Annals of Library and Information Studies 52, 2; 2005; 51-55

THE SCATTERING OF ARTICLES OVER A JOURNAL SYSTEM AS OBSERVED FROM THE VIEWPOINT OF BOSE-EINSTEIN STATISTICS

Bidyarthi Dutta[1]
Librarian
St. Xavier's College
30 Park St.; Kolkata 700 016
E-mail: bidyarthidutta@rediffmail.com

B K Sen
80, Shivalik Apartments
Alaknanda
New Delhi 110 019
E-mail: bksen@ndb.vsnl.net.in

Rank vs. Number of articles distribution pattern of a journal system has been investigated from the viewpoint of Bose-Einstein statistics assuming a subject as equivalent to a phase space, a specific or microsubject as equivalent to a phase cell and corresponding journals as the Bose-Einstein particles.

INTRODUCTION

The journal is a looking glass through which the actual status of a subject can be understood. A journal may also be looked upon as an information shell in a subject space comprising of two axis: i) the intense or depth of the subject and ii) the extension or spread of the subject. These two parameters are mutually exclusive and thus cannot be reckoned simultaneously by any particular journal, although exceptions might be there.

The study of scattering of scientific papers in journals is an interesting phenomenon in scientometrics. Assuming a subject as a characteristic of an article, the scattering or distribution is shown to follow a clustering process of ranked groups, which was first noted by Bradford.

Bradford's Law describes an empirical relationship that depicts the distribution of scholarly articles, in any particular discipline, among journals, which are ranked in terms of the number of articles (in that discipline) published by them. Bradford found that a small core of journals publishes the bulk of articles related to a particular discipline. By ranking the journals in decreasing order of productivity, he was able to show that there exists an approximate linear relationship between the total number of articles published by journals up to some rank r and the natural logarithm of r. He observed that the ratio of the number of journals in the nucleus or core zone, allied zone and alien zone is 1: n: n2, where n is an integer. However, the journals of very high and very low productivity appeared to deviate from the simple linear form [1, 2].

This law actually deals with the relationship between sources and items (items being produced by sources), for example, articles 'produced' by journals or authors, citations 'produced' by articles and so on. All other classical informetric laws [Lotka (1926), Zipf (1949), Mandelbrot (1954,1977), and so on] deal with the same relationship, viz. source-item relationship.

Bradford's Law points out the similarity of journal productivity distribution with some well-known distributions and laws such as those described by Pareto (1895), Lotka (1926) and Zipf (1949). All these laws depict variations in populations of some sources and their activities producing some results or items distributed in a highly skewed manner where most of the items are concentrated in a small population of sources; the remaining items are spread over the rest of the population.

In practical terms, the Law may provide an important heuristic tool by which libraries and information services can decide the extent of journal coverage they wish to incorporate into their services in a cost effective manner.

The first theoretical expression for the scatter of journals was given by Leihmkuhler in 1967. It was subsequently simplified by Brookes to the form [3, 4]:

$$X(r) = a + k \ln r$$

1 Address for correspondence: 73/20, Golf Club Road, Kolkata 700 033

Where X(r) represents the total number of articles published in journal upto rank r; a and k are constants. This is the most widely used formulation of Bradford's Law, and describes the central linear portion of the curve X(r) vs. lnr.

JOURNAL SCATTERING AND BOSE-EINSTEIN STATISTICS: BACKGROUND

This statistics was first introduced by S.N Bose to explain the photon statistics. Subsequently, Albert Einstein developed this statistics to include the particles having integral or zero spin. The particles obeying this statistics are known as bosons.

