Library and Information Services in Astronomy IV
July 2-5, 2002, Prague, Czech Republic
B. Corbin, E. Bryson, and M. Wolf (eds)

A Comparative Study of the Distribution of Authors of the "Revista Mexicana de Astronomía y Astrofísica" with the Application of the Lotka Law

Veronica Mata-Acosta

Instituto de Astronomía, UNAM, Apartado Postal 70-244, C.P. 04510, México, D.F., México. veronica@astroscu.unam.mx

Maria Elena Jiménez-Fragozo

Instituto de Astronomía, UNAM, Observatorio Astronómico Nacional, Apartado Postal 877, Ensenada, B.C., México. jimenez@astrosen.unam.mx

Abstract. With the application of Lotka law, we sought information regarding the publications in *RevMexAA* from the years 1979-2001 and 1974-1998.

1. Introduction

We have used the PERIODICA database (database compiled by the Dirección General de Bibliotecas of the UNAM, that offers a published bibliographic register of documents of around 1300 Latin-American journals specialized in science and technology). From this database, we took from Vol. 4, 1979 to Vol. 36, 2001, and also the study compiled by the Institute of Astronomy librarians from Vol. 1, 1974 to Vol. 34, 1998. The comparison of the same journal, with the help of two different databases, gave us two different results, since the data provided by the PERIODICA database, covered a period of 22 years, resulting in 1176 articles written by 640 different authors, whereas the 24 year database at the Institute of Astronomy was comprised of 1665 articles and 786 authors.

2. Lotka's Model and Applications

Starting from the methodology proposed by Lotka, the frequencies of authors and of their contributions were identified and calculated according to a model for both databases. In the PERIODICA database, the results were as follows:

In Table 1 there are nine columns in which the frequency of authors with their observed contributions are distributed. In addition, the columns calculate the $s(x)$ and $f(x)$ in accordance with the accumulated rate of authors who have observed. According to the methodology of Kolmogrov-Smirnov statistical graph, the columns are ordered as follows:

Table 1.

A CONTRIB.	B AUT. OBSER.	C ACU. OBSER.	D $s(x)$	E $1/(n)^2$	F TEOR. ACU.	G $f(x)$	H $f(x) - s(x)$	I $\|f(x) - s(x)\|$
1	422	422	0.659375	422.00	422.00	0.65937500	–	–
2	115	537	0.84037559	105.50	527.50	0.82550861	-0.01486698	0.01486698
3	44	581	0.91065831	46.89	574.39	0.90029606	-0.01036224	0.01036224
4	17	598	0.94025157	26.38	600.76	0.94459731	0.00434574	0.00434574
5	16	614	0.96692913	16.88	617.64	0.97266754	0.00573841	0.00573841
6	4	618	0.97630332	11.72	629.37	0.99425926	0.01795594	0.01795594
7	5	623	0.99046105	8.61	637.98	1.01427402	0.02381297	0.02381297
8	3	626	1	6.59	644.57	1.02966790	0.02966790	0.02966790
9	3	629	1.00963082	5.21	649.78	1.04298874	0.03335792	0.03335792
10	4	633	1.02427184	4.22	654.00	1.05825564	0.03398379	0.03398379
11	2	635	1.03420195	3.49	657.49	1.07082994	0.03662799	0.03662799
12	1	636	1.06354515	2.93	660.42	1.10438151	0.04083636	0.04083636
14	2	638	1.09810671	2.15	662.57	1.14040138	0.04229467	0.04229467
15	1	639	1.18994413	1.88	664.45	1.23733475	0.04739061	0.04739061
17	1	640	1.51658768	1.46	665.91	1.57798333	0.06139565	0.06139565
TOTALES	640							

