TOWARDS INFORMETRICS: HAITUN, LAPLACE, ZIPF, BRADFORD AND THE ALVEY PROGRAMME

B. C. BROOKES
Department of Information Science, The City University, London

Haitun has recently shown that empirical distributions are of two types—'Gaussian' and 'Zipfian'—characterized by the presence or absence of moments. Gaussian-type distributions arise only in physical contexts: Zipfian only in social contexts. As the whole of modern statistical theory is based on Gaussian distributions, Haitun thus shows that its application to social statistics, including cognitive statistics, is 'inadmissible'.

A new statistical theory based on 'Zipfian' distributions is therefore needed for the social sciences. Laplace's notorious 'law of succession', which has evaded derivation by classical probability theory, is shown to be the 'Zipfian' frequency analogue of the Bradford law. It is argued that these two laws together provide the most convenient analytical instruments for the exploration of social science data.

Some implications of these findings for the quantitative analysis of information systems are briefly discussed.

1. INTRODUCTION

THOUGH DISCLAIMING ANY expertise in statistical theory, Geoffrey Woledge always listened (it seemed to me) with intuitive grasp of the main issues when statistical papers which had evoked conflicting reports from the referees had to be sorted out by the Editorial Board of this Journal. And, in particular, I always felt his sympathetic interest when papers on the Bradford law, some of them mine, called for such critical discussion.

As Bradford had been a librarian and bibliographer and as I defended Bradford's work on the grounds of its innovatory importance and relevance to quantification in the *social sciences,* it could be that Geoffrey Woledge's sympathetic interest stemmed from his own experience of the social sciences and of the statistics of documentation while he was Librarian of the London School of Economics.

So it is with pleasure that I can now report some recent developments which give authoritative support to arguments which I slowly developed during the keen discussions of the Bradford law arising from papers published by this Journal. The saga of the Bradford law as there reported can now be rounded off and it is fitting that this final report on it—at least from *me*—should be a contribution to this issue of the Journal commemorating Geoffrey Woledge's long service to its Editorial Board.

2. HAITUN'S TWO TYPES OF STATISTICAL DISTRIBUTION

The Russian statistician S. D. Haitun[1] recently published a three-part comprehensive review of all the empirical frequency distributions that have been reported in the literature of bibliometrics and related fields. With confident use of modern statistical theorems he has shown that all empirical distributions can be divided into two main types. One of these types is characterized by the fact that the

Journal of Documentation, Vol. 40, No. 2, June 1984, pp. 120–143.

distributions have as many higher moments as modern statistical theory demands; Haitun defines these as the 'Gaussian' type. The other type have no moments whatever and Haitun defines these as the 'Zipfian' type. I shall henceforward refer to these two types of distributions as being of the 'G-' and 'Z-' types respectively.

Haitun thus noted that modern statistical theory is, in effect, the theory of moments and that all distributions and techniques of that theory are derived from distributions whose population moments can be estimated with reasonable accuracy from sample values.

The moments of hypothetical G-type populations can be calculated with reasonable accuracy from a finite sample. All distributions of this broad class can be related directly to the Gaussian (or Normal) distribution. And all of them originated in and have been very successfully applied to problems in the *physical* sciences.

Z-type distributions are very skew; they have long tails which break down into small frequency classes of 1's or 2's separated by increasing runs of zero frequencies. Ever since Zipf,[2] working on the statistics of vocabularies, formulated his law in 1935, the statistics of such distributions have been disputed by theoreticians among whom Benoit Mandelbrot[3] and Herbert Simon[4] made useful clarifications. But the main issues remain unresolved. It is very difficult for an orthodox statistician not to believe that the long-tailed distributions provided by vocabularies or bibliographies cannot be captured and sorted out theoretically by the highly sophisticated techniques he commands. But it has not happened yet and Haitun has shown that the application of G-type statistical techniques to Zipfian distributions is 'inadmissible'. Modern statistical theory, the theory of moments, cannot be validly applied to distributions *which have no moments at all.*

When Zipfian distributions are said to have 'no moments at all', it has to be understood that this description is being used in a technical sense. Any finite empirical distribution, of course, has a mean, a variance and as many higher moments as one cares to calculate. But the moments of a Zipfian distribution grow inexorably with the size of the sample. Though it is not impossible to say that the hypothetical population from which the sample was drawn has moments which are 'infinite', the theoretician prefers to say of such distributions that their moments 'do not exist'. In such a case the theory of moments is useless.

Haitun was able to find only *one* Zipfian distribution which did not arise from social contexts. That exception is the distribution of the energies of the particles which reach Earth from outer space. Haitun relates this phenomenon to the slow evolutionary degradation of cosmic radiation during the expansion of the universe which has made it entropically possible for living forms to emerge in the physical substrate of the cosmos. This is an intriguing speculation (in conformity with Marxist-Leninist materialism) but this is not the place to pursue its implications. The most important result of Haitun's comprehensive and critical review is that he points to a great void in modern statistical theory: *it has nothing to offer the social sciences.* As I have said before: 'Humans do not behave like physical particles',[5] but it is taking a long time for this idea to be reflected in the techniques applied to social statistics.

(A paper on Bradford's law I submitted to this Journal some years ago was rejected by one of the referees—a 'Cambridge statistician' unknown to me—on the grounds that I had fitted some data—Zipfian of course—to a distribution without applying the appropriate techniques of orthodox statistics, i.e. by using the method of moments. My Zipfian distribution had no moments. But at that time I could not prove nor even cite the proof that it was not possible to fit the data by the method of moments. So Haitun's discovery has been a great encouragement.)

3. LAPLACE'S 'LAW OF SUCCESSION'

Some years ago I was, for a period, a member of the Committee and, for a time, Secretary of the British Society for the Philosophy of Science when LSE was powerfully represented on it by Sir Karl Popper, Imre Lakatos, John Wisdom and J. W. N. Watkins. The monthly meetings of the Committee, regarded by its anxious Secretary as the only means of getting decisions on urgent mundane matters, were liable to become immersed in highly illuminating (but secretarially irrelevant) issues of intense philosophical interest to everyone else. One of the issues often mentioned at that time concerned induction: How many positive exemplifications of an empirical datum were needed before it could be asserted that it was True? The very question was of course dismissed by Popper as futile: scientific conclusions could *never* be claimed as *True*—they were always open to overthrow or modification in the light of further evidence.

But induction was an issue which had long engaged philosophers and was still being debated. One of the results at the centre of this issue had been derived by the French mathematician Laplace in 1774 and was known as his 'law' or 'rule' or 'principle' of 'succession'. As often used in the perennial controversy, Laplace's law was improperly reduced to statements of the form: If an event, such as the rising of the sun, has been known to occur for n successive days without exception, what is the probability that it would arise on the $(n + 1)$th day? And the answer given by the law is $(n+1)/(n+2)$ and the probability that it will *not* rise is therefore $1/(n+2)$. As n increases, $(n+1)/(n+2)$ gets closer to 1 but never quite attains it. Strictly, therefore, we can never assert that the sun *will* rise tomorrow with a truth value of 1 precisely.

The influence of Sir Karl Popper, combined perhaps with growing uncertainties about other aspects of human life, has dampened the search for Truth along this line of thought. But Laplace's law is still with us and has relevance to my present problem. And it is still misunderstood and misapplied.

Laplace was a self-taught mathematical genius who moved quickly over the fields that interested him, leaving the pedants on his tracks gasping as he leapt lightly from one mathematical conclusion to the next. He wrote a five-volume *Traité de mécanique céleste* which was published volume by volume over the period 1799–1825. Laplace said of this celestial world that if there were a *mind* (we would now say *computer*) capable of grasping all the data of the motions of all the bodies of the cosmos at some specific instant, then it could predict its future and, if the sign for t (for time) were reversed, it could similarly reveal how it all began. So his celestial world was wholly *deterministic*. Yet during this same period Laplace was also working on the theory of probability and made important advances in it. He published his *Théorie analytique des probabilités* in 1812 and everything he dealt with in this work was reduced to *chance*. So Laplace was simultaneously describing *two* worlds—one wholly deterministic and the other wholly probabilistic. I feel that we have not yet quite caught up with him!

