


## Subject: Library and Information Science

Production of Courseware

 **-Content for Post Graduate Courses**



**Paper No : 10** Informetrics and Scientometrics

**Module : 06** Bradford Distributions: An Overview



### Development Team

Principal Investigator  
&  
Subject Coordinator

Dr. Jagdish Arora, Director  
INFLIBNET Centre, Gandhinagar

Paper Coordinator

Dr I K Ravichandra Rao  
Retd Professor, Documentation Research and  
Training Centre

Content Writer

Prof B K Sen  
Retd Professor, National Institute of Science Communication and  
Information Resources

Content Reviewer

Dr I K Ravichandra Rao  
Retd Professor, Documentation Research and  
Training Centre

## **UNIT 6**

### **Bradford Distributions: An Overview**

#### **I. Objectives**

After going through this case study you will come to know about the following:

- To study and understand Derivation of equations for Bradford distribution by various bibliometricians.
- To discuss viewpoints of some bibliometricians on the law.
- To study the Ambiguity between verbal and graphical representations of Bradford distribution.
- To discuss Bradford-Zipf distribution
- To study the Characteristics of bibliometric distribution, etc.

#### **II. Learning Outcome**

After completion of this module, you will be certainly knowledgeable with regard to Bradford distribution and related work. At the end of this module, you gained knowledge on various aspects of Bradford's law -- Bradford-Zipf distribution, ambiguity between verbal and graphical interpretation of Bradford's law, Leimkulher distribution; computational aspects of Bradford's law.

#### **III. Module Structure**

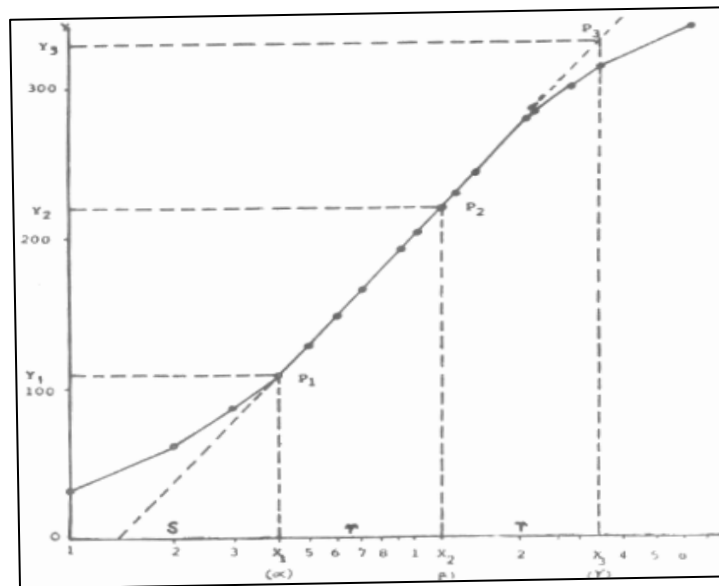
1. Introduction
2. Cole's Formulation
3. Leimkuhler's Formulation
4. Brookes Formulation
5. Naranan's Viewpoint
6. Bookstein's Viewpoint
7. Bradford Multiplier
8. Ambiguity between Verbal and Graphical Statements
9. Bradford-Zipf Distribution
10. Characteristics of Bibliometric Distribution
11. References

## 1. Introduction

The law of scattering was propounded by Samuel Clement Bradford (1878-1948), a British librarian, mathematician and document artists at the Science Museum in London after a laborious study of scientific literature in mid-1930s. After examining the distribution of scientific literature in periodicals and their coverage in abstracting and indexing periodicals he realized that the distribution of literature follow a particular pattern [1, 2]. He opined that ‘the nucleus of periodicals devoted to the given subject must contain, individually, more articles on that subject than periodicals dealing with related subjects’ [3]. ‘In consequence, it is possible to arrange periodicals in zones of decreasing productivity, in regard to papers on a given subject, and the numbers of periodicals in each zone will increase as their productivity decreases’ [3]. He described a scattering pattern of journals in the area of applied geophysics and lubrication. He plotted the partial sums of references against the natural logarithm of the partial sum of numbers of journals, and he noticed that the resulting graph is a straight line. On the basis of this observation, he suggested the following linear relation to describe a scattering phenomenon [2]

