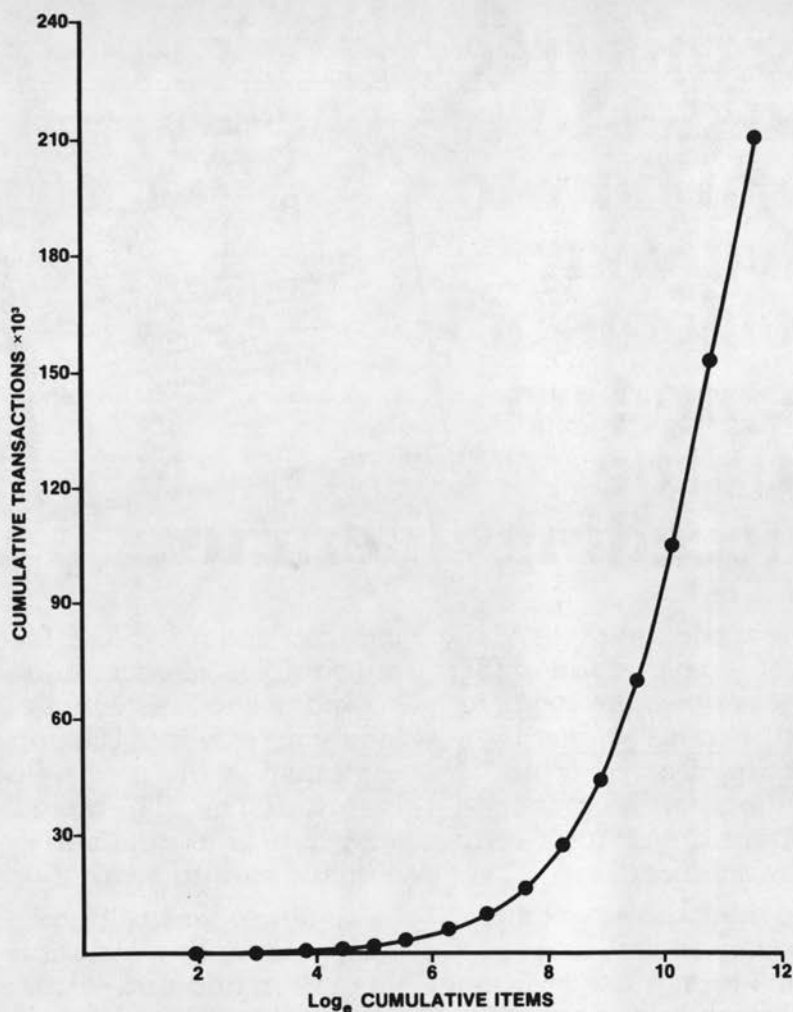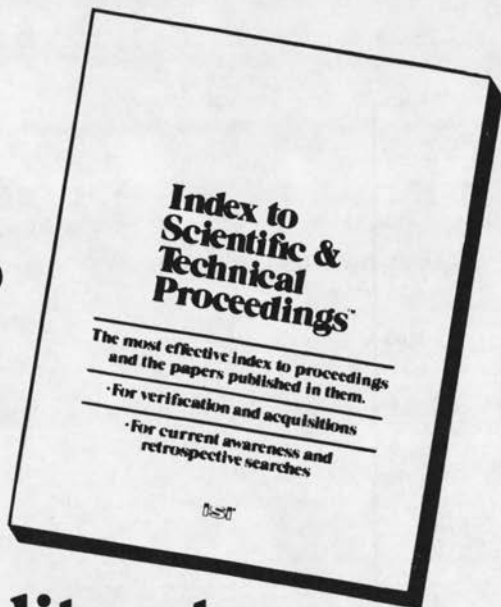
Fig. 1
Bradford-Zipf Graph of Data in Columns C and E
of Table 1

5. Vickery, "Bradford's Law of Scattering," p.202.

6. B. C. Brookes, "Bradford's Law and the Bibliography of Science," *Nature* 224:953–56 (1969).

7. F. F. Leimkuhler, "The Bradford Distribution," *Journal of Documentation* 23:197–207 (1967).

8. Brookes, "Bradford's Law and the Bibliography of Science," p.954.

9. Wilkinson, "The Ambiguity of Bradford's Law," p.124.

10. Vickery, "Bradford's Law of Scattering."

11. Leimkuhler, "The Bradford Distribution." Wilkinson shows that Leimkuhler uses the verbal formulation.

12. Brookes, "Bradford's Law and the Bibliography of Science." B. C. Brookes, "Numerical Methods of Bibliographic Analysis," *Library Trends* 22:18–43 (1973).

13. Wilkinson, "The Ambiguity of Bradford's Law."

14. Leimkuhler, "The Bradford Distribution."