Library and Information Science No. 24 1986

Bradford の法則の数式表現——その歴史的展開

Mathematical Formulations of Bradford's Law; a Discussion of its Historical Development

海　野　　　敏
Bin Umino

Résumé

In this paper, the author reviews the diverse studies on mathematical formulations of Bradford's law, which is a representative distribution law in bibliometrics.

Since S.C. Bradford presented the law in 1934, there have been three trends of the study. In the first trend, researchers tried to formulate a law following the proportional expression originally shown by Bradford. The studies of Vickery, Leimkuhler, Wilkinson, and others are classified in this trend. In the second trend, researchers tried to lead a formula according to the shape of graphs describing observed values. The studies of Brookes, Haspers, Asai, and others are classified in this trend. And in the third trend, researchers have attempted to produce a formula from various probabilistic models. The author shows Simon's model, Onodera's model, and Naranan's model as sophisticated mathematical models.

These formulations are so diverse that there is still no formula which can be recognized as a universal one. The author concludes that to test the fitness of each formula with observed data stastistically and to establish the most desirable formulation of Bradford's law are urgent tasks in bibliometric studies.

海野　敏：東京大学教育学研究科修士課程，東京都文京区本郷 7-3-1
Bin Umino, Graduate School of Education, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo.

I. はじめに

Prichard が“図書や他のコミュニケーション・メディアに対して数学と統計的方法を適用する”[1] 研究テーマに‘bibliometrics’なる名称を与えてから15年余りが経過した。この間多種多様な研究がこの名のもとに展開され，図書館情報学の理論と応用両面に多くの成果が蓄積されてきた。その結果現在では，「ビブリオメトリクス」は，記録情報の蓄積・組織から提供・利用に至るプロセスの中で出現する様ざまな数量データを，数学的・統計学的な方法を用いて分析する図書館情報学の研究領域，あるいはその研究手法に対して与えられた呼称として，一般に広く認められるようになっている[2]。

この領域には，書誌的な数量の分布を記述したいくつかの法則が存在する。中でも最も著名なのは雑誌論文の分布に関する Bradford の法則と，科学者の生産性に関する Lotka の法則であり，性格は異なるが単語の出現頻度に関する Zipf の法則もよく知られている。これらのオリジナルな法則自体は，ビブリオメトリクスなる用語が提示される数十年前に発見されたものであり，しかもごく単純な経験則にすぎない。しかし，後の多くの研究者の研究対象となって洗練かつ一般化され，ビブリオメトリクスの発展に一貫して大きな刺激を与えてきたという点で，これらの分布法則はいわばビブリオメトリクスの代名詞ともいえるものである。

さて，Bradford の法則は50年以上も前にすでに発見されていたものであり，図書館における雑誌選択や，主題書誌作成の方針の理論化に有効な手段を与えるものとして，多くの研究者に様ざまな形でとりあげられてきたものである。ところが，この法則に関してはひとつの意外な事実が存在している。それは，この法則の内容を数式で記述するにあたって，50年を経ていまでも，一般的に認められた表現方法が確立していないという事実である。Bradford の法則を研究の対象とする研究者たちは，それぞれ独自の数式表現を考案，採用して自らの論を進めており，「これが Bradford の法則の望ましい数式表現である」いとう了承を広く得た表現式は，今なお見出だすことができない。

このような現状を省みて，筆者は従来提示されてきた多様な数式表現を，あらためて整理しなおすことを試みた。すなわち本稿は，ビブリオメトリクスの代表的な分布法則である Bradford の法則が，その発見当初から現在に至るまで各研究者によってどのような形で数式表現されてきたのか，その表現方法にいかなる展開，進展が見られたのかを歴史的にあとづけることをねらいとするものである。本稿では，II章からIV章までがこの考察にあてられている。またV章では，まず社会科学の分野の法則に関して，望ましい数式表現の一般的な条件をいくつかあげ，つぎにそれらの条件に照らして Bradford の法則に対する従来の数式表現を検討しなおしている。これは，いまだに見いだされていない「Bradford の法則の望ましい数式表現」を確立するための，準備作業のつもりで試みたものである。

II. Bradford による法則表現とその数式化

A. Bradford による法則の記述

Bradford の法則の数式表現がどのように展開してきたかを探るにあたって，まずはじめに確認しなければならないのは，発見当初，法則は Bradford 自身によってどのように表現されたのかということである。彼が法則をはじめて提示したのは 1934年の論文であるが[3]，この論文で彼が論証しようと試みたのは，特定の主題の論文が多数の雑誌の中に散らばっている，その分布の状態に関する，ある経験的な仮説であった。すなわち，特定主題の論文が，その主題を専門に扱っている少数の雑誌に，１誌あたり数多く掲載されているのは当然として，それ以外の多分野にわたる雑誌にも少しずつ広く分散して掲載されているという仮説である。この経験的に推測

される事実を，Bradford はデータの収集とその計量的分析によって確認し，定式化しようとしたのである。

彼はこの仮説を確かめるために，応用地球物理学と「潤滑」という2つの主題領域に対して独自の文献調査を行い，まず各雑誌タイトルが関係論文（特定主題の論文）をいくつ含んで(掲載して)いるかを計数している。つぎに彼は，雑誌を，含んでいる関係論文の数の多い順，いいかえればその主題についての生産性の高い順に並べかえ，全雑誌をその順位にしたがって，ほぼ同数の関係論文を含む3つの集合に分割するという作業を行なっている。その結果得られたのは，応用地球物理学に関する論文をそれぞれ429，499，404ずつ含む，9，59，258の3つの雑誌集合と，「潤滑」に関する論文をそれぞれ110，133，152ずつ含む，8，29，127の3つの雑誌集合である。ここにおいて Bradford は，この3つの集合の大きさの比がいずれもほぼ $1:5:5^2$ (=25) になっていることに着目し，この発見を一般化してつぎのように記述したのである。

> “科学雑誌における，ある主題に関する論文の分布の法則は，次のように述べることができる。もしも科学雑誌が，ある主題に関する論文の生産性の高い順に並べられたとすれば，それらは，その主題をより特別に扱っている核 (nucleus) と，その核と同数の論文を含むいくつかのグループ，あるいはゾーンに分けることができる。このとき，核とそれに続くゾーンにそれぞれ含まれる雑誌の数は，$1:n:n^2\cdots$ になる”[3]。

これがいわゆる Bradford の法則のオリジナルな表現である。雑誌群における特定主題の論文の集中と分散の状態は，この表現においては同数の論文を含む一連の雑誌集合（核とゾーン）の大きさの変化によって記述され，彼の経験的仮説は，ゾーンの大きさの幾何級数的な増大という形で確認され定式化されているのである。なお，定数は後の研究者たちによって「Bradford の乗数」(Bradford's multiplier) と名づけられている[4]。

Bradford は，発見した法則をこのように文章と比例式で表現しているのだが，同一論文中の別の箇所では，法則をグラフによっても表現している。彼が作成したのは，雑誌の累積度数の常用対数を横軸，論文の累積度数を縦軸にとった片対数軸のグラフである。雑誌は含んでいる関係論文の数の多い順に並べられているので，雑誌の累積度数とは雑誌の生産性の順位に相当する値である。彼はこれがはじめのたわんだ部分のずれをのぞいてほぼ直線になることを示し，これを，“関係する雑誌が生産性の高い順に並べられたとき，論文の累積数は，生産性の高いはじめの部分を除けば，雑誌数の対数に比例している”[3] と言い表わしている。この表現において彼の経験的仮説は，雑誌の累積数の増加に対して関係論文の総数は対数関数的にしか増加しない，すなわち急速に増加率が減少するという形で確認され定式化されているのである。

以上が Bradford 自身による法則の2つの表現方法である。これらの表現方法は次節で述べるように区別される必要があるので，本稿では前者を「比例式による表現」，後者を「グラフによる表現」と呼ぶことにする[5]。

Bradford による法則表現には，以降の研究者の数式表現との関係上，注目すべき特徴が4つある。それは，(1)同一の現象を，雑誌群のグループ分けにもとづく比例式と，片対数軸をもつグラフという全く異なった2通りの方法で表現したこと[6]，(2)実際には雑誌群を3つにしかグループ分けしていなかったにもかかわらず，比例式による表現では任意の数へのグループ分けに一般化して表現したこと，(3)グラフによる表現は，生産性の高い部分を除外した上での表現であったこと，(4)分布を記述するための変数として，雑誌を生産性の順に並べたときの累積数，すなわち生産性の順位を採用していることである。(1)と(2)については次節で，(3)についてはⅢ章で，(4)についてはⅣ章で，それぞれ論ずることにする。

B. Vickery の指摘

Bradford の法則の数式表現に関する最初の重要な研究論文は，Vickery によって1948年に発表されている[7]。彼はこの論文で，Bradford 自身の示した法則の2つの表現，すなわち比例式による表現とグラフによる表現は，実は数学的に異なった内容をもつものなのに，Bradford および彼の後継の研究者たちがこれらを混同して論じていることを指摘している。

2つの表現の数学的な不一致は，つぎのように証明されている。今，雑誌を特定主題について生産性の高い順に並べたとき，関係論文を合わせて s 含む最も生産性の高い雑誌の数を R_s としよう。R_s は雑誌の累積数であり，また雑誌の生産性の順位をも表わしている。Bradford のいう核の雑誌群に含まれる論文数 c をとすれば，核の雑誌数は R_c である。この表記にしたがえば，各ゾーンに含まれる雑誌数は雑誌の累積数の差として求められるから，比例式による表現は，

$$R_c : R_{2c}-R_c : R_{3c}-R_{2c}=1 : b : b^2 \qquad (2.1)$$

と表わすことができる。ただし，b は定数であり，比例式による表現では Bradford の乗数に相当するものである。一方，グラフによる表現を同じように表わせば，グラフの直線性，すなわち傾き一定なる性質から，

$$R_c : R_{2c} : R_{3c}=1 : b : b^2 \qquad (2.2)$$

と表わされるはずである。(2.1)と(2.2)は明らかに異なった式であるから，Bradford の２つの表現は数学的には同一のものではない。

Vickeryはさらに，比例式による表現から法則のある一般的な特性を導いてみせている。いま，関係論文を合わせて ic 含む最も生産性の高い雑誌の数 R_{ic} は，$p=R_c$ とおけば，(2.1) から，

$$R_{ic}-R_{(i-1)c}=pb^{i-1}$$
$$\therefore \quad R_{ic}=p(b^i-1)/(b-1) \qquad (2.3)$$

