OPERATIONS RESEARCH CENTER

working paper

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

EXACT SOLUTION FOR THE BRADFORD DISTRIBUTION AND ITS USE IN MODELLING INFORMATIONAL DATA

by

Philip M. Morse
and
Ferdinand F. Leimkuhler

OR 068-77 November 1977

Exact Solution for the Bradford Distribution and its Use in Modelling Informational Data.

by Philip M. Morse
Operations Research Center
Mass. Inst. of Technology

and Ferdinand F. Leimkuhler
School of Industrial Eng.
Purdue University

ABSTRACT.

The Bradford distribution requires a logarithmic relationship between the cumulative distribution F_n, the fraction of items (periodicals, books, information banks, et al.) having "productivity" (number of articles on or references to a given subject, rate of use of the item, et al.) n or more, and the total cumulative "production" G_n of these items. An exact solution is obtained for this distribution for all integer values of $n \geqslant 1$. The result is compared with the Zipf and the modified geometric distributions, also of use in describing library and other informational data. Examples are given of the operational value of representing informational data by one of these distributions. Simple tests are developed to see how well a given sample of data conforms to one or the other distribution, and are applied to two samples; first the number of O/R articles appearing in various journals in a given time interval and, second the rate of use of various physics periodicals in a science library. It is shown that the first fits the Bradford distribution quite well, whereas the second fits the modified geometrical distribution. There is a discussion of some of the implications of this difference.

Definitions.

Librarians and other documentation analysts have utilized a number of probability distributions to describe their data; among them being the Bradford[1], the Zipf[2] and the modified geometric distributions[3]. Since some of these distributions are not familiar to many operations researchers and since, in at least one case, the mathematical properties of the distribution ha not been completely worked out, this note will define and compare the three, complete the mathematical analysis and indicate how they are used in comparing, predicting and acting on the collections measured by the distributions.

We start with a collection of A items; books in a library, technical journals, articles in a journal, citations of an article in later issues of journals, or cities in a given country, to go farther afield. Each of these items has a given <u>productivity</u> n; yearly circulation or other use of a book or journal, number of articles on a given field in the given journal, number of references to the given article in subsequent journals, or the population of a given city, for example. Each of the distributions is defined by segregating the items according to their productivity, the number of items with productivity n being Af_n and the cumulative number of items with productivity n or greater being

$$AF_n \equiv A \sum_{m \geq n} f_m \; ; \quad F_1 = 1 \qquad (1)$$

A = Total number of productive items

so that f_n is the probability that an item has productivity n and F_n is the cumulative distribution function.

Connected with this is the cumulative production, the total production of all items with productivity n or greater,

$$AG_n \equiv A \sum_{m \geq n} m f_m \quad ; \quad G_1 \equiv \bar{n} \equiv \text{mean productivity of collection} \qquad (2)$$

Note that in each of these distributions, the lower limit of n is unity; items with zero productivity are excluded. Also, in many cases, the distribution diverges for very large n's, so an upper limit must be imposed.

Informational data have been compared to one or more of these distribution functions, for several reasons. In some cases[1,2,4,5] the agreement between the data and one of the distributions has been used to attribute a certain psycho-stochastic or socio-stochastic behavior to the individuals who generated the data, which behavior could be used to derive the distribution. However, as Kendall[6] has gently suggested, such attributions are usually too vague to suggest action, corrective or otherwise.

A more practical use of the distribution is to simplify and quantify the comparison between data from different universes (different libraries, different technical fields or journals, or successive year's records, for example) in an effort to discover trends on which to base action. For such purposes it is necessary to understand the mathematical (as opposed to the philosophical) properties of each distribution.

Therefore this paper will first compare the mathematical properties of the three distributions under discussion, first in the case when n is large enough to be considered as a continuous variable. Next the more complicated (but more applicable) case of discrete n is treated and the (hitherto untreated) details of the discrete Bradbury distribution are worked out. Finally, a few examples of the various applications of the distributions are noted, together with comments as to how the results may be used in practice.

Continuous Variable n.

We start by considering the case where the more important range of the productivity n is greater than about 20, so that n can be considered to be a continuous variable. In this case n is to be restricted to the range between a lower limit M and an upper limit N, above which there are only a scattered number of items, the properties of which can be lumped into the limiting quantities F(N) and G(N). Within the allowed range the density function f(n) is related to cumulative distribution and the cumulative productivity function G(n) by the relations

$$f(n) = -\frac{dF}{dn} = -\frac{1}{n}\frac{dG}{dn} \qquad (M < n < N) \qquad (3)$$

The Zipf distribution[2] assumes that F varies inversely with n;

$$F(n) = M/n \;\; ; \;\; f(n) = M/n^2 \qquad (M \leq n \leq N)$$
$$G(n) = \bar{n} - M\ln(n/M) = M[\ln(N/n) + (n_c/N)] \qquad (4)$$