Laplace dedicated the first edition of his treatise on probability to Napoleon, who had already made him a Count. But by the time the second edition was published, in 1814, Napoleon was incarcerated on Elba. So Laplace deleted the dedication and added a note saying that those who understood the laws of probability (as described in his treatise) would have been able to predict the fall of an Emperor so ambitious. In 1817, when Napoleon was in St Helena and the French monarchy had been restored under Louis XVIII, Pierre-Simon Laplace of humble origin duly became the Marquis de Laplace. As he had lived through the Revolution, Laplace was well aware of the vagaries of human life!

When philosophers of induction, complaining that they can find no exemplification of Laplace's law in 'Nature', seek such examples by contriving experiments with the hardware of classical probability theory, i.e. with dice, cards, roulette wheels, urns full of coloured balls and so forth, they have so far failed to find what they seek. Thus Kneale[6] writes:

> 'If this rule were sound it would be extemely useful. Nor is authoritative support lacking. De Morgan, Jevons and Karl Pearson have all given it their blessing. But even last century, when Laplace's reputation as a writer on probability was higher than it is now, there were critics such as Boole. And today it is generally admitted that this rule must be a mistake, since it leads to absurdity and self-contradiction' (p. 203, *op. cit.*).

I do not understand how it is that even the most humane philosophers still regard 'Nature' as inanimate, devoid of life, purely *physical.* The derivation of the Laplace law is still sought through carefully contrived experiments with physical objects which behave according to the laws of mechanics or on a theory of probability which is also based on games of chance with physical objects. Yet, since Einstein's theory of relativity and Planck's quantum theory were elucidated in the early years of this century, even theoretical physicists have had to concede a central role to the human observer of physical 'Nature'.

So Laplace's law presents two difficulties: first, no example of it has been found in the physical world or in physical experiments and, secondly, it cannot be derived from the classical theory of probability. But that, too, has a physical basis and has been eminently successful when applied to the physical sciences. Both experiment and theory have so far rigorously excluded any human role.

However, as Kneale (*op. cit.*) and many others have demonstrated, there is no difficulty whatever in deriving Laplace's law if only one can swallow the 'principle of indifference'. As Laplace derived his law while analysing his concept of *categorization,* which seems to me to be a *mental* rather than a *physical* activity, it does not surprise me that the law is *not* manifested in the *physical* sciences.

But that is a fact of great interest because examples of the Laplace law abound in the *social* sciences, as I show when I have explained what to look for. So the negative findings of the philosophers of induction give positive support to Haitun's conclusion that application of modern statistics to the social sciences is 'inadmissible'.

4. THE LAPLACE DISTRIBUTION

In everyday life we calculate when we need to be exact, and know how to perform the calculation. Even if some quantitative data are available we may not be able to make a decision wholly by calculation because many aspects of daily life—convenience, comfort, taste—cannot be quantified. So we guess, estimate, make a *judgement.* The principle of indifference provides a rule for making *judgements.*

The principle would be well understood by punters who bet on horses in the world of racing. If a horse has won, say, all 5 of its first 5 races, it clearly promises well when it runs its 6th race; it is more likely to win (*ceteris paribus*) than to lose. The principle indicates that the odds on it winning are (5 + 1):1, i.e. it allows for what is already known (its 5 wins) and then 'splits the difference' (here expressed as the two 1's) about the remaining doubt. So the odds on the horse winning its 6th race would be 6:1 and its probability of winning would therefore be 6/7. Similarly, if the horse had won 2 and lost 3 of its first 5 races, the principle takes note of the wins and losses and again splits the difference over the remaining doubt to give the odds of the horse winning its 6th race as (2 + 1):(3 + 1) or 3:4 and its probability of winning would therefore be 3/7. If the horse wins all of its first n races, the probability of winning its (n + 1)th race would be (n + 1)/(n + 2) and the probability of it losing would be 1/(n + 2). As punters well know, even the best horses sometimes fall down; the fraction 1/(n + 2) never attains the value of zero however large n may be. Laplace had similar reservations about ambitious and successful Emperors.

One cannot apply the principle of induction to games of chance. If, with a normal 6-sided die one throws three sixes in a row, it would be unwise to assume that the probability of getting a 6 with the next throw is 4/5—it remains at 1/6.

Laplace, however, applied the principle to a highly abstract problem. If a collection of 'A-things', i.e. things which are clearly distinguishable from all things which are *not* 'A-things', becomes large enough to warrant the need, it may be decided to categorize the A-things further—into the categories we will label C1, C2, C3 and so on. For such a purpose we need to verify that all the A-things share some secondary characteristic which varies in form or quantity among the A-things. Imagine that all A-things are brought together and, to simulate an experiment in classical probability, are put in an *urn*. The categorizer focuses his mind on one of the sub-categories and asks himself the odds of the first A-thing he draws out being a member of that category.

In this first case we have no *a priori* knowledge of what to expect. So the odds that the first A-thing drawn conforms with the particular category we have in mind, according to the principle of indifference, are as 1:1 and the probability is 1/2. It does not matter whether the first A-thing—A1—belongs to the sub-category we expected or not: it has to be categorized and this first category can be labelled C1. We now *know* something—that the set of A-things contains at least one C1. For the second drawing from the urn we now allow for this knowledge by changing the odds that A2 will be another C1 to (1 + 1):1. So the probability that A2 is a C1 is 2/3; that it demands a new sub-categorization is 1/3.

So the rule is established. The sub-categorization can continue indefinitely on the same lines. Considering the total process, as the number of A-things drawn from the urn increases from 1 to n, the ratios of (the probability that the next A-thing will be one of the sub-categories already established) to (the probability that it will need a new categorization) are successively as follows:

$$(1/2:1/2),\ (2/3:1/3),\ (3/4:1/4)\ \ldots\ldots\ldots\ n/(n+1):1/(n+1)$$

and, finally, on the (n + 1)th drawing, the ratio becomes (n + 1)/(n + 2) to 1/(n + 2) in conformity with the rule for all values of n. The distribution I sought is therefore represented by the series:

$$(1 - 1/2) + (1/2 - 1/3) + (1/3 - 1/4)\ \ldots\ldots\ldots\ (1/(n+1) - 1/(n+2)).$$

We can therefore write

$$p(m) = \frac{k}{m} - \frac{k}{m+1} \ldots\ldots m = 1, 2, 3, \ldots n, \ldots \quad (1)$$

As the series above 'telescopes' neatly to $1-1/(n+1) = n/(n+1)$, the constant $k = (n+1)/n$ for finite *n*. This constant tends to 1 as *n* tends to infinity.

This highly general and highly abstract probabilistic procedure for categorizing *any* collective of A-things looks absurdly improbable and impractical. Every new A-thing examined as it emerges from the urn generates the *possibility* that the *n* A-things will differ from all others in respect of the secondary characteristic. On the other hand, it allows for the *possibility* that the *n* A-things examined will all be of the same kind. It therefore allows for there to be *n* different categorizations of the *n* A-things or just 1—or, of course, for any intermediate result to arise. Anything could happen! But we know that already. On the face of it, the idea looks so impractical as to be unworthy of further consideration.

I would have dismissed it on sight except for the fact that it has been reported in the literature. It was found by Maurice Kendall[7] (then Professor of Statistics at LSE., who acknowledges the bibliographical help of Geoffrey Woledge) to fit the data of the bibliography prepared for the Operations Research Society of America and he remarked on the surprisingly good fit it gave. But he did not recognize it as related to Laplace's law. Since that time (1960) papers on this same set of data have been published in this Journal. The first was by Leimkuhler[8] (1980). The second was by Brookes[9] (1981) who fitted the data to what I now recognize to be a more general form of the distribution above, and which I derived from the Bradford law.[10]

The distribution described above, as derived from Laplace's law of succession (the first such derivation as far as I know) is not yet well known. Though Laplace's name has long been associated with other distributions on which he worked—such as the Gaussian or Normal—I henceforward refer to the distribution derived from his law of succession as the *Laplace distribution of categories* or, more shortly in the present context, as the *Laplace law*.

However, because of the doubts about its validity arising from the use of the principle of indifference in its derivation, the Laplace distribution must be regarded as empirical until a satisfactory derivation is found. Its claim to fit empirical data will be demonstrated in Section 5.