Here, the main objective is to interpret Bradford's scattering phenomenon with Bose-Einstein statistics [5]. The main assumption made to set the proposition is:

- A journal system behaves like the boson particles; that's why a subject has been assumed to be equivalent with a phase-space; a specific or microsubject has been assumed to be equivalent with a phase-cell and journals are similar to boson particles. The energy of a boson particle is assumed to be equivalent with the productivity of the journal and the temperature of a boson particle is assumed to be equivalent with the rank of the journal. The temperature of a particle represents its state or level and the energy tells us about the internal potential of the particle in the phase space. Similarly, the productivity of a journal presents its internal potential while the rank designates the level or the actual position of the journal in the concerned subject domain.
- After making this assumption let us now stipulate the fundamental assumptions of Bose-Einstein statistics along with the analogies with the journal-system for a subject-domain which are given in the adjoining parentheses:
- The particles are independent, identical and indistinguishable from one another (The journals pertaining to a subject-domain are also identical; may or may not be independent depending on several conditions of the authors involved and the affiliating institutions; distinguishable as indistinguishability occurs only at micro level, but not at macro level)
- The total number of particles (N) in the system is constant

  (The total number of journals N_1 in a particular journal- system is also constant in a particular subject-domain for a particular time)
- A phase-cell can contain any number of particles ranging from zero to N.

  (A specific subject-domain can contain any number of journals ranging from zero to N_1)
- The total energy of the system is constant

  (The total productivity of the entire journal-system is constant)

Attempts have been made to find a theoretical foundation for the literature scattering phenomenon. The dynamic process of allocation of articles among the journals has been modeled in 1970 by Naranan [6] in terms of a power law distribution under the assumption that articles and journals grow exponentially in time. Karmeshu [7] et al put two models to interpret Bradford's law: one based on the random subdivision of the papers over the field of journals, and the other model based on Shockley's ideas of individual scientific productivity. The scattering phenomenon was also interpreted by Yablonsky [8] using stable non-Gaussian distribution. Sen [9, 10] gave an interpretation of bibliographic scattering in a different way, viz. from a generalized source approach based on Bose-Einstein statistics. Basu [11] derived a theoretical expression for the distribution of articles in journals, based on a model of the random and unequal partitioning of the total number of articles. A qualitative shift from the conventional conceptual framework within which the Bradford's distribution has usually been considered already starts by attempting to put it within the frame of some other non-informetric laws [12]. The highly skewed, non-Gaussian nature of distribution of Bradford's pattern has recently been attempted to be clarified on the

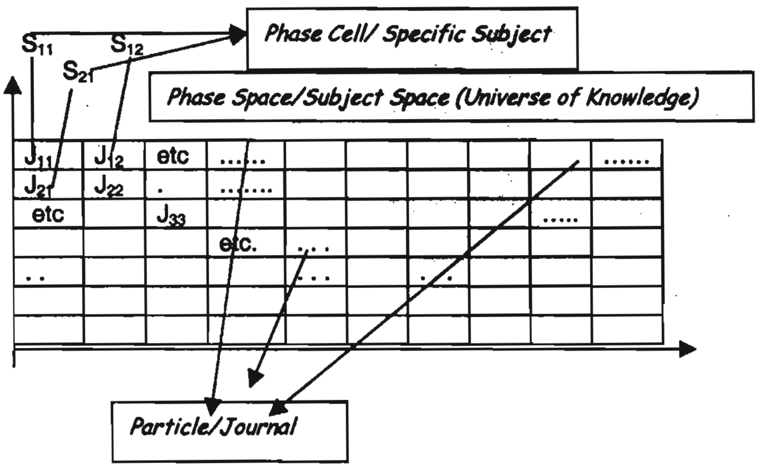

Figure 1 — Journal Scattering and Bose-Einstein Statistics: Mathematical Derivation

basis of Stankov's 'Universal law' and a new wave concept of scientific information has been propounded [13].

The analogy between the Bose-Einstein phase space and a subject domain is shown in the Figure 1.

From Bose-Einstein distribution, we know that the average number of particles n_{av} with an average energy E_{av}, at a temperature T is given by the following equation [14]:

$$n_{av} = 1/\{\exp[(E_i - E_{av})/kT] - 1\} \qquad \text{........ (1)}$$

Where, E_i is the kinetic energy of the i^{th} particle and k is Boltzman's constant.