The value (M/N) of F at the upper limit hopefully takes into account the scattered number of items having productivity greater than the upper limit N. The ratio $G(N)/F(N) \equiv n_c$ is the mean productivity of these hyperterminal items, whereas $G(M) \equiv \bar{n}$ is the mean productivity of the whole collection. We note that F is an exponential function of G,

$$F = \exp\left[\frac{G(n) - \bar{n}}{M}\right] = \frac{M}{N}\exp\left[\frac{G(n)}{M} - \frac{n_c}{N}\right] \qquad (5)$$

The Zipf distribution is often too restrictive to fit many cases encountered in practice, because it is seldom that the hyperterminal items add up to equal (M/N) exactly. The examples Zipf shows that fit reasonably well are those where N is large enough so that the discrepancy between F(N) and (M/N) is negligible. A modification, having much wider applicability, is obtained by allowing F(N) to have an arbitrary value,

$$F(n) = \frac{a}{n} + b \; ; \quad b = 1 - \frac{a}{M} \; ; \quad f(n) = \frac{a}{n^2} \quad (M \leq n \leq N)$$

$$G(n) = \bar{n} - M \ln\left(\frac{n}{M}\right) = n_c\left(\frac{a}{N} + b\right) + M \ln\left(\frac{N}{n}\right) \tag{6}$$

$$F(G) = 1 - \frac{a}{M} + \frac{a}{M} \exp\left[\frac{G(n) - G(M)}{M}\right] = b + \frac{a}{N} \exp\left[\frac{G(n) - G(N)}{M}\right]$$

$$G(F) = M \ln\left[\frac{a}{a - M(1-F)}\right]$$

This is just the Bradford distribution (for continuous n) as clarified by Vickery[7] and utilized by Leimkuhler[8]. The difference, as pointed out by Kendall[6], is in the point of view. Instead of looking primarily at the large number of items with small values of n, Bradford was interested in the small number of high-productivity items. He has called that small number of high-productivity items (those with n larger than N in our notation) the <u>core</u> of the collection. Of course the exact value of N is rather arbitrary; one picks the group of largest-n items responsible for an appropriate fraction of the total production, below which the distribution of Eqs.(6) hopefully holds. Bradford has given cases where, for the items outside the core ($n < N$) the logarithmic relationship between F and G holds remarkably well.

To express this relationship in the form given by Bradford, we set N (i.e., we adjust the size of the core) so that $A\,G(M) = A\,\bar{n}$, the total production of the collection, is some integer R times the production of the core, $A\,G(N)$. Then the quantity $r = G(n)/G(N)$ runs from 1 to R as n goes from N to M and

$$F(r) = 1 - \frac{a}{M} + \frac{a}{N}\exp\left[\frac{(r-1)G(N)}{M}\right] = 1 - \frac{a}{M} + \frac{a}{N}\,\alpha^{r-1} \qquad (7)$$

where $\alpha^{R-1} = (N/M)$

Each integer step of r corresponds to a successive <u>zone</u>, as defined by Bradford, each generating a production equal to that of the core, but requiring a geometrically increasing number of items to produce it. The number of items in the r'th zone is

$$AF(r+1) - AF(r) = \frac{aA}{N}\left(1 - \frac{1}{\alpha}\right)\alpha^{r} \quad (r = 1,2,3,\cdots,R) \qquad (8)$$

<u>Discrete Variable n.</u>

The foregoing sketch displays the asymptotic behavior of the Bradford distribution; the equations are good approximations for n greater than about $\bar{n} = 20$. But much of the data to be modelled by this distribution extends down to $n = 1$, where the asymptotic formulas (3) to (7) are far from correct. It is the chief purpose of this note to generate a distribution that satisfies Bradford's criterion, Eq.(8) or the last two of Eqs.(6), exactly over the whole range of integral values of n equal to or greater than unity.

We start by defining the distribution function and the production function for discrete n's,

$$F_n = \sum_{m=n}^{N-1} f_m + F_N \; ; \quad f_n = F_n - F_{n+1} \; ; \quad F_1 = 1$$
$$G_n = \sum_{m=n}^{N-1} mf_m + G_N \; ; \quad nf_n = G_n - G_{n+1} \; ; \quad G_1 = \bar{n} \qquad (9)$$

where f_n is the probability that an item of the collection has productivity equal to the integer n. Next we rewrite Bradford's criterion (8) in their terms;

$$f_n = F_n - F_{n+1} = C(e^{\beta G_n} - e^{\beta G_{n+1}}) = Ce^{\beta G_n}(1 - e^{-\beta n f_n}) \qquad (10)$$

where β and C are to be determined. To ensure asymptotic approach to Eqs.(6), we define a function Y_n (to be called the Bradford function), such that

$$F_n = (Y_n/\beta) + b \; ; \; b = 1 - (Y_1/\beta) \; ; Y_n = \sum_{m=n}^{\infty} y_m = \beta C e^{Z_N} \longrightarrow \frac{1}{n}$$
$$y_n = (f_n/\beta) \longrightarrow 1/n^2 \; ; \; \beta G_n = Z_n = \sum_{m=n}^{N-1} z_m + Z_N \qquad (11)$$
$$z_n = (nf_n/\beta) \longrightarrow 1/n \; ; \; Y_n = Y_1 e^{-V_n} \; ; \; V_n = Z_1 - Z_n = \sum_{m=1}^{n-1} z_m$$