5. SOME PROPERTIES OF THE LAPLACE DISTRIBUTION

Categorization is a *discrete* process—items can be categorized only one by one. But exploration of the properties of the Laplace distribution is needed in order to find out how to apply it to practical cases. This exploration is greatly simplified if we consider a distribution with *n* so large that it can be treated as though it were *continuous.*

Consider the distribution with the probability density function k/x^2. The area under the graph of this function between the abscissae *m* and $(m+1)$ is given by:

$$p(m) = \int_m^{m+1} k/x^2\,dx = \frac{k}{m} - \frac{k}{m+1}, m = 1, 2, 3, \ldots n, \ldots \quad (2)$$

which is the function derived in eqn (1). This presents the Laplace distribution as based on a continuous inverse square probability density. The inverse square reminds one of the Lotka law which Lotka[11] regarded as a *discrete* function (but I can show that the Laplace law of eqn (2) gives better fits to Lotka's data than Lotka's own modifications of his law). The discreteness we need is obtained by summing the areas under the graph of $y = k/x^2$ over discrete intervals of x.

The *mean* or first moment of the distribution eqn (2) is given by:

$$\int_1^{n+1} \frac{k}{x^2} x\,dx = k \ln(n+1) \qquad (3)$$

As $n \rightarrow \infty$, so also does $\ln(n+1)$. This distribution is therefore said to have 'no mean'—and therefore no higher moments.

The Laplace distribution is thus of Haitun's Z-type. If it has not been noticed before, it is because it has no place in physical contexts, no role in the G-type statistics of modern statistical theory. I note that it is excluded from Karl Pearson's categorization of useful statistical distributions.[12]

Before applying the Laplace distribution to empirical data, we need to know what has to be counted and how the results are to be arranged. As derived in the previous section, the more probable categories are likely to emerge first and appear at the head of the distribution. But statisticians would count the categories in terms of their sizes of membership, giving priority to those which have only 1, then 2, 3 . . . members and so on in order to put the data in the form of an orthodox frequency distribution. So we have to reverse the order we already have.

We therefore have to repeat the categorizing process—not with A-things this time—but with categories C1, C2, C3 . . . into which they have now been formed. The secondary characteristic of the C-things to be examined is the number of A-things each can claim. In the previous section I marvelled at the abstract generality of Laplace's law, but that generality has its advantages; one is that we already know that the probability distribution of this categorization of the C-things, as of everything else, should conform to the Laplace distribution.

It is time to see how it fits empirical data.

6. LAPLACE ANALYSIS OF SOME BOOK-INDEX DATA

My readiest sources of empirical data sets on which to explore the uses of the Laplace distribution are book indexes. A book index categorizes the topics the author discusses: those topics the author touches only lightly have only one or two page references, those treated more fully have more. A quick scanning of the book index therefore indicates how the author treats his subject.

A large treatise on a scientific subject often has entries in which two or more levels of sub-categorization appear; they are shown by deeper and deeper insets under the main heading. But in this initial exploration of the possible applications of the Laplace distribution one has to begin with the simplest cases.

Of the books within reach, the first to provide the simple index I sought was Ernest Gellner's *Legitimation of belief* (C.U.P. 1974), on philosophical aspects of sociology, whose only index listed the names of those mentioned in the text. There were 123 such entries, many with only one page reference and a few with rather more.

In counting such references one has to make decisions on what to count and how to count them. For example, a reference to, say, pp. 10–15 would count as 6 page references; a reference to 'p. 8 and n', i.e. to the main text and to a footnote on the same page, would count as 1 rather than as 2 because when a name occurs more than once in the main text on the same page it earns only one entry. Having made such rules, it is important to apply them consistently even though they are arbitrary and may be inappropriate for the next index count.

The frequency distribution of pages references obtained from Gellner's book is displayed in Table 1. It is typically very skew with a long straggling tail.

TABLE 1. *A frequency distribution of book index page references*

m	1	2	3	4	5	6	7	8	9	10	Totals
f	60	23	7	7	7	3	3	1	1	—	112
m	11	12	13	14	15	16	17	18	19	20	
f	1	1	1	—	—	2	—	1	—	—	118
m	21	22	23	24	25	26	27	28	29	30	
f	—	2	1	—	—	—	—	1	1	—	123

To fit the probability distribution of eqn (2) to the data only one parameter has to be evaluated—that for sample size. As the sum of the frequencies is 123 and as the sum to 29 terms of p(m) is $n/(n+1)$ with $n = 29$ in this case we have the required factor, k, given by $123{\cdot}30/29 = 127{\cdot}24$. The frequencies are then calculated from eqn. (2). They are shown as a compacted distribution in Table 2 in which the frequency classes have been grouped to ensure frequencies of at least 5 so that the χ^2 test for goodness of fit can be applied. (For this test I prefer larger values of the frequencies than 5, but here the frequencies in the tail, as in all such distributions, are very small.)

TABLE 2. *The book-index data compared with calculated values*

m	1	2	3	4	5	6–10	11–29	Total
f: data	60	23	7	7	7	8	10	123
f: calc	63·6	31·2	10·3	6·4	4·2	9·6	7·3	122·9

When the χ^2 test is applied to the 7 groups, the χ^2 sum is 4·37. As only one parameter has been calculated from the data, the number of degrees of freedom for this case is $(7 - 1) = 6$. For the 6 d.f. the value of χ^2 ranges from 1·6 (95%) to 12·6 (5%) and as 4·37 lies in the middle of the range, the data can be said, by orthodox standards, to fit the Laplace distribution. So here at last is one empirical manifestation of that distribution and there are many others like it to be found in one-level book indexes.

But I have to point out that the χ^2 test is derived from G-type statistics and therefore, according to Haitun, its application to the Z-type Laplace distribution is 'inadmissible'. There are other reasons for reservations about the χ^2 test as applied to Z-type distributions which always have long straggling tails in which the highest frequencies arise. In *social* contexts these high frequencies relate to the most important, most prolific or most productive members of the categorized collective. Yet the form of the distribution is determined by the large numbers of the least productive members. So disputes have arisen about the exact form of these long tails and in such analyses the χ^2 test is useless—but then it was not intended for use with Z-type distributions. We need new techniques for measuring the goodness of fit in Z-type analysis.

7. THE INEFFICIENCY OF FREQUENCY DISTRIBUTION ANALYSIS

When two teams A and B meet to play there is usually a decision to be made: which of the two goals do they first defend? The decision is usually made by the tossing of a coin by the referee or umpire. But whether the decision is made by fiat or by chance, the information implicit in the decision is measurable (in terms of information theory) as $\log_2 2$ or 1 bit.

If it has to be decided in what order three athletes should take turns in a jumping event, there are more possibilities: the three athletes C, D, and E can jump in any one of the 6 orders: C,D,E; C,E,D; D,E,C; D,C,E; E,C,D; or E,D,C. Again the issue can be decided by fiat or by the drawing of three straws from a hand which holds three straws of unequal length but presents only tips of equal length to the athletes. The information needed in this case is $\log_2 6 = 2{\cdot}58$ bits.

Similarly, if there are n orders to be determined, the information measure of the n choices is $\log_2 n!$ bits. (If logs to the base 2 are not available, then it is possible to use $\log_2 n = \log_{10} n/\log_{10} 2$.)

The index of Gellner's book consisted of 123 names in alphabetical order together with the corresponding numbers of page references. The statistical information the index provides for analysis is therefore $\log_2 123! = 681{\cdot}3$ bits. When the frequency classes are formed some of this information is lost by the grouping into frequency classes. Thus in forming Table 1 the amount of statistical information lost in this way is

$$\log_2 60! + \log_2 23! + 3 \log_2 7! + 2 \log_2 3! + 2 \log_2 2!$$

and this sum totals 390·7 bits.

More information is discarded when the data of Table 1 are compacted into Table 2 for the application of the χ^2 test. The information lost in this way is

$$\log 8! + \log 10! - 2\log 3! - 2\log 2!$$

which amounts to 29·9 bits.

The fit to the Laplace distribution therefore applies to only $681{\cdot}3 - 390{\cdot}7 - 29{\cdot}9 = 260{\cdot}7$ bits of the original 681·3 bits, i.e. to only 38·3% of the data available. That something has been lost can be verified without calculation: it is not possible to reconstitute the order of the index from Table 1, nor even Table 1 unambiguously from Table 2.

Similar criticism can be made of any other theoretical distribution that might be fitted to the index data if that distribution is of the orthodox frequency kind. The fault lies in the *frequency* form of the distribution not in the Laplace distribution in particular.

When analysing the world of inanimate objects as, for example, mass-produced motor cars of the same type, whose *individualities* are of no social interest, such discarding of empirical information is of no account. Though the object *could* be distinguished by marking them, as cars in use could be identified by their number plates, in many contexts such distinctions are of no social interest—any one car from a mass-production line is regarded as equivalent to any other of the same type.