と表わせる。ここで，$d=p/(b-1)$ とおくと，(2.3) から，

$$R_{ic}=d(b^i-1) \qquad (2.4)$$

となり，さらに log を自然対数として，$\log b=\alpha c$，$I=ic$ とおくと，

$$R_{ic}+d=db^i=de^{\alpha ic}$$
$$\therefore \quad R_C+d=de^{\alpha I} \qquad (2.5)$$

となる。ここから Vickery は，横軸に関係論文の累積数，縦軸に雑誌の累積数の対数をとったグラフの曲線に関して[8]，つぎのような特性があると結論を述べている。

“乗数を b のかわりに b^k とみなせば，グループの大きさを c のかわりに kc としても，最後の式［(2.5)のこと—筆者注］と全く同じ式が得られる。したがって，b のある値に対して，事実と十分に適合するような理論的な曲線が得られるならば，他のどんな b の値に対しても，事実と同じように適合するこれと同一の曲線を得ることが可能なのである”[7]。

すなわち彼は，比例式によって記述された法則がいったん成立すれば，ゾーンの数，ゾーンの大きさにかかわらず，$1 : b : b^2 : \cdots$となるような b の値が存在することを証明してみせたのである。

しかし，この証明は誤っている。b に b^k，c に kc を代入したとき，(2.5) の式が同じく成り立つためには，等しい α，d が導かれなければならないはずだが，α は等しくなるが，d が等しくなるという保証はない。d が等しくなるためには，上で結論されたような性質が R にあらかじめ仮定される必要があり，トートロジーに陥ってしまう。どうやら，彼の考えていた理論的な曲線には，ゾーンの数や大きさによらず成立するという条件がすでに含まれていたように思われるのである。

このように若干の誤りはあるものの，結果的に彼は，Bradford があいまいにしか表現しなかった，雑誌群の任意の数のゾーンへの分割を示唆し，Bradford の法則を，ゾーンの数と大きさによらず成立するものとしてあらたに提起しなおしたものとして評価できる。Bradford の行なった２表現が同一ではないという指摘に加えて，この点でも，Vickery は法則の数式表現の洗練化に少なからぬ貢献をしたのである。

C. 比例式による表現の忠実な数式化

Bradford の法則は，本来特定主題の論文の雑誌群への分布状態に関する記述である。ところが比例式による表現は，その分布状態を一見して明らかにするものとはいいがたい。複数のゾーンへの雑誌群の分割というアイデアはユニークではあるが，法則の扱いをむずかしくする大きな原因となっている。そこで，比例式による表現をわかりやすくかつ扱いやすいものとするために，これを２変量の共変関係を示す数式に表現しなおす試みが，いく人かの研究者によって行われている。

その最も初期のものは，すでに紹介した Vickery の (2.3) の式である。この式の内容を改めて考えてみれば，これは生産性の高い i 個のゾーンを構成している雑誌の総数を，核を構成している雑誌の数と Bradford の乗数で表わしたものであるといえる。

同様の表現は，Goffman と Warren の69年の論文にも見出だすことができる[4]。彼らは，特定主題の論文を含んでいる雑誌群を，生産性の順に並べ，同数の関係論文を含む k 個のゾーンに分割したとき，各ゾーンを構成する雑誌の数をそれぞれ J_1，J_2，…，J_k とすれば，Bradford の法則は，

$$J_i=b_k J_{i-1}=b_k^{i-1}J_1 \qquad (2.6)$$

と表わすことができると述べている。ただし，パラメータ b_k は，雑誌群を k 個に分割したときの Bradford の

乗数である。これは，i 番目のゾーンを構成する雑誌数を，核を構成する雑誌数 J_1 で表わした式である。b に添え字をつけて，これが分割するゾーンの数によって変化することを明示していることからもわかるように，彼らは Vickery と同じく，Bradford の法則がゾーンの数にかかわらず成立することを前提として，その比例式による表現を数式化したのである。

Vickery の式も Goffman らの式も，Bradford 自身の表現よりは数学的に洗練されてはいるものの，論文の雑誌群への分布を表わす式としては，やはりわかりにくい。その原因は，いずれの式も雑誌群の核とゾーンへの分割という考え方に捕らわれているために，論文数と雑誌数の関係を直接に示す式にはなっていないからである。彼らと同じく比例式による表現から出発して，Bradford の法則を改めて雑誌数と関係論文数の関係として数式化したのは，Leimkuhler である[4]。

Leimkuhler も，Bradford の法則がゾーンの数にかかわらず成立することを前提として論を進めた点は Vickery, Goffman らと同様である。しかし彼らと異なり，各ゾーンの雑誌数そのものではなく，雑誌の総数に対する各ゾーンの相対度数に着目したところに，Leimkuhler が独自の数式を展開できた理由がある。いま，雑誌群を論文数が等しい k 個のゾーンに分割したときの Bradford の乗数を b_k とし，k 個のゾーンのうち生産性の上位の i 個のゾーンに含まれる雑誌数の全雑誌数に対する相対度数を R_{ik} と表わしたとき，Leimkuhler はまず比例式による表現から，

$$R_{ik}=(b^{i/k}-1)/(b-1) \qquad (2.7)$$

なる式が導けることを証明している。ただし，$b=b_k{}^k$ である。ここで，$r_0(0\leqq r_0\leqq 1)$ を上位雑誌数の全雑誌数に対する相対度数とし，$G_0(0\leqq G_0\leqq 1)$ を r_0 に対応する雑誌に含まれている論文数の全論文数に対する相対度数とすれば，$i/k=G_0$，$R_{ik}=r_0$ と考えることができる。したがって (2.7) は，

$$r_0=(b^{G_0}-1)/(b-1) \qquad (2.8)$$

と表わせる。(2.8) より，G_0 を r_0 の関数として表現すれば，

$$G_0(r_0)=\log(1+br_0-r_0)/\log b$$
$$\therefore\ G_0(r_0)=\log(1+\beta r_0)/\log(1+\beta) \qquad (2.9)$$

ただし，$\beta=b-1$ である。

(2.9) の式において，G_0 は，r_0 を確率変数とする累積相対度数分布，すなわち分布関数を表わしている。かくして Leimkuhler は，Bradford の法則を論文数と雑誌数の関係を示す関数の形ではじめて表現したのみならず，法則を統計学的分布の記述として認識した表現にはじめて成功したのである。Brookes は，Leimkuhler の提示した数式が，ドキュメンタリストが伝統的な方法とは異った新たな方法を用いて情報システムの設計，情報サービスの計画を行うことを可能にしたと評価している[9]。

さて，Leimkuhler の67年の論文では，このように雑誌と論文の相対度数が法則を表現する変量として採用されているが，彼の77年の論文では，同じく比例式による表現から，雑誌と論文の度数の関係を直接に示す式が，67年とほぼ同じ手順で導かれている[10]。いま，生産性の高い r の雑誌に含まれている関係論文数の総数を G_r としたとき，

$$G_r=B\log(1+\beta r)/\log(1+\beta) \qquad (2.10)$$

なる式が，77年の表現式である。r は雑誌の生産性の順位とも考えられることに注意しておいてもらいたい。ここでは，分布が相対度数に標準化されていないために，新たにBというパラメータが導入されているが，形式的には (2.9) と本質的な差異はない。

比例式による表現から導かれた法則の数式表現をもうひとつ紹介しよう。Wilkinson が72年の論文の中で示した式がそれである[11]。彼女は Leimkuhler とよく似た手続きで，

$$G(r)=c\log\{r(b-1)/a+1\}/\log b \qquad (2.11)$$

なる式を導いている。ここで G, r は (2.10) と同じものを表わし，a は核を構成する雑誌数，b は Bradford の乗数，c は核の雑誌に含まれる論文数を表わしている。これは，$j=c/\log b$, $t=a/(b-1)$ なるパラメータを用いれば，

$$G(r)=j\log(r/t+1) \qquad (2.12)$$

となる。これらの式は Leimkuhler の式とみかけ上異なっているが，(2.12)が (2.9) の式と形式的に対応するものであることは，Wilkinson 自身の手によって証明さ

れているし，彼女の論文よりも後に提示された (2.10) とも対応している。

ところで，Fairthorne は Wilkinson が上述の式を提示する以前の69年に，Bradford の法則を

$$G(r)=k\log(1+cr) \tag{2.13}$$

と表わしているが，これは(2.12)の式と極めて似た形をもった式である[12)]。彼自身は，この式をどのようにして導いたのか，2つの定数が何を表わしているのか一切明らかにしていないが，式の形からみて，彼もまた比例式による表現から表現式を導出したのではないかと推測することは十分に可能であろう。

III. 観測データにもとづく数式表現の修正

A. 理論値と観測値との食いちがい

前節で紹介したいくつかの数式は，「Bradford の法則が分割するゾーンの数によらず成立する」という法則の一般化は施されているものの，基本的には Bradford の比例式による表現をそのまま忠実に数式化したものであった。それでは，比例式による表現とは異なる内容をもつ「グラフによる表現」を忠実に数式化すれば，どのような式が得られるのであろうか。

