

林业大学学报论文第1作者和合作者的洛特卡分布对比

1926年,美国统计学家洛特卡在《科技生产率的频率分布》中首次揭示了表明科学论文作者与论文数量之间关系的洛特卡定律。自那以后,国外不少学者运用洛特卡定律继续研究科技论文作者的分布规律问题,如: Dresden 对美国数学家的研究; Dufrenoy 和 Williaus 对美国生物学家的研究; Hubert 对加拿大数学家的研究。在我国,王崇德和刘东维也先后对我国情报科学期刊论文的作者分布进行了研究。这些研究都属于对科技论文第1作者分布问题的研究。随着科技发展的日新月异和研究规模的日趋扩大,对日趋广泛的合作者的研究也已提到了议事日程之上。国外学者 Pao 对此已进行了研究,并对论文的第1作者和合作者进行了对比分析。本文试图作为国内这一研究的首次尝试,并希望能将这一有意义的研究持续进行下去。

一、数据来源

我们以1979~1989年以下3个大学的学报数据作为分析、对比的依据:一是北京林业大学(以下简称“北林大”)学报,包括论文537篇,论文第1作者325人,论文合作者511人;二是南京林业大学(以下简称“南林大”)学报,包括论文552篇,第1作者331人,合作者669人;三是东北林业大学(以下简称“东林大”)学报,包括论文734篇,第1作者463人,合作者979人。

二、数据处理

对搜集的数据,我们首先舍取数据组数 $N \geq 8$ 的数据,计算出处理数据中所包含的作者百分比;再利用最小二乘法计算出 n 值,得到回归方程,并计算出相应论文的估计作者数;然后用相关指数 (r^2) 和剩余标准差 (S) 公式计算出论文第1作者和合作者的相关指数 (r^2) 和剩余标准差 (S);接着利用 Pao 的 C 值计算公式求出论文第1作者和合作者的 C 值和理论频率分布;最后进行 K-S 检验。

三、结果分析

(一) 三种林业大学学报科技论文作者科学生产率的频率分布比较(见表1)。

由统计数据可见,科技论文第1作者数与合作者数相距甚大。其中,北林大学报两者相距为186人,南林大学报为338人,东林大学报为516人。由此可见,三种林业大学学报论文合作者数与第1作者数之差相对于第1作者数的增长趋势是:北林大学报合作者增长57.23%;南林大学报合作者增长102.00%;东林大学报合作者增长11.45%。这一情况充分证明,林业学科科技论文作者研究不仅要考虑第1作者的贡献,也应对合作者的贡献加以重视。

(二) 三种林业大学学报论文第1作者数与合作者数的对比(见表2)。

表1 中国三所林业大学科学生产率频率分布

论文数量(X)	北京林业大学				南京林业大学				东北林业大学			
	Y _{x,1}	Y _{x,1} / ΣY _{x,1}	Y _{x,合}	Y _{x,合} / ΣY _{x,合}	Y _{x,1}	Y _{x,1} / ΣY _{x,1}	Y _{x,合}	Y _{x,合} / ΣY _{x,合}	Y _{x,1}	Y _{x,1} / ΣY _{x,1}	Y _{x,合}	Y _{x,合} / ΣY _{x,合}
1	216	0.6646	326	0.6380	219	0.6616	416	0.6218	313	0.6760	649	0.6629
2	66	0.2031	93	0.1820	61	0.1843	121	0.1809	95	0.2052	176	0.1737
3	16	0.0492	37	0.0724	21	0.0634	53	0.0867	25	0.0540	74	0.0756
4	15	0.0462	28	0.0548	14	0.0423	28	0.0418	14	0.0302	43	0.0439
5	6	0.0185	10	0.0196	10	0.0302	24	0.0359	7	0.0151	19	0.0194
6	2	0.0062	7	0.0137	2	0.0060	3	0.0045	2	0.0043	6	0.0061
7	2	0.0062	5	0.0098	3	0.0091	6	0.0090	4	0.0086	4	0.0041
8	0	0	2	0.0039	0	0	5	0.0075	2	0.0043	9	0.0092
9	0	0	0	0	1	0.0030	3	0.0045	1	0.0022	3	0.0031
10	0	0	0	0	0	0	0	0	0	0	2	0.0020
11	1	0.0031	2	0.0039	0	0	1	0.0015	0	0	0	0
12	0	0	0	0	0	0	1	0.0015	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0
14	1	0.0031	1	0.0020	0	0	1	0.0015	0	0	0	0
15	0	0	0	0	0	0	2	0.0030	0	0	0	0
总计	325		511		331		669		463		979	