But a book index presents not only statistical information but much other information of cognitive interest to the user of the book. The names which attract the most page references are discernible at a glance: on skimming over the index to Gellner's book I note at once that the names most often cited are those of Chomsky, Hume, Kant and Popper. Those names indicate to the reader familiar with the topic the scope of Gellner's argument.

A crucial weakness of G-type statistics is thus made apparent: Z-type distributions present more *statistical* information than G-type analysis can analyse efficiently. Can the information available in social statistics be used more efficiently?

8. LAPLACE'S LAW AND THE LAWS OF BRADFORD AND ZIPF

Bradford, the professional bibliographer, and Zipf, the professional linguist, were both more interested in their respective professional fields than in statistics. But both appreciated that the statistical regularities they had observed in their fields might reveal aspects of fundamental interest. In striving to capture these regularities as precisely as possible, both Bradford and Zipf abandoned the frequency distributions of orthodox statistics for the unconventional frequency-rank distributions.

In a scientific bibliography, or computer print-out in response to a query to a database, the richest sources of relevant papers are those journals which specialize in the particular topic. In learning the vocabulary of a new language, the words of most immediate interest are those which occur most frequently in texts in that language. There is a rough and ready rule-of-thumb—the 80/20 rule—which states that 80% of the bibliography or of the language text is provided by the most productive 20% of the sources.

So both Bradford and Zipf, in the same year (1935) but quite independently, adopted the unorthodox statistical technique of *ranking* their sources, beginning with the most productive. Bradford cumulated the frequencies of his ranked items; Zipf did not. The advantage of ranking is that it brings to the forefront of the distribution those items of greatest professional interest and relegates to the distant tail those items of rare occurrences which are relatively difficult to find and identify—thus reversing the procedure imposed by *frequency* distributions.

As both Bradford and Zipf were concerned with *categorization,* it seemed to me to be of interest to see whether the Laplace law was related to them.

The number of items, and therefore the number of entities to be ranked, in the tail of the Laplace distribution from $x = m$ to its end point at $x = n+1$ is given by

$$r = \frac{k}{m} - \frac{k}{n+1}.$$

As both k and $(n+1)$ are constants, we can put $k/(n+1) = w$ and re-write this relation as

$$k/m = r + w \ldots\ldots \qquad \ldots\ldots\ldots\ldots\ldots\ldots \qquad (4)$$

The number of items embraced by the Laplace law over this same range, m to $(n+1)$ is given by

$$G(r) = \int_{m}^{n+1} \frac{k}{x^2} . x\,dx = k\,1n\,(n+1) - k\,1n\,m$$

$$= k\,1n(k/w) - k\,1n\,k/(r + w), \qquad \text{from eqn (4)},$$

$$= k\,1n(1 + r/w) \ldots\ldots\ldots\ldots \qquad (5)$$

Eqn (5) is formally identical to the formulation of the Bradford law as derived by Leimkuhler[13] from Bradford's somewhat ambiguous formulation.

As Zipf did not cumulate the frequencies of his f/r data, the Zipf law is given by

$$g(r) = \frac{dG(r)}{dr} = k/(r + w) \ldots\ldots\ldots\ldots \qquad (6)$$

which is one of the forms proposed by Zipf.[2]

(I use f(x) and F(x) for *frequency* distributions and g(r) and G(r) for the equivalent functions of f/r distributions.)

Though I have reservations to make about all three formulations of these three laws, it is evident that the laws of Laplace, of Bradford and of Zipf are all manifestations of the same basic, but still *empirical,* law of categorization.

The fact that it has not been possible to derive the Laplace law from classical probability theory, now explains why it has not been possible to derive the Bradford or Zipf law in that way either. As both of them were expressed in the unorthodox form of *ranked* distributions, that fact, too, compounded the difficulty of making them acceptable to orthodox statisticians. But as the Laplace law is of the Z-type, so, too, are the laws of Bradford and Zipf; they are therefore beyond the reach of modern G-type statistics.

It also becomes clear that these three laws—or *four* if Lotka's law is included (since that can be equally well described by the Bradford law)—which permeate many other social contexts apart from those in which they were first found, must play a fundamental role in the development of the Z-type statistical theory needed for the social sciences. These four laws—or should it not be this *one* law?—appear to be as fundamental to Z-type statistics as the Gaussian distribution is to the analysis of the physical domain.

9. THE BRADFORD NUCLEUS

To emphasize Bradford's contribution to the present issues, I have to repeat (once again!) what Bradford actually wrote:[14]

> '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 numbers of articles as the nucleus, when the number of periodicals in the nucleus and succeeding zones will be as $1 : n : n^2 : \ldots\ldots$'

In this statement, Bradford introduced the term *nucleus* and centred his formulation on it. He explained what he meant by the term and related it to the periodicals outside it. He also drew a diagram in which the nucleus is clearly marked and which exactly conforms with the statement. It therefore seems to me that the concept of the *nucleus* is central to Bradford's formulation of his law.

I have found the pattern Bradford described in all the bibliographies I have examined except one possibly—the well-known ORSA bibliography in which the nucleus reduces to the single specialist journal in its field—and as I found and reported many similar patterns in empirical distributions drawn from other social contexts, I find Bradford's concept of the *nucleus* to be essential to the understanding of the statistical patterns of the distributions which arise from the social sciences.

Bradford also made it clear that he is *ranking* the periodicals 'in order of decreasing productivity of articles on a given subject' and thus describing his law as what I now call a *frequency-rank* distribution. I am therefore disturbed when, as often happens, I see in the literature Bradford's law reduced to the form $f(m) = k/m^2$ with no mention of the nucleus. This form is an approximation to the Laplace *frequency* distribution but is often confidently cited as 'the *Bradford* distribution'. In this crude approximation the rich nucleus has been despatched to the tail of the frequency distribution where it is reduced to an untidy random scatter of detached fragments which, it is erroneously assumed, can be safely ignored.

This unscholarly misnaming of well-defined empirical laws is not only deplorable in itself but it has also deflected research effort from the detailed analyses required to clarify the Z-type distributions of the social sciences into premature and futile searches for distributions of high generality. I call this the 'one true law syndrome'—the obsessive pursuit of a formula which can capture all the statistical manifestations and all the variants of the simple regularities revealed by Bradford, Zipf and others in one comprehensive embrace. The basic and serenely unquestioned assumption of all this work is that the empirical data sets derived from diverse sources are *homogeneous,* i.e. that in Bradford's terms they have no nucleus.

As Bradford's carefully worded statement of his law implicitly recognized the fact that the proportion of the nucleus to the periphery is variable, and as distributions can even be *multiply* composite, in all kinds of proportions, it is not surprising that the thirty-year search for the 'one true law' has led to heated disputes at times but to no convincing discovery.

Thus Mandelbrot[3] derived a formulation of Zipf's law, $g(r) = k/(r + w)^v$, from arguments based on information theory. Mandelbrot's formula conforms with eqn (6) derived from the Laplace law except for his additional parameter v which is needed only for the attempted capture of the whole family of composite distributions.

Herbert Simon,[4] a Nobel laureate in economics, developed a stochastic model of vocabulary usage which could assume several forms but he, too, sought to relate his models to homogeneous sets. One of Simon's models was introduced into bibliometrics by Derek de Solla Price,[15] who called it the Cumulative Advantage Distribution and applied it to large-scale sets of citation data.

Then Simon and Mandelbrot engaged in disputation. With hindsight it can now be seen that Mandelbrot, working with *frequency-rank* distributions and empirical data, and Simon, working with *frequency* distributions and computer models, had different objectives and were never quite on 'the same wave-length'.

As orthodox statisticians, dependent on G-type statistics, became interested in the statistically peculiar form of vocabulary distributions, the proposed solutions became increasingly esoteric. Thus Sichel[16] proposed an ingenious formulation, expressed in terms of Bessel functions of the second order, which reduced to more familiar G-type distributions when its three parameters were given selected values. More recently, it has been proposed that the one true solution might emerge from the statistics of quantum mechanics.

Even Haitun, for whose clarification of the statistical problem I have the greatest respect, similarly seeks the one true solution but is forced to take note of what he calls 'rank distortion' to explain some misfits. I regard 'rank distortion' as a euphenism for 'misfit' and as an empiricist I have to say, as Gertrude Stein might have done, that a misfit is a misfit is a misfit.