前述したように，Bradford は雑誌の累積度数の常用対数を横軸，論文の累積度数を縦軸にとった片対数軸のグラフを描き，その直線性を発見して“雑誌が生産性の高い順に並べられたとき，論文の累積数は，雑誌数の対数に比例している”[3)] と述べている。いま，生産性の高い r の雑誌に含まれている関係論文の累積数を $G(r)$ とおいて，彼の表現をそのまま数式化すれば，

$$G(r)=k\log r \tag{3.1}$$

となるはずである。ここで k は比例定数を表わすパラメータである。ところが後の研究で，この式を Bradford の法則を表現するものとして採用しているものは，ほとんど見あたらない。次節が述べるように，Brookes が，グラフによる表現から導かれる式とは自覚せずに，同じ形の式を示しているぐらいである[9)]。

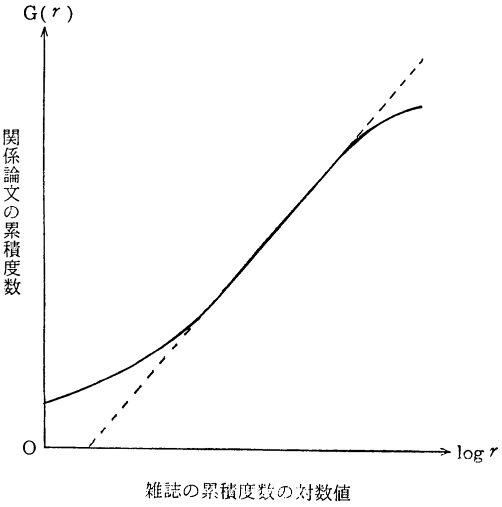
(3.1) の式が法則を表現する数式として採用されがたい理由は，これが現実に観測されたデータとあまりに食いちがった値を示すからである。もし仮に (3.1) の式が関係論文数と雑誌数の関係を正しく記述するものだとすれば，片対数軸のグラフは原点を通り，一貫して傾きの一定な直線にならなければならない。しかし，観測データから実際に作成されるグラフは，これとははなはだ異なった様相を呈するのである。

(3.1) の式が示すグラフと観測データが示すグラフの食いちがいは，3つの形で現われている。(以下第1図を参照。) 第1の食いちがいは，グラフの下方のたわみである。Bradford が，グラフの直線性を生産性の高いはじめの部分を除いた部分で成立するものと条件づけしたのは，このたわみに気づいていたからである。この食いちがいは，(1)生産性の最も高い部分に属する各雑誌に含まれている関係論文数が，直線グラフの示す理論より大きいこと，あるいは(2)生産性の最も高い部分に属する雑誌数が，理論値より小さいことを示すものといえる。

第2の食いちがいは，グラフ上方のたれ下がりである。この食いちがいは Groos によってはじめて指摘されたので，「Groos のたれ下がり」(Groos droop) とも呼ばれている[13)]。この食いちがいは，(1)生産性の最も低い部分に属する各雑誌に含まれている関係論文数が，直線グラフの示す理論的な値より小さいこと，あるいは(2)生産性の最も低い部分に属する雑誌の数が，理論的な値より大きいことを示している。Groos は，このたれ下がりが示している，法則の観測データに対する不適合を重くみて，“Bradford は最少数の関係論文を含む雑誌数のパーセンテージを過少に評価し，かくして根拠の薄い比例式を導き出したのだ”[13)] と厳しい批判を行なっている。

下方のたわみと上方のたれ下がりの部分を除けば，片対数のグラフが直線で近似できることは，Bradford 以下多数の研究者が実際に収集したデータから描いたグラフを見れば，直観的にほぼ明らかである[14)]。しかしこの直線は，下方に延長したときに横軸に正の切片，縦軸に負の切片をもち，原点は通らない。これが第3の食いちがいである。この食いちがいはグラフの下方にたわみが存在していることから説明することも可能だが，グラフを数式化するにあたって直線部分が原点を通らないということは重要なポイントとなるので，これを3番目の食いちがいとしてあげておくことにする。

以上のような理論値と観測値の食いちがいの存在は，これを克服するような法則の数式表現を探る一連の研究を促した。これらの研究は，理論値と観測値のずれを認識しつつも基本的には片対数軸のグラフの直線性を認めて，グラフの形状から観測値をよりよく記述しうる数式



第1図 観測データから描いたグラフの概形

表現を直接に導きだそうとするものである。これらの研究を行なった研究者たちは，もはや Bradford による法則のオリジナルな表現にこだわっていない。彼らは，観測データから描いたグラフの曲線をよりよく近似する数式表現を追求することで，法則の表現のみでなく，その内容までも修正，洗練しようと試みたのである。またグラフから導いたがゆえに，彼らの式はいずれも，関係論文の累積数とそれを含む雑誌の累積数の関係を記述するものとなっている。このような一連の研究をつぎにみてみたい。

B. グラフによる数式表現の修正

ここで，すでに使用してきたものも含めて，いくつかの記号をまとめて定義しておくことにしよう。いま，特定主題に関係する論文をひとつ以上含む雑誌を，その含んでいる関係論文数の多い順，すなわち生産性の高い順に並べたとき，以下のように記号を定めておく。

N：関係論文をひとつ上以含む雑誌の総数

P：関係論文の総数

r ：生産性の高い方からの雑誌の累積度数，あるいは雑誌の生産性の順位

r_0 ：生産性の高い方からの雑誌の累積相対度数 (r/N)

G：関係論文の累積度数

G_0：関係論文の累積相対度数（G/P）

また以下 log は，ことわりのないかぎり e を底とする自然対数である。これら以外の記号は，そのつど定義してゆくことにする。

グラフの形状から直接に数式を導き出す手法は，まず前述した Vickery の論文に見出だすことができる[7]。彼の示したのは

$$G=k\log r+a \tag{3.2}$$

なる式である。これは，x 軸 $\log r$，y 軸 G の片対数軸のグラフが，傾き k，y 切片 a の直線になることを表わしている。前に，ゾーンへの分割という考え方にこだわらず，法則をはじめて雑誌数と関係論文数の関係を示す関数の形で数式化したのは Leimkuhler だったと述べたが，ゾーンへの分割にこだわらずに2変量の関係を数式化する試みは，すでにここで実現されていたといえよう。

同じく Leimkuhler 以前に，Cole も，グラフの直線性からつぎの4つの式を導いている[15]。

$$G_0=a+k\log r_0 \tag{3.3}$$

$$\boldsymbol{r_0}=\mathrm{antilog}\,\{(G_0-a)/k\} \tag{3.4}$$

$$P=k'\log N-a' \tag{3.5}$$

$$N=\mathrm{antilog}\,\{(P+a')/k'\} \tag{3.6}$$

ただし，antilog とは e を底とする指数関数である。実のところ Cole の式は，自ら収集した観測データを記述するために提示されたもので，上の式のパラメータ a，k，a'，k' には具体的な数値（すべて正の値）が入っており，このように一般化した式がはっきり示されているわけではない。しかし，彼が法則を，これらの式の示しているような2変量間の関係を記述するものとしてとらえて数式化を行なっていることは確かである。

さて，Brookes は，68，69年の3つの論文で，Bradford の法則を表現する数式として，以下に説明するような5つの式を提示している[9][16][17]。これらは，当初の単純な数式を，グラフの曲線をよりよく近似するように段階的に修正を加えていった結果得られた一連の式である。68年の論文で，まず彼は Bradford 自身の法則表現から，

$$G(r)=G(r^2)-G(r)=G(r^3)-(r^2)=\cdots \tag{3.7}$$

なる式を導き，これを

$$G(r^i)=iG(r) \qquad (i=1,\ 2,\ 3,\ \cdots) \tag{3.8}$$

と一般化している。ところが (3.8) を満たすよう単調増

加の関数Gは，

$$G(r)=k\log r \qquad (k \text{ は定数}) \tag{3.9}$$

しか存在しない。そこで Brookes は，これを第1の数式表現として採用している。実は Brookes は，(3.7)の式を Vickery[7] が最初に導いたものとして紹介しているのだが，これは彼の誤解である。Vickery はこれに相当する式をどこにも示していない。Brookes が導いたのは，結果的に (3.1) に等しい式であり，彼はそうと気づかずにグラフによる表現を数式化したのである。

さて，いま生産性の順位が r の雑誌1誌に含まれている関係論文数を $g(r)$ とすれば，g は累積度数関数Gに対応する度数関数であるから，(3.9) より

$$\begin{aligned} g(r)&=G(r)-G(r-1)\\ &=-k\log(1-1/r) \end{aligned} \tag{3.10}$$

である。生産性の最も低い雑誌，つまり第N位の雑誌には，関係論文がただ1編だけ含まれていると仮定すれば，$g(N)=1$ であり，これを (3.10) に代入すれば，

$$-k\log(1-1/N)=1 \tag{3.11}$$

となる。$\log(1-1/x)$ は x が十分に大きいとき，$-1/x$ で近似できるから，(3.11) より，$k=N$ と近似できる。これを (3.9) に代入すれば，

$$G(r)=N\log r \tag{3.12}$$

が得られる。これが Brookes の示した第2の数式表現である。

この式では雑誌の総数という実にわかりやすい数値が，グラフの傾きを近似するものとして用いられているが，実際には「第N位の雑誌には，関係論文がただ1編だけ含まれている」という仮定は非現実的である。なぜなら，現実には関係論文をただ1編だけ含んでいる雑誌の数はかなり多いため，そのような雑誌の順位はNより大分小さくなるからである。しかし，ここではとりあえず Brookes の論を認めて，その続きをみてみることにする。

(3.12) の式は，片対数軸のグラフ上では，当然原点を通る直線として描かれる。ところがすでに述べたように観測データから描いたグラフは原点を通らない。この食いちがいを解消するために，彼は a および s というパラメータを導入して再び数式を修正している。すなわち，