Then Eq.(10) becomes

$$y_n = Y_n - Y_{n+1} = Y_n(1 - e^{-z_n/n}) \quad \text{or} \quad z_n = nY_n(1 - e^{-z_n/n}) \qquad (12)$$

Asymptotic series for z_n and Y_n can be obtained by setting $y_n \simeq (1/n^2) + (c/n^3) + (d/n^4)$, from which it follows that

$$Y_n \simeq S_2(n) + cS_3(n) + dS_4(n) \; ; \quad S_m(n) = \sum_{k=n}^{\infty} (1/k^m) \qquad (13)$$

Asymptotic series for the sums $S_m(n)$ can be worked out. When all these series are inserted in Eq.(12) the coefficients c and d can be calculated. It turns out that, for an error less than $\pm$ .0000003, for $n \geqslant 20$, c is effectively zero and $d \simeq -0.1845$. Thus asymptotic series, valid to six decimal places, for the functions involved in Eq.(11), are

$$z_n \simeq \frac{1}{n} - \frac{0.1845}{n^3} \quad ; \quad y_n \simeq \frac{1}{n^2} - \frac{0.1845}{n^4} \tag{14}$$
$$Y_n \simeq \frac{1}{n} + \frac{1}{2n^2} + \frac{1}{12n^3} - \frac{1}{10n^4} \qquad (n \geqslant 20)$$

To proceed for n smaller than 20 we introduce an adjoint function λ_n and divide Eq.(12) into two parts for ease in calculation and preservation of accuracy during the stepwise process of computing Y_n and z_n from $n = 20$ to $n = 1$;

$$\lambda_n = \frac{e^{z_n} - 1}{z_n} \qquad \text{or} \qquad z_n = \ln(1 + \lambda_n z_n) \tag{15}$$
$$\lambda_{n-1} = \frac{e^{z_{n-1}}}{(n-1)Y_{n-1}} = \frac{\exp(z_{n-1} + V_{n-1})}{(n-1)Y_1} = \frac{n}{n-1} e^{-z_n} \lambda_n \tag{16}$$

Knowing z_n, Y_n and thus λ_n, we use Eq.(16) to calculate λ_{n-1}, then use Eq.(15) to obtain a self-consistent value of z_{n-1}, and so on. Values of Y_n, z_n, y_n, V_n and $Y_1 - Y_n$ are tabulated in Table 1. Values for the higher values of n not given in the Table may be obtained from the asymptotic formulas of Eq.(14) plus the following,

$$V_n \simeq 0.402533 + \ln n - \frac{1}{2n} + \frac{1}{20n^2} + \frac{1}{30n^3} \qquad (n \geqslant 20) \tag{17}$$

We see thet Y_n differs greatly from $(1/n)$ when n is less than 5. The asymptotic formulas for Eq.(6) do not hold for n small.

Nevertheless, since Y_n has been generated to satisfy Eq.(10), the F_n of Eq.(11) satisfies Bradford's criterion of Eq.(10) for n clear down to $n = 1$. We have

$$F_n = \frac{1}{\beta}(Y_n - Y_1) + 1 \ ; \ F_n - F_{n+1} = \left(\frac{Y_1}{\beta}\right) e^{\beta(G_n - G_1)}(1 - e^{-z_n}) \tag{18}$$

If the total production is divided into R parts, with the portion of highest productivity AG_N equalling $AG_1/R = A\bar{n}/R$ being the core, then the r'th zone, corresponding to the portion with

G_n between $(r-1)\bar{n}/R$ and $r\bar{n}/R$, consists of $A[F(r) - F(r-1)]$ units, where

$$F(r) - F(r-1) = \frac{1}{\beta}Y_1(1 - e^{-\beta G_1/R})\exp[-\beta G_1 + (\beta G_1 r/R)] = D\alpha^r$$
$$D = \frac{1}{\beta}Y_1\,\alpha^{-R}(1 - \frac{1}{\alpha}) \quad ; \quad \alpha = e^{\beta\bar{n}/R} \tag{19}$$

Note that dividing total productivity into an integer, R, number of parts will cut through some groups of items with a given productivity n. For example, if $(r\bar{n}/R)$ comes between G_n and G_{n+1}, so that $(r\bar{n}/R) - G_{n+1} = (G_n - G_{n+1})\omega$, with $0 < \omega < 1$, then the $Af_n = A(F_n - F_{n+1})$ items with productivity n are be divided, with the portion $A(F_n^{\omega}F_{n+1}^{1-\omega} - F_{n+1})$ being allotted to the r'th zone and the portion $A(F_n - F_n^{\omega}F_{n+1}^{1-\omega})$ allocated to the (r+1)'st zone.

