No. 342

Rising Tail in Bradford Distribution: Its Interpretation and Application

by

Hajime Eto

October 1987

Rising Tail in Bradford Distribution : Its Interpretation and Application

Abstract
The right tail of the Bradford distribution has been considered to be straight or drooping. This paper reports the cases in which the right tail is rising upward, explains and verifies conditions of its occurrences, interpretes it and proposes its application to evaluation and forecasting of technological development at the basic research stage.

Key Words : Bradford distribution; Diffusion of technology; Cumulative advantage ; Technological forecasting; Basic research; Evolving technology ; Evaluation of firms

§1. Introduction

In the latest two decades a number of works have been devoted to the Bradford distribution as a distribution of scientific informations. In 1934 and 1948 Bradford empirically discovered what is now called the Bradford law or the Bradford distribution of scientific articles on a topic across journals [Elvin 1978]. The Bradford distribution has usually been interpreted as a result of diffusion of a topic across journals (referred to as the diffusion interpretation hereafter) with the cumulative advantage effect [e.g., Brookes 1969B ; de Sola Price 1976].

The Bradford distribution may be described as follows. Journals are ranked in the decreasing order of the number of their articles on a topic and their ranks are arranged along the horizontal axis in the common logarithmic scale. The cumulated number of articles is arranged along the vertical axis. Thus the coordinates is semilogarithmic (semilog) with the logarithmic horizontal axis. The cumulated numbers for the middle ranked journals lie on a straight line in the rank semilog coordinates. Those for the high ranked journals are usually above the line and those for low ranked journals are either on the line or below it. The section of the graph for high ranked journals which is above the line (cases in which it is below the line have always been neglected) is called the core or nucleous. The section of the graph for low ranked journals, if it is below the line, is called the droop. In mathematical terminilogy, the core section of the graph is convex, the middle section is linear and the droop section (if exists) is concave in the rank semilog coordinates. The graph with the three sections was called standard [Eto and Candelaria 1987] while the graph without droop will be called classic in sequel. It is mathematically proved [Egghe 1984 ; Eto and Candelaria 1987] that the rank semilog linearity is necessary and sufficient for the geometric zoning which was the principal law verbally stated by Bradford himself.

The original description by Bradford himself was verbal and graphic. Afterwards mathematical formulations have been attempted. The interest was first in representing the rank semilog linearity, i.e., the middle section alone [e.g., Kendall 1960 ; Brookes 1968 ; Brookes 1969A]. Then formulas were presented for the convex curve rapidly approaching to the linearity in the rank semilog coordinates to represent both the rank semilog linearity in the middle and the convexity in the core sections uniformly in a single formula [e.g., Leimkuhler 1967 ; Asai 1981] or separately in two formulas to capture their subtle characteristics [Brookes 1969B]. Truncating the core and representing its characteristics by a parameter, a unified single formula in the conventional size-frequency form represents the middle and low productive sections which correspond to the middle and low ranked journals in the Bradford rank form [Morse and Leimkuhler 1979 ; Morse 1981]. The breakdown of graph into zones (will be referred to as the zone index approach) gives a single formula with a parameter of zones, generating the piecewise semilog linearity for each zone [Maia and Maia 1984]. A single formula can represent all the three sections in a flexible way [Egghe 1985] but still may not be enough to represent a vareity of the graph. So far, only the breakdown of graph into each interval between consecutive ranks (will be referred to as the rank index approach) represents every variant of the graph [Praunlich and Kroll 1978 ; Eto and Candelaria 1987 ; Chen and Leimkuhler 1987]. It may be right to say that no unified single expression is successful in representing all the three sections so far [Brookes 1981] and to conjecture that no single formula can embrace the Bradford phenomena [Brookes 1979].

It seems that Bradford paid no attention to the droop. The first attention was paid in 1967 by Groos to what is called the droop. The droop has been interpreted to indicate that the collection of articles is incomplete [Brookes 1968 ; Brookes 1969A ; Brookes 1969B ; Aiyepeku 1977] despite a counter argument against this incompleteness interpretation [Drott and Griffth 1978 ; Drott 1981]. The diffusion interpretation combined with the incompleteness interpretation of the Bradford distribution leads to a conjecture that the complete collection of articles on an immatured topic tends to result in a graph with a droop because the complete collection on an insufficiently diffused topic may give the same effect as the incomplete collection of a sufficiently diffused one. The conjecture was verified at least for cases of distribution across countries : For countries instead of journals, the distribution of the number of articles across countries was found to have an extremely heavy droop for developing countries, indicating the slow diffusion of R&D activities to the developing countries [Eto and Candelaria 1987]. The diffusion interpretation on which our discussion just above is based may be defended against the counter argument above by Drott and his coworker as follows : Their counter argument is based on the truncated data that the top four journals and all of the one-article journals are truncated while many discussions on the Bradford distribution including ours are interested in the both extremes.