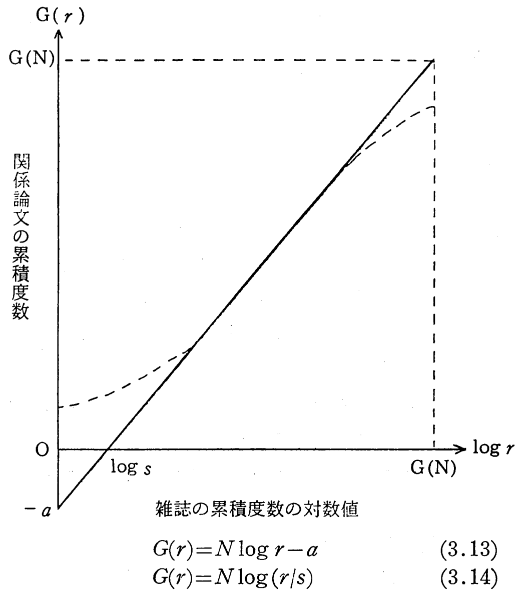
$$\begin{aligned} G(r)&=N\log r-a \qquad &(3.13)\\ &=N\log(r/s) \qquad &(3.14) \end{aligned}$$

が彼の第3，第4の数式表現である。ただし，$a=N\log s$ である。この2つの式から得られる片対数軸のグラフは，$-a$ の y 切片と $\log s$ の x 切片をもっている。Brookes はパラメータを導入して直線のグラフを軸に沿って平行移動することによって，観測データから描かれるグラフの形状に近づけようとしたわけである（第2図を参照）。

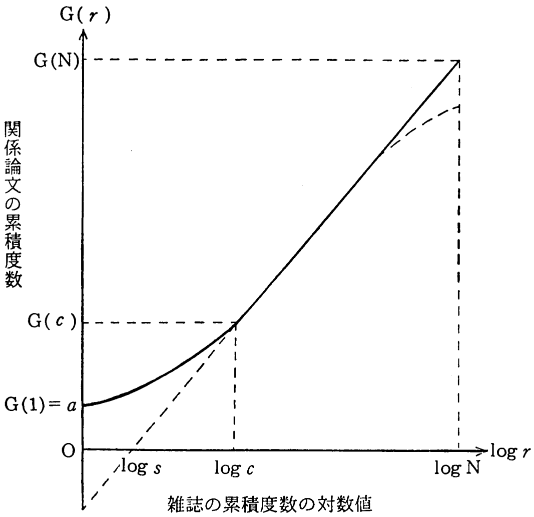
第3，第4の数式表現は，グラフの直線部分はかなりよく近似しているものの，グラフの下方のたわみについてはいまだに食いちがいが大きい。Brookes の69年の論文ではこの点を解消すべく，数式表現に修正を加えている。彼の示した第5の数式表現は，

$$G(r)=\begin{cases} ar^{\beta} & (1\leqq r\leqq c)\\ N\log(r/s) & (c\leqq r\leqq N) \end{cases} \tag{3.15}$$

なるものである[18]。（第3図を参照。）ただし，α，β は定数であり，c の値は，グラフが曲線から直線に変わる変曲点の位置を示すものである。r が c 以下の式は，グラフにおいては下に凸のたわんだ曲線を表わし，r が c



第2図 Brookes の第3，第4の数式表現によるグラフ



$$G(r)=\begin{cases} ar^{\beta} & (1\leqq r\leqq c) \\ N\log(r/s) & (c\leqq r\leqq N) \end{cases} \qquad (3.15)$$

第3図 Brookes の第5の数式表現によるグラフ

以上の式は (3.14) と同一である。また，$r=c$ において2式の示すグラフは連続かつ接していなければならないから，$r=c$ における2式の値，および2式の微分係数の値は等しくなければならない。これより (3.15) を補足する条件として，$1/\beta=\log(c/s)$ なる関係をあげる必要がある。

つぎに Haspers の数式表現をみてみよう[19]。Brookes は，片対数軸で直線のグラフを軸に沿って平行移動することで観測データとの食いちがいを小さくしようとして，数式表現の修正を行なった。しかしそれだけではグラフ下方のたわみを記述できないので，この部分だけは別の式を用いて表現したわけである。これに対して Haspers の式は，横軸のめもり自体を修正するという発想にもとづいて，下方のたわんだ部分をも同じ式で表現しようと試みたものである。彼の示した式は，つぎのようなものである。

$$G(r)=k\log(r/u+1)+G(0) \qquad (3.16)$$

k, u, $G(0)$ は定数である。この式は，Brookes らのように横軸に r の対数をとった片対数軸のグラフにおいては直線を示さず，横軸に $r+u$ の対数をとったグラフにおいてはじめて直線を示すものである。従来の片対数軸でこの式をグラフ化すると，r が小さいときにはグラフはパラメータ u の影響を受けて下に凸になるが，r が u に対して大きくなるにつれてグラフは限りなく直線に近づいてゆく。かくして(3.16)は，下方にたわみをもち，次第に直線となってゆくグラフを記述しうるのである。

Asai はその81年の論文で，Bradford の法則の数式表現をグラフからのアプローチによって試みた，6人の研究者による8つの数式をすべて包括する一般式として，

$$G_0=a\log(r_0+c)+b \qquad (3.17)$$

なる式を提示している[20]。彼は8つの数式表現を，この式の a, b, c 3つのパラメータに相当する値がそれぞれ未知か既知かによって5つのタイプに分類し，各タイプの数式がいずれも3つのパラメータをもつ式で表現しうることを証明している。彼の示した5つのタイプとは，(1) a のみ未知のもの (Cole の (3.3))，(2) a と b が未知のもの (Brookes の (3.14))，(3) c のみ未知のもの (Leimkuhler の (2.9) および Brookes と Griffith が78年に示した式[21])，(4) a と c が未知のもの (Fairthorne の (2.13), Wilkinson の (2.12), および Leimkuhler の (2.10))，(5) a, b, c いずれも未知のもの (Haspers の (3.16)) である。彼の分類を補足すれば，Brookes の (3.9), (3.12) は第1のタイプに，Vickery の (3.2), Brookes の (3.13) は第2のタイプに加えることができるであろう。

Asai のこの論文にはいくつかの誤解があるように思われる。そのひとつは，彼が第3，第4のタイプとしてあげた5つの数式は，グラフによるアプローチによって導かれたとはいいがたいことである。特に Leimkuhler と Wilkinson の式は，明らかに比例式による表現から導出されたものである。もうひとつの誤解は，Cole の (3.3) の式において $a=1$ として論じている点である。1という数値は Cole の収集したデータからたまたま算出されたものであり，Cole は a を既知の値として論じているわけではないから，彼の式は本来第2のタイプに分類すべきである。しかし，比例式による表現から導かれた数式もグラフの形状から導かれた数式も，結局は (3.17) の形の式で一般化できるという Asai の主張は，重要な指摘である。

ところで，Brookes の (3.15) と Haspers の (3.16) の式は，片対数軸のグラフの下方のたわんだ部分をも記述しようと試みたものであった。一方，グラフの上方の

たれ下がった部分をも記述しようと試みた研究者はほとんどみあたらない。そのひとつの理由は，Brookesをはじめとする多くの研究者は，グラフ上方のたれ下がりの原因をデータ収集の不十分さにあると考えていたので，この部分をも記述する必要性を認めなかったからである[17]。この点で例外的なのは，Praunlich と Kroll の研究である[22]。彼らはグラフのたれ下がりが雑誌群への関係論文の分布の状態そのものに由来するものと考え，たれ下がりの部分をも記述しうる数式表現を模索している。しかし，彼らの数式は複雑かつ数学的モデルとして十分に洗練されているとはいいがたいので，本稿ではそのような試みがなされたことをここに記するにとどめておく。

IV. 確率的モデルからの数式表現の導出

A. 多様な分布法則間の類似性の指摘

II章とIII章では，Bradford の法則を，そのオリジナルな法則の記述から，よりわかりやすく扱いやすい形の数式に表現しなおそうと試みた研究の，2つの流れをたどった。第1は，Bradford 自身によって行われた比例式による表現を，数学的に洗練しなおすことによって数式表現を導こうとする流れであり，第2は，観測データから描いたグラフの曲線の形状を，直接に数式表現しようとする流れである。

Bradford の法則の数式表現の研究には，さらにもうひとつの流れが存在している。それは，Bradford の法則が記述しているような2変量の関係の背後に，ある一般的な確率分布の存在を予測し，この分布を生成させるよう数学的モデルを提示することによって法則を表現かつ説明しようと試みるものである。いいかえれば，確率的なモデルの構築から，数式表現を導こうとする流れである。

このような確率的モデルの研究の契機となったのは，似かよった形式をもつ統計的分布が全く異なった分野に広く出現していることを指摘した，一連の研究である。特に，ビブリオメトリクスの分野で最もよく話題にされる3つの分布法則が，実は本質的に同一の統計的分布を異なった方法で表現したものにすぎないという指摘は，これらの分布法則を包括するような確率的モデルの構築を強く促す原因となった。この3つの法則とは，研究者の論文生産性に関する Lotka の法則，テキスト中の単語の出現頻度に関する Zipf の法則，および Bradford の法則である。

Lotka の法則と Zipf の法則の類似性を最も早く指摘したのは，Zipf 自身である[23]。彼はその著作において，自分の発見した単語の出現頻度に関する法則と同じ数量関係が，都市人口の分布から新聞の発行部数の分布に至るまで，人間の社会的活動の実に多様な分野において見出だせることを，多くの例を引いて説明しており，その中で Lotka の法則にも言及している。

Simon も，Lotka, Zipf 両法則の表わしている分布，および経済学，生物学などの分野でみられる分布のあわせて5つの分布が，いくつかの共通の性質を備えた類似の分布であることを指摘している[24]。C節で詳しく述べるように，彼はこれらの分布を包括して記述する統計的分布を確率的モデルから導き，これを「Yule 分布」(Yule distribution) と名づけている。そして Kendall は Yule 分布を支持し，Bradford の法則もこの一般的分布から説明しうることを示している[25]。ビブリオメトリクスの3つの分布法則がはじめて結びつけられたのは，この Kendall の60年の論文においてであったといえよう。さらに60年代の終わりには，Fairthorne, Brookes など，ビブリオメトリクスの代表的な研究者によって，この3つの法則の類似性が論じられている[12][17]。