My own approach has been to seek the simplest possible distribution which could serve as an 'analytical instrument' to sort out the complexities and subtleties of real data sets and to work on very simple examples, collecting the data myself from sources accessible to others who may be interested to check my analysis. The problem is not purely numerical—the question of goodness of fit—but also a matter of interpretation of the outcome of the analysis. Thus I have found, by applying the *Bradford* technique to vocabulary data, that vocabulary distributions, like bibliographies, have a nucleus but, not surprisingly, are *multiply* composite.[17]

The unsuccessful search for 'the one true law' expressible as a G-type distribution has at least provided circumstantial evidence in support of Haitun's conclusion that G-type statistics applied to Z-type data is 'inadmissible'.

10. BRADFORD ANALYSIS OF THE BOOK-INDEX DATA

Figure 1 shows the histogram of the Laplace *frequency* distribution of the Gellner book-index data; its skewness and long straggling tail are displayed. Figure 2 shows graphs of the Bradford *frequency-rank* distribution of the same data. The curvilinear graph is obtained by plotting the cumulated frequencies, G(r), against $\ln(1 + r/w)$ with $w = 1$; this graph is drawn as a means of finding the general shape of the graph, hoping that it will indicate what analysis might be useful. There is a possible discontinuity between $r\ (8) = 16$ and $r\ (9) = 13$. Does this dicontinuity mark the upper bound of the nucleus? It is possible that it does but we need to apply Wilkinson's technique[18] to evaluate the parameters w and k for each of the two truncated f/r distributions.

The equations derived by Wilkinson's technique are shown in Table 3, which also compares the calculated with the data frequencies. Towards the end of the peripheral component the calculated and data frequencies diverge consistently enough to invite the identification of a possible third component—but I do not pursue the possibility. The Bradford graph fitted to the nuclear and peripheral components become the two straight lines shown on the graph (Figure 3) as AB_1 and B_2C. The discontinuity between the two components is now evident.

The main point to be made in comparing the Bradford with the Laplace analysis of the same data is that the Bradford analysis, using *all* the statistical information in the data, is more fruitful than the Laplace distribution which begins by discarding more than 60% of it. The success of the χ^2 test implied that the Laplace *frequency* distribution is *homogeneous*: the Bradford *frequency-rank* analysis demonstrates that the eight most frequently cited names constitute a nuclear component which differs significantly in its statistical characteristics from the remainder of the index entries, and that therefore the data are *not* homogeneous.

The Laplace analysis therefore poses no questions: the data conform with expectations. But the Bradford analysis immediately poses the question: why does a *nucleus* appear in the context? The answer to that question can be found, not

TABLE 3. *The book-index data as a f/r distribution*

r = 1 to 123: r_1 = 1 to 8, r_2 = r − r_1 = 1 to 115
For calculated values of G(r): $G(r_1) = 251 \ln(1 + r_1/8)$ – the *nucleus*,
$G(r_2) = 110 \ln(1 + r_2/8{\cdot}5)$ – the *periphery*

r_1	1	2	3	4	5	6	7	8
$g(r_1)$	29	28	23	22	22	18	16	16
$G(r_1)$ data	29	57	80	102	124	142	158	174
$G(r_1)$ calc.	30	56	80	102	122	140	158	174

r_2	1	2	3	10	20	30	50	60	80	115
$g(r_2)$	13	12	11	6	4	3	2	1	1	1
$G(r_2)$ d.	13	25	36	86	135	170	212	227	247	282
$G(r_2)$ c.	12	23	33	86	133	166	212	230	258	294

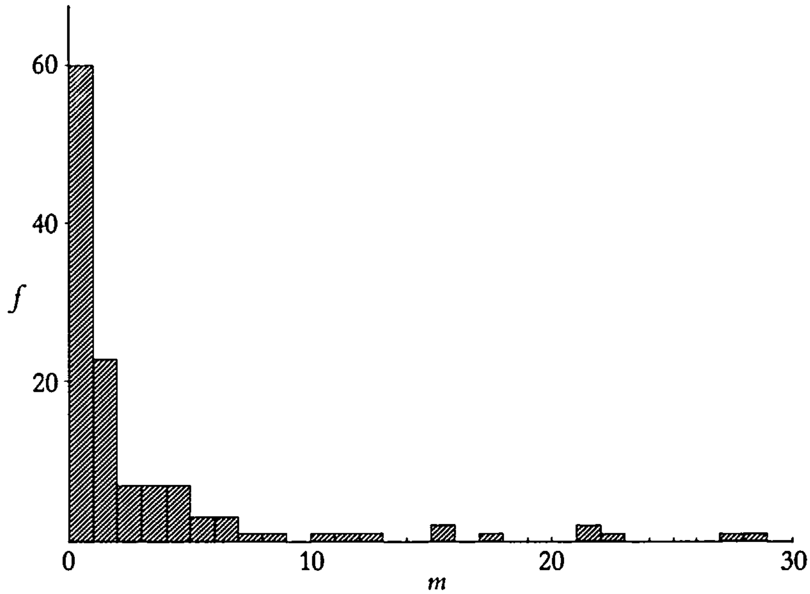

FIG. 1. *Book-index data: Laplace distribution*

by further statistical analysis, but only by examining the data, its method of collection and the source itself. And such work usually evokes further questions.

In the index itself, only *one* page reference is given if the specified name is mentioned more than once on that page. This 'page effect', as it might be called, bears more heavily on those names that appear most frequently than on those which occur only 1, 2, 3, . . . times throughout the whole book. On checking, for example, the frequency of the name *Bertrand Russell* or *Russell* in the pages of text (i.e. apart from the footnote references to his publications), I found that his original score of 16 would have had to be increased to 32. But then new counting decisions have to be made: does one count the pronoun 'he' when it refers to Russell or the pronoun 'I' when Russell refers to himself in a citation from his