Mathematical conditions were explicitly presented that the section of the graph associated with the low ranked journals is not concave but, at least partly (locally), convex [Eto and Candelaria 1987 ; Chen and Leimkuhler 1987]. Such a convex subsection of the droop had been known to exist [Brookes 1973 ; Summers 1983] but no particular interpretation has been attempted. The average slope of a convex subsection of the droop is available from a mathematical formula of the droop [Praunlich and Kroll 1978] by replacing their "concavity" with our "convexity". The convex subsection can occur when the number of journals with the same or the nearly same number of articles is highly numerous because the horizontal axis is logarithmic [Praunlich and Kroll 1978 ; Eto and Candelaria 1987]. However, as the total number of scientific journals is limited, many journals rarely have the (nearly) same number of articles in practice. Therefore the convex subsection is usually small enough to neglect within a reasonable allowance of errors [Eto and Candelaria 1987]. Yet theoretically, the innegligible convexity is possible. Extending this possibility leads to another possibility that the section of the graph associated with the low ranked journals does not droop down at all but wholly rises up above the line, or mathematically, that the whole graph is convex (this phenomenon will be referred to as the rising tail in sequel). A case of slightly rising tail was observed and a mathematical formula was presented for it [Maia and Maia 1984], but it may be within an allowance of errors. Mathematically, a special case of the Lotka's law generates the wholly convex graph but this is considered to be quite rare [Egghe 1985]. This paper reports a number of real cases of the aloftly rising tail, identifies a condition of its occurrence, interprets it and presents its application.

This paper especially examines the applicability of the phenomenon of rising tail to technological forecasting and the evaluation of technological performances of firms in immatured technology fields. The existing methods for technological forecasting such as the extrapolation method, the envelop curve method, the Delphi method, the technological substitution theory and so forth [e.g., Bright 1968 ; Jones and Twiss 1978 ; Martino 1983] are mainly for technologies at the development or even production stages while newly evolving technologies at the basic research (BR) stage are important today.

The section 2 discusses under what condition the rising tail occurs and interprets it. The section 3 examines whether the rising tail merely indicates the invalidity to apply the Bradford distribution to our data. The section 4 presents a method to apply the phenomenon of rising tail to technological forecasting and evaluation of technological performance of firms at the BR stage. The section 5 states quantitative relations between parameters of the Bradford distribution. Finally the section 6 states the conclusion.

§2. Rising Tail and its Interpretation

The graph of the Bradford distribution for low ranked journals is locally convex where many journals have the same number of articles. As the total number of scientific journals is limited, this rarely occurs (precisely, it occurs but within an allowance of errors). But it is possible that middle ranked journals do not have many articles and that relatively many journals have one or two articles on a topic. The number of one- or two-articles journals can be highly numerous as compared to the total number of journals when a hitherto highly specialized topic for which relavely small number of journals have already several contributors of articles suddenly becomes diffused simultaneously to many journals which have thus far almost no contributor of articles on the topic. In such an explosive and young diffusion many latecoming or peripheral journals have only one or two articles while the total number of relevant journals and the number of articles by middle ranked journals remain small yet (Appendix 1).

Indeed the rising tail is observed for the topic of Expert system from the scientific database COMPENDX. A couple of journals specialized in computer science form the core on this topic followed by several middle ranked journals with relatively small numbers of articles while many journals in mechanical engineering, civil engineering, control engineering, chemical engineering and so forth have recently started to publish one or two articles on the topic. The long-lasting technological obstacles of this topic have kept the number of specialists, therefore the total number of relevant journals and the numbers of articles by middle ranked journals at low levels. Similar phenomena are observed for similar topics of Artificial intelligence, Knowledge database, Knowledge engineering from relevant scientific databeses.

The most extreme case is seen for the topic of Ciper from SSCI and SCIO-SCI from 1961 to 1986 which list 28 different articles on this topic (the first article was published in 1970). Only one journal has two articles and the other 26 articles are published by 26 different journals (one article by one journal). Naturally its Bradford distribution graph is wholly convex. Less extreme but similar phenomena are seen for the nearly synonymous topics of Authentication, Cryptology, Cryptogra and Cryptosystem all from SSCI and SCIO-SCI from 1960 to 1986. All of these topics are related to the recent concern with the information security or the privacy protection in information systems. For example of Authentication, only one journal has six, five, four and three articles, respectively, and only two journals has two articles while 18 journals has one article each. In other words, 18 articles out of 40 articles are published by 18 different journals though only 24 journals are relevant. Naturally its Bradford distribution graph is wholly convex.

The above theoretical discussions and empirical facts lead to the following proposition.

Proposition 1 (Sudden and simultaneous participation of journals). The rising tail tends to occur for topics on which articles suddenly started to appear in many journals which have published almost no article on the topics till recently while the total number of relevant journals and the numbers of articles by middle ranked journals have been kept at relatively low levels.

The idea underlying the Bradford distribution may be applicable to other items of science and technology. The scientific development is associated with the diffusion of knowledge among scientists [e.g., de Sola Price 1965], and the technological development process is often associated with the diffusion [Mahajan and Wind 1986 ; Homer 1987]. Further, the scientific and technological diffusion can be modeled in association with the cumulative advantage effect, and this model fits the real data of Japanese industrial R&D [Eto and Makino 1983A], world technologies [Eto and Makino 1983B] and regional high technologies [Eto and Fujita 1987]. Theoretically, the Bradford distribution is related to the cumulative advantage effect via its "reproducibility" [Morse 1981] and via its asymptotic approach to the Yule distribution [kendall 1960] which is a steady state distribution of a cumulative advantage process [Ijiri and Simon 1977].