Another discrete distribution, also of use in the library field is the modified geometric (omitting $n = 0$)

$$f_n = (1-\gamma)\gamma^{n-1} \quad ; \quad F_n = \gamma^{n-1} \quad ; \quad G_n = \left[\frac{1}{1-\gamma} + n - 1\right]\gamma^{n-1}$$
$$G_1 = \bar{n} = \frac{1}{1-\gamma} \quad ; \quad \gamma = \frac{\bar{n}-1}{\bar{n}} \tag{20}$$

It of course does not satisfy the Bradford criterion. In fact

$$F_n = \frac{G_n}{G_n + 1 - n} \quad ; \quad G_n = \left[\frac{1}{1-\gamma} + \frac{\ln(F_n)}{\ln\gamma}\right]F_n \tag{21}$$

This distribution fits some data better than does the Bradford distribution, as will be shown in the next section.

Examples of Applications.

A familiar task of fitting a model to some operational system is to try fitting appropriately chosen data to some known distribution. The validity and utility of the model will then depend on the accuracy of the fit, but also on the amenability of the distribution to easy comparison with those for related systems. Also of importance is the degree of interest attached to different ranges of the variable involved. For example, with informational data, the geometric distribution is more useful if one focusses on the large number of small-n items, whereas the Bradford distribution is more useful (if it applies) when one wishes to concentrate on the small number of high-production items. The two distributions can differ considerably for certain ranges of n, of course, so some data will turn out to fit well only one of the two forms. Whether this implies certain psycho-stochastic behavior of the people generating the data, may or may not be a moot question.

To show how one most easily fits data to a Bradford distribution, we take two different sets of data. The first is taken from the analysis by Kendall[6] of articles on operations research in various journals, as reported in the ORSA Bibliography of Operations Research[9], published in 1958. In all, 370 journals were listed, 203 having but one article apiece, 10 having 5 articles apiece, on up to one having 242 OR articles in the time interval covered. The complete distribution and its analysis are given in Table 2; the second column, AF_n, being the number of journals having n or more OR articles, and column 8 listing the corresponding total OR articles produced, AG_n, by the AF_n journals.

The second collection of data is that gathered by Chen[10] giving the total use of various physics journals in the MIT Science Library from March 15 to June 31, 1971. Here the productivity n of a given journal series is the number of times a volume (or unbound issue) of the series was taken from the shelves and borrowed or read in the library or was used to make a photocopy of an article. Again AF_n in Table 3 is the number of different journal series with n or more uses and AG_n is the cumulative number of uses of all the AF_n journals.

To see how well each distribution satisfies the Bradford distribution of Eqs.(11) and (12), we plot $A(1-F_n)$ against $Y_1 - Y_n$ or else tabulate the ratio between the two quantities. According to Eq.(11), $A(1-F_n)$ should be equal to $(A/\beta)(Y_1 - Y_n)$, so the ratio should remain fairly constant, if the data fits the Bradford distribution. We see from column 6 of Table 2 that this is reasonably the case for Kendall's data on OR articles in different journals (the ratio changes only from 234 to 244 as n goes from 1 to 20). But, looking at column 6 of Table 3, we see that it is far from the case with Chen's data on journal use in a library (the ratio changes from 28 to 66 over the same range of n). This is shown graphically in Fig. 1 (circles are Kendall's data, crosses are Chen's). Kendall's data (a) approximates a straight line, whereas Chen's (b) has a continually changing slope.

To determine the constants for Kendall's data, we use the ratio between $A(1-F_n)$ and $Y_1 - Y_n$ listed in column 6 of Table 2. We choose the ratio 245, corresponding to $n = 15$, thus lumping together the items above $n = 15$ to constitute the core, the 20 journals that are responsible for about half the articles on OR. Since $A = 370$, the choice of $A/\beta = 245$, shown by the dashed line in Fig.1, corresponds to $\beta = 1.51$ and $AY_1/\beta = 366.4$ (to be used later). A choice of $A/\beta = 240$ (dotted line) would result in a better fit for the lower values of n, but the Bradford distribution is more useful in regard to the large values of n.