Let us now consider a collection of N journals ranked from 1 to r. Now the average number of journals (N_{av}) having rank r_i, instantaneous productivity μ_i and average productivity over the collection μ, from the analogy of Bose-Einstein distribution can be given by:

$$N_{av} = 1/\{\exp[(\mu_i - \mu)/r_i] - 1\} = 1/[\exp(M/r_i) - 1] \qquad \text{...(2)}$$

Where $M = \mu_i - \mu$ = Deviation of productivity from the average value. The number of journals is altered by journal-rank, or it can be said that the number of journals is a function of journal-rank. In Bose-Einstein system, the average number of particles changes with the absolute temperature of the system, similarly in a journal-system the average number of journals also changes with the rank of the system. Again, the number of journals is a function of the number of articles contained in the journals under consideration; and thus the number of articles contained in the journals is a function of journal-rank also. The equality of journal-rank with the journal-productivity deviation may occur under certain circumstances.

Taking derivative of average number of journals with respect to journal-rank, we get: -

$$(dN_{av}/dr_i) = [M \exp(M/r_i)] / \{(r_i)^2 [\{\exp(M/r_i)\} - 1]^2\} \qquad \text{...(3)}$$

Now, $(dN_{av}/dr_i) = (0.9194/r_i) \sim (1/r_i)$, if $M = r_i$ ...(4)

i.e. fractional change in average number of journals with respect to corresponding change in journal-rank is almost equal to inverse of journal-rank provided the journal-rank is equal to the deviation of journal-productivity from the average value. The deviation of journal-productivity from the average value for an arbitrary journal gives the deviation in mean citation age among the said journals, which in turn may provide some idea about the rank of that journal [15].

Now, the average number of journals in a collection $[N_{av}(x)]$ is a function of the number of articles in the journals under consideration for a given subject is well described by a simple power law frequency function [16]. Thus we can write:

$$N_{av}(x) = A(x)^{(-\gamma)} \quad \text{....(5)}$$

Where, x = Total number of articles in a collection of N journals and 'A' and 'Y' are constants.

Substituting equation (5) in equation (4) we get:

$$d\,[Ax^{(-\gamma)}]/dr_i = (1/r_i)$$

$$\text{or, } [(-A\gamma)x^{(-\gamma-1)}]\,dx = dr_i/r_i \quad \text{...(6)}$$

Integrating both sides we get:

$$X^{(-\gamma)} = (\ln r + b)/A \quad \text{...(7)}$$

Where, b is the constant of integration and r is the rank.

Equation (7) presents a relationship between the number of articles in a collection of journals and the rank of the journals. The form of equation (7) is alike to the equation:

$$X(R) = a + k\ln R \quad \text{...(7a)}$$

Where k = 1/A, a = b/A and X(R) has been replaced by $x^{(-\gamma)}$. We thus have arrived Brooke's simplified equation through Bose-Einstein distribution equation. In a collection of N journals, it is possible to find out the rank of a certain journal after knowing the total number of articles in the said collection. Let us now consider an arbitrary journal collection belonging to a particular subject where number of articles contained by the first ranked (Rank=1) core journal is N (say) and the lowest rank (Rank=R) of the alien journal containing only single article (X=1) is R (say);

i.e. $(x)_{max} = N$ when $(r_i)_{min} = 1$ (this case corresponds to core journal) and

$(r_i)_{max} = R$ when $(x)_{min} = 1$ (this case corresponds to alien journal)

Substituting maximum and minimum values of x and r_i in eq. (7) and then dividing we get

$$\ln R = b\,(N^{\gamma} - 1) \quad \text{...(8)}$$

As L.H.S of eq. (8) is a positive quantity (since R>1), therefore either $N^{\gamma}>1$ and b>0; or $N^{\gamma}<1$ and b<0. For fairly large number of core journal articles, $N^{\gamma} >> 1$ and we get:

$$\ln R = b\,N^{\gamma} \quad \text{...(9)}$$

$$\text{or, } \gamma = [\ln (M/b)/ \ln N] \quad \text{...(10)}$$

$$\text{Where } M = \ln R$$

Eq (10) reveals the fact that γ will be negative only when M < b; or, $\mathbf{N}^{\gamma} < 1$; which contradicts the assumption on which eq (9) has been derived. Hence ã can't be a negative quantity. Hence the possibility of $N^{\gamma}<1$ and b<0 has been excluded. Thus $N^{\gamma}>1$ and b>0, i.e. b is also a positive quantity just like γ.