The R&D activity level of firms can especially be considered as a result of cumulative advantage process because, once R&D becomes highly active in a firm, the firm attracts R&D personnels and the manager tends to feel confident of the effectiveness of R&D investment in competition with other firms [Eto and Makino 1983A]. It is also considered as a result of diffusion process because many firms had made no R&D in old days and their R&D was diffused from technologically leading firms. Thus it is expected to follow the Bradford distribution. In fact the Bradford distribution holds for R&D input in Japanese firms with very heavy droops, reflecting a young history of R&D in Japanese firms [Eto 1984]. Besides the diffusion and the cumulative advantage effect interpretations both of which hold also for natural events, another aspect may need to be added as far as the Bradford distribution is claimed to be inherent to human phenomena [Haitun 1982A ; Haitun 1982B]. A claim that the Bradford distribution is associated with competitive events [Brookes 1979] (referred to as the competition interpretation of the Bradford distribution in sequel) may be accptable in this context. As R&D in firms is now in a severe competition, the Bradford distribution may reasonablly be applicable to it as a combination of diffusion, cumulative advantage and competition interpretations (abbreviated as the DCC interpretations). Extending Proposition 1 to other items in science and technology, the following hypothesis may be claimed on articles by scientists in firms.

Hypothesis 1 (Sudden and simultaneous participation of firms). The rising tail tends to occur for topics on which many firms recently started their R&D while the total number of interested firms and the numbers of articles by middle ranked firms have been kept at relatively low levels.

To verify Hypothesis 1, articles on 52 evolving technological topics in various ranges are taken from BA (Biological Abstract), CANCER (Cancernet) and PASCAL (Pascal Oncology) as the successor of CANCER, COMPENDX (Computerized Engineering Index), EM (Excepta Medica), ENERGY (Energyline), ENV (Enviroline), IPA (International Pharmaceutical Abstract) and POLL (Pollution Abstract). The topics and ranges are selected on the following indisjoint criteria : (1) Observations are limited to articles by scientists in Japanese firms for many topics (41 out of 52 topics) because Japanese industrial R&D is relatively young except for leading firms and many firms started R&D recently ; (2) Topics are mainly in biotechnology (42 out of 52) because it is young especially in Japanese inustrial R&D except for pharmaceutical firms and many firms outside the parmaceutical sector started R&D in this field recently. The two selection criteria both satisfy the condition for Hypothesis 1. Another selection criterion is to see the stability of distribution shapes for a variety of databases (data on the same topic from different databases), regions (Japanese firms vs the world firms on the same topic), years (shorter and longer time ranges on the same topic) or the width of topics (one topic is included in another topic) for the comparison purpose (16 out of 52). All the 52 topics satisfy at least one of the three selection criteria.

The rising tail is observed for 14 topics out of 52 topics (27%). Table 1 shows a couple of its examples. In addition the local convexity, i.e., the occurrence of convex subsections is observed for 35 topics, yielding totally 49 topics out of 52 topics (94%). Table 2 shows a couple of its examples. As the rising tail has been considered to be quite rare, this figure is very high. Further, the rising tail phenomenon is stable across the differences in databases, regions, years or width of topics except only one case (this only one exception is on the same topic of Enzymes with differences in databases and regions where the graph for Japanese firms from a biological database BA has convex subsections in the droop while the graph for world firms from an engineering database COMPENDX has a rising tail). This stability of rising tail across these differences indicates that the rising tail phenomenon depends little on these differences but mainly on the current state of technological development of topic. Thus it may be allowed to claim that Hypothesis 1 is verified. However, it might be argued that the rising tail occurs at such a high rate because it is invalid to apply the Bradford distribution to the data on these topics. It is difficult to refute this argument in a direct way but an indirect examination will be made as to whether interpretations other than the DCC interpretations are valid.

§3. Examination of Other Interpretations

3.1 Subdivision interpretation

Conceive that the total number of articles is a priori determined (perhaps by God). The total articles are divided by two journals and this division is repeated. If the division is done at random, the number of articles in each journal follows the log-normal distribution [Karameshu and Cano 1984]. If data plotted on the log-normal nomograph forms a straight line with a good approximation, this interpretation is verified. For our data the fit is quite poor especially for peripheral firms. Only 4 cases out of 52 cases show good fits. Thus this interpretation is refuted for our data.

3.2 Powder engineering reconsideration of subdivision interpretation

Theoretically the subdivision yields the log-normal distribution. But experiences in powder engineering recommend an empirical distribution called the Rosin-Rammler distribution which usually yields better fits to the particle size than the log-normal distribution. The Rosin- Rammler distribution generates a good approximation of the log-normal distribution for a particular value of parameter (Appendix 2). The Rosin- Rammler nomograph is also available in which a straight line is generated if the data follows the Rosin-Rammler distribution. Plotting of our data on the nomograph fails to generate a straight line, showing poor fits again. The fits are especially poor for the core section. All of the 52 cases show poor fits in the core section while the middle section shows fairly good fits for all the 52 cases. The reason of poor fits is clear. Engineers make grinding efforts not to leave extremely large particles unbrokendown while our data contains core firms with extremely large numbers of articles. Thus this interpretation is refuted.