Another way of testing goodness of fit with the Bradford distribution is to see whether G varies logarithmically with F. We rewrite Eqs.(11) as follows;

$$A(G_1 - G_n) \equiv AV_n = A\ln(Y_1/Y_n) = A\ln\left[\frac{\mu}{\mu - A(1-F_n)}\right] \qquad (22)$$

where $\mu = (AY_1/\beta)$. Taking the data on $A(1-F_n)$ from the fourth column of Table 2 and inserting it in Eq.(22), using $\mu = 366.4$, gives us the next to last column, which is to be compared with the preceding column of data. The fit is reasonably good. Percentagewise , a comparison between AG_n and $AG_1 - AV_n$ from Eq.(22) would look even better. But the use of Eq.(22) is neither as sensitive nor as simple a test for fit as is the comparison of the ratio between $A(1-F_n)$ and $Y_1 - Y_n$ for different values of n.

In the end, we have reduced the double columns of data to a trio of numbers, $A = 370$, $\beta = 1.51$ and $\bar{n} \equiv G_1 = 1763/370 = 4.76$ And, when we wish to compare these data with another set (such as with the references to OR at a later date or the references to another field in the same journals) we need only compare the values of β and $\bar{n}$, instead of comparing two pairs of columns or looking at two graphs of the sort of Fig.1. For instance the second example of Kendall, the distribution of articles on statistical methodology among 184 journals also fits the Bradford distribution well, with a $\beta = 2.43$ and $\bar{n} = 8.93$ Since the α of Eq.(19) equals $\exp(\beta\bar{n}/R)$, the degree of concentration of productivity is measured by the product $\beta\bar{n} = \beta(AG_1/AF_1)$. This product is 7.2 for the OR articles; it is 21.7 for those on statistical methodology. Therefore if, in either case, the journals were divided into a core plus 4 zones ($R = 5$), the α in the former case would be 4.2, the α in the latter case is 77. Thus the articles on statistical methodology are much more concentrated in a few journals than are the articles on OR; the values of $\beta\bar{n}$ provide the quantitative comparison.

Leimkuhler[11] discusses other applications of the Bradford distribution in literature search and file organization.

As for the data of Chen, it is futile to try to fit the crosses of Fig.1 to a straight line through the origin, as would be required for a Bradford distribution. If, however, we wished to fit the highest productivity, we could try to fit by running a straight line through the $n = 50$ cross, choosing A/β to be 77 (making the core the 6 journal series responsible for 44 percent of the use. Table 3, columns 7 and 10, compared with columns 2 and 9, show how poor a fit it is for lower values of n.

On the other hand, Chen's data fits the modified geometric distribution of Eq.(20) much better than Kendall's does. Here the simplest test is to plot AF_n against n on a semilog plot, as in Fig.2 (or else to tabulate the ratio between $n - 1$ and $\ln(F_n)$ for different n's; call its mean q; then $\gamma = e^{1/q}$). Chen's data, plotted as crosses, fits the dot-dashed line, representing $AF_n = 138(0.93)^{n-1}$, fairly well (see also the third column of Table 3). The circles are Kendall's data and the dashed line for $\gamma = 0.79$ fits nearly as poorly as Chen's data fits the

Bradford distribution.

Other library-use data [12] also fit the modified geometric distribution. The comparison of their values of γ can suggest action or further investigation. The value of γ for Chen's data can be compared with the corresponding γ for the use of physics periodicals more than 5 years old, for example, to see when to retire some journals to less accessible shelves.

Thus we seem to have uncovered a generic difference between use data (such as Chen's), which seem to follow the geometric distribution, and the distribution of articles over a collection of journals (such as Kendall's data), which seem to follow the Bradford distribution more closely. The two distributions differ sufficiently so that distinction can be made. Fig.3, which gives plots of several geometric distributions against $Y_1 - Y_n$, shows that none of them could be mistaken for a Bradford distribution. It is not clear why use data seem to fit the geometric, and reference data seem to fit the Bradford distribution. One might devise socio-stochastic reasons for the difference, perhaps along the lines suggested by Kendall[6] and by Bush et al[5], if the question ever became important.

However, no matter what these reasons might turn out to be, either distribution, if it fits the data, can be used in many operational ways. Either distribution can be used, for example, to pick out the fraction of most productive items, a matter of some consequence in a budget-limited library. Moreover, a comparison of the values of the parameters γ, or $\bar{n}$ and β, for different, related sets of data, can help in reaching other administrative decisions. For example, a comparison of the γ for the use of newly bought books with the γ for books 5 years old will indicate how fast books of that class can be retired to less accessible shelves. For another example, a comparison of the values of β and $\bar{n}$ for articles on different subjects in a collection of periodicals might indicate the most convenient place to shelve the collection (near the books on physiology or near those on psychiatry, for instance). When large numbers of items are involved, the parameters of a distribution (provided, of course, the data fit the distribution) are easier to compare and to reach decisions on, than the full tabulation of all the data.

References.