$$F(x) = a + b \log x.$$

$F(x)$  is the cumulative number of references contained in the first  $x$  most productive journal;  $a$  and  $b$  are constants. The following figure is a hypothetical, but typical, log-linear curve (as described by Bradford) showing aggregates of articles on a given subject corresponding to the number of journals.



**Fig. 1: Bradford Bibliograph(17)**

This type of a curve is usually called a Bradford curve; In X-axis: Partial sum of Journals (in log scale). In Y-axis: Partial sum of articles contained in X top most journals (in linear scale)

$P_1$  in the figure is the point at which the straight line part of the curve begins. Draw  $Y_1P_1$ ,  $Y_2P_2$ , and  $Y_3P_3$  such that they are parallel to the X-axis and  $OY_1 = Y_1Y_2 = Y_2Y_3$ . Draw  $P_1X_1$ ,  $P_2X_2$ , and  $P_3X_3$  such that they are parallel to Y-axis. Since  $P_1P_3$  is a straight line and since  $Y_1Y_2 = Y_2Y_3$ ,  $X_1X_2$  and  $X_2X_3$  are equal, say  $r$  units. Let the distance between  $O$  and  $X$  is  $s$  units. Thus, if  $\alpha$ ,  $\beta$  and  $\gamma$  are the positive real numbers corresponding respectively to the logarithmic abscissa  $OX_1$ ,  $OX_2$  and  $OX_3$ , we have,  $\log \alpha = s$ ,  $\log \beta = s+r$ , and  $\log \gamma = s+2r$ .

That is,  $\alpha = 10^s$ ,  $\beta = 10^{r+s} = 10^s \cdot 10^r$ , and  $\gamma = 10^{s+2r} = 10^s \cdot 10^{2r}$

Substituting  $n = 10^r$ , we see that the natural numbers  $\alpha$ ,  $\beta$ , and  $\gamma$  are related to each other as  $1:n:n^2$ . On the basis of this relationship and also since  $OX_1$  represents a number of periodicals in a subject area Bradford stated his law as

“Bradford stated that “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 subject, they and several groups of zones containing the same number of articles as the nucleus, when the zones will be  $1:n:n^2 \dots$ ”

This is called Law of Scattering or Bradford's law.

Since then a great deal of work has been done with this law [17,9,10, 13]. An attempt is made in this module to highlight some of those works.

## 2. Cole's Formulation

Cole [8] also experimented with the law and named the slope of the curve as the reference scattering co-efficient and concluded that the coefficient might be the characteristic of the subject field. For petroleum literature, Cole obtained the relationship

$$F(x) = 1 + b \log_{10} x \quad (x > c)$$

where  $F(x)$  stands for the cumulative number of papers contained in the  $x$  number of most productive journals, and  $c$  represents the number of journals figuring in the nuclear zone. For petroleum literature Cole found the value of  $b$  as 0.43.

## 3. Leimkuhler's Formulation

Ferdinand F. Leimkuhler [13] Professor of Industrial Engineering of the Purdue University of the United States analyzed all the published data on Bradford distribution and derived the following equation applying statistical techniques.