3.3 Multiplication of factors interpretation

A number of factors is needed for a success of R&D. Assume that the value of each factor follows the normal distribution and that a success is determined by the multiplication of the values of factors. Then the success follows the log-normal distribution [Karameshu and Cano 1984]. As an article represents a success of R&D, the number of articles also follows the log-normal distribution. This interpretation is refuted for the same reason as the above subdivision interpretation.

3.4 Summation of factors interpretation

Suppose the same situation as above except that the success is determined by the sum instead of the multiplication of factors. Then, by the wellknown central limit theorem, the normal distribution is yielded. Plotting our data on the normal distribution nomograph fails to form a straight line, proving poor fits. Hence this interpretation is refuted.

3.5 Maximal factor interpretation

Suppose the same situation as above except two points that the value of factors follows the exponential distribution and that the success is determined by the strongest factor. The exponential distribution assumption is more natural than the normal distribution assumption for inequitably distributed factors, and the strongest factor assumption may be natural for success of R&D as a fruit of genius works in that the strongest factor represents some genius factor. By a theorem in the extremes statistics [Gumbel 1958], the maximal values in exponential distributions generate the double exponential distribution (Appendix 2). Plotting of data on the double exponential distribution nomohraph forms a straight line if the data follow the double exponential distribution. But plotting of our data yields straight lines only for 1 case out of 52 cases. Thus this interpretation is refuted.

Replacing the exponential distribution with some other distributions (the Weibull, gamma, log-normal, normal distributions) in the above discussion theoretically or approximately yields also the double exponential distribution which our data do not fit. Thus this replacing is also refuted.

3.6 Minimal value interpretation

Today R&D, especially in firms, is done in a team work in which the success may be determined by the weakest factor. Assuming the same as above with the maximum replaced by the minimum yields the Weibull distribution as is wellknown in reliability engineering (Appendix 2). Plotting of data on the Weibull distribution nomograph forms a straight line if the data follows the Weibull distribution. But plotting of our data forms straight lines only for 3 cases out of 52 cases. Thus this interpretation is refuted.

Thus the non-DCC interpretations are refuted for our data. This may indirectly suggest the applicability of the Bradford distributions to our data upon the basis of the DCC interpretations. It is interesting to test whether the bibliometric data in general follows some of the above distributions, but this test seems too costly to attempt here.

§4. Application of Rising Tail Phenomenon

It may be assumed that the number of articles published by scientists in a firm represents the BR performances of the firm. For individual scientists and accademic institutions, the BR performances are evaluated usually by (the number of) articles [Moed, Burger, Frankfort and van Raan 1985 ; Hicks, Martin and Irvine 1986]. One may doubt to extend this staraightly to the evaluation of firms because firms are not interested in the publication activity itself. However, there are reasons to justify this extension to firms, at least for their BR performances as an extension of the idea of cumulative advantage effect. The BR oriented firms need to attract BR personnels in the labour market in competition with accademic institutions. As scientists tend to join institutions with high performances and as the performance is usually evaluated by (the number of) articles, firms tend to encourage publishing (many) articles for attracting BR personnels. Indeed, the number of eminent scientists in a firm is more correlated to the publications than to the patenting in the firm [Halperin and Chakrabarti 1987]. Further, in the Japanese patent system, a firm which (or a person who) has published a paper on a topic is given priority in patenting on this particular topic. Suppose that a firm published a paper on a topic at time t but did not apply patents on this topic till another firm applied a patent on this particular topic later at t+s without publishing a paper before t. Appealing to the patent authority, the former firm can win the latter in patenting. This system motivates firms to publish articles. As an single patent is usually insufficient to develop a new product or a new process, an innovative firm needs a number of patents which may not be covered by a single article. This situation encourages firms to publish numerous articles. Hence it may be reasonable to evaluate the BR performances of firms by the number of articles. Further, if a firm is strong in BR in a field, it is reasonable to expect that the firm will be strong at the later stages of applied research (AR), development and production in this field. This may be explained as follows. Once a firm builds a strong step in a field over competitors at the BR stage, it tends to allocate more resources to this field. Thus it is reasonable to expect that the firm continues to dominate competitors at the following stages in the same field. Thus, the number of articles of a firm in a BR-supported field may be interpreted to indicate the performance and strategy of the firm [Steck, Cox and Hagmeyer 1981].

Applying the Bradford distribution to journals (or firms), they are classified into core, middle and peripheral journals (or firms) according to their correspondence to the three sections of the graph. This classification is useful in the evaluation of firms' technological performances for investing and granting [Eto 1984] as well as the library management of journals [Goffman and Morris 1970 ; Brookes 1971]. As was discussed above, this classification may be applied to an evaluation of BR performances of firms as core firms, middle firms and peripheral firms for the similar use.