1 Bradford, S.C., Documentation, Crossby Lockwood, London (1948).

2 Zipf, G.K. Human Behavior and the Principle of Least Effort, Addison-Wesley, Reading, MA, (1949).

3 Morse, P.M.,"Demand for Library Materials", Collection Management, 1, (1977).

4 Simon, H.A., "On a class of Skew Distribution Functions", Biometrika 4, 198 (1948).

5 Bush,G.C.,Galliher,H.P.,and Morse,P.M., "Attendance and Use in the Science Library at MIT", American Documentation, 7, 87 (1960).

6 Kendall, M.G., "The Bibliography of Operational Research", Opnl.Res.Q. 11, 31 (1960).

7 Vickery,B.C., "Bradford's Law of Scattering", Jour. of Documentation, 4, 198 (1948).

8 Leimkuhler, F.F., "The Bradford Distribution", Jour. of Documentation 23, 197 (1967).

9 A Comprehensive Bibliography on Operations Research, through 1957, ORSA Publication No.4, John Wiley Sons, N.Y. 1958.

10 Chen, C.C.,"The use Patterns of Physics Journals in a Large Academic Research Library", Jour.Am.Soc.Info.Sc,, July-Aug. 1972, page 266-8.

11 Leimkuhler, F.F., "A literature Search and File Organization Model", American Documentation 19, 131 (1968).

12 Morse, P.M. Library Effectiveness, MIT Press, Cambridge, MA, 1968, pages 157-165.

TABLE 1.

The Bradford Function.

see Equations (11).

n	Y_n	z_n	y_n	V_n	$Y_1 - Y_n$	n
1	1.495479	0.867172	0.867172	0.000000	0.000000	1
2	.628308	.475665	.237832	.867172	.867172	2
3	.390476	.325254	.108418	1.342836	1.105003	3
4	.282058	.246357	.061589	1.668090	1.213421	4
5	.220469	.198146	.039629	1.914447	1.275010	5
6	.180840	.165512	.027585	2.112593	1.314639	6
7	.153254	.142193	.020313	2.278105	1.342225	7
8	.132941	.124477	.015560	2.420298	1.362538	8
9	.117381	.110820	.012313	2.544775	1.378098	9
10	.105069	.099713	.009971	2.655595	1.390410	10
11	.095097	.090766	.008251	2.755308	1.400382	11
12	.086846	.083145	.006929	2.846074	1.408633	12
13	.079917	.076860	.005912	2.929219	1.415562	13
14	.074005	.071300	.005093	3.006079	1.421474	14
15	.068912	.066620	.004441	3.077379	1.426567	15
16	.064471	.062413	.003901	3.143999	1.431008	16
17	.060570	.058786	.003458	3.206412	1.433909	17
18	.057112	.055500	.003083	3.265198	1.438367	18
19	.054029	.052604	.002769	3.320698	1.441450	19
20	.051260	.049977	.002499	3.373302	1.444219	20
25	.040805	.039988	.001600	3.601431	1.454674	25
30	.033892	.033327	.001111	3.787079	1.461587	30
35	.028981	.028567	.000816	3.943607	1.466498	35
40	.025314	.024997	.000625	4.078921	1.470165	40
50	.020201	.019999	.000400	4.304562	1.475278	50
60	.016806	.016666	.000278	4.488548	1.478673	60
70	.014388	.014285	.000204	4.643888	1.481091	70
80	.012578	.012500	.000156	4.778312	1.482901	80
90	.011173	.011111	.000123	4.896789	1.484306	90
100	.010050	.010000	.000100	5.002705	1.485429	100

TABLE 2

O/R Articles in Journals.
From Kendall[6], Table 1.

n	Data AF_n	Theor Eq.20 AF_n	P_n Data $A-AF_n$	Q_n Tbl.1 Y_1-Y_n	$P_n/Q_n \simeq A/\beta$	Theor Eq.18 AF_n	Data AG_n	Data AV_n	Theor Eq.22 AV_n	n
1	370	370	0	0		370	1763	0	0	1
2	167	292	203	0.867	234	159	1560	203	198	2
3	113	230	257	1.105	232	100	1452	311	296	3
4	84	183	286	1.213	236	74	1365	398	372	4
5	67	144	303	1.275	238	58	1297	466	430	5
6	57	114	313	1.315	238	49	1247	516	472	6
7	51	90	319	1.342	238	41	1211	552	501	7
8	43	71	327	1.363	240	36	1155	608	546	8
9	35	56	335	1.378	243	32	1091	672	602	9
10	31	44	339	1.390	244	30	1055	708	635	10
15	20	14	350	1.427	245	20	932	831	761	15
20	13	4	357	1.444	247	16	820	943	897	20
30	8	0	362	1.462	248	12	700	1063	1083	30
50	6	0	364	1.475	247	7	660	1103	1231	50
100	4	0	366	1.485	246	6	458	1305	1671	100

TABLE 3

Library Use of Physics Periodicals.
From Chen[9], Appendix 1.