一方，Bradford の法則に代表される統計的分布が，図書館情報学の領域に限ってさえ，様ざまな場面で出現しうることも，多くの研究者によって指摘されている。たとえば Haspers は図書館における雑誌タイトルの貸し出し回数の分布に[19]，Bulick は図書館における図書の貸し出し回数の分布に[26]，それぞれ Bradford の法則がよくあてはまると指摘している。また Naranan は，後述する彼の「累乗の法則」が，Bradford の法則の示す雑誌数と論文数の関係のみならず，雑誌間，論文間の引用—被引用関係を表わすいく通りかの数量の関係にも適合するものであると主張している[27]。

B. ビブリオメトリクスの3つの分布法則の近似関係

Lotka, Zipf, Bradford の3つの法則は，みかけ上全く異なった数式によって表現されるものである。Lotka の法則は，ある研究者の集合において s 編の論文を発表した研究者の人数を f 人としたとき，

$$f=k/s^2 \qquad (4.1)$$

と表わされ，Zipf の法則は，あるテキスト中での出現

頻度の順に単語を並べて，r 位の単語の出現回数を s 回としたとき，

$$r \cdot s = k \tag{4.2}$$

と表わされるものである。ただし，k はそれぞれの式の定数である。これらはいままでみてきた，主に対数を用いて表現された Bradford の法則の数式表現とは明らかに違った形の式である。そこでこの節では，みかけ上異なった3つの分布法則が，実は数学的には同等あるいは近似の関係にあることを説明しておくことにする。それによって Bradford の法則の数式表現のもつ，ひとつの重要な性格が明らかにされるであろう。

ここで，3つの分布を同時に扱うために，これらをいずれも生産者と生産物の関係を表わしているものであると考えることにする。すなわち Bradford の法則においては雑誌と特定分野の論文，Lotka の法則においては研究者と研究論文，Zipf の法則においては単語とその出現を，それぞれ生産者と生産物に置きかえて考えるのである。するとまず気がつくのは，Lotka が，生産物の数と生産者の度数を直接に結びつけて法則化しているのに対し，Zipf と Bradford は，生産者をそのおのおのが産出した生産物の数の多い順に順位づけし，その順位と生産物の数を結びつけて法則化しているということである。生産性の順に雑誌が並べられているとき，雑誌の累積数は雑誌の順位でもあることは，すでに示唆した通りである。

ここで，生産者を生産物の数の「大きさ」(size) にしたがって配列し，「順位」(rank) をつけると，ある順位 r の生産者の，生産物の数の大きさ s を定めることができる。この関係を Rapoport は「ランク—サイズ」(rank-size) の関係と呼んでいる[28]。これに対し，生産物の数の大きさと生産者の度数の関係は，従来からのいわゆる度数分布として表わされるが，彼はこれを前者と区別して「サイズ—度数」(size-frequency) の関係と呼んでいる。Brookes, Griffith, Hubert は前者を「度数—ランク」(frequency-rank) と呼び，Hubert は後者を「度数—サイズ」(frequency-size) と呼んでいるが[21)29)30]，これは論を進める上で誤解を与えやすい。よってここでは Rapoport の呼び方にならって，順位による大きさの分布を「ランク—サイズ分布」，大きさによる度数の分布を「サイズ—度数分布」と呼ぶことにする。

この2つの概念を導入すれば，Bradford, Zipf の2法則はランクサイズ関係の記述であるのに対し，Lotka の法則はサイズ—度数関係の記述であると説明することができる。前章までの様ざまな数式表現は，Vickery の式を除いてすべてランク—サイズ分布を記述する試みである。Vickery の (2.3) のみは，サイズ—ランク分布とでもいうべき関係を表現したものである。

つぎに，ランク—サイズ分布とサイズ—度数分布の間の数学的関係を，Hubert[30] と Rapoport[28] の論をふまえた上で，筆者なりに説明してみることにする。いま，総数Nの生産者の集合において，ちょうど s_i 個の生産物を産出した生産者，すなわち大きさ s_i の生産者の度数が $f(s_i)$ なる度数関数で表わされるとする。このサイズ—度数分布の累積度数分布を考えれば，「たかだか s_i 個」の生産物を産出した生産者の総数 $F(s_i)$ は，

$$F(s_i) = \sum_{t=1}^{s_i} f(t) \tag{4.3}$$

である。ここで，生産者を生産物の個数の多い順に並べたときの順位 r を考えると，ちょうど s_i 個の生産物を産出した生産者の順位 r_i は，「少なくとも s_{i+1} 個」を産出した生産者の数に1を加えた値に等しい。1位の生産者の産出量，すなわち s_i の最大値を $s_{\max}$ とすれば，順位 r_i は，

$$r_i = \sum_{t=s_{i+1}}^{s_{\mathrm{mxa}}} f(t) + 1 \tag{4.4}$$

と表わされる。(4.4) は度数，順位，大きさの関係を表わす数式であり，とりもなおさずサイズ—度数分布とランク—サイズ分布の関係を示すものである。

ところで，(4.3) と (4.4) の式は s_i と r_i が離散型の変数であることを前提としたものだが，これらを連続型の変数とみなしても近似が可能な場合には，以下のことが成り立つ。連続型変数 s に対して，そのサイズ—度数関係を表わす密度関数を f_0 とすれば，その分布関数 F_0 は，

$$F_0(s) = \int_{-\infty}^{s} f_0(t)dt \tag{4.5}$$

である。連続型変数 s においては $s_i = s_{i+1}$ とみなせるから，

$$\begin{aligned} r &= \int_{s}^{\infty} N f_0(t)dt \\ &= N[F_0(t)]_s^{\infty} \end{aligned} \tag{4.6}$$

と表わせる。“このようにして，ランク―サイズ曲線は，本質的にサイズ―度数曲線の積分なのである”[28]。

以上の論にもとづけば，ビブリオメトリクスの3つの分布法則が，数学的には近似の関係にあることは，つぎのように説明できるだろう。ある連続型の確率分布がサイズ―度数分布として考えられたとき，その密度関数 f_0 が，

$$f_0(s)=k/s^2 \qquad (k\ \text{は定数}) \tag{4.7}$$

で与えられるならば，その分布関数 F_0 は (4.7) を積分してやれば，

$$F_0(s)=\int_{-\infty}^{s}(k/t^2)\,dt=-k/s \tag{4.8}$$

と求められる。ここでこの分布をランク―サイズ分布とみなすと，順位 r は，

$$\begin{aligned} r&=\int_{s}^{\infty}Nf_0(t)dt\\ &=Nk/s=g^{-1}(s) \end{aligned} \tag{4.9}$$

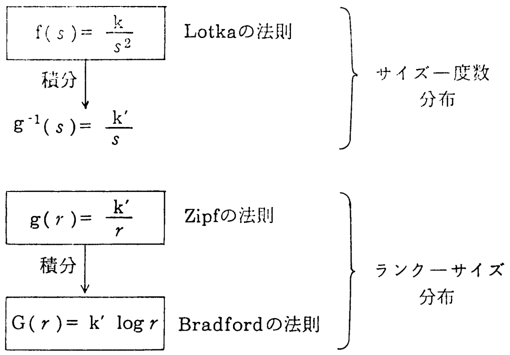
と表わされる。ランク―サイズ分布の度数関数は $g^{-1}: s\to r$ の逆関数 $g: r\to s$ で与えられるから，$k'=kN$ とおけば，

$$g(r)=k'/r \tag{4.10}$$

となる。その分布関数Gは (4.10) を積分してやれば，

$$\begin{aligned} G(r)&=\int_{1}^{r}g(t)dt\\ &=k'\log r \end{aligned} \tag{4.11}$$

と求められる。この確率分布を表現するにあたり，(4.7) の形を採用すれば Lotka の法則，(4.10) の形ならば Zipf の法則が得られる。また (4.11) の形は，Bradford のグラフによる表現から得られた (3.1) の式に一致しており，Bradford の法則を近似する式とみなせるものである。以上をまとめると，3法則の関係は第4図のように表わせるだろう。



第4図　ビブリオメトリクスの3つの分布法則の近似関係

C. Bradford の法則を導く確率的モデル

同一の形式をもつ分布が多くの分野において見出だしうること，つまり同一の分布現象が広い範囲に存在しているという事実は，多くの研究者に，この分布の遍在を理論的に説明する試みを促した。そのために研究者たちが用いた方法は，この遍在する分布現象の本質を抽出し，これを生成するような一般的な「確率的モデル」を構築することであった。

統計的な分布は，いくつかの変数によって記述される。分布の様相は，これらの変数の生起する確率が周囲の環境に影響を受けて，どのように変化するかによって決定される。本稿でいう確率的モデルとは，環境が変数に及ぼす影響をいくつかの数学的な仮定へ抽象化して表現し，これより演繹的に統計的分布を導き出すような数学的モデルのことである。

以下この節では，Bradford の法則に代表される統計的分布を導く確率的モデルとしてしばしば引用され，モデルとしても比較的洗練されている2つのモデルを，少し詳しく論じてみたい。それらは Simon と Kendall のモデル，および Naranan のモデルである。これらに関連して，他のいくつかのモデルも紹介する。

Simon はすでに述べたように，異なった5つの分布現象が共通の特徴をもつことに着目し，このような特徴をもつ分布を生成する確率的モデルを求めようと試みた[24]。Simon 自身は Bradford の法則に言及していないが，Kendall が，このモデルの Bradford の法則に対する有効性を指摘している[25]。Simon は，Zipf の法則における単語とその出現頻度の関係を例としてモデルを展開しているが，ここではそれを Bradford の法則の雑誌とその掲載論文数の関係におきかえて論ずることにする。