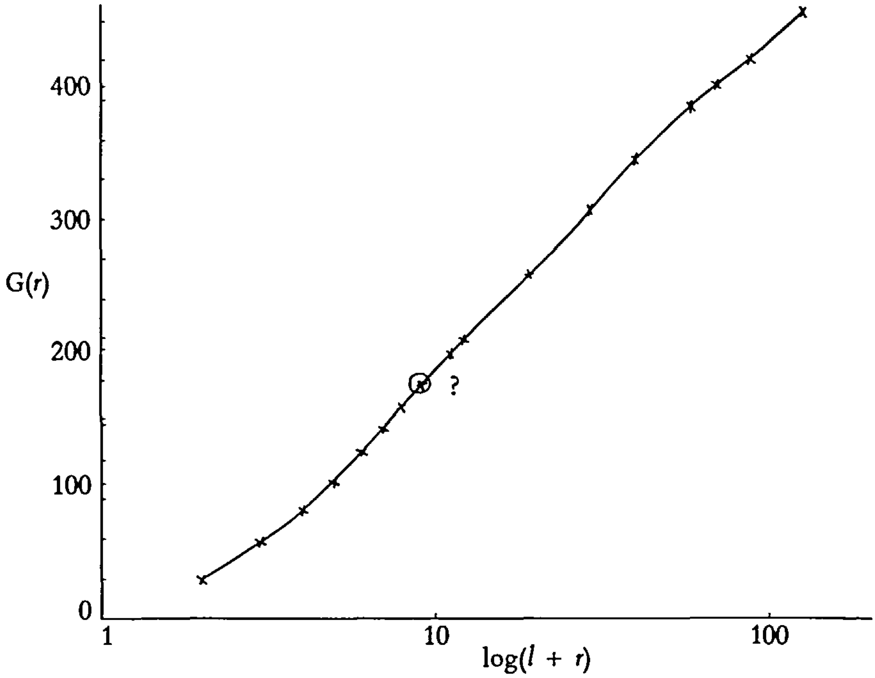

FIG. 2. *Book-index data: initial f/r distribution*

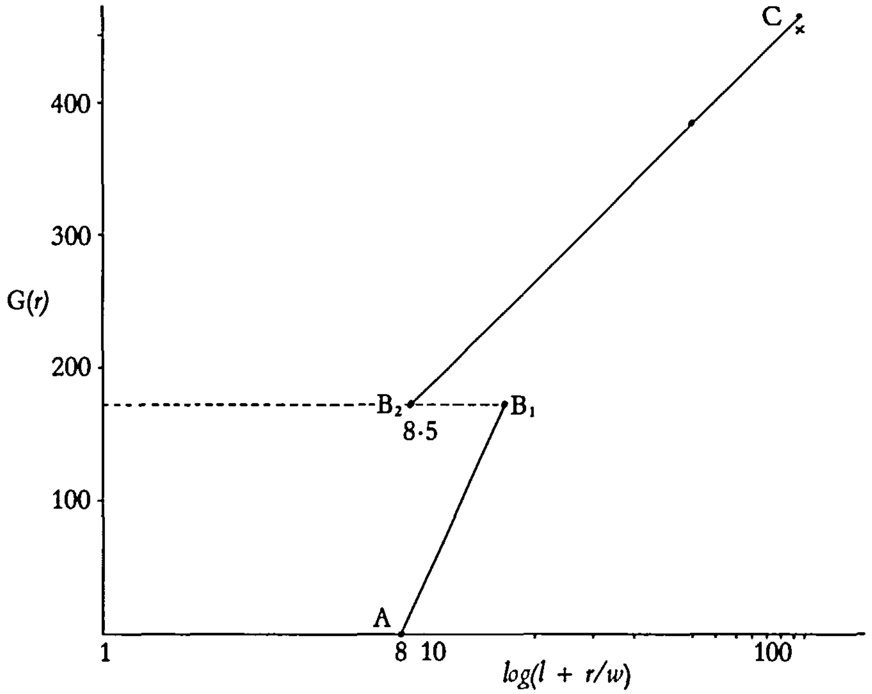

FIG. 3. *Book-index data: the Bradford nucleus and periphery*

works? Such counting, too, would bear most heavily on the most frequently cited authors.

But I do not here pursue these matters. All I am concerned to emphasize is that the book-index data were therefore *not* homogeneous. And though the 'page effect' is obvious as soon as one sees a nucleus emerge, the Laplace analysis did not even reveal that this effect was present though it was strong enough to disturb the homogeneity significantly.

The intriguing aspect of this analysis is, nevertheless, that the Laplace distribution is so abstract and general that it must apply to all *simple one-level* categorizations. Here it was applied to an index of names only; but it applies to concept or subject indexes if they are indexed at only *one* level of categorization also. Theoretically, it should not matter how many levels of categorization are used in an index since all levels should conform with the Laplace law. The theory, however, assumes that all distributions extend without limit, that all subcategorizations are completed and that no 'page effects' or other such limitations are present. The totality of all such effects quickly obscures the simplicities of the primary categorization when Laplace analysis is applied to very large and complex indexes.

11. NATURAL CATEGORIZATION

As an author writes his manuscript he does not consciously scatter the names or concepts he mentions in any pre-conceived pattern. Nor, as he writes, does he select words from his personal vocabulary except to say as precisely as he can what he intends to say. Yet, in both cases, in a long enough text, the distributions of both names or topics mentioned, or of words used, conform with the Laplace/Bradford/Zipf distribution. These effects are wholly unplanned.

The derivation of the Laplace distribution in Section 4 is dubious not only because of its dependence on the principle of indifference but also because it is unrealistic: it describes a *deliberate* process of categorization. In real life we make new discriminations, one by one, only when it helps to do so. But there are exceptions. The classifications adopted by librarians and the categorizations adopted by administrators to bring order to their respective domains are rational schemes worked out *a priori* and then *imposed* on the entities concerned. If we look for exemplifications of the Laplace law in contrived situations of that kind, it is most unlikely that they will be found.

The Laplace/Bradford/Zipf distribution therefore lies deeper. It describes some fundamental aspect of the way in which we mentally organize the world around us for the purposes of social discourse. It describes the patterns of the distinguishable elements used in that discourse—the words, the topics, the names, the numbers, the papers, the journals, . . . and many others. The fact that such patterns, all similar, emerge from so many facets of social life, but that they are never found in the physical world, must derive, it seems to me, not from the physical world itself but only from our egocentric subjective views of it.

12. RELATIONS BETWEEN THE LAPLACE AND BRADFORD DISTRIBUTIONS

The inverse square formulation of the Laplace distribution, $f(m) = k/m^2$, is very simple; it invites obvious transformations.

For example, the cumulative sum of the first moment about the origin is

$$F(m) = \int_{m}^{m+1} \frac{k}{x} dx = k \ln(m + 1), m = 1, 2, 3 \ldots .$$

This is similar to the Bradford law except that here the *frequency* classes rather than the *ranked* frequencies are the basic elements. But this function, too, should give a linear graph when plotted with *m* on a log scale. An advantage of this form is that it does give due weight to the remote high frequencies in the tail.

When the book-index data of Table 1 are plotted in this way, the erratic graph of Figure 4 does not show the single linearity expected from a homogeneous distribution but *two* possible straight lines AB and BC. The upper straight line BC corresponds with the eight terms of highest frequencies; it is therefore the frequency analogue of the Bradford nucleus. Unfortunately, this technique is not easy to apply in analysis because in the graph the zero class frequencies make the values of F(m) jump very erratically as *m* increases. But the reappearance of the Bradford nucleus in this simpler graph emphasizes its reality and can be of help in directing the more detailed Bradford analysis.

In a previous paper published in this Journal,[9] I applied to the analysis of the ORSA bibliography a frequency distribution which I had derived from the Bradford law. Its formulation (in my current notation) was

$$f(m) = \frac{k}{j+m-1} - \frac{k}{j+m}, m = 1, 2, 3, \ldots \ldots \qquad \ldots \ldots \tag{7}$$

so that

$$F(m) = \frac{k}{j} - \frac{k}{j+m} \ldots \ldots \tag{8}$$

where *j* is a second parameter whose value must be greater than zero. These equations revert to the simple Laplace form when $j = 1$. This form gives a wider generality to the Laplace distribution which has been of use in the analysis of composite forms in which truncated distributions arise.

There are other variants and transformations of the Laplace and Bradford laws, some of which are surprising and of interest, and the fact that the two laws transform into each other exactly, though with the data in reversed order—the 'tail' of one becoming the 'head' of the other—has made it possible to explore some problems in the tails of both of them. But these matters take space to explain and exemplify so I shall be reporting them more fully elsewhere.

13. THE BRADFORD LAW AND HOUSEHOLD INCOMES

I have claimed that the Laplace/Bradford/Zipf laws and their variants permeate social statistics (when one looks for them) but in this paper I have demonstrated their application only to a small sample of book-index data. Another example of more general interest is provided by the published data on household incomes in the UK. The 'household' is not an ideal statistical entity because it varies widely in number; 'household income' can apply to the lonely pensioner living by himself or to a family of ten or more members, some of whom may be supplementing the household income provided by the head of the family. Nevertheless, the

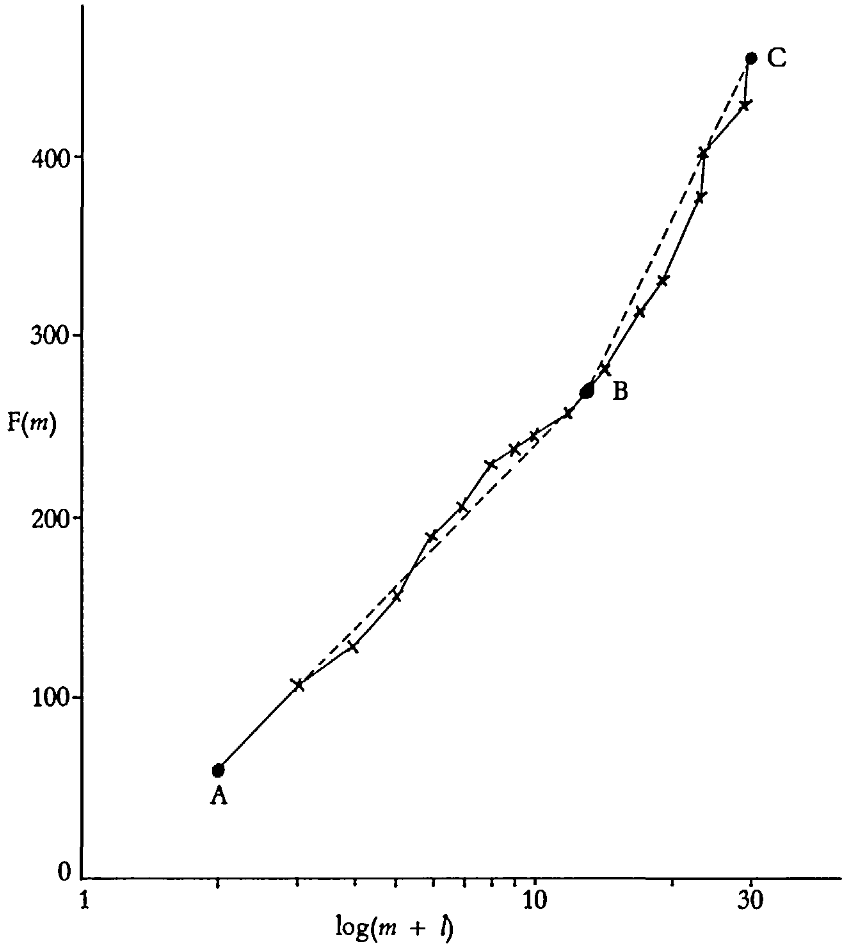

FIG. 4. *Book-index data: cumulated first moment of Laplace distribution*

L/B/Z laws are robust; they can accommodate large deviations from the statistical ideal of homogeneity.