n	Data AF_n	Theor Eq.20 AF_n	P_n Data $A-AF_n$	Q_n Tbl.1 Y_1-Y_n	P_n/Q_n ≅ A/β	Theor Eq.18 AF_n	Data AG_n	Data AV_n	Theor Eq.22 AV_n	n
1	138	138	0	0		138	4292	0	0	1
2	114	128	24	0.867	27.7	71	4268	24	18	2
3	101	119	37	1.105	33.5	53	4242	50	30	3
4	91	111	47	1.213	38.7	45	4212	80	40	4
5	87	103	51	1.275	40.0	40	4196	96	44	5
6	78	96	60	1.314	45.7	37	4151	141	55	6
7	75	89	63	1.342	46.9	35	4133	159	58	7
8	69	83	69	1.363	50.6	33	4091	201	67	8
9	66	77	72	1.378	52.2	32	4067	225	71	9
10	62	72	76	1.390	54.7	31	4031	261	77	10
15	50	50	88	1.427	61.7	28	3881	411	99	15
20	43	35	95	1.444	65.8	27	3766	526	130	20
30	37	17	101	1.462	69.1	26	3619	673	174	30
50	25	4	113	1.475	76.6	25	3158	1134	252	50
100	6	0	132	1.485	88.9	24	1876	2460	2460	100

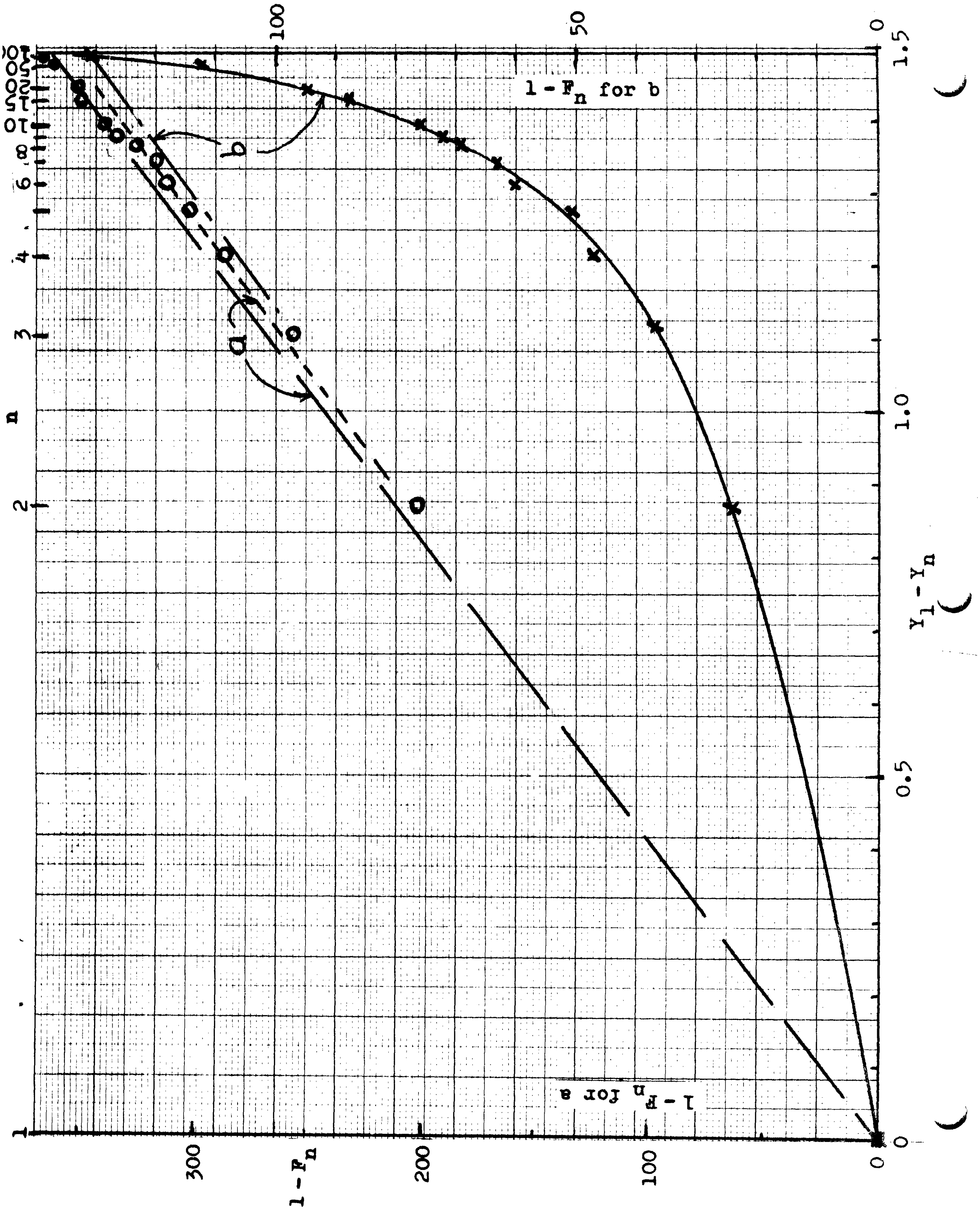
Figure Captions.

