

로트카 법칙과 학술정보의 생산성 연구*

한 복 희**

목 차

- | | |
|-------------------|------------------------|
| 1. 서론 | 4. 수확분야 학술정보의 생산성 분석 |
| 2. 계량정보학의 법칙과 모형 | 5. 기계공학분야 학술정보의 생산성 분석 |
| 3. 로트카 법칙 | 6. 결론 및 제언 |
| 3.1 개요 | |
| 3.2 로트카 법칙을 다룬 문헌 | |

1. 서론

계량정보학이란 용어는 최근에 만들어 졌지만 이에 대한 연구는 1920년대로 거슬러 올라간다. 더욱이 최근 30년간 이 분야 연구는 계속 증가하고 있다.

최근에 계량정보학을 ‘측정(Measurement), 정보현상에 관한 제변수간의 수리 모형화, 그리고 정보의 축적과 검색을 다루는 학문’으로 정의하였다.¹⁾ 계량경제학이나 계량경영학 등 타 인문사회과학분야에 비하여 문헌정보학 연구의 계량분석분야는 아직 미성숙단계에 있으며 이 분야의 명칭도 ‘계량서지학’, ‘계량과학학’ 혹은 ‘계량정보학’ 등 명칭사용의 문제와 용어가 포괄하는 내용등에 관한 문제들이 제기되고 있다고 하였다.²⁾

*충남대학교 문헌정보학과 부교수

**이 논문은 1991년도 충남대학교 자체연구비 지원에 의하여 연구되었음.

1) L.Egghe & R.Rousseau, Introduction to Informetrics:Quantitative Methods in Library, Documentation and Information Science, Amsterdam, Elsevier, 1990, 1.

2) 金賢姬 & 金容璇, 計量情報學, 서울, 구미무역(주) 출판사, 1993, 5.

역사적으로 계량정보학의 선행 용어라 할 수 있는 계량서지학은 주로 서양에서 발전하였으며 서지의 통계연구에서 시작되었다. 계량서지학은 도서와 다른 커뮤니케이션 미디어에 수학 및 통계방법을 적용하는 학문으로 정의되어 사용되었고 계량과학은 주로 동구에서 사용되었으며 과학과 기술발달의 추정을 연구하는 학문이다. 본 연구에서는 계량정보학으로 통일하여 쓰고자 한다.

로트카 법칙과 학술정보의 생산성 연구라는 제목의 연구를 수행하기 위하여 계량정보의 연구범위와 연구내용 그리고 계량정보학의 주요 법칙들 중에서 특히 로트카 법칙을 다룬 연구논문들을 조사하였다. 로트카 법칙을 간단히 말하면 특정 주제분야 저자들의 학술정보 생산성에 관한 법칙으로 n 개의 논문을 쓴 저자들의 수는 1편의 논문을 쓴 저자의 $1/n^2$ 이라는 것이며 또 1편의 논문을 쓴 저자들의 비율은 전체 저자의 약 60%라는 것이다.³⁾

이 로트카 법칙을 우리나라의 연구 출판물 상황에도 응용될 수 있다고 생각하고 이를 분석하기 위하여 자연과학과 공학분야 중 학회의 활동과 자료의 수집 등을 고려하여 수학과 기계공학 분야를 택하였다. 1946년부터 1979년까지 34년간의 수학 분야의 논문과 1957년부터 1983년까지 27년 간의 기계공학 분야 논문을 컴퓨터에 입력하여 분석하였다.

2. 계량정보학의 법칙과 모형

법칙이란 규칙적이고도 되풀이되어 발생하는 패턴을 서술하는 것이며 또한 법칙은 특수한 패턴이 어떻게 발생하는지에 대한 이유를 예측하고 설명해 줄 수 있는 이론의 정립을 돕는다. 법칙은 이론적인 이해와 경험적인 적용의 기회를 제공하기 때문에 많은 연구자들이 관심을 기울이는 것이다. 모형(Model)은 구조적인 면에서 뿐만 아니라 기능적인 면에서도 실상과 동상유질(同像類質)의 관계를 지니고 있어야 한다. 모형은 실상을 단순화시킨 것으로써 실상의 대표성을 지니며

3) Alled J. Lotka, "The Frequency Distribution of Scientific Productivity", Journal of Washington Academy of Sciences 16 (19 June 1926), 323.

이론의 약도와 같은 것이다. 모형의 대표적인 기능은 사물이나 현상에 대한 인식을 가능하게 해주는 것이다.⁴⁾

계량정보학의 내용은 사서나 문헌정보학연구자들에게는 어렵게 여겨지고 있다. 따라서 계량정보학이 실제로 이용될 수 있고 접근하기 쉽도록 하기 위해서는 계량정보학 자체를 서술하는 것에 그치는 것이 아니라 이러한 서술을 통해서 정보생산성의 현상을 예견하고 설명할 수 있어야 한다. 계량정보학이 주로 서지적인 현상을 묘사하였을 뿐 이러한 현상들을 예견하고 설명하지 못하였으므로 이론을 형성하기보다는 단지 방법론에 그치고 있다.⁵⁾ 따라서 연구자들은 계량정보학이 다루는 정보의 생산성에 영향을 주는 내재한 원인들에 대해 관심을 기울여야 할 것이다.

계량정보연구에는 이론과 사실의 두가지 근본적인 요소가 있으며 계량정보학의 주된 관심은 이 두 요소를 결합하는 데 있다. 즉 이론은 계량정보학의 목적을 수행할 수 있도록 개발되어야 하며 사실은 실제 일어나는 정보현상의 데이터를 말한다. 따라서 계량연구를 수행할 때 이론을 계량모형의 형태로 개발하고 사실들을 자료화한 다음 이 두 요소를 결합하고 계량모형을 이용하여 '구조분석', '예측', 그리고 '정책수립 또는 평가'를 할 수 있게 된다.

계량분석은 모형의 설정, 모형의 추정, 추정치의 평가, 추정된 모형의 예측력 평가등의 4단계로 이루어진다.⁶⁾ 계량정보학의 연구영역은 통계학, 운영연구, 계량정보법칙, 인용분석, 대출이론, 정보이론, 정보검색의 이론적 측면 등 7개의 하위 주제분야들로 구성되어 있다.

Library Trends는 1981년 여름호를 계량정보학 특집호로 출판하였다. 계량정보학의 3가지 주요법칙으로는 브래드포드 법칙 (Bradford's law),⁷⁾ 로트카 법칙 (Lotka's Law),⁸⁾ 지프의 법칙 (Zipf's Law)⁹⁾ 이 있다. 계량정보학의 일반적 모형¹⁰⁾ 으로는 프라이스 모형 (Price Model), 북스테인 모형 (Bookstein Model),

4) 金光雄, 社會科學研究方法論, 서울: 博英社, 1977, 117-121.

5) William G. Potter, "Bibliometrics Introduction", Library Trends, 1981, 30(1), 5-6.

6) 유지성, 計量經濟學言論, 서울, 博英社.

7) Carl M. Drott, Bradford's Law : Theory, Empiricism and The Gap Between Library Trends, 1981, 30(1), 41-52.

8) William G. Potter, "Lotka's Law Revisited", Library Trends, 1981, 30(1), 21-39.

9) Ronald E. Wyllys, Empirical and Theoretical Bases of Zipf's law, Library Trends, 30, (1), 1981, 53-64.

10) John J. Hubert, General Bibliometric Models, Library Trends, 30, (1), 1981. 65-81.

부룩스 모형(Brookes Model)등이 있다.

이외에도 계량정보학의 주요 연구주제로서 인용분석(Citation analysis),¹¹⁾ 정보의 쇠퇴현상(Obsolescence),¹²⁾ 지수함수적 성장법칙(The Law of exponential Growth)¹³⁾ 등이 있다. 본 연구는 로트카의 법칙을 대상으로 그 이론적 발전과 경험적 실재를 조사 연구하고자 한다.

3. 로트카 법칙

3.1 개 요

로트카법칙은 1926년 로트카의 '학술논문 생산성의 빈도분포'라는 논문에서 시작되며 그것은 n개의 저작을 생산해내는 저자들의 수는 1편의 논문을 쓰는 저자들의 약 $1/n^2$ 이 된다는 것이며 그리고 한편의 논문을 생산하는 저자는 전체저자의 약 60%를 차지한다는 것이다. 1926년 로트카의 원저논문의 출판이후 다양한 분야에서 저자의 생산성에 관한 많은 연구가 수행되었다. 로트카의 '역제곱법칙'은 생산성의 일반적이고 이론적인 예측이라고 할 수 있다.

로트카는 1907-1916년까지 Decennial Index to Chemical Abstracts에서 화학분야 6891명의 저자와 Auerbach's Geschichtstafeln der Physik에서 물리학분야 1325명의 저자를 추출하여 각 분야에서 저자들의 논문발표수를 조사하였다. 그는 일련의 공식을 작성하였다. 즉 X개의 논문을 생산한 사람과 빈도수 Y간에 존재하는 관련성에 대한 일반적인 공식은

$$x^n y = \text{일정하다} \quad (1)$$

특별한 경우에 즉 $n = 2$ (학술정보의 생산성의 역자승 법칙)가 될때

(1)에서 일정수의 가치는 다음과 같다.

11) Linda C. Smith, Citation Analysis, Library Trends, 30, (1) 1981, 83-106.

12) Kaye K. Gapen & Sigrid P. Milner, Obsolescence, Library Trends, 30, (1) 1981, 107-124.

13) Tague, Jean, Beheshti Jamshid & Rees-Potter L., The Law of Exponential Growth: Evidence, Implications and Forecasts, Library Trends, 30 (1), 1981, 125-149.

$$y_1 = C / 1^2 \quad (2)$$

$$y_2 = C / 2^2 \quad (3)$$

$$y_n = C / n^2 \quad (4)$$

$$\sum_1^{\infty} y = C(1/1+1/2+1/3\cdots) \quad (5)$$

$$= C \sum_1^{\infty} 1/n^2 \quad (6)$$

$$\sum_1^{\infty} y = C \pi^2 / 6 \quad (7)$$

$$C = 6 / \pi^2 \sum_1^{\infty} y \quad (8)$$

그러나 y 가 빈도수이기 때문에, 합계 $\sum_1^{\infty} y$ 는 일치된다.

그러므로

$$C = 6 / \pi^2 \quad (9)$$

$$= 6 / 9.87 \quad (10)$$

$$= 0.6079 \text{ 또는 } 60.79 \text{ 퍼센트} \quad (11)$$

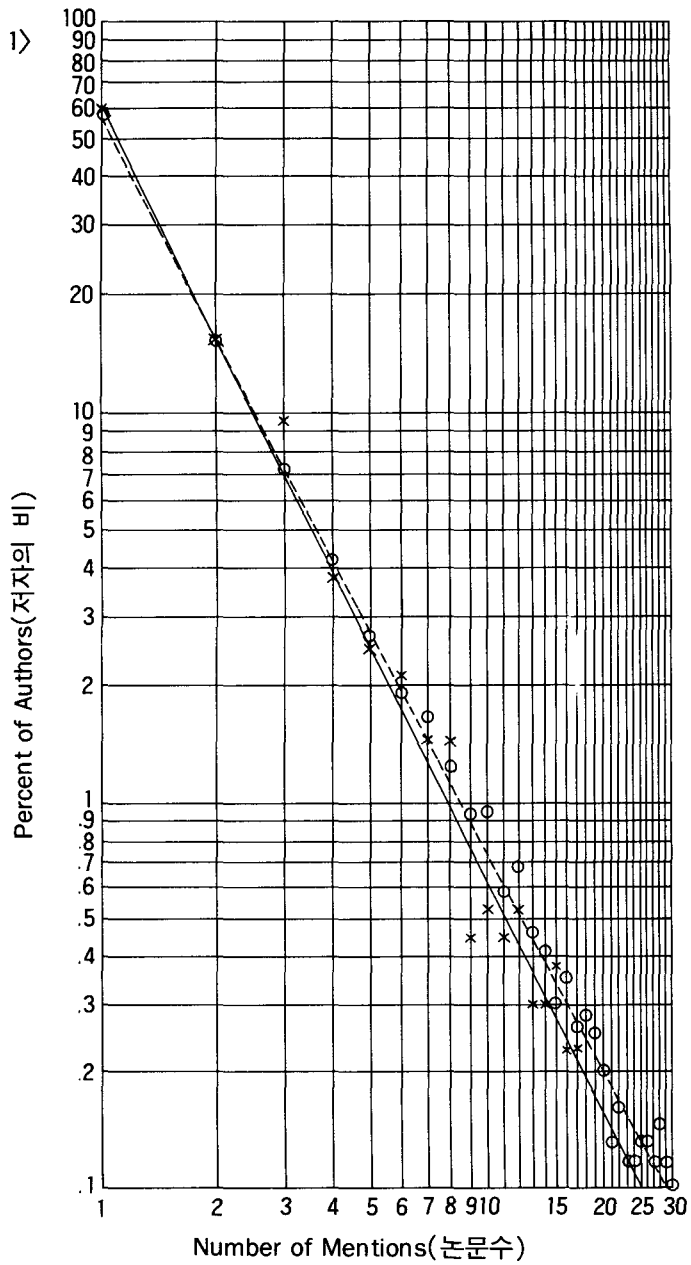
따라서 역자승법칙에 따르면, 한편의 논문을 생산하는 사람은 모든 논문저자의 비율이 60%가 넘는 정도이다. 이 연구에서 물리 분야의 경우 1325명의 저자들 중 59.2%가 1편의 저술을 했으며 화학 분야의 경우 6891명의 저자 중 57.9%가 1편을 생산한 것으로 조사되었다.

그리고 조사대상이 된 모든 저자들이 쓴 1편, 2편, n 편 등 논문의 생산성을 체크하여 로그좌표에 표시하여 그림을 작성하였다.¹⁴⁾ 그림은 반대경사인 직선을 나타낸다. 조사결과 2편의 논문을 쓴 사람은 1편의 논문을 쓴 사람의 약 1/4이고 3편을 낸 사람은 약 1/9 등 n 개의 논문을 쓴 사람은 1편의 논문을 쓴 사람의 약 $1/n^2$ 이라는 것이 발견되었다. (그림 1참조) 로트카는 두개의 실험자료를 가지고 이러한 연구를 한 것이다.

포터(W. G. Potter)는 저자데이터를 선정할때 주제, 언어, 출판물의 형태가 로

14) A.J. Lotka, 앞글, 1926, 322.

<그림 1>



트카 법칙의 일치를 결정하는 중요한 변수가 된다고 하였다.¹⁵⁾

일련의 데이터에는 로트카 법칙의 적합성을 테스트하기 위하여 통계적 처리를 하였다. Chi-square 방법은 적합치 않았으며 the Kolmogorov-smirnov(K-S) 통계로 처리하였다. K-S 테스트는 최대편차(Maximum deviation, D)를 결정한다.

$$D = \text{Max } |F_o(X) - S_n(x)| \text{이다.}$$

$F_o(X)$ 는 이론상의 누적빈도 기능이고

$S_n(X)$ 는 관찰된 표본의 누적빈도 기능이다.

유의도 0.01은 K-S 에서 $1.63/n^2$ 과 같다. 만약에 D가 K-S 통계보다 크다면 그 표본분포는 이론적 분포와 맞지 않는 것이다. K-S 통계는 표본분포가 법칙에 적합한가를 검증하기 위하여 이용되었다. 그 결과 화학은 적합하지 않고(<표 1> 참조) 물리학은 적합한 것(<표 2> 참조)으로 나타났다.

<표 1> 화학분야 저자비교

논문 수	저자수	관찰된 빈도	관찰된 누적 빈도 $S_n(X)$	기대치	이론적 누적 빈도 $F_o(X)$	최대편차 $ F_o(X) - S_n(X) $
1	3991	0.5792	0.5792	0.6079	0.6079	0.0287
2	1059	0.1537	0.7329	0.1520	0.7599	0.0270
3	493	0.0715	0.8044	0.0675	0.8274	0.0230
4	283	0.0416	0.8460	0.0380	0.8654	0.0194
5	184	0.0267	0.8727	0.0243	0.8897	0.0170
6	131	0.0190	0.8917	0.0169	0.9066	0.0149
7	113	0.0164	0.9081	0.0124	0.9190	0.0109
8	85	0.0123	0.9204	0.0095	0.9285	0.0081
9	64	0.0093	0.9297	0.0075	0.9360	0.0063
10	65	0.0094	0.9391	0.0061	0.9421	0.0030
.
.

15) W.G. Potter, 앞글, 1981, 37.

$$\text{최대편차 } D = \text{Max } |F_o(X) - S_n(X)| = 0.0287$$

$$\text{유의도 } 0.01 \text{에서 } K-S \text{ 통계} = 1.63\sqrt{6891} = 0.0195$$

$$D > 0.0195$$

그러므로 화학분야의 자료는 로트카법칙을 따르지 않는다.

이 경우에는 좀더 망라적인 표본집단을 대상으로 작성할 필요가 있다.

〈표 2〉 Auerbach의 물리학분야 저자비교

논문 수	저자수	관찰된 빈도	관찰된 누적 빈도 $S_n(X)$	기대치	이론적 누적 빈도 $F_o(X)$	최대편차 $ F_o(X) - S_n(X) $
1	784	0.5917	0.5917	0.6079	0.6079	0.0162
2	204	0.1540	0.7457	0.1520	0.7599	0.0142
3	127	0.0958	0.8415	0.0675	0.8274	0.0141
4	50	0.0377	0.8792	0.0380	0.8654	0.0138
5	33	0.0249	0.9041	0.0243	0.8897	0.0144
6	28	0.0211	0.9252	0.0169	0.9066	0.0186
7	19	0.0143	0.9395	0.0124	0.9190	0.0205
8	19	0.0143	0.9538	0.0095	0.9285	0.0253
9	6	0.0045	0.9583	0.0075	0.9360	0.0223
10	7	0.0053	0.9636	0.0061	0.9421	0.0215
.
.

$$\text{최대편차 } D = \text{Max } |F_o(X) - S_n(X)| = 0.0253$$

$$\text{유의도 } 0.01 \text{에서 } K-S \text{ 통계} = 1.63\sqrt{1325} = 0.0448$$

$$D < 0.0448$$

그러므로 Auerbach의 물리학자료는 로트카 법칙을 따른다.

3. 2 로트카 법칙을 다룬 문헌

로트카법칙을 응용하여 유사한 연구가 화학자, 물리학자, 수학자, 수리경제학자

등 다양한 주제분야에서 생산성 패턴에 대한 연구가 수행되었다. 과학자들의 생산성에 대한 연구는 로트카의 선구자적인 조사연구 이후 계량정보학의 주요 관심주제가 되었다. 로트카의 학술생산성의 '역제공법칙'은 다양한 시기와 다양한 학문분야의 연구에 의해 입증되었다.

로트카법칙을 다룬 논문들은 저자들 나름대로 로트카의 방법론을 이용하여 연구를 수행하였으나 로트카의 원저논문과 비교하는 작업은 이루어지지 않았다. 1973년 머피(Larry J. Murphy)는 로트카 법칙과 인문과학 분야의 데이터의 수집을 시도하였다.¹⁶⁾

그외에도 쇼오(Alan E. Schorr), 부스(Henry Voos)를 비롯한 많은 학자들이 로트카 법칙을 다양한 학문분야에 적용하여 연구를 수행하였다.¹⁷⁾ 그러나 망라적인 서지작성의 문제로 로트카의 원전 연구와의 비교는 어렵게 되었다. 블라키(Jan Vlachy)는 로트카 법칙과 관련된 현상에 관한 서지를 작성하였다.¹⁸⁾ 몇몇 연구들은 로트카의 원저논문의 이러한 요인들과 맞지 않은 것으로 나타났다. 코일(R. C. Coile)은 로트카 법칙에 통계적인 비교를 하기 위해서는 가능한한 로트카가 연구한 그 방법을 따르도록 강조하였다.¹⁹⁾

대부분의 저자들의 생산성에 관한 연구는 저자들을 구성하는 학문분야의 설정이 중요하다. 또한 연구자들은 특정분야에 관심을 가지고 있으며 주제분야에 따라 학술잡지의 기여도가 다르기 때문에 몇 해 동안 몇몇 잡지의 서지조사를 통하여 연구된 논문은 저자들의 생산성을 비교적 정확하게 표출해 낼 수 없다. 로트카 법칙과 비교할 만한 두편의 연구가 최근에 이루어졌다. 이 연구들은 학술잡지 기사가 아니라 단행본을 연구대상으로 삼았다. 그것은 미국의회도서관(LC)이 MARC

16) L.J. Murphy, "Lotka's Law in the Humanities? JASIS, 1973, 24(6), 461-462.

17) A.E.Schorr, "Lotka's Law and Map Librarianship," JASIS, 1975, 26(3), 189-190.

T. Radhakrishnan and R. Kernizan, "Lotka's Law and Computer Science Literature", JASIS, 1979, 30(1), 51-54.

A.E. Schorr, "Lotka's Law and History of Legal Medicine", Research in Librarianship, 1975, 30(5), 205-209.

A.E. Schorr, "Lotka's Law and Library Science", Reference Quarterly, 1974, 14(1), 32-33.

Miranda L. Pao, "An Empirical Examination of Lotka's Law", JASIS, 1986, 37(1), 26-33.

18) J.Vlachy, "Frequency Distributions of Scientific Performance A Bibliography of Lotka's Law and Related Phenomena, "Scientometrics, 1978, 1, 109-130.

19) Russell C. Coile, "Lotka's Frequency Distribution of Scientific Productivity", JASIS, 1977, 28(6), 369-370.

테이프의 모든 저자표제를 조사한 것²⁰⁾ 과 일리노이 대학도서관 카드목록으로 2,500,000 표제의 기록을 가지고 연구한²¹⁾ 것을 들 수 있다. 일리노이대학의 연구는 무작위로 2345명의 저자를 추출하여 조사한 결과 로트카의 이론적 추세(곡선의 경사)와 일치하였으며 K-S 테스트는 이론적으로 적합한 것으로 나타났다(<표 3>참조).

<표 3> 일리노이대학도서관 저자비교

저자/ 제목	이론적 빈도 (Lotka)	이론적 누적 빈도Fo(X)	관찰된 빈도 (Illinois)	관찰된 누적 빈도Sn(X)	최대편차 Fo(X)-Sn(X)
1	0.6079	0.6079	0.6350	0.6350	0.0271
2	0.1520	0.7599	0.1463	0.7813	0.0214
3	0.0675	0.8274	0.0682	0.8495	0.0221
4	0.0380	0.8654	0.0392	0.8887	0.0233
5	0.0243	0.8897	0.0188	0.9075	0.0178
6	0.0169	0.9066	0.0149	0.9224	0.0158
7	0.0124	0.9190	0.0115	0.9339	0.0149
8	0.0095	0.9285	0.0077	0.9416	0.0131
9	0.0075	0.9360	0.0051	0.9467	0.0107
.
.

$$\text{최대편차 } D = \text{Max } |F_o(X) - S_n(X)| = 0.0271$$

$$\text{유의도 } 0.01 \text{ 에서 } K-S \text{ 통계} = 1.63\sqrt{2345} = 0.0337 \quad D < 0.0337$$

그러므로 일리노이대학도서관의 자료는 로트카법칙을 따른다.

20) Sally H. McCallum and James L. Godwin,

"Statistics Headings in the MARC File", Network Development Office, Library of Congress, unpublished paper, 5 Jan, 1981.

21) William C. Potter, "When Names Collide: Conflict in the Catalog and AACR 2," Library Resources & Technical Services 24(Winter 1980), pp.3-16.

MARC테이프에 대한 LC연구는 1969~1979년의 10년동안 1,336,182건의 기계가독형 목록기록을 포함하고 있다. 이 연구의 결과는 로트카의 이론적 법칙에 적합하지 않은 것으로 나타났다. 그 몇가지 원인으로 LC 기계가독형 목록자료에서 저자와 함께 주제도 저자로 간주되었을 가능성과 일리노이 대학의 경우 목록작성 초기부터 현재까지 망라적인 저자를 포함한다에 대하여 LC의 경우는 10년간의 자료이기 때문이 아닌가 생각된다. 이것은 로트카의 연구에서도 화학분야는 연구대상기간이 10년간이었고 물리학분야는 1900년까지의 모든 자료를 포함하였기 때문에 물리학분야는 적합하고 화학분야는 적합하지 않은 결과를 가져온 것과 유사한 결과를 보인것이다. 여러 연구의 결과 광범위한 저자군과 오랜기간 동안의 망라적인 서지자료를 대상으로 할 때 로트카법칙과 적합한 것으로 나타나고 있다.

그러나 특기할 사항은 LC와 일리노이 연구는 AACR2의 발간계획과 연관되어 실제적인 관리문제를 위하여 편집 수행되었다는 것이다. 다른 계량정보학 법칙들이 실무계획에 이용되는 경우로서는 브래드포드 분산법칙이 있으며 이 법칙은 연속간행물 수서정책수립에 활용되고 있다. LC와 일리노이 연구는 로트카법칙이 실무의 계획단계에서 이용되었다는 것을 최초로 알려주게 된 연구이다.

대부분 한국문헌을 분석한 연구들은 로트카의 $n=2$ 의 로트카 법칙이 적용됨을 증명하고 있다.²²⁾

4. 수학분야 학술정보의 생산성 분석

우리나라 수학분야 학자들의 학술정보 생산성을 분석하기 위하여 학술원에서 출판된 학술총람²³⁾의 수학분야 1946~1979년까지의 망라적인 서지사항을 컴퓨터에 입력하여 표 4를 작성하였다. (<표 4> 참조). 표 4에 의하면 34년동안 생산된

22) 최희윤, 경영학문헌의 계량서지학적 분석에 관한 연구, 연세대학교 대학원 석사학위논문, 1983.

최승주, 전자계산학문헌에 대한 계량서지학적 연구, 중앙대학교 대학원 석사학위 논문, 1984.

최정혜, 도서관학문헌의 계량서지학적 연구, 숙명대학교 대학원 석사학위논문, 1987.

박성미, 한국통계학문헌의 계량서지학적 분석, 숙명대학교 석사학위논문, 1987.

23) 學術院, 學術總覽 第2編, 數學篇(1946-1979), 1982, 472.

수학분야의 총 논문편수는 2006편이며 논문생산에 참여한 저자는 800명이었다.

〈표 4〉 수학분야 저자들의 논문 생산성

논문수	저자수	저자비(%)	논문 편수
1	465	58.13	465
2	131	16.38	262
3	59	7.38	177
4	35	4.38	140
5	30	3.75	150
6	19	2.37	114
7	15	1.88	105
8	13	1.63	104
9	6	0.75	54
10	5	0.63	50
11	4	0.50	44
12	2	0.25	24
13	6	0.75	78
14	3	0.37	42
16	1	0.12	16
17	2	0.25	34
22	1	0.12	22
31	1	0.12	31
46	1	0.12	46
48	1	0.12	48
계	800	100.00	2006

〈표 5〉 수확분야 저자들의 논문생산성 분포

논문 수	저자수	관찰된 빈도	관찰된 누적 빈도Sn(X)	이론적 빈도 (Lotka)	이론적 누적 빈도Fo(X)	최대편차 Fo(X)-Sn(X)
1	465	0.5813	0.5813	0.6079	0.6079	0.0266
2	131	0.1638	0.7451	0.1520	0.7599	0.0148
3	59	0.0738	0.8189	0.0675	0.8274	0.0085
4	35	0.0438	0.8627	0.0380	0.8654	0.0027
5	30	0.0375	0.9002	0.0243	0.8897	0.0105
6	19	0.0237	0.9239	0.0169	0.9066	0.0173
7	15	0.0188	0.9427	0.0124	0.9190	0.0237
8	13	0.0163	0.9590	0.0095	0.9285	0.0305
9	6	0.0075	0.9665	0.0075	0.9360	0.0305
10	5	0.0063	0.9728	0.0061	0.9421	0.0307
.
.

$$\text{최대편차 } D = \text{Max } |F_o(X) - S_n(X)| = 0.0307$$

$$\text{유의도 } 0.01 \text{ 에서 } K-S \text{ 통계} = 1.63\sqrt{800} = 1.63/28.28 = 0.0576$$

$$D < 0.0576$$

그러므로 수확분야 저자들의 생산성은 로트카 법칙을 따르는 것으로 입증되었다.

5. 기계공학분야 학술정보의 생산성 분석

기계공학분야는 1985년에 생산된 학술총람²⁴⁾ 을 참조로 1957년~1983년까지의

24) 學術院, 學術總覽 第28輯, 機械·造船工學篇(1957-1983), 1984, 285.

망라적인 서지사항을 컴퓨터에 입력한후 자료를 전산처리하여 <표 6>을 작성하였다. 27년동안 기계공학분야의 총논문편수는 1826편이며 연구에 참여한 저자는 696명으로 나타났다.

<표 6> 기계공학분야 저자들의 논문 생산성

논문수	저자수	저자비(%)	논문 편수
1	429	61.64	429
2	99	14.22	198
3	42	6.03	126
4	33	4.74	132
5	21	3.01	105
6	9	1.29	54
7	11	1.58	77
8	13	1.87	104
9	6	0.86	54
10	4	0.57	40
11	1	0.14	11
12	5	0.72	60
13	4	0.57	52
14	1	0.14	14
15	3	0.43	45
16	3	0.43	48
17	3	0.43	51
18	1	0.14	18
20	2	0.29	40
21	3	0.43	63
24	1	0.14	24
25	1	0.14	25
46	1	0.14	46
계	696	100.00	1,826

<표 7> 기계공학분야 저자들의 논문 생산성 분포

논문 수	저자수	관찰된 빈도	관찰된누적 빈도Sn(X)	이론적 빈도 (Lotka)	이론적 누적 빈도Fo(X)	최대편차 Fo(X)-Sn(X)
1	429	0.6164	0.6164	0.6079	0.6079	0.0085
2	99	0.1422	0.7586	0.1520	0.7599	0.0013
3	42	0.0603	0.8189	0.0675	0.8274	0.0085
4	33	0.0474	0.8663	0.0380	0.8654	0.0009
5	21	0.0301	0.8964	0.0243	0.8897	0.0067
6	9	0.0129	0.9093	0.0169	0.9066	0.0027
7	11	0.0158	0.9251	0.0124	0.9190	0.0061
8	13	0.0187	0.9438	0.0095	0.9285	0.0153
9	6	0.0086	0.9524	0.0075	0.9360	0.0164
10	4	0.0057	0.9581	0.0061	0.9421	0.0160
.
.

최대편차 $D = \text{Max } |Fo(X) - Sn(X)| = 0.0164$

유의도 0.01에서 $K-S \text{ 통계} = 1.63 \sqrt{696} = 1.63 / 26.38 = 0.0618$

$D < 0.0618$

그러므로 기계공학 분야 저자들의 학술정보 생산성분포는 로트카 법칙을 따르는 것으로 입증되었다.

6. 결론 및 제언

로트카법칙은 화학과 물리학분야 저자들의 표본조사를 통해 저자들의 생산성을 측정할 연구의 결과로 얻어진 법칙으로서 원저논문에서 두개의 표본조사를 통해 얻어진 일반화법칙이라고 할 수 있다. 즉 로트카법칙은 '역제곱법칙'에 기초한 저

자의 생산성에 관한 법칙으로서 n 개의 논문을 쓴 저자의 수는 1편의 논문을 쓴 저자들의 $1/n^2$ 이라는 것이다.