$\ln(1+\beta x)$

$$F(x) = \frac{\ln(1+\beta x)}{\ln(1+\beta)}, \quad (0 \leq x \leq 1)$$

In the equation,  $F(x)$  stands for the cumulative fraction of the references,  $x$  for the corresponding fraction of the most productive journals, and  $\beta$  is a constant related to the document collection. Brookes [5] opined that the equation is but a compromise since Leimkuhler accepted the empirical data as both complete and exact. Brookes further observed that “Unfortunately, though Leimkuhler’s formulation can be used theoretically without difficulty, it has some disadvantages for the practical document a list. The numerical evaluation of the key parameter  $\beta$  requires tedious statistical computation and the solving of an implicit equation by approximation methods....In fact it was the exasperation evoked by an attempted practical application of Leimkuhler’s formulae that led the author of this paper to seek a simpler formulation of the Bradford distribution” [5: p249]

#### 4. Brookes Formulation

Brookes [5, 6] formulation of the Bradford distribution follows. Suppose  $R(n)$  is the cumulative total of relevant papers found in the first  $n$  journals when all the journals are ranked in order of decreasing productivity, the Bradford’s law requires that

$$R(n) = R(n^2) - R(n) = R(n^3) - R(n^2) = \dots$$

for all integral values of  $n$  greater than unity.

$$\begin{aligned} \text{We have} \quad R(n) &= R(n^2) - R(n) && \text{Eq. 1} \\ \Rightarrow R(n) + R(n) &= R(n^2) \\ \Rightarrow 2R(n) &= R(n^2). \end{aligned}$$

$$\begin{aligned} \text{Again, we have} \quad R(n^2) - R(n) &= R(n^3) - R(n^2) && \text{Eq. 2} \\ \Rightarrow R(n^2) + R(n^2) - R(n) &= R(n^3) \\ \Rightarrow 2R(n^2) - R(n) &= R(n^3) \\ \Rightarrow 2 \cdot 2R(n) - R(n) &= R(n^3) \quad [R(n^2) = 2R(n) \text{ from Eq. 1}] \\ \Rightarrow 4R(n) - R(n) &= R(n^3) \\ \Rightarrow 3R(n) &= R(n^3) \end{aligned}$$

$$\begin{aligned} \text{We have} \quad R(n^2) &= 2R(n) \\ R(n^3) &= 3R(n) \end{aligned}$$

Hence, we can generalize  $R(n^x) = xR(n)$ , where  $x$  is a positive integer

$$\Rightarrow R(n) = \frac{1}{x} R(n^x)$$

$$\begin{aligned} \text{Putting } 1/x = a, \text{ and } R(n^x) = n^b \text{ we can write the equation as} \\ \Rightarrow R(n) = an^b \quad [1 \leq n \leq c] &&& \text{Eq. 3} \end{aligned}$$

where  $c$  is the value of  $n$  at the point where the straight portion of the curve begins [Fig. 1]

The only function that fully satisfies this condition is

$$R(n) = k \log n, \text{ where } k \text{ is a constant} \quad \text{Eq.4}$$

Brookes has provided another model

$$R(n) = N \log \frac{n}{s} \quad [c \leq n \leq N] \quad \text{Eq.5}$$

where  $N$  = Total number of periodicals expected to publish paper on the subject, and  
 $s$  = Value of  $n$  at the intersection of the straight portion of the curve with the X axis [Fig. 1].

Now, let us take Eq. 3 and find out the values of  $a$  and  $b$

In Table 1, we find that the first 20 periodicals account for 590 articles, and the first 30 periodicals account for 689 articles. In the first case, putting these values into the Eq. 3 we get

$$590 = a \cdot 20^b \quad \text{Eq. 6}$$

$$689 = a \cdot 30^b \quad \text{Eq. 7}$$

Dividing Eq. 7 by Eq. 6, we get

$$689/590 = 30^b/20^b$$

Taking log we get  $\log 689 - \log 590 = b \log 30 - b \log 20 = b (\log 30 - \log 20)$

$$\Rightarrow 2.8382 - 2.7709 = b (1.4771 - 1.3010)$$

$$\Rightarrow 0.0673 = b (0.1761)$$

$$\Rightarrow b = 0.382$$

Now, putting the value of  $b$  in Eq. 6, we get

$$590 = a \cdot 20^{.382}$$

Taking log we get  $\log 590 = \log a + .382 \log 20$

$$\Rightarrow 2.7709 = \log a + .382 \times 1.3010$$

$$\Rightarrow 2.7709 = \log a + .4970$$

$$\Rightarrow \log a = 2.7709 - .4970 = 2.2739$$

Taking antilog of 2.2739 we get 187.9 i.e. 188

The values we get are  $a = 188$ , and  $b = .382$

With the values of  $a$  and  $b$  we can now determine the value of the cumulative total of references for the periodical of any rank.