いま，総数Nの雑誌群に，特定主題に関係する論文が

すでに n 編含まれている状態を考えたとき，Simon のモデルはつぎの４つの仮定を含むものである。

(1) N の雑誌のうち，ちょうど s の関係論文を含んでいる雑誌の数は，n と s との関数 $f(s, n)$ で表わすことができる。

(2) $n+1$ 番目に掲載される関係論文が，すでにちょうど s の関係論文を含んでいる雑誌に掲載される確率は，$sf(s, n)$，つまり，「すでにちょうど s の関係論文を含んでいるすべての雑誌に含まれている関係論文の総数」に比例する。

(3) $n+1$ 番目に掲載される関係論文が，N の雑誌以外の雑誌，すなわちいままでひとつも関係論文を含んでいなかった雑誌に掲載される確率 α は，つねに一定である。

(4) ちょうど s の関係論文を含んでいる雑誌の数は，つねに n に比例する。すなわち

$$f(s, n+1)/f(s, n)=(n+1)/n$$

が成り立つ。

Simon はこれらの仮定から，まず s の関係論文を含んでいる雑誌の相対度数 $f_0(s, n)$ が n からは独立で，s のみに従属することを示し，これをもとに数式を展開させて，

$$f_0(s)=\frac{\Gamma(s)\Gamma(\rho+1)}{\Gamma(s+\rho+1)}f_0(1) \tag{4.12}$$

なる式を導いている。ただし $\rho=1/(1-\alpha)$ である。$f_0(1)$ は定数だから，これをAとおき，ガンマ関数をベータ関数におきかえて表現すれば，(4.12) は，

$$f_0(s)=AB(s, \rho+1) \tag{4.13}$$

となる。これが，Simon が提示した確率的モデルから導出された分布の確率関数である。

さて，ここで $\rho=1$ の場合を考えると，(4.13) は，

$$f_0(s)=A/s(s+1) \tag{4.14}$$

となるが，これは s が１より十分に大きければ，

$$f_0(s)\fallingdotseq A/s^2 \tag{4.15}$$

と近似することができ，(4.7) の式と同等の式になる。

Simon のモデルが提示されたおよそ20年後，Price は同じくベータ関数で表現される確率分布のモデルを，Simon とは若干異なった仮定から，似通った数式の展開によって組み立てている[31]。Price はこのモデルを，個個の人または物 (individual) が成功 (success) を重ねてゆく過程をイメージして論じているが，ここでもやはり雑誌と論文数の関係におきかえてみれば，彼のモデルはつぎの３つの仮定を含むものである。

(1) Nの雑誌のうち，ちょうど s の関係論文を含んでいる雑誌の数は，n と s との関数 $f(s, n)$ で表わすことができる。

(2) $n+1$ 番目に掲載される関係論文が，すでにちょうど s の関係論文を含んでいる雑誌に掲載される確率は，s に比例する。

(3) n と s との関数 $f(s, n)$ は，

$$f(s, n)=\Psi(n)\cdot\phi(s)$$

のように，s に独立な n の関数と n に独立な s の関数の積で表現できる。

この３つ仮定のうち，第１の仮定は Simon の第１の仮定と全く同一のものであり，第３の仮定は Simon の第４の仮定をより一般化したものである。彼は第２の仮定で示されるような，より多く割り当てられているところほどつぎに割り当てられる確率の高い過程を，“成功が成功を生む”[31]過程だと表現している。

これらの仮定から Price の導き出したサイズ一度数分布の確率関数は，

$$f_0(s)=(m+1)B(s, m+1) \quad (m \text{ は定数}) \tag{4.16}$$

である。彼は自分のモデルから導かれる分布を「成功の累積される分布」(Cumulation Advantage Distribution)，略して「CAD」と名づけ，この分布が，Lotka, Bradford, Zipf のいずれの分布法則に対しても有効な概念的基盤を与えるものであり，しかも社会科学の分野に広範に存在する確率的メカニズムを解明するものだと主張している。しかし彼のモデルは，Simon のモデルに比較して数式の展開が不明快な部分があり，数学的に十分に洗練されたものとはいいがたい。

これに対して小野寺の「玉入れモデル」は，Price と同じ３つの仮定にもとづいて展開されているものだが，数式の展開が厳密で，数学的な洗練度は Price のモデルよりも高いものである[32]。小野寺は Bradford その他の

法則を導く確率的モデルを，ボールを箱の中へ投げ込んでゆく過程をイメージして論じ，

$$f(s,\ n)=\Psi(n)B(s,\ \rho)\quad(s\geqq 1) \tag{4.17}$$

なるサイズ一度数分布の確率関数を導出している。この式は大きなに s 対しては，

$$f(s,\ n)\fallingdotseq\Gamma(\rho)\Psi(n)s^{-\rho} \tag{4.18}$$

と近似される。彼はまた，第2の仮定で示されるような過程を「付和雷同型挙動」と名づけ，このような過程を確率モデルの構築の際に導入することは，偶然よりも人為的要素が左右する社会科学的現象を，確率論的な立場から説明するにあたって有効な方法であると主張している。

つぎに Naranan のモデルをみてみたい[33]。Naranan も，Bradford の法則その他の分布法則が近似の関係にあることに注目し，これらがいずれも，

$$f(s)\propto s^{-r}\quad(r\ は定数) \tag{4.19}$$

なる式で表現しうることを示し，これを「累乗の法則」(power law) と名づけている。この式は，(4.7) の式を一般化したものである。彼は累乗の法則を導くような確率的モデルを，つぎの3つの仮定をもとに展開している。

(1) 特定主題に関わる雑誌の数は，時間の経過につれて指数関数的に増加する。

(2) 同時に各雑誌も，その主題についての関係論文の数を，時間の経過につれて指数関数的に増加させている。

(3) 各雑誌に含まれている論文数の増加の速度は，すべての雑誌において等しい。

Naranan は，(1)，(2) の仮定が実際の観則データとも一致するものであることを，Price の研究を例にあげて主張している。

(1)の仮定より，時間 t が経過したときの関係雑誌の総数 N は，t の関数 $N(t)$ で表わすことができる。時間 t が経過するたびに N が e 倍になるような定数 t_N を考えると，これは，

$$N(t)=N(0)e^{t/t_N} \tag{4.20}$$

と表わせる。$N(t)$ はまた，時間 t が経過したときに加わった雑誌数を累積，あるいは積分したものと考えられる。いま，時間 t_0 が経過した時点で，ある雑誌が関係雑誌の集合に加わってから，すなわち関係論文をはじめて掲載してから経過した時間を，その雑誌の「年齢」と定義する。すると，雑誌が年齢 τ の雑誌数は，(4.20) を t で微分したときの，$t=t_0-\tau$ における微分係数 $N'(t_0-\tau)$ として求められ，

$$N'(t_0-\tau)=\frac{N(0)}{t_N}\,e^{(t_0-\tau)/t_N} \tag{4.21}$$

となる。さらに雑誌の年齢 τ による雑誌数の分布を表わす密度関数 $f_0(\tau)$ は，(4.21) と $\int_0^\infty f_0(\tau)d\tau=0$ より，$t_0=0$ とおけば，

$$f_0(\tau)=e^{-\tau/t_N}\frac{1}{t_N} \tag{4.22}$$

と求められる。一方，(2)の仮定より，各雑誌において時間 τ が経過したときの関係論文の総数 s は，τ の関数 $s(\tau)$ で表すことまできる。時間 t が経過するたびに s が e 倍になるよう定数 t_s を考えると，

$$s(\tau)=s(0)e^{\tau/t_s} \tag{4.23}$$

と表わせる。ここにおいて t_s は，(3) の仮定よりすべての雑誌において等しい値をもつ。(4.22)，(4.23) より τ を消去すれば，雑誌の相対度数 f_0 を論文数 s の関数として表わすことができる。Naranan はこれを，

$$f_0^*(s)=\frac{t_s}{t_N}\,s(0)^{t_s/t_N}\cdot s^{-(1+t_s/t_N)} \tag{4.24}$$

と解いている。この式は，$r=1+t_s/t_N$ とおけば，(4.19) と全く同等の式であり，結果的に「累乗の法則」が導かれたことが確認されよう。以上が彼のモデルである。

ところで，Naranan のモデルには，最後の (4.24) の式を導く部分の数式展開に若干の問題がある。筆者の計算によれば，(4.22)，(4.23) より τ を消去した式は，

$$f_0^*(s)=\frac{1}{t_N}\,s(0)^{t_s/t_N}\cdot s^{-t_s/t_N} \tag{4.25}$$

と求められる。これは (4.24) と一致しない。Naranan は計算上の単純な誤りを犯しているものと思われる。とはいうものの，$r=t_s/t_N$ とおけば，(4.19) と全く同等の式であることにはかわりない。

Bookstein は，Naranan とは別のモデルから，(4.19) と同じ数式を導き，これを「科学者の生産性と社会変化のパターン」を表現する式として提示している[34]。しかし，彼のモデルの前提となっている「ビブリオメトリクス的な対称性」(bibliometric symmetry) なる概念はきわめて抽象的かつあいまいであり，数式を展開するにあたっての仮定も，数学的に十分明示されているとはいいがたい。したがって，彼のモデルを Bradford の法則を導く数学的モデルとして採用することには，ためらいがある。

V. Bradford の法則の望ましい数式表現

A. 望ましい数式表現の条件

前章までに，法則の発見から近年の研究に至るまで，Bradford の法則の数式表現を研究の 3 つの流れに沿っておよそ20あまり論じてきた。この章ではこれらの表現を，数式表現としての望ましさという観点から簡単に論じ返してみたい。