Though 'household income' is a concept that is generally understood, some appreciation of the range of possibilities must temper interpretation of the published data. For example, a high proportion of the 'rich' households are likely to consist of relatively large households as compared with the 'poor' households which will include single pensioners in relatively high proportion.

The published data on this subject are rather sparse and are not well adapted to Bradford analysis. However, one form of presentation includes a *ranking* of household incomes, beginning with the poorest; the incomes are then cumulated successively. The graph of such data produces a curve which rises slowly at first and then ever more rapidly as the richer incomes are included in the total. The data are then reduced to percentages—to percentages of households and to percentages of the total income for the whole of the UK.

A typical graph of this kind is shown in Figure 5. To measure the effects of the redistribution of income arising from taxation on the one hand and social security on the other, the Gini coefficient of concentration is used. If all incomes were exactly equal, the resulting graph would be the diagonal OA. The Gini coefficient measures the difference of area (as a proportion of the triangle below the diagonal) between the graph and the diagonal. The Gini coefficient would therefore be zero only if all household incomes were exactly equal.

The data of Table 4 are derived from the published data for 1980[19] and Figure 5 shows the usual graphs of original and final incomes. The two Gini coefficients are 45·9% for the *original* incomes and 32·3% for the *final* incomes; the redistribution thus improves the original inequalities.

To see whether such data conform with the Bradford law I have to reverse the published data so that the Bradford graph begins with the richer household incomes. I then have to find the values of the two parameters needed, w and k, before I can draw the graph of

$$I\% = k \ln(1 + bd)/, \ d = 1, 2, 4, 6, 8, 9, 10.$$

where I% is the cumulative sum of the incomes expressed as a percentage of the total for the UK corresponding to the deciles 1, 2, 4, 6, 8, 9 and 10, the only data provided by the official statistics, which represent the first 10%, 20%, 40% . . . and so on of the total number of households. In the formula $b = 1/w$, the usual Bradford parameter, for convenience in this case.

The value of b has to be found, rather clumsily, from the fact that at the end of the second decile the proportion of total income is 38·8%. We therefore have to solve the equation

$$\ln(1 + 2b) = 0{\cdot}388 \ln(1 + 10b)$$

which gives $b = 0{\cdot}506$. Since $k \ln(1 + 10b) = 100$, we also have $k = 55{\cdot}5$.

In drawing the graph, the first step is to mark off $1 + b, 1 + 2b, \ldots 1 + 10b$ along the log scale of the base. The percentages of total income are marked off along a linear scale for the ordinates.

TABLE 4. *Distribution of household incomes in the UK (1980)*

Decile	*Original*	*cum.*	*Final*	*cum.*	*Final Bradford*	*cum. calc.*
1	27·7	—	23·2	—	22·7	—
2	17·8	45·5	15·4	38·8*	16·1	38·8*
4	26·9	72·4	24·1	62·9	22·6	61·4
6	18·6	91·0	18·0	80·9	16·0	77·4
8	8·5	99·5	12·3	93·2	12·5	89·9
9	0·5	100·0	4·1	97·3	5·3	95·2
10	0·0	100·0	2·7	100·0	4·8	100·0

*Datum points

The Bradford equation $I = 55{\cdot}5 \ln(1 + 0{\cdot}506d)$ is drawn in Figure 6 as the dotted diagonal OA and the *final* data points are marked. They fit the Bradford equation well at first but deviate from it at the upper end to the relative disadvan-

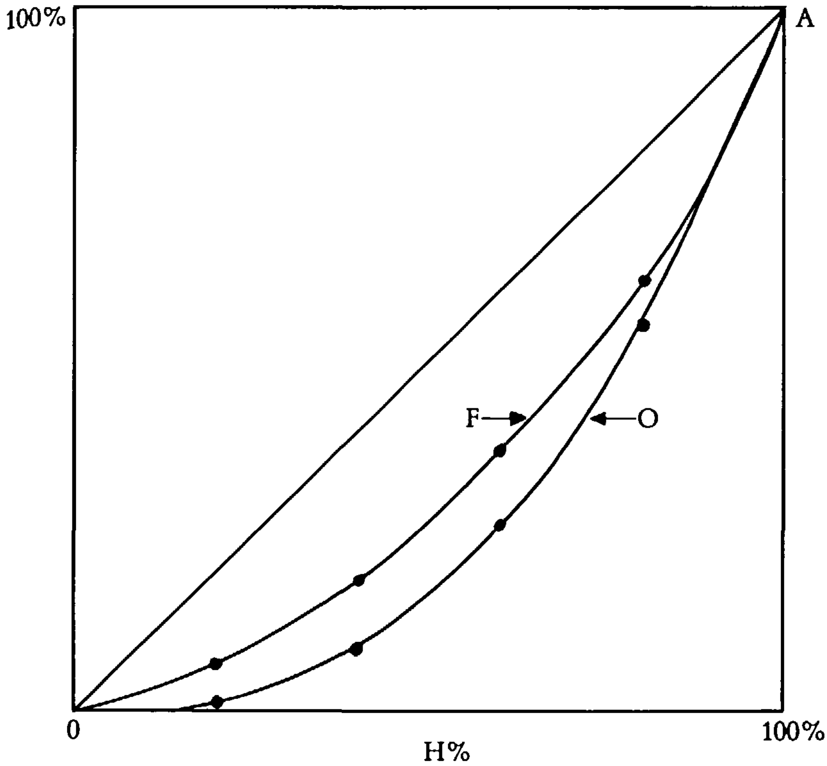

FIG. 5. *Household income data: conventional graph of original and final incomes*

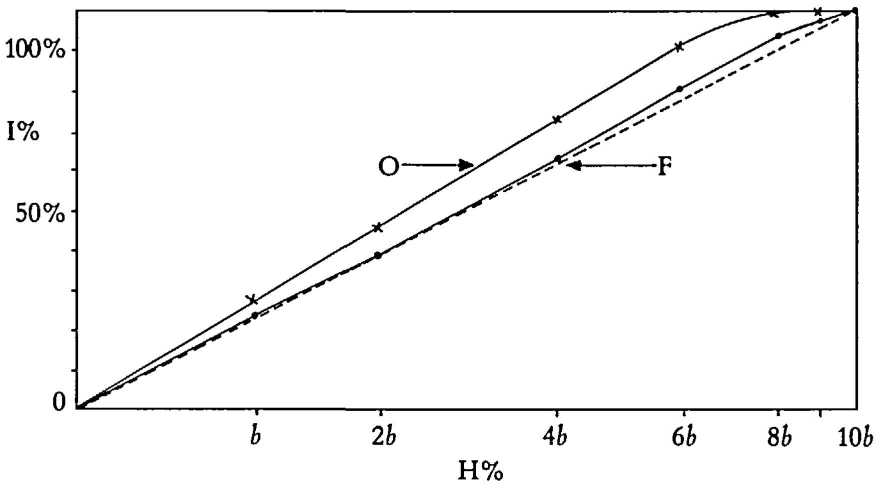

FIG. 6. *Household income data: Bradford graphs of original and final incomes*

tage of the poorer households. The second graph is that of the *original* incomes using the same value of *b*. This graph displays a nucleus. It conforms with the Bradford equation $I = 65{\cdot}1 \ln(1 + 0{\cdot}506d)$ up to $d = 6$ but then deviates as the poorer incomes are reached.