For testing, let us take 45th rank.

We have  $R(n) = an^b$   
 Putting the values , we get  $R(n) = 188 \times 45^{.382}$   
 $= 188 \times 4.276$   
 $= 803.888$   
 $= 804$ , which is quite close to the observed value of 802.

Bradford considered the Bibliograph to be a straight line which has resulted in two different formulations, one is verbal and the other is graphical. The algebraic expressions for the two

formulations given by Brookes are:

$$R(n) = j \log (n/t + 1) \text{ for the verbal formulation, and}$$

$$R(n) = k \log n/s \text{ for the graphical formulation}$$

## 5. Naranan's Viewpoint

In 1970 Naranan [14] opined that (i) Bradford's law of bibliography of scientific literature is explainable in terms of an underlying power law distribution of the number of articles in scientific journals; (ii) the law emerges as a natural consequence of exponential growth of scientific literature and journals at comparable rates; and (iii) a model like this predicts a strong correlation between the age of a journal and the number of articles it carries. The author was hopeful that the proposed mechanism might find wider application in many other fields of science.

Brookes [7] pointed out that Naranan's analysis was not valid for Bradford. However, his paper provided a plausible model of Lotka's law with suitable verbal amendments. The comments of Brookes as to the paper are being reproduced verbatim. "The inverse square law of scientific authorship has hitherto been regarded as an inexplicable and useless scientific oddity. Naranan's model of it is therefore welcome. And, together with other measures of scientific productivity, Lotka's law has recently been applied by Dobrov and Korennoi in determining the optimum size of research institutes in USSR".

Hubert [ 11] was of the view that Naranan interpretation of the original form of Bradford's law does not follow a stochastic argument based on his assumptions.

## 6. Bookstein's Viewpoint

In his paper published in 1976, Bookstein [4] analyzed the distributions of Lotka, Zipf, Bradford and Leimkuhler and adopted a point of view that allows us to understand that these distributions are in fact the different versions of a single theoretic distribution. He generalized these distributions with the following words. "All of these distributions are almost equivalent . . . In each case we have a set of entities (for example, chemists, words) producing events (publications, occurrences) over some dimension of extension (time, length of text) and in each case the distribution describes the number of occurrences of events over a fixed interval of that dimension. Under these conditions it is possible to describe the same distribution in at least four

distinct ways; these modes of description are represented above by the distributions of Lotka, Zipf, Bradford, and Leimkuhler”.

Physicists all over the world have tried to unify four natural forces, i.e. electromagnetic force, gravitational force, weak nuclear force, and strong nuclear force for the last hundred years or so. Bookstein has done the same thing for bibliometric distributions. He has shown that basically all the four bibliometric distributions are different versions of the same distribution.

### 7. Bradford Multiplier

The number of periodicals in the three zones of Bradford distribution generally follows the ratio  $1:n:n^2$ , where  $n$  is the Bradford multiplier. Ravichandra Rao [16] analyzed the Bradford multiplier with a small sample of 12 datasets using t test. An attempt has also been made to identify a suitable model to explain the law of scattering. Among the various methods tried log normal fits much better than many models including the log linear model.