With the aid of the DCC interpretations, the three sections in the Bradford distribution are interpreted as providing the following informations (Fig. 1).
Core : When BR is "improportionately" concentrated on a small number of firms, the core is formed where "proportion" corresponds to the linearity in the rank semilog coordinates. When the opposite, the graph associated with the "core" is below the straight line in the rank semilog coordinates. When "proportionate", the core is not formed and the graph associated with the "core" is a straight line in the rank semilog coordinates.
Middle : When the diffusion is in "steady state", the graph forms a long straight line in the rank semilog coordinates. Although the precise

definition of "steady state" is still an open question, the linearity in the rank semilog coordinates is proved equivalent to the geometric zoning property [Eto and Candelaria 1967] which is a mathematical version of the verbal expression of the Bradford law by Bradford himself. (Note that the approximation of the rank semilog linearity is sometimes different from the approximation of the geometric zoning [Aiyepeku 1977 ; Wilkinson 1972] presumably by approximation errors). When the diffusion is not in "steady state", the line is short. When the productivity level of the middle ranked firms is high, the slope of the line is steep in the rank semilog coordinates.
Droop : When the diffusion to peripherals is not sufficient, the heavy droop is observed. When sufficient, the diffusion is in "steady state" and the droop disappears. When the new participation suddenly starts after the "steady state", the right tail rises upward and possibly above the line in the rank semilog coordinates. The degree of rise of the tail represents the degree of new participation as compared to the total size.

The abovestated interpretation of the rising tail as representing the sudden and simultaneous participation and the real occurrences of rising tail in distributions of articles by (mainly Japanese) firms in evolving technological fields are consistent to the recent behaviour of (especially Japanese) firms in these fields. In this sense, this interpretation may be considered as valid.

Now the graph of the Bradford distribution is classified into several patterns, and interpretations are given them.
Pattern 1 (Immatured diffusion). As was described as standard in the

section 1, the core section is usually convex, the middle section is a long straight line and the droop section is slightly concave. This graph indicates the insufficient diffusion and the possibility of future diffusion.

Pattern 2 (Matured diffusion). As was discribed as classic in the section 1, the core section is usually convex and the rest is a straight line. This graph indicates the sufficient diffusion and the stationarity of R&D on a topic possibly till a future breakthrough.

Pattern 3 (Closed and immatured diffusion). The core section is convex and shares a large number of articles. The middle section is a short straight line. A large number of firms belong to the droop section which is heavily drooped. This graph indicates the inbalanced development with a heavy concentration on a few leading firms [Eto 1984] or a few leading countries [Eto and Candelaria 1987], namely, the domination by a few.

Pattern 4 (Sudden and simultaneous participation). The droop section is rising possibly above the straight line. This indicates the sudden participation of many firms in R&D on a topic simultaneously with a faster speed than the total progress of R&D. In an extreme case the whole graph is convex. This indicates non-existence of leading (i.e., core) firms which is equivalent to saying oppositely that every firm is leading (i.e., every firm is in the core).

Reversing the above logic, the pattern of the Bradford distribution may be taken as suggesting the stage of technological development. Extending this idea might suggest a hypothesis on the "stability" of the patterns of distributions. In fact, though in a different context, it is already found that the degree of dispersion of articles across journals in a field is correlated to the obsolescing speed of the field [Buckland 1972].

Hypothesis 2 (Stability of distribution patterns). Pattern 4 will not last long (is unstable) and will shortly "evolve" to other patterns. On the other hand, Pattern 2 is stable.

This hypothesis could be verified by tracing the "trajectories" of the patterns of the Bradford distribution of several topics in the past though it is quite costly. If Hypothesis 2 is supported with some plausility, it would provide a useful suggestion in identifying the current and future stages of technological development of a topic.

§5. Parameters of Bradford Distribution

Discussions have been made about the s-value which denotes the interception of the semilog line with the horizontal axis (Fig. 2). The core size (the number of core firms) is claimed to be correlated to the s-value [Brookes 1969B] and this claim is theoretically explained and weakly evidenced [Eto and Candelaria 1987]. For our 52 topics, the correlation is 0.94 with the 5% significance, strongly supporting the claim.

The interpretation of the s-value is controversial. A claim that the s-value represents the width of a topic [Brookes 1969B] is supported [Eto 1984] and refuted [Drott and Griffth 1978 ; Eto and Candelaria 1987]. For the purpose of this examination, the selection of our 52 topics were designed so that, for 8 pairs of topics, a topic is included in the other topic with the both being from the same database. For 6 pairs out of the 8 pairs, the s-value is larger for wider topics, supporting the claim.

As an extension of the above interpretation of the s-value, it may be claimed that the s-value is larger for data of longer years because the longer duration causes the widening effect of a topic. For the purpose of this verification, the selection of our 52 topics were designed so that, for 14 pairs of topics, a set of data on a topic from a database covers shorter years than the other set of data on the same topic from the same database. For 10 pairs out of 14 pairs, data covering longer years has a larger s-value. This may support the claim but may indicate that a longer duration increases the number of core firms which is correlated to the s-value.

Table 3 shows other relations between parameters including the s-value.

§6. Conclusion

The phenomenon of rising tail in the droop section in the Bradford distribution graph was reported. The condition of its occurrence was explained and evidenced as a (series of) tie value (values) in the number of articles among many journals (or firms). It was interpreted as representing the sudden and simultaneous participation of many journals (or firms) in the research of a topic. This interpretation was supported for the selected 52 topics in evolving technology fields. Reversely, it was proposed to estimate the current and future stages of technological development from the rising tail phenomenon. Additionally, quantitative relations between parameters of the Bradford distribution was reported with emphasis on the interpretation of the horizontal axis interception (i.e., the s-value).