It seems as though those who manage these matters unconsciously regard the Bradford law as providing a norm or 'natural' objective towards which redistribution of incomes should conform, even though the poorest households do not get their Bradford shares in this case. In an early paper,[10] I described the Bradford law as showing a 'success-breeding-success' social mechanism. I am not surprised to find that incomes conform with the law but cannot commend it as a social ideal. Yet I have to admit that it describes competitive social affairs realistically. A stronger redistribution of incomes would be indicated by a smaller value of the parameter *b*.

The prevalence of such 'log law' effects on social affairs implies that we measure our social differences in terms of *ratios* rather than of the arithmetical differences we use in practical work with physical objects. Thus it is now conventional to demand higher incomes in terms of *percentages* of the present incomes. Such demands show that the 'log' scale of Z-type statistics has been implicitly adopted by those concerned.

14. PHYSICAL V. COGNITIVE PROBLEMS IN THE ALVEY PROGRAMME

The physicalist approach has been very successful in exploring the physical world because subjective observations of it can be directly related to physical entities in the same physical space we all share. But applied to the cognitive world this approach has been less successful because descriptions of purely abstract matters relate to entities observable only in different mental subjective spaces. It is thus more difficult to ensure that our several mental images are congruent—that we are talking 'about the same thing'—especially if the mental images have only very remote or no discernible physical correlate whatever. We cannot kick a mental entity and say: '*That* is what I am talking about'.

The first analytical approach to this problem was to reduce matters of cognitive interest to their *physical* correlates and to talk of them as though they were the *cognitive* entities of interest. Thus what was called an 'information' need was regarded as satisfied if *documents* believed to be relevant to that need were provided. One advantage of this approach was that documents are also physical objects which could be kicked if they were not relevant.

It was therefore not unreasonable to assume that as documents are physical objects they could be dealt with by what we can now call G-type statistics—the calculus of the physical domain. When the next problems arose at the level of the index terms assigned to scientific papers collected in a database, the same calculus was applied to index terms, i.e. to *words*. Index terms expressed as printed words or as coded signals within the computer are also physical correlates of cognitive entities. So the problems of information retrieval from scientific databases were analysed by G-type statistics also.

What Haitun has shown is that processes involving cognitive elements such as documents and words throw up distributions which lie beyond the analytical reach of G-type statistics but which do conform with Z-type statistics. The Bradford law is a typical Z-type distribution. But the Bradford law, too, relates to physical entities, i.e. to documents and papers. But the documents of a bibliography are not selected at random from a large population of documents; they are related by the fact that that they share the property of *relevance* to some specified topic. And *relevance* is not a physical entity or relationship but a reflection of a cognitive grouping of things for attention. Is it this mental grouping of physical entities which modifies G-statistics into Z-statistics for analysis of information problems?

I believe it to be helpful, as Sir Karl Popper has proposed,[20] to regard mental events as occurring in a world different from that of the physical world. As Z-type statistics also conform with Shannon information theory, there appears to be an *information* space different from physical space. In an earlier paper I likened our individual information spaces to *landscape* views and described the sciences as the systematic transformation of our subjective landscape views of the world around us into objective *maps* we all share if we wish to correct the subjective distortions that perspective imposes on us.

These ideas make me a dualist, but that is not a popular position to adopt in the present world of confident high technology. This issue, however, will soon be resolved *empirically*.

The world is awash with information, readily accessible but incoherent. Much of it is of no concern to most of us because we cannot make personal use of it; information of no immediate use is better described as *noise*. A problem of increasing importance is to select from all this noise the information that concerns us and to group it as efficiently as possible into coherent knowledge of use.

This idea seems to be the target of the Alvey programme.[21] An 'intelligent knowledge-based system' should be capable of selecting from the bases to which it has access that information which can be grouped together and organized into coherent knowledge directly presented to its users. On the one hand we therefore have the highly sophisticated physical/G-type/cardinal apparatus of modern technology applied to the design of machines intended to operate with the primitive cognitive/Z-type/ordinal apparatus which applies to the many particulars of human knowledge.

Present machines are purely physical: they operate on physically generated coded pulses. Information is not a physical entity; it is we humans who endow the inputs and outputs of computers with the information they are said to process. The next stage of technological development in this field is therefore a direct confrontation between the physical and the cognitive worlds in the design of systems which simulate human thought. I have no doubt that it will be confidently assumed that the G-type calculus which produces the machines will be equally effective when brought to bear on the Z-type phenomena of human discourse on which the machines are designed to operate.

I await the outcome of this confrontation with keen interest.

15. THE FUTURE ROLE OF THE BRADFORD LAW

It is exactly fifty years since Bradford first adumbrated his law in 1934 but only two years since Haitun clarified the relationship of 'Zipfian' distributions to modern statistical theory and showed the need of a wholly new statistical theory for the social sciences.

In the development of that theory the Bradford law will, I believe, have a central role. The frequency-rank distribution Bradford introduced offers a more penetrating analysis of social affairs than the frequency groupings of conventional statistics. The Bradford law is simpler to use as an exploratory instrument than any of its possible rivals. And Bradford's concept of the nucleus warns against assumptions of homogeneity.

There are some problems which need attention before a Z-type statistics can be firmly based: the present statistical convention that equates the ordinal *nth* to the cardinal number *n* is not well-founded for general use and needs to be clarified and formalized; and the fact that Z-type distributions cannot be derived from classical probability theory points to some serious fault in that theory.

So the saga of the Bradford law continues, but now on a different plane.

ACKNOWLEDGEMENT

I gratefully acknowledge the support of a Leverhulme Emeritus Research Fellowship while preparing a book which expands ideas outlined in this paper.

REFERENCES

1. HAITUN, S. D. Stationary scientometric distributions. Part I. The different approximations. *Scientometrics, 4*(1), 1982, 5–25. Part II. Non-Gaussian nature of scientific activities. *Scientometrics, 4*(2), 1982, 89–104. Part III. The role of the Zipf distribution. *Scientometrics, 4*(3), 1982, 181–94.
2. ZIPF, G. K. *Psycho-biology of language.* Houghton Mifflin, 1935.
3. MANDELBROT, B. An information theory of the statistical structure of language. *Proceedings of the symposium on the applications of communication theory. London, 1952.* Butterworths, 1953, pp. 486–500.
4. SIMON, H. A. On a class of skew distribution functions. *Biometrika, 44,* 1955, 435–40.
5. BROOKES, B. C. People versus particles. *Theory and applications of information research. Proceedings of the Second International Research Forum on Information Science. Copenhagen, 1977.* London: Mansell, 1980, pp. 106–18.
6. KNEALE, W. *Probability and induction.* Oxford: Clarendon Press, 1949.
7. KENDALL, M. G. The bibliography of operational research. *Operations Research Quarterly, 2,* 1960, 21–36.
8. LEIMKUHLER, F. F. The Bradford distribution. *Journal of Documentation, 23* (3), 1967, 197–207.
9. BROOKES, B.C. A critical commentary on Leimkuhler's 'exact' formulation of the Bradford law. *Journal of Documentation, 37*(2), 1981, 77–88.
10. BROOKES, B. C. Bradford's law and the bibliography of science. *Nature, 224,* 1969, 953–5.
11. LOTKA, A. J. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences. 10,* 1926, 217–23.
12. KENDALL, M. G. *The advanced theory of statistics.* Vol. 1. Fifth ed. London: Griffin, 1952.
13. LEIMKUHLER, F. F. An exact formulation of Bradford's law. *Journal of Documentation, 36*(4), 1980, 285–92.
14. BRADFORD, S. C. *Documentation.* Crosby Lockwood, 1948.
15. PRICE, D. DE SOLLA. A general theory of bibliometrics and other cumulative advantage distributions. *Journal of the American Society for Information Science, 27*(2), 1976, 292–307.
16. SICHEL, H. S. On a distribution law for word frequencies. *Journal of the American Statistical Association, 70*(352), 1975, 542–47.
17. BROOKES, B. C. Quantitative analysis in the humanities: the advantage of ranking techniques in *Quantitative Linguistics,* 16. *In:* GUITER, H. *and* APAROV, M. V., *eds. Studies in Zipf's law.* Bochum: Brockmeyer, 1982, pp. 65–115.

18. WILKINSON, E. M. The ambiguity of Bradford's law. *Journal of Documentation, 28*(2), 1972, 122–30.
19. ECONOMIC TRENDS. London: HMSO, January 1983, p. 103.
20. POPPER, K. R. *Objective knowledge: an evolutionary approach.* Oxford: Clarendon Press, 1972.
21. *A programme for advanced information technology: the Report of the Alvey Committee.* London: HMSO, 1982.