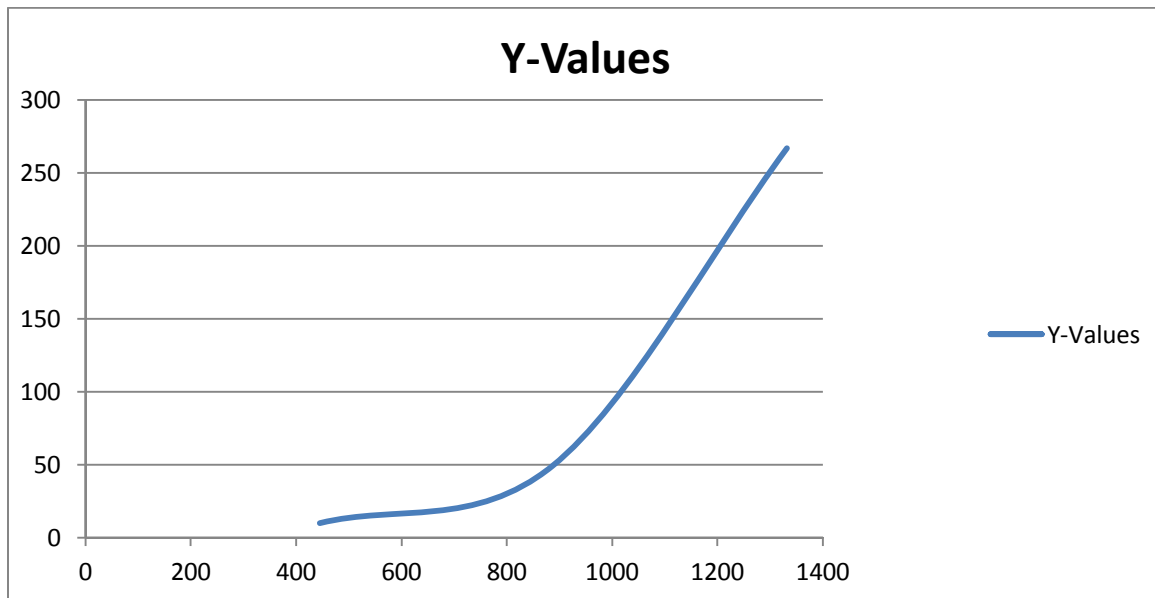
### 8. Ambiguity between Verbal and Graphical Statements

Bradford's law can be looked into two different ways -- graphically and verbally. This was first observed by Vickery [18]. Bradford’s verbal formulation of the law is recorded as “If scientific journals are arranged in order of decreasing productivity of articles on a given subject, they may be divided into a nucleus of periodicals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus, when the number of periodicals in the nucleus and succeeding zones will be as  $1: n: n^2$ ”.[2: p. 154]. In 1948, Brian C. Vickery [18] contributed an important paper on Bradford’s law. He analyzed about 1600 journal references and compared his results with Bradford’s and found an inconsistency. He remarked – “We can ... regard the theoretical distribution of papers on a given subject in scientific periodicals as derived by Bradford, as fully corroborated by the distributions observed in the sample investigations. The rectilinear relation . . . incorrectly assumed by Bradford to be identical with his theoretically derived relation, fits only the upper portion of the observed curve (Figs. 2 and 3). The theoretical relation itself , however, enables us to predict the whole curve”. Vickery showed that if  $n_m$  journals contribute a cumulative  $m$  papers, and  $n_m$  is greater than the nucleus, then the verbal formulation is equivalent to the expression  $n_m : n_{2m} - n_m : n_{3m} - n_{2m} : \dots :: 1: a_m: a_m^2: \dots$ . The graphical formulation is equivalent to the expression  $n_m : n_{2m} : n_{3m} : \dots :: 1: b_m: b_m^2: \dots$ . This apart, when the graph is plotted with the data of verbal formulation it takes a different shape compared to the shape of the graph with complete data set. In the verbal expression, the data in Table 1 will take the following shape and generate a curve as given in Fig. 2.