Appendix 1 : Mathematical expression of the rising tail [Eto and Candelaria

1987]

Let journals be ranked : $j = 1,2, \ldots, J$ with J denoting the total number of journals and the number of journals published in the j-th journal be $a(j)$. Thus, $a(i) \geq a(j)$ for $1 \geq i > j \geq J$.

Let $q(j)$ denotes the increment or the piecewise linearized right-side slope of the graph in the rank semilog coordinates. Then, for $1 < j < J - 1$,

$$q(j) = a(j + 1)/\log\{(j + 1)/j\}$$

If $q(j) < q(j+1)$, then the graph is convex. If $q(j) > q(j+1)$, then the graph is concave. If $q(j) = q(j+1)$, then the graph is a straight line.

If $a(j) = a(j+1)$, then $q(j) < q(j+1)$. Thus the graph is convex. It often happens in the right tail that $a(J - k) = \ldots = a(J - 1) = a(J) = 1$. If k is large, then the convexity here is innegligible. Thus the rising tail happens. The same phenomenon also happens when the right-hand side above takes 2 or more.

Appendix 2 : Rosin-Rammler, Double exponential and Weibull distributions

<u>The Rosin-Rammler distribution</u>

Let d denote the size of particles (or the number of articles in our case), $r(d)$ denote the ralative frequency (%) of particles with size d and $R(d)$ denote the cumulative ralative frequency upto d so that $R(\infty) = 100$ %. Then the Rosin-Rammler formula is (b and n constant),

$R(d) = 100 * \exp(-bd^n)$

Define the absolute size constant D as $1/D^n = b$. Then the Rosin-Rammler-Bennet formula is

$R(d) = 100 * \exp\{-(d/D)^n\}$

Letting D = d yields R = 100/e = 36.8 (%).

The Rosin-Rammler nomograph plots values of d along the horizontal axis and values of R (%) along the vertical axis on which the 36.8 (%) line is marked to read the value of D. Either of the both axes is in a specially designed scale. This distribution can be approximated by the log-normal distribution when $n > 1$ while its density r monotone decreases when $n \leq 1$.

The double exponential distribution

The density f(x) and the cumulative distribution F(x) are as follows.

$f(x) = \exp\{-x - \exp(-x)\}$

$F(x) = \exp\{-\exp(-x)\}$

The double exponential distribution nomograph plots the cumulative relative frequency (%) along the horizontal axis and the observed values (the number of articles in our case) along the vertical axis, both in specially designed scales.

The Weibull distribution

The cumulative distribution F(x) is given as

$F(x) = 1 - \exp\{-x^m/T\}$

where m denotes the shape parameter and T denotes the scale parameter. The Weibull distribution nomograph plots x along the horizontal axis and the F(x) along the vertical axis, both in specially designed scales.

Acknowledgement

The author heartily thanks Professor K. Nakayama and Mrs. H. Abe at the Science Information Center at the University of Tsukuba for their invaluable help in the usage of scientific databases. He also thanks Miss Y. Tsukada (now with Asahi Chemical Corporation) and Mr. K. Mase, both being graduate students at the University of Tsukuba, for their data collection and calculation activities.

References

W.O. Aiyepeku, The Bradford Distribution Theory : The Compounding of Bradford Periodical Literatures in Geography, Journal of Documentation 33 (1977) 210 - 219

I. Asai, A General formulation of Bradford's Distribution : The Graph-oriented Approach, Journal of the American Society for Information Science 32 (1981) 113 - 119

J.R. Bright(ed), Technological Forecasting for Industry and Government : Methods and Applications, Prentice Hall, Englewood Cliffs 1968

B.C. Brookes, The Derivation and Application of the Bradford-Zipf Distribution, Journal of Documentation 24 (1968) 247 - 265

B.C. Brookes, The Complete Bradford-Zipf 'Bibliograph', Journal of Documentation 25 (1969A) 58 - 60

B.C. Brookes, Bradford's Law and Bibliography of Science, Nature 224 (1969B) 953 - 956

B.C. Brookes, Optimum p% Library of Scientific Periodicals, Nature 232 (1971) 458 - 461

B.C. Brookes, Numerical Methods of Bibliographic Analysis, Library Trends July 1973, 18 - 43

B.C. Brookes, The Bradford Law : A New Calculus for the Social Sciences ? Journal of the American Society for Information Science 30 (1979) 233 - 234

B.C. Brookes, A Critical Commentary on Leimkuhler's 'Exact' Formulation of the Bradford Law, Journal of Documentation 37 (1981) 77 - 88

M.K. Buckland, Are Obsolescence and Scattering Related ?, Journal of Documentation 28 (1972) 242 - 246

Y.S. Chen and F.F. Leimkuhler, Bradford's Law : An Index Approach, Scientometrics 11 (1987) 183 - 198

D.J. de Sola Price, Networks of Scientific Papers, Science 149 (1965) 510 - 515

D.J. de Sola Price, A General Theory of Bibliometric and Other Cumulative Advantage Processes, Journal of the American Society for Information Science 27 (1976) 292 - 306

M.C. Drott, Bradford's Law : Theory, Empiricism and the gap Between, Library Trends, Summer 1981 41 - 52