まず，一般に社会科学の分野における法則の望ましい数式表現の条件として考えられるものを，4 つほどあげてみることにする。条件の第 1 は，その表現が正確なものであること，すなわちその数式が，法則の対象となっている社会現象を十分に記述しうることである。法則が法則たるために，実際の観測値が十分に説明できることは最低の条件である。ただし社会科学においては自然科学と異なり，現象を 100% 記述すること，すなわち観測値を予想しきることは普通不可能と思われるから，どの程度まで観測値を近似しうるかが望ましさの基準となるであろう。また，どの程度記述しうれば十分であるかは，その法則が何のために用いられるのか，どの様な応用がなされるのかによるものである。

条件の第 2 は，法則の意味内容がわかりやすいこと，すなわち記述している現象の特性が数式から読みとりやすいことである。同等の内容をもつ数式を作ることは数学的にはいくらでも可能であろうが，人間が見て，現実に生起する現象がどのようにふるまうかが想像しやすい式の方が，優れた表現だといえよう。あるいは，数式のわかりやすさは，数式の現実を解釈する力を計るひとつの基準であるともいえるだろう。

条件の第 3 は，その法則を実際に利用して何かを行うにあたって，数式が扱いやすいことである。この条件は，第 2 のわかりやすさという条件と深くかかわるものであるがより実践的な意味を含んでいる。社会科学の法則は，単に現実の現象を記述するためのものではなく，それを予測し，評価するために用いられるべきものである。したがって，法則は現象の予測と評価が行いやすい形で表現されていなければならない。これが，ここでいう扱いやすさの意味である。

条件の第 4 は，その表現が数学的に十分洗練されたモデルから導かれていることである。モデルが数学的に洗練されているということは，上の 2 つの条件，すなわちわかりやすく扱いやすいという条件を含むものであるが，それ以前に数学的に明示された仮定と正しい数式展開から成り立っていることが前提である。もちろん，経験的に発見された法則は，このような仮定の明示された数学的モデルがなくとも成立しうるし，現象の予測と評価に十分な効果を与えうるかもしれない。しかし，経験則がひとつの一般的な理論へ高められるためには，このようなモデルの存在は不可欠である。

以上の 4 つの条件は，これだけで望ましい数式表現の十分条件とは言えないが，それぞれ必要条件であることは間違いないだろう。次節では，これら 4 つの条件に照らして，従来の Bradford の法則の数式表現に存在するいくつかの問題点を指摘してみたい。

B. 4 つの条件をめぐる考察

数式表現の正確さをめぐる問題としては，法則の適合度検定の問題がある。表現の正確さは，結局理論値と観測値の適合度の問題に帰着する。この適合度を高めるために数式表現を修正していった研究者たちについては，Ⅲ章で紹介した通りである。しかし，この点に関して従来の研究が決定的に問題なのは，数式表現からえられる数値と観測データとの適合度の統計学的に満足のゆく検定がいまだかつて行われていないということである。

ほとんどの研究者は，数式表現と観測データの適合度を直観的あるいは視覚的に確認しているにすぎない。数式表現から得られる数値が，観測データから得られる数値にほぼ近いものならば，あるいは，数式表現から描いたグラフが，観測データから描いたグラフにおよそ似通った形をしているならば，その数式は観測値を近似しうるものだという判断が下されているのである。しかもこのような判断に用いられるデータは，ひとつかふたつの限られた主題分野に対する調査から得られたものである場合がほとんどである。若干の数学的手法を使ったものに，Wilkinson の検定があるが，これも，観測データ

の中から累積論文数と累積雑誌数のわずか2組のデータのみを選んで行うという，きわめて信頼度の低い方法で行われている[11]。

例外的に多くの主題分野に対する観測データを収集して検定を行なっているものとして，Asai の研究[20]，および Drott と Griffith の研究がある[35]。前者は11件の調査，後者は23件の調査をもとに検定を行なった研究で，データ数はまず満足のゆくものである。しかし，Asai が行なったのは，自分の分類したどのタイプの数式がより適合度が高いかを相対的に比較したもので，各表現に対する適合度の絶対的な評価は行なっていない。また，Drott らは適合度を計るものとして Pearson の相関係数を用いているが，これは有意水準が明示された検定ではないし，検定を行うにあたってグラフの直線部分からずれそうなデータをあらかじめ除外している点にも疑問が残る。

同じビブリオメトリクスの分布法則である Lotka の法則には，カイ2乗検定法を使った Rao の研究[36]，Kolmogorov と Smirnov の検定法を使った Pao の研究[37]がある。これらの研究にみられる検定手法に問題がないとはいえないが，Bradford の法則においてもこのような信頼度の高い統計学的手法を多くのデータ・セットに適用した検定を，今後の大きな研究課題とすべきである。

数式表現のわかりやすさをめぐる問題としては，パラメータの意味づけの問題がある。数式に現われるパラメータに何らかの意味づけをすることによって数式に付加価値を与える試みは，多くの研究者によって行われている。たとえば，Goffman と Warren[4] は (2.6) の b_k の最小値に，Cole[15] は (3.3) の k の値に，いずれも「その主題領域における関係論文の雑誌群へのちらばりの程度」を示す値という意味を与えている。また Leimkuhler は，(2.10) の B, β それぞれに，「その主題に関係する論文群の生産性」と「関係論文のちらばりの程度」を示す値と意味づけしている[4]。Brookes もパラメータの意味づけには熱心で，(3.13) の a に生産性の最も高い部分の雑誌における「飽和欠損」(saturation deficiency) の度合いという意味を，さらに (3.15) の α, β, c, s の4つのパラメータに，それぞれ生産性1位の雑誌に含まれている関係論文数の数，グラフの下方のたわみの「曲度」(curvature)，核のゾーンの雑誌数，主題領域の広さという意味を付与している[9)16]。

このような試みは，たしかに数式の内容を豊かにし，表現を感覚的にわかりやすくするものではある。しかしここまでまず第1に問題なのは，各研究者はここでもまた，十分な統計学的検定を経ずして，直観的にパラメータの意味を定めているにすぎないという点である。パラメータに安易に意味を見出だすことの危険性は，いく人かの研究者の指摘するところである[35)38]。また第2に，実際的にパラメータの意味づけがどれほど有効かという点も問題である。「飽和欠損」の度合いを示す値が正確に求められたからといって，それが果して何の役に立つのかを考える必要があろう。さらにもし有効なパラメータの意味づけがしうるとして，その値を観測データからどのように算出すべきなのか，その方法も問題となる。

つぎに，数式表現の扱いやすさについてだが，これは多くの研究者によって大いに改善されてきたと評価することができよう。(1) そもそも Bradford 自身の比例式とグラフによるあいまいな表現を数式で表現する試み，(2)雑誌群の核とゾーンへの分割という考えにこだわらずに，2変量間の関係として表現する試み，(3) さらに2変量間の関係を関数を用いて表現する試み，(4) 2変量間の関係を統計的分布を記述するものと認識し，確率関数や分布関数で表現する試みなどは，いずれもより扱いやすい表現方法を追求するものであったといえるだろう。

数式表現の扱いやすさをめぐる問題として，ランク—サイズ分布とサイズ—度数分布とどちらの分布による表現が扱いやすいかという問題がある。これに関して Brookes と Griffith は，書誌情報その他の社会的な関係を表現した分布を研究するにあたっては，ランク—サイズ分布の方がサイズ—度数分布よりも有効であると主張している[21]。その根拠として彼らがあげているのは，(1) ランク—サイズ分布の方が含んでいる情報量が多い，(2) ランク—サイズ分布の方が従来の理論から自由であるという2点である。

しかし，従来の統計学でほとんど対象とされていない分布だけに，仮に法則を数学的なモデルから導こうとするならば，ランク—サイズ分布による表現は非常に扱いにくいであろう。一方，この法則が実際の情報管理業務において，雑誌の選択方針の決定や，作成した主題書誌の網羅性の評価などに応用されていることを考えると，実践面においてはやはりランク—サイズ分布の方が適当だと考えられる。なぜなら，ランク—サイズ分布では，雑誌の累積数と論文の累積数の関係が表現されており，すべての関係論文に対してある割合の関係論文を含む雑

誌の数が，理論的に予測可能だからである。さらにいえば，Bradford の法則がサイズ—度数分布を記述したものであったなら，このような予測による応用は簡単には行われなかったのではないだろうか。

最後に，Bradford の法則を導く数学的モデルについて考えてみたい。ここでいう数学的モデルとは単に変量間の関係を数式で表現するだけでなく，そのような関係がどのようにして生じたのか，そのメカニズムを解明するようなモデルのことをさしている。このようなモデルの例として，Ⅳ章では Simon と Naranan のモデル，およびこれらに関連するモデルを紹介したが，これら以外にもいままでに多くの研究者によって様ざまなモデルが提示されている。

しかし，遺憾なことにこれらの多くは数学的モデルと呼ぶにふさわしくない。第1に，モデルの前提となる数学的仮定が明確に提示されているものが少なく，第2に，数式の展開に誤りがあるのに気づかずに論を進めているものがかなりある。仮定が明示されており，比較的洗練されていると評価できる Naranan のモデルでさえ，すでに述べたように数式の展開に誤りを含んでいる。このよう状況は，法則の検定が十分に行われていないこともあわせて考えると，図書館情報学の分野に，いまもって数学的・統計学的手法が定着していないことを示すものといえるかもしれない。また，従来提示された様ざまなモデルは，個個ばらばらに研究されてきたもので，モデル間の関係に考察を加えた研究があまり見あたらないのも問題である。いくつかのモデルの比較を試みたものとして Hubert の論文があるが，これも単に並列的な紹介にとどまっており，十分な比較とはいいがたい[39]。この点では，小野寺が自分のモデルと Simon のモデルの比較を行なっている記述が，両者のモデルの数学的な連続性と相違点を明確にしており，示唆深いものである[32]。いずれにせよ Bradford の法則を導く数学的モデルに関しては，より優れたモデルを構築するための積みあげがまだ不足しているように思われる。