Zone	No. of Periodicals	Cumulative number of Articles
1 <sup>st</sup> Zone	10	445
2 <sup>nd</sup> Zone	49	886
3 <sup>rd</sup> Zone	267	1332

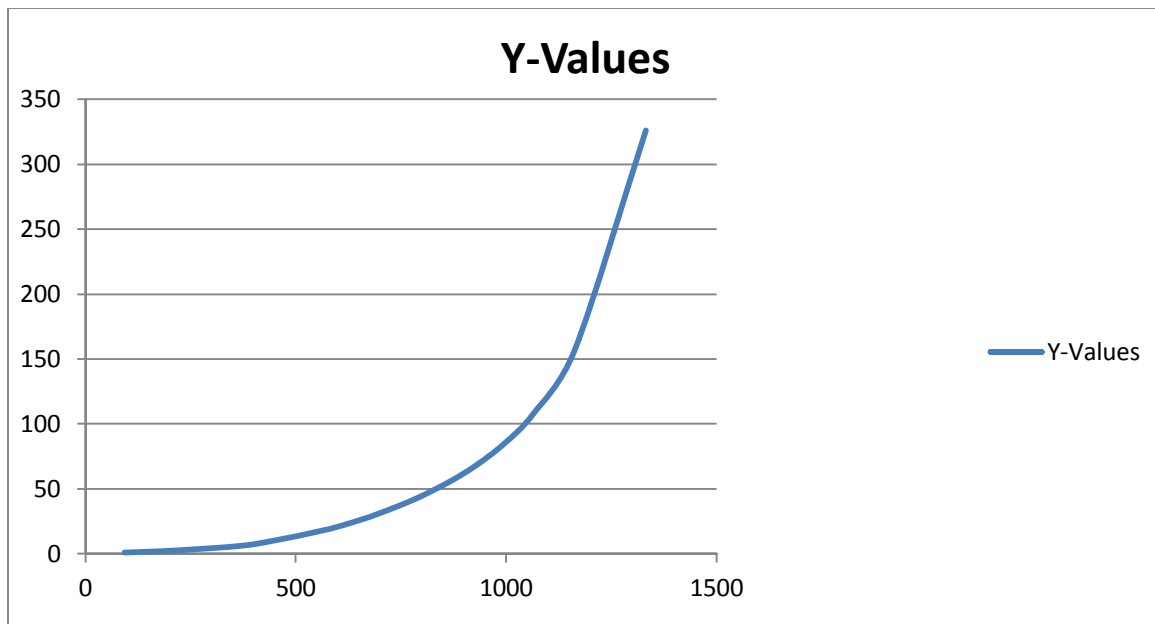
**Table 2 – Distribution of Articles according to Zones**





**Fig. 2 Bradford's Bibliograph with verbal formulation**  
**Y axis – Cumulative number of periodicals; X axis – Cumulative number articles**

When the graph is drawn with entire data set we find finer details which are lost in the graph of verbal formulation. Now let us see the graph with the entire data of Table 1



**Fig 3 Bradford's Bibliograph with complete data**  
**Y axis – Cumulative number of periodicals; X axis – Cumulative number articles**

Comparing the two bibliographs we find the following:

- i. In the verbal formulation, the entire data is not available. What we get is practically a summary of the entire data set.
- ii. The Bibliograph in Fig 2 is incomplete, inasmuch as it does not indicate the starting point of the curve.
- iii. The last portion of the graph in both Figs. 2 and 3 is a straight line.
- iv. In the graphical presentation of some Bradford distributions, a droop is observed at the end of the graph, which is not seen with the data of verbal formulation.

With these, the distinction between verbal formulation and the graphical presentation becomes quite clear and the shortcomings of the verbal formulation apparent. Brookes have provided equations both for verbal formulation as well as the graphical formulation. The equations are given under Brookes' formulation.

### 9. Bradford-Zipf Distribution

Kendall [12], a statistician by profession, also studied Bradford distribution using 1,763 references on operational research pertaining to 370 journals. For the sake of comparison '1465 references to statistical methodology (covering the period 1925-39)' were used. The graph plotted following Bradford's method produced a curve which was remarkable for its linearity. He also noticed that the Law is similar to, but not identical with the Zipf's law. Let us consider the data given in Table 4. The data set provides the typical Bradford distribution.