M.C. Drott and B.C, Griffth, An Empirical Examination of Bradford's Law and the Scattering of Scientific Literature, Journal of the American Society for Information Science 29 (1978) 238 - 246

L. Egghe, Consequences of Lotka's Law for the Law of Bradford, Journal of Documentation 41 (1985) 173 - 189

F.J. Elvin, Bradford's Law, Journal of Documentation 34 (1978) 246 - 247

H. Eto, Bradford Law for R&D Expending of Firms and R&D Concentration, Scientometrics 6 (1984) 183 - 188

H. Eto and P.M. Candelaria, Applicability of the Bradford Distribution to International Science and Technology Indicators, Scientometrics 11 (1987) 27 - 42

H. Eto and M. Fujita, Regularities in the Growth of High Technological Industries in Regions, Discussion Paper No. 331(1987), The University of Tsukuba, Institute of Socio-Economic Planning

H. Eto and K. Makino, Stochastic Model for Innovation and Resulting Skew Distribution for Technological Concentration with Verification in Japanese Industry, Scientometrics 5 (1983A) 219 - 243

H. Eto and K. Makino, Theoretical and Empirical Analysis of the Differentiation Process in the Technology Gap between Developed and Developing Nations, in W. Isard and Y. Nagao (eds) : International and Regional Conflict : Analytic Approaches, Chapter 9 (pp 149 - 159), Ballinger, Cambridge, Mass., 1983B

E.J. Gumbel, Statistics of Extremes, Columbia University Press, New York 1958

W. Goffman and T.G. Morris, Bradford's Law and Library Acquisitions, Nature 226 (1970) 922 - 923

S.D. Haitun, Stationary Scientometric Distributions Part II. Non-Gaussian Nature of Scientific Activities, Scientometrics 4 (1982A) 89 - 104

S.D. Haitun, Stationary Scientometric Distributions Part III. The Role of the Zipf Distribution, Scientometrics 4 (1982B) 181 - 203

M.R. Halperin and A.K. Chakrabarti, Firm and Industry Characteristics Influencing Publications of Scientists in Large American Companies, R&D Management 17 (1987) 167 - 173

D. Hicks, B.R. Martin and J. Irvine, Bibliometric Techniques for Monitoring Performance in Technologically Oriented Research ; The Case of Integrated Optics, R&D Management 16 (1986) 211 - 223

J.B. Homer, A Diffusion Model with Application to Evolving Medical Technologies, Technological Forecasting and Social Change 31 (1987) 197 - 218

Y. Ijiri and H.A. Simon, Skew Distributions and the Sizes of Business Firms, North-Holland, Amsterdam 1977

H. Jones and B.C. Twiss, Forecasting Technology for Planning Decisions, Macmillan, London 1978

N.C.L. Karmeshu and V. Cano, Rationales for Bradford's Law, Scientometrics 6 (1984) 233 - 241

M.G. Kendall, The Bibliography of Operational Research, Operational Research Quarterly 11 (1960) 31 - 36

F.F. Leimkuhler, The Bradford Distribution, Journal of Documentation 23 (1967) 197 - 207

V. Mahajan and Y. Wind (eds), Innovation Diffusion Models of New Product Acceptance, Ballinger, Cambridge, Mass., 1986

M.J.F. Maia and M.D. Maia, On the Unity of Bradford's Law, Journal of Documentation 40 (1984) 206 - 216

J.P. Martino, Technological Forecasting for Decision Making, North-Holland, Amsterdam 1983

H.F. Moed, W.J.M. Burger, J.G. Frankfort and A.F.J. van Raan, The Use of Bibliometric Data for the Measurement of University Research Performance, Research Policy 14 (1985) 131 - 149

P.M. Morse, Implications of the Exact Bradford Distribution, Journal of the American Society for Information Science 32 (1981) 43 - 50

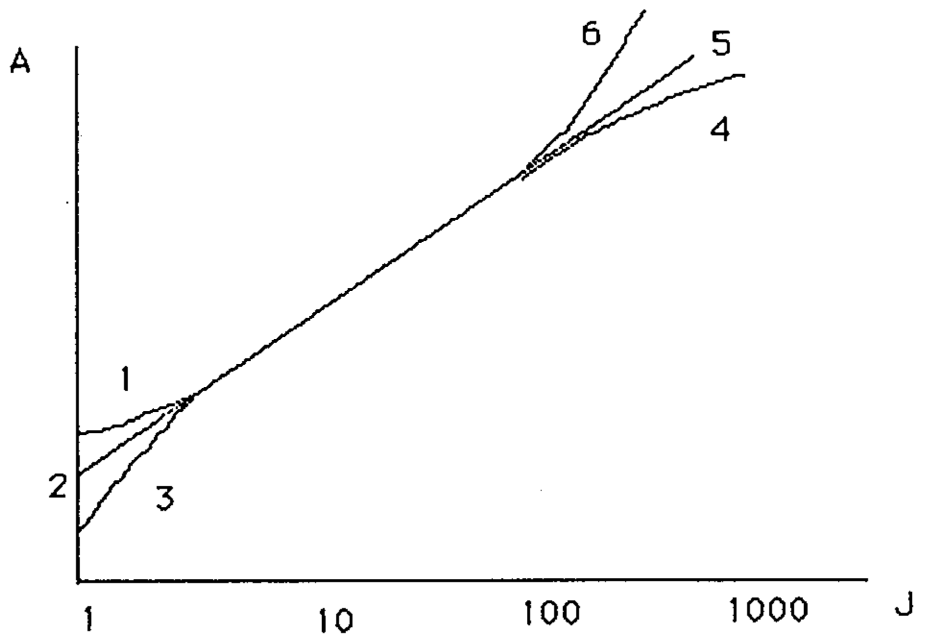
P.M. Morse and F.F. Leimkuhler, Exact Solution for the Bradford Distribution and its Use in Modeling Informational Data, Operations Research 27 (1979) 187 - 198