VI. おわりに

Bradford の法則について，法則の発見から今日に至るまで，その表現方法にいかなる展開，進展が見られたのかを歴史的にあとづけた結果，数式表現に関する多様な研究は大きく3つの流れに整理できることがわかった。第1の流れは，Bradford の比例式による表現を忠実に数式化することによって，表現を導こうとするものである。このような試みは，Bradford 自身によって行かれた法則表現のあいまいさに端を発するものである。法則を統計的分布の記述として明確に意識した Leimkuhler の研究を，この流れの画期をなすものとしてあげることができる。第2の流れは，観測データから描いたグラフの曲線の形状を直接に数式化することによって，表現を求めようとするものである。このような試みは，理論値と観測値の食いちがいをできるかぎり小さくし，法則の適合度を高るめことをその動機としている。Brookes の一連の研究を，その代表にあげることができるだろう。第3の流れは，特定の分布を生成するような確率的なモデルの構築によって，表現を導こうとするものである。このような試みは，似かよった形式をもつ統計的分布が全く異なった分野に広く出現していることを指摘した一連の研究を契機としてはじめられたものである。本稿では，Simon と Naranan の2人の研究を中心に紹介した。

Bradford の法則を含む社会科学分野の法則に対する望ましい数式表現の条件として，(1) 観測値との適合度が十分に検定されていること，(2) 法則の意味内容が把握しやすいこと，(3) 実際に法則を応用する際に扱いやすいこと，そして (4) その数式を理法的に導く数学的モデルが確立されていることをあげた。しかし Bradford の法則に関しては，これらの条件をすべて満たすような数式表現はいまのところ見あたらなかった。これらの条件をめぐっていくつかの問題点を指摘したが，中でも各表現に対して十分に信頼しうる統計学的な検定が行われていないことは最大の問題であった。

はじめに述べたように，Bradford の法則について一般的に認められた表現方法は，発見から50年を経ていまだに確立していない。その原因としては，2つの事実をあげることができるだろう。ひとつは，発見当初の法則が数学的に洗練されていなかったために，その後の研究の展開において多様な数式表現が並立しえたという事実であり，もうひとつは，表現の多様さにもかかわらず，現在までに望ましい数式表現の条件を十分満足する表現式が存在していないという事実である。

ビブリオメトリクスの分布法則に関しては，(1) 観測データからの法則性の発見とその記述，(2) 法則の検定，(3) 法則の理論化，(4) 法則の実践的な応用方法などが従来研究されてきたが，Bradford の法則の場合，(1) の研究が先行しすぎて (2)，(3)，(4) の研究が十分に展開され

ていない。それゆえに Bradford の法則の研究における焦眉の課題は，まず十分な量の観測データから統計学的に信頼しうる検定を行なって，望ましい数式表現を確立することである。そのうえで法則を説明する理論の構築や，法則の図書館業務その他の情報管理業務への実践的な応用が行われなければならない。そうしてはじめて，Bradford の法則が図書館情報学にとって真に価値あるものかどうかを評価することができるであろう。

最後になったが，本稿の執筆にあたってご指導下さった東京大学教育学部の長澤雅男教授に感謝の意を表したい。

1) Prichard, Alan. "Statistical bibliography or bibliometrics?". Journal of Documentation. Vol. 25, No. 4, p. 348-349 (1969).
2) わが国では 'bibliometrics' に「計量文献学」,「計量書誌学」,「計量情報学」などの訳語を与えている研究者もある。
3) Bradford, Samuel C. "Sources of information on specific subjects". Engineering. Vol. 137, p. 85-86 (1934).
4) 'Bradford's multiplier' という呼称は，たとえばつぎのような論文に現われている。
Leimkuhler, Ferdinand F. "The Bradford distribution". Journal of Documentation. Vol. 23, No. 3, p. 197-207 (1967).
Goffman, William.; Warren, Kenneth S. "Dispersion of papers among jounals based on a mathematical analysis of two diverse medical literatures". Nature. Vol. 221, No. 5187, p. 1205-1207 (1969).
5) ここでいう「比例式による表現」を,「文章による表現」(verbal formulation) と呼んでいる研究者が多いが,「グラフによる表現」も文章で記述されているので，この呼び方はまぎらわしい。
6) Bradford は，2つの表現が同じことを意味しているとは言っていないが，2つの表現の関係についての言及も全くない。
7) Vickery, B.C. "Bradford's Law of scattering". Journal of Documentation. Vol. 4, No. 3, p. 198-203 (1948).
8) このグラフは，Bradford のグラフの縦軸と横軸を反転させたものである。
9) Brookes, Bertram C. "The derivation and application of the Bradford-Zipf distribution". Journal of Documentation. Vol. 24, No. 4, p. 247-265 (1968).
10) Leimkuhler, Ferdinand F. "Operational analysis of library systems". Information Processing and Management. Vol. 13, No. 2, p. 79-93 (1977).
11) Wilkinson, Elizabeth A. "The ambiguity of Bradford's law". Journal of Documentation. Vol. 28, No. 2, p. 122-130, 232. (1972).
12) Fairthorne, Robert A. "Progress in documentation; Empirical hyperbolic distributions (Bradford-Zipf-Mandelbrot) for bibliometric description and prediction". Journal of Documentation. Vol. 25, No. 4, p. 319-343 (1969).
13) Groos, O.V. "Bradford's law and the Keenan-Atherton data (Brief communications)". American Documentation. Vol. 18, p. 46 (1967).
14) 直観的あるいは視覚的ではない，統計学的検定の不在については，V章で論ずる。
15) Cole, P.E. "A new look at reference scattering". Jounal of Documentation. Vol. 18, No. 2, p. 58-64 (1962).
16) Brookes, Bertram C. "The complete Bradford-Zipf 'Bibliography'". Journal of Documentation. Vol. 25, No. 1, p. 58-66 (1969).
17) Brookes, Bertram C. "Bradford's law and the biliography of science". Nature. Vol. 224, No. 5223, p. 953-956 (1969).
18) Brookes の69年の *Journal of Documentation* の論文では，(3.15) の上の式が

$$\alpha\ r\ \beta$$

と印刷されているが，これは前後の文脈から判断して明らかな誤植である。
19) Haspers, Jan H. "The yield formula and Bradford's law". Journal of the the American Society for Information Science. Vol. 27, No. 5-6, p. 281-287 (1976).
20) Asai, Isao. "A general formulation of Bradford's distribution: the graph-oriented approach". Journal of the American Society for Information Science. Vol. 32, No. 2, p. 113-119 (1981).
21) Brookes と Griffith が下記の論文で示したのは，

$$G_0=\log_b\{(a+r)/a\}$$

なる形の数式である。ただし，a, b は定数。
Brookes, Bertram C.; Griffiths, J.M. "Frequency-rank distributions". Journal of the American Society for Information Science. Vol. 29, No. 1, p. 5-13 (1978).
22) Praunlich, Peter; Kroll, M. "Bradford's distribution: a new formulation". Journal of the American Society for Information Science. Vol. 29, No. 2, p. 51-55 (1978).
23) Zipf, George Kingsley. Human Behavior and the Principle of Least Effort. Cambridge, Addison-Wesley Press, 1949, 573 p.
24) Simon, Herbert A. "On a class of skew distribution functions". Biometrika. Vol. 42, p. 425-440 (1955).

25) Kendall, M.G. "The bibliography of operational research". Operational Research Quarterly. Vol. 11, No. 1-2, p. 31-36 (1960).
26) Bulick, Stephen. "Book use as a Bradford-Zipf phenomenon". College and Research Libraries. Vol. 39, No. 3, p. 215-219 (1978).
27) Naranan, S. "Power law relations in science bibliography: a self-consistent interpretation". Journal of Documentation. Vol. 27, No. 2, p. 83-97 (1971).
28) Rapoport, Anatol. "Rank-size relations". International Encyclopedia of Statistics. Kruskal, William; Tanur, Judith, eds. New York, Free Press, 1978, p. 847-854.
29) Brookes, Bertram C. "Ranking techniques and the empirical log law". Information Processing and Management. Vol. 20, No. 1-2, p. 37-46 (1984).
30) Hubert, John J. "A relationship between two form of Bradford's Law". Journal of the American Society for Information Science. Vol. 29, p. 159-161 (1978).
31) Price, Derek J. de Solla. "A general theory of bibliometric and other cumulative advantage processes". Journal of the American Society for Information Science. Vol. 27, No. 5, p. 292-306 (1976).
32) 小野寺夏生. "'Bibliostatistics'—情報現象の統計学的説明—". 情報管理, Vol. 21, No. 10, p. 782-802 (1979).
33) Naranan, S. "Bradford's law of bibliography of science: An interpretation". Nature. Vol. 227, No. 5258, p. 631-632 (1970).
34) Bookstein, Abraham. "Patterns of scientific productivity and social change: a discussion of Lotka's law and bibliometric symmetry". Journal of the American Society for Information Science. Vol. 28, p. 206-210 (1977).
35) Drott, M. Carl; Griffith, Beler C. "An empirical examination of Bradford's law and the scattering of scientific literature". Journal of the American Society for Information Science. Vol. 29, No. 5, p. 238-246 (1978).
36) Rao, I.K. Ravichandra. "The distribution of scientific productivity and social change". Journal of the American Society for Information Science. Vol. 31, No. 2, p. 111-122 (1980).
37) Pao, Miranda Lee. "An empirical examination of Latka's law". Journal of the American Society for Information Science. Vol. 37, No. 1, p. 26-33 (1986).
Brookes, "The complete Bradford-Zipf 'Bibliography,'" op. cit.
38) 原田 勝. "ビブリオメトリクスの方法と応用". Library and Information Science, No. 12, p. 109-141 (1974).
39) Hubert, John J. "General bibliometrics models". Library Trends. Vol. 30, No. 1, p. 65-82 (1981).