表2 中国三种林业大学学报论文第1作者数与合作者数的对比

期刊	作者数	ΣY _x	论文总数	包含的作者(%)	r ²	剩余标准差	n	C	K-S值(0.01显著性水平)	Dmax
北林大学学报	第1作者数	325	537	99.4	96.1	0.1631	2.4932	0.7439	0.0904	0.0793
	合作者数	511	537	99.0	98.7	0.0951	2.1698	0.6622	0.0721	0.0440
南林大学学报	第1作者数	331	552	99.7	94.6	0.1816	2.3242	0.7043	0.0869	0.0427
	合作者数	669	552	98.1	90.1	0.2535	2.3473	0.7169	0.0630	0.0951
东林大学学报	第1作者数	463	734	99.4	95.5	0.1807	2.5439	0.7547	0.0758	0.0783
	合作者数	979	734	98.6	95.4	0.1861	2.5904	0.7641	0.0521	0.1012

以北林大学学报论文第1作者回归直线的相关指数(r²)和剩余标准差(S)数据的处理过程为例:

相关指数(r²)和剩余标准差(S)公式为:

$$r^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y}_i)^2},$$

$$S = \sqrt{\frac{\sum(Y_i - \hat{Y}_i)^2}{N-2}}$$

$$\text{而} \sum(Y_i - \bar{Y}_i)^2 = \sum Y_i^2 - \frac{1}{N}(\sum Y_i)^2$$

式中: $\sum(Y_i - \hat{Y}_i)^2$ 为剩余平方和;
 $\sum(Y_i - \bar{Y}_i)^2$ 为总平方和(也称总离差平方和);(N-2)为剩余平方和的自由度;Y_i为

实测作者数;Y_i 为根据回归方程求得相应论文的估计作者数; $\bar{Y}$ 为平均作者数。

由于北林大第1作者的回归方程 $Y = 2.44934 \sim 2.49321X$, $Y = \log Y_x$, $Z = \log X$, 所以 Y² 和 S 公式中的变量均要作相应变换, 则:

$$r^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y}_i)^2},$$

$$S = \sqrt{\frac{\sum(Y_i - \hat{Y}_i)^2}{N-2}},$$

$$\sum(Y_i - \bar{Y}_i)^2 = \sum Y_i^2 - \frac{1}{N}(\sum Y_i)^2 \text{ (处理数据见表3)}.$$

表 3 北林大学学报论文第 1 作者回归直线 r^2 与 S 计算所需数据处理情况

X	Y _x	Y _i (= log Y _x)	Y _i ²	$\hat{Y}_x$	$\hat{Y}_i$ (= log $\hat{Y}_x$)	Y _i - $\hat{Y}_i$	(Y _i - $\hat{Y}_i$) ²
1	216	2.33445	5.44967	281.40785	2.44934	-0.11489	0.01320
2	66	1.81954	3.31074	49.98106	1.69881	0.12073	0.01458
3	16	1.20412	1.44990	18.18751	1.25977	-0.05565	0.00310
4	15	1.17609	1.38319	8.87718	0.94828	0.22781	0.05190
5	6	0.77815	0.60552	5.08930	0.70666	0.07149	0.00511
6	2	0.30103	0.09062	3.23030	0.50924	-0.20821	0.04335
7	2	0.30103	0.09062	2.19953	0.34233	0.04130	0.00171
Σ	323	7.91441	12.38025				0.13295

$$\therefore \Sigma(Y_i - \hat{Y}_i)^2 = 0.13295,$$

$$\therefore S = \sqrt{\frac{0.13295}{7-2}} = 0.1631$$

$$\Sigma(Y_i - \bar{Y}_i)^2 = \Sigma Y_i^2 = \frac{1}{N} (\Sigma Y_i)^2$$

$$= 12.38025 - \frac{1}{7} (7.91441)^2$$

$$= 3.43198$$

$$r^2 = 1 - \frac{0.13295}{3.43198} = 1 - 0.03874$$

$$= 0.96126 = 0.961 = 96.1\%$$

包含的第 1 作者 (%)

$$= 323 / 325 = 0.9938 \approx 99.4\%$$

由表 2 数据可见:

1. 截删作者数在处理数据中是一个关键

问题。1926 年, 洛特卡在研究《化学文摘》作者时舍去了 $N > 30$ 的点, 在研究《物理学史一览表》作者时舍去了 $N > 17$ 的点, 从而得到 $n=2$ 的结论, 因此, 我们在研究三种林业大学学报论文作者时, 第 1 作者数和合作者数均舍去 $N > 7$ 的点。

2. 相关指数 (r^2) 和剩余标准差 (S) 决定回归方程拟合程度。根据洛特卡 $Y_x = C/X^n$ 公式, 两边取对数为 $Y_x = \log C - n \log X$, 令 $Y = \log Y_x, Z = \log x$, 则 $Y = \log C - nX$ 。对三所林大学报 $N=1 \sim 7$ 的数据用最小二乘法去计算时, 分别得到论文第 1 作者数和合作者数的回归方程 (见表 4)。决定表 4 回归方程拟合程度的好坏, 需用相关指数 (r^2) 和剩余标准差 (S) 的大

表 4 中国三所林大学学报论文第 1 作者数和合作者数的回归方程

期 刊	第 1 作者回归方程	合作者回归方程
北林大学报	$Y = 2.44934 \sim 2.49321X$	$Y = 2.58212 \sim 2.16984X$
南林大学报	$Y = 2.42533 \sim 2.32415X$	$Y = 2.74838 \sim 2.34730X$
东林大学报	$Y = 2.59775 \sim 2.54393X$	$Y = 2.97287 \sim 2.59036X$

小来决定。当 r^2 越大, r^2 越接近 1 时, 回归直线效果就越好, 而 S 值却越小 (S 值是回归方程估计作者数精确度的标准)。这说明实测作者数与估计作者数的差距越小, 估计作者数越精确, 回归效果就越好。表 2 数据得到: 三种林大学学报论文第 1 作者 r^2 值

在 94.6~96.1 之间, 合作者 r^2 值在 90.1~98.2 之间, r^2 值均接近 1。其中北林大学学报论文第 1 作者 $r^2 = 96.1$, 合作者 $r^2 = 98.7$, 均为最大。另外, 三种林业大学学报论文第 1 作者 S 值在 0.1631~0.1816 之间, 合作者 S 值在 0.095~0.2535 之间, 它们的 S 值均

较小。其中北林大学学报论文第 1 作者 $S=0.1631$, 合作者 $S=0.0951$, S 值均最小。这些数据表明三种林业大学学报以北林大学学报论文第 1 作者数与合作者数的回归效果为最佳。

3. n 值和 C 值均随论文作者数的变化而变化。表 2 数据得到: 三种林业大学学报论文第 1 作者 n 值在 2.3242~2.5439 之间, C 值在 0.7043~0.7547 之间。它们的合作者 n 值在 2.1698~2.5904 之间, C 值在 0.6622~0.7641 之间。这一情况表明 n 值与 C 值均与作者数有关, 它们与作者数成正比关系。

4. $K-S$ 检验与合作者多寡有关。根据 $K-S$ 检验法, 当显著性水平为 0.01 时, 其临界值等于 $1.63/\sqrt{\Sigma Yx}$, (ΣYx 为作者数的总和)。在表 2 数据中, 北林大学和南林大学学报论文第 1 作者均通过 $K-S$ 检验。这说明洛特卡定律研究第 1 作者时, 在多数情况下是能通过 $K-S$ 检验的。而合作者则只有北林大学学报论文合作者才通过了 $K-S$ 检验。分析其原因, 主要是与合作者的多寡有关, 如南林大学学报论文合作者相对第 1 作者净增 102.00%, 东林大学学报论文合作者相对第 1 作者净增 111.45%。