Fig. 1 Test of fit with Bradford Distribution: $1 - F_n$ from data plotted against $Y_1 - Y_n$ from Table 1.

Fig. 2 Test of fit with Modified Geometric Distribution; F_n from data plotted on semi-log paper.

Fig. 3 Unconformity of the distributions; $1 - F_n$ of geometric plotted against $Y_1 - Y_n$ of Bradford. Result is not linear.

Fig. 1



1 - F_n for b
1 - F_n for a
1 - F_n
n
Y_1 - Y_n
a
b
100
50
0
300
200
100
0
1.5
1.0
0.5
0
1
2
3
4
6
8
10
15
20
50
100

Fig. 2

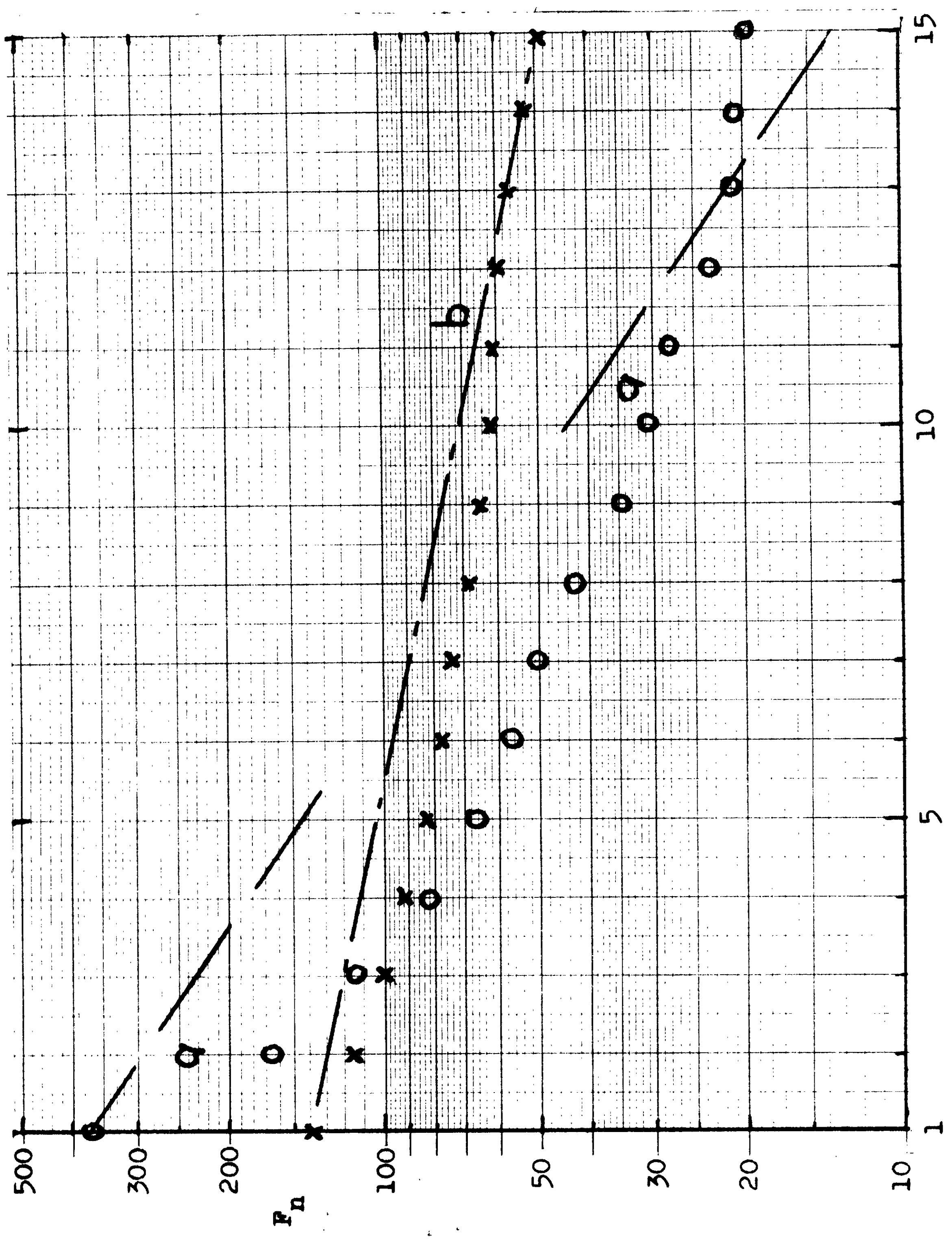


Fig. 3