Equation (7) may be rewritten as:

$$X = B\,[1 + (\ln r/b)]^{(-1/\gamma)} \quad \text{..(7b)}$$

$$\text{Where, } B = (b/A)^{(-1/\gamma)}$$

$$\text{Or, } X = B\,[1 - (\ln r/\gamma b)] \quad \text{...(7c)}$$

Expanding binomially for ln r < b. Derivation of equation 7(c) from equation 7(b) is therefore constrained by an imposed condition, i.e. ln r < b.

$$\text{Or, } r < \text{Exp}(b) \quad \text{...(11)}$$

Equation (7c) exactly represents Brooke's simplified equation (equation (7a)), where, a = B and k = -(B/γ b). Brooke's simplified equation is derivable from Bose-Einstein distribution function with some conditions. The journal system obeys Brooke's simplified equation for rank (r) less than Exp (b), which corresponds a perfect consonance with the said equation for some limited alien journals only, as alien journals hold higher ranking values.

CONCLUSION

In this paper, we have derived a theoretical expression for the distribution of articles in journals, based on Bose-Einstein statistics and we have derived Brooke's simplified equation from Bose-Einstein Statistics. Brooke's simplified equation is thus a limiting case of Bose-Einstein Statistics for rank (r) < Exp (b). Journals belonging to a particular subject are assumed to be equivalent to Boson particles. The articles are assumed to be related with the journals through the 'Power law'. An attempt for setting up of a theoretical background of Bradford's distribution will enable to examine critically and more scientifically other classical informetric laws.

ACKNOWLEDGEMENT

Valuable suggestions from Dr. K.P. Majumder, Head, Department of Library and Information Science, Jadavpur University, Kolkata and friendly assistance from Mr. Anup Kumar Das, Research Scholar, Department of Library and Information Science Jadavpur University, Kolkata are gratefully acknowledged.

REFERENCES

1. BRADFORD (S C). Sources of information on specific subjects. *Engineering.* 137; 1934; 85-86.
2. BRADFORD (S C). Documentation. London: Crosby Lockwood; 1948.
3. BROOKES (B C). The derivation and application of the Bradford-Zipf Distribution. *Journal of Documentation.* 24; 1968; 247-65.
4. BROOKES (B C). Theory of the Bradford law. *Journal of Documentation.* 33, 180-209, (1977).
5. Huang (K). Statistical mechanics. New Delhi: Wiley Eastern Ltd., 1999.
6. NARANAN (S). Bradford's law of bibliography of science: an interpretation. *Nature.* 277; 1970; 631-632.
7. KARMESHU (N C), Lind (V C). Rationales for Bradford's law. *Scientometrics.* 6; 1984; 233-241.
8. YABLONSKY (A I). Stable non-Gaussian distributions in bibliometrics. *Scientometrics.* 7; 1985; 459-470.
9. SEN (S K). Bibliographic scattering: a generalised source approach. *Scientometrics.* 17; 1989; 197-204.
10. SEN (S K). A note on theoretical correlation between Bradford's law and recently proposed linear equation of the type R(r) = a.r – b. *Scientometrics.* 17; 1989; 205-210.
11. BASU (A). Hierarchical distribution and Bradford's law. *Journal of the American Society for Information Science.* 43; 1992; 494-500.
12. VUKOVIC (V O). Bradford's distribution: from the classical bibliometric "Law" to the more general stochastic models. *Journal of the American Society for Information Science.* 48; 1997; 833-842.
13. IVANCHEVA (L E). The non-Gaussian nature of bibliometric and scientometric distributions: A new approach to interpretation. *Journal of the American Society for Information Science and Technology.* 52; 2001; 1100-1105.
14. REIF (F). Fundamentals of statistical and thermal physics. Singapore: Mcgraw-hill Book Company, 1965.
15. EGGHE (L). A noninformetric analysis of the relationship between citation age and journal productivity. *Journal of the American Society for Information Science and Technology.* 52; 2001; 371-377.
16. NARANAN (S). "Power Law" version of Bradford's law: statistical tests and methods of estimation. *Scientometrics.* 17; 1989; 211-226.