表 5 Pao 处理论文(期刊论文)第 1 作者和合作者数据的对比情况

期 刊	作者数	论文数	包含的作者 (%)	r^2	剩余标准差	n	C	$K-S$ 值(0.01 显著性水平)	Dmax
Map librarianship(第 1 作者)	306	463	98.4	97.2	0.1547	2.8370	0.8079	0.0932	0.0399
Map librarianship(合作者)	326	463	98.5	97.0	0.1653	2.9292	0.8213	0.0903	0.0698
Legal medicine(第 1 作者)	997	1487	99.0	100.0	0.0126	2.4737	0.7397	0.0516	0.0228
Legal medicine(合作者)	1010	1487	99.3	98.7	0.1031	2.7478	0.7932	0.0513	0.0338
History of technology(第 1 作者)	164	231	100.0	90.5	0.2695	2.6161	0.7691	0.1273	0.0354
History of Technology(合作者)	170	231	100.0	90.8	0.2678	2.6469	0.7750	0.1250	0.0333
Library science-LQ(第 1 作者)	198	229	100.0	96.9	0.1752	3.0964	0.8446	0.1158	0.0645
Library Science-LQ(合作者)	210	229	100.0	96.8	0.1811	3.1352	0.8494	0.1125	0.0649
Computational musicology(第 1 作者)	458	970	99.1	91.4	0.2313	2.1680	0.6616	0.0762	0.0238
Computational musicology(合作者)	544	970	99.1	88.9	0.2726	2.2725	0.6909	0.0699	0.0310
American Revolution(第 1 作者)	764	1075	100.0	92.0	0.2708	2.6321	0.7722	0.0590	0.0262
American Revolution(合作者)	790	1075	99.2	95.7	0.1951	2.8364	0.8078	0.0580	0.0158
American Revolution-shy(第 1 作者)	1316	2043	97.8	99.2	0.0712	2.4822	0.7416	0.0449	0.0197
American Revolution-shy(合作者)	1356	2043	97.9	98.6	0.0964	2.5469	0.7551	0.0443	0.0203
American Rev-shy(62-71)(第 1 作者)	416	587	100.0	92.0	0.2695	2.7772	0.7981	0.0799	0.0217
American Rev-shy(62-71)(合作者)	413	587	100.0	92.1	0.2725	2.8107	0.8036	0.0785	0.0426
Ethnomusicology(第 1 作者)	2269	4434	98.4	99.4	0.0620	2.2851	0.6942	0.0342	0.0237
Ethnomusicology(合作者)	2422	4655	98.5	99.6	0.0511	2.2914	0.6959	0.0331	0.0291

四、与国外研究成果的比较

Pao 截删作者数采用洛特卡作图观点与我们处理方法一致。Pao 处理了 9 种期刊论文的第 1 作者和合作者, 第 1 作者 r^2 为 90.5~100.0, S 为 0.0126~0.2708, 合作者 r^2 为 88.9~99.6, S 为 0.0511~0.2726。充分说明它们的回归方程的拟合程度都比较好(见表 5)。从统计数据可以看到, 期刊论文第 1 作者 n 值为 2.1680~3.1352, C 值为 0.6616~0.8446。期刊论文合作者 n 值为 2.2725~3.1352, C 值为 0.6909~0.8494。它们的 n 值和 C 值均随作者数的递增而增加, 这一结果与我们的研究结果是相同的。Pao 处理 K-S 检验时, 期刊论文第 1 作者与合作者数据之差相对第 1 作者数之比在 1.30%~18.78%之间, 故它们的第 1 作者和合作者均通过 K-S 检验符合洛特卡分布。而我们进行 K-S 检验时, 期刊论文第 1 作者与合作者数据之差相对第 1 作者数之比在 57.23%~111.45%之间, 便出现有部分第 1 作者和合作者不能通过 K-S 检验的