P. Praunlich and M. Kroll, Bradford's Distribution : A New Formulation, Journal of the American Society for Information Science 29 (1978) 51-55

P. Steck, J.S.G. Cox and F.W. Hagemeyer, Literature Indexing Systems for Corporate Strategy : A Case-Study in the Pharmaceutical Industry, R&D Management 11 (1981) 97 - 101

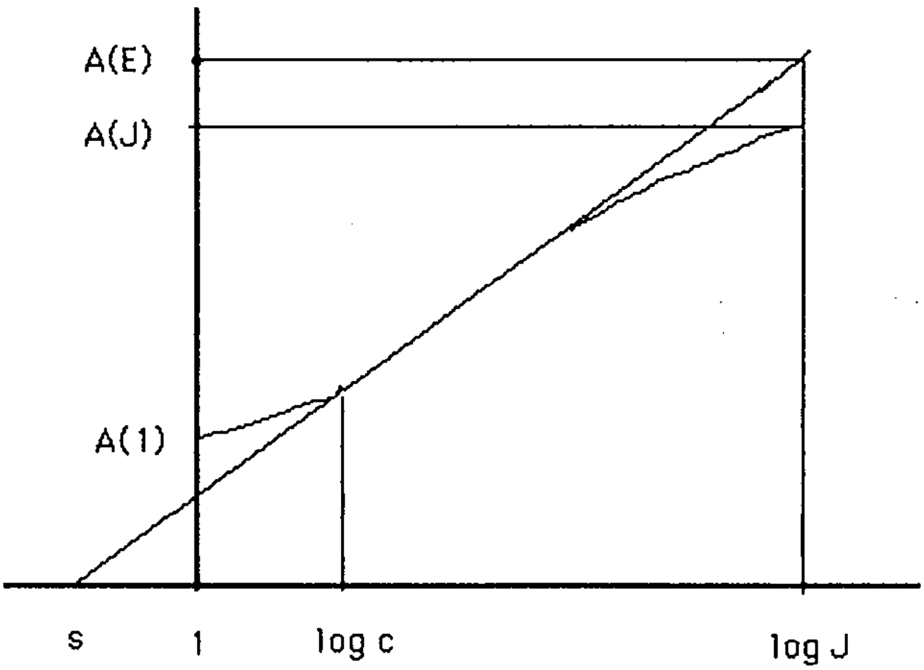
E.G. Summers, Bradford's Law and the Retrieval of Reading Research Journal Literature, Reading Research Quarterly XIX/I (Fall 1983) 103 - 109

E.A. Wilkinson, The Ambiguity of Bradford's Law, Journal of Documentation 28 (1972) 122 - 130



1 : Core is above the line
2 : Core does not exist
3 : "Core" is below the line
4 : Droop is below the line
5 : Droop does not exist
6 : Rising tail

Fig. 1 Various shapes of graph of Bradford distribution



$$A(j) = g \log j + h$$

$$A(E) = g \log J + h$$

$$A(D) = A(E) - A(J)$$

Fig. 2 Parameters of Bradford distribution

Table 1. Examples of topics causing the rising tail

Enzymes (world) COMPENDX			Fermentation (world) COMPENDX		
Rank	No. of articles	No. of firms	Rank	No. of articles	No. of firms
1	9	1	1	9	1
2,3	8	2	2,3	7	2
4,5	6	2	4	5	1
6,7	5	2	5,7	4	3
8	4	1	8,12	3	5
9,18	3	10	13,34	2	22
19,35	2	17	35,129	1	95
36,123	1	88			
Total	203	123	Total	194	129

Table 2. Examples of topics causing conves subsections

Enzyme inhibitors(Japan) EM			Amorphous (Japan) COMPENDX		
Rank	No. of articles	No. of firms	Rank	No. of articles	No. of firms
1	15	1	1	19	1
2	13	1	2	16	1
3	9	1	3	8	1
4,7	7	4	4,5	6	2
8,9	6	2	6	5	1
10	5	1	7,8	4	2
11,15	4	5	9,10	3	2
16,18	3	3	11,13	2	3
19,24	2	6	14,21	1	8
25,46	1	22			
Total	145	46	Total	88	21

Table 3. Correlation between parameters

	A(D)	c	s	g	A(1)
A(1)/A(J)	0.20	-0.40*	-0.42*	-0.20	0.44*
A(1)	0.46*	-0.14	-0.22	0.63	
g	0.54*	0.21	0.16		
s	-0.39*	0.94*			
c	-0.46*				
A(D)					

See Fig. 2 for parameter names

* : 5 % significant