Rank	No. of Periodical/s	No. of article/s	Cumulative total
1	1	20	20
2	1	14	34
3	1	12	46
4	1	11	57
5	1	10	67
6	1	9	76
9	3	8	100
10	1	7	107
12	2	6	119
14	2	5	129
15	1	4	133
25	10	3	163
40	15	2	193
84	44	1	237

**Table 4– A data set following Bradford distribution.**

Inverting the columns 1 and 3 of Table 4 and multiplying the numbers of each row we get the following result (Table 5). The number in the second column may be considered as frequency.

<b>Rank</b>	<b>No. of article/s, i.e. Frequency</b>	<b>Rank x Frequency</b>
84	1	84
40	2	80
25	3	75
15	4	60
14	5	70
12	6	72
10	7	70
9	8	72
6	9	54
5	10	50
4	11	44
3	12	36
2	14	28
1	20	20

**Table 5 – Partly inverted form of Table 4**

The figure in the third column clearly indicates that they by and large follow Zipf's law. The two distributions are in fact very close, hence they are often referred to as Bradford-Zipf distribution. The linearity of the Bradford Bibliograph indicates a true Zipf situation.

### **10. Characteristics of Bibliometric Distribution**

- Bibliometric distributions can generally be expressed through algebraic expressions.
- On graphical presentation, they form different types of curves.
- All these distributions have given rise to well-established laws which have found applications in journal selection, ranking of authors, ranking of words for keyword generation, and so on.
- The classical laws of Bibliometrics generally follow power law distribution.
- All these laws are basically different versions of a single bibliometric distribution.

## 11. References

1. Bradford, Samuel Clement. Wikipedia<en.wikipedia.org>. Web. 1.5.2013.
2. Bradford, S. C. (1934) Sources of information on specific subjects. *Engineering*, 26: 85-86.
3. Bradford, S.C. (1953) The documentary chaos. In Bradford S C. *Documentation*. Crossby Lockwood. London. Ch. IX, p. 144-59.
4. Bookstein, Abraham. "Bibliographic distribution". *Library Quarterly* 46(1976): 416-23.
5. Brookes, B. C. "The derivation and application of the Bradford-Zipf distribution". *Journal of Documentation* 24 no.4 (1968): 247-265.
6. Brookes, B. C. "Bradford's law and the bibliography of science". *Nature* 224, Dec 6 (1969): 956
7. Brookes, B. C. Correspondence. "Scientific bibliography". *Nature* 227 Sept. 26 (1970):1377
8. Cole P F. "A new look at reference scattering". *Journal of Documentation* 18 no.2 (1962): 58
9. Egghe, Leo and Rousseau, Ronald . *Introduction to Informetrics : Quantitative Methods in Library, Documentation and Information Science*. Amsterdam : Elsevier Science Publishers, 1990.
10. Hertzfel, Dorothy H. "Bibliometrics, history of the development of ideas in statistical bibliography, or bibliometrics". *Encyclopedia of Library and Information Science* 42(1987): 144-219.
11. Hubert, John J. "On the naranan interpretation of Bradford's law". *Journal of the American Society of Information Science* 27 no. 5 (1976): 339-3.
12. Kendall M G. "The bibliography of operations research". *Operations Research Quarterly* 2(1960): 31-6.
13. Leimkuhler F F. "The Bradford distribution". *Journal of Documentation* 23 no. 3 (1967): 197-207
14. Naranan, S. "Bradford's law of bibliography of science: an interpretation". *Nature* 227, Aug. 8(1970): Power law. Wikipedia <[https://en.wikipedia.org/wiki/Power\\_law](https://en.wikipedia.org/wiki/Power_law)>'

15. Ravichandra Rao, I. K. An analysis of Bradford multipliers and a model to explain the law of scattering. *Scientometrics* 41 no.1 (1998 ), 93-100.
16. Ravichandra Rao, I. K . *Quantitative Methods for Library and Information Science*. New Delhi: Books in my Basket, 2003. Xii, 271.
17. Vickery, B. C. “Bradford’s law of scattering”. *Journal of Documentation* 4(1948): 198-202.