情况。

五、结论

(一) 三种林业大学学报科技论文第 1 作者和合作者研究适用于洛特卡一般公式 $Y_x = C/X^n$ 。

(二) n 值的大小随学科而异, 也随作者数的增加而增加。

(三) 论文第 1 作者数据多数通过 K-S 检验, 符合洛特卡分布。而 K-S 检验论文合作者的情况则比较复杂: 当合作者数与第 1 作者数之差越小时, 合作者数据就能通过 K-S 检验, 反之就不通过。这是值得进一步研究的。

(四) 我国三所林业大学科技论文合作者现象较为普遍, 在研究第 1 作者与合作者分布对比时, 对国外学者的结论, 应慎重对待。

(作者单位: 赵世华、袁浩: 北京林业大学; 张秦生: 北京市科技情报学会。来稿时间: 1991.3。编发者: 丘峰。)

(上接 32 页)

参考文献

- (1) 杨智辉. 论图书馆时空. 云南民族学院学报, 1988, (2)
- (2)、(5) 鲍振西. 国外图书馆事业见闻与观感. 黑龙江图书馆, 1987, (1)
- (3) 鲍振西. 澳大利亚图书馆见闻与观感. 图书馆工作与研究, 1988, (4)
- (4)、(11) 王赣华、黄健. 我国公共图书馆事业发展水平预测. 图书馆界, 1988, (3)
- (6) 桑健. 关于文献情报资源开发利用问题的思考. 黑龙江图书馆, 1987, (1)
- (7) 鲍振西. 坚持改革, 加强协作, 发展和建设我国图书馆事业. 图书馆学通讯, 1987, (4)
- (8) (美) 玛利·J·德卫勒著, 贾晓东译, 林祖藻

校. “最佳规模”的公共图书馆. 图书馆研究与工作, 1987, (2)

(9) 黄纯元. 论我国图书馆事业的结构. 图书馆学通讯, 1988, (4)

(10)、(13) 沈国弟. 浅论分馆制. 图书馆研究与工作, 1988, (1)

(12) 中国统计年鉴(1988). 北京: 中国统计出版社, 1988

(14) 董纯元. 我国图书馆事业发展战略的若干思考. 图书馆学通讯, 1986, (3)

(15) 美国图书馆协会资源技术服务部藏书发展委员会颁布, 肖自力译. 藏书发展方针规范指南. 黑龙江图书馆, 1982, (2)

(16) 肖自力. 试论藏书结构. 图书情报工作, 1981, (1)

(作者单位: 金陵图书馆。来稿时间: 1991.5。编发者: 丘峰。)

A Comparison of Lotka Distribution of the First Authors and co-authors of Articles Published in Journals of Universities of Forestry / Zhao Shihua, Yuan Hao and Zhang Qinsheng / / Bulletin of the Library Science in China / China Society of the Library Science. -1992,18(1). -62~66

In 1926, Lotka, an America Statistician, established for the first time the Law of Lotka which indicated the relationship between authors of scientific treatises and the quantity of them. The article takes three titles of journals published by universities of forestry as an example and makes an expository comparison which cover four parts as data source, data processing, result analysis, and research achievements of foreign countries. 1 illus. 5 tables.

Law of Lotka - Studies

Library science and information science - Theories

Scientific treatises - Authors

G256

Three-dimensional Thinking and Library Science Research / Li Guiyong / / Bulletin of the Library Science in China / China Society of the Library Science. -1992,18(1). -67~69

The other name of the "three-dimensional thinking mode" is "integral thinking mode". The present-day library undertaking is a three-dimensional network crisscrossed with various factors which help each other forward and are implicative of each other. To apply the three-dimensional thinking to observe and study library science will surely grasp the whole of the study object. It should be the document information communication carried out by the library, the library undertaking and its related factors.

Library science - Studies

Scientific research methods - Three-dimensional thinking

G250

Considerations Concerning the Forming of Great Information Science / Chen Dahui / / Bulletin of the Library Science in China / China Society of the Library Science. -1992,18(1). -70~74

The Study object of the great information science is the brain system of man. The content of the study is the structure, the moving mechanism and the integral function of the system. The forming of the great information science does not mean that theories of library science, information science, bibliography, archivistics and publication and distribution lose their value of existence. The forming of the great information science lays a more solid foundation for them. 2 illus. 11 references.

Great information science - Approaches

Brain system of man - Theories

G350