본 연구는 계량정보법칙중 로트카법칙을 중심으로 그 연구내용과 발전과정을 조사하였다. 그리고 한국의 경우 수학과 기계공학분야의 망라적 서지자료를 컴퓨터에 입력하여 자료를 처리하여 로트카의 학술정보 생산성의 빈도분포를 적용하였다. 검증한 결과 수학과 기계공학 두 분야 모두 로트카의 법칙을 따르는 것으로 입증되었다. 이제까지 이루어진 연구들중 로트카 법칙에 적합하지 않은 경우는 통계 처리에서 로트카의 방법을 따르지 않거나 너무 적은 잡지 종수, 대상기관과 주제의 선정, 저자집단의 많고 적음, 출판물의 형태 등으로 인하여 적절한 자료 표본집단 구성에 문제가 있기 때문인 것으로 본다.

파오(M. Pao)는 로트카 법칙을 검증하기 위하여 다음과 같은 연구 절차 방법을 권고하였다.

1. 주제안에서 저자와 생산된 논문에 관한 적절한 데이터 수집
2. 생산된 논문수와 저자들의 빈도수
3. 로트카 법칙을 위한 경사도 n 값을 구한다.
4. $C = 1/X^n$ 이며 $X=1, 2, 3, \dots, 10$ 인데 이때의 C 값을 구한다.
5. 로트카의 이론 분포에 대해 K-S test를 한다. (Kolmogrov - Smirnov test of goodness-of-fit)

앞으로 도서관정보관리 업무의 실무에 이러한 법칙이 활용되고 이 계량 모형을 이용하여 정보의 구조분석, 예측 그리고 정책수립 또는 평가를 수행할 수 있기를 기대한다.

참 고 문 헌

- 金光雄, 社會科學研究方法論, 서울:博英社, 1977.
- 金賢姬 & 金容琬, 計量情報學, 서울, 구미무역(주) 출판사, 1993. 5.
- 유지성, 計量經濟學原論, 서울:博英社, 1985, 9-16.
- 學術院, 學術總覽 第28緝, 機械. 造船工學篇(1957~1983), 1984,
- 學術院, 學術總覽 第2緝, 數學篇(1946-1979). 1982.
- Beheshti Jamshid Tague, Jean, & Rees-Potter L., The Law of Exponential Growth : Evidence, Implications and Forecasts, Library Trends, 1981(30) (1), 125-149.
- Coile Russell C., "Lotka's Frequency Distribution of Scientific Productivity," JASIS, 1977, 28(6), 369-370.
- Drott, Carl M., Bradford's Law : Theory, Empiricism and The Gap Between, Library Trends, 1981, (30), (1) 41=52.
- Egghe L. & R.Rousseau, Introduction to Informetrics:Quantitative Methods in Library, Documentation and Information Science, Amsterdam, Elsevier, 1990, 1.
- Gapen, Kaye K. & Sigrid P. Milner, Obsolescence, Library Trends, 1981 (30), (1), 107-124.
- Hubert, John J., General Bibliometric Models, Library Trends, 1981, (30) (1), 65-81.
- Lotka, Alfred J., "The Frequency Distribution of Scientific, Productivity," Journal of Washington Academy of Sciences 16, (19 June 1926).
- McCallum, Sally H. and James L. Godwin, "Statistics Headings in the MARC File," Network Development Office, Library of Congress, unpublished paper, 5 Jan, 1981.
- Murphy L. J. , Lotka's Law in the Humanities? JASIS, 1973, 24 (6), 461-462.

- Potter, Gray W., Lotka's Law revisited, *Library Trends*, 1981, (30), (1), 21-40.
- Potter, William G., "When Names Collide : Conflict in the Catalog and AACR 2," *Library Resources & Technical Services* 24 (Winter 1980).
- Potter, William G., "Bibliometrics Introduction," *Library Trends*, 1981, 30 (1), 5-6.
- Potter, William G., "Lotka's Law Revisited," *Library Trends*, 1981, 30 (1), 21-39.
- Schorr A. E. , "Lotka's Law and Map Librarianship," *JASIS*, 1975, 26 (3), 189-190.
- Smith, linda C., *Citation Analysis*, *Library Trends*, 1981, (30) (1), 83-106.
- Vlachy J. , "Frequency Distributions of Scientific Performance A Bibliography of Lotka's Law and Related Phenomena," *Scientometrics*, 1978, 1, 109-130.
- Wyllys, Ronald E., *Empirical and Theoretical Bases of Zipf's law*, *Library Trends*, 30, 1 summer 1981.

ABSTRACT

Lotka's Law and the Frequency Distribution of Scientific Productivity of Mathematicians and Mechanical Engineers.

Bock-Hee Hahn

In 1926, Alfred Lotka examined the frequency distribution of scientific productivity of chemists and physicists. He observed that the number of persons making n contributions is about $1/n^2$ of those making one and the proportion of all contributions that make a single contribution is about 30%.

Investigator studying the applicability of "Lotka's Law" to Mathematics and Mechanical engineers have fitted Lotka's Law and concluded that the law applied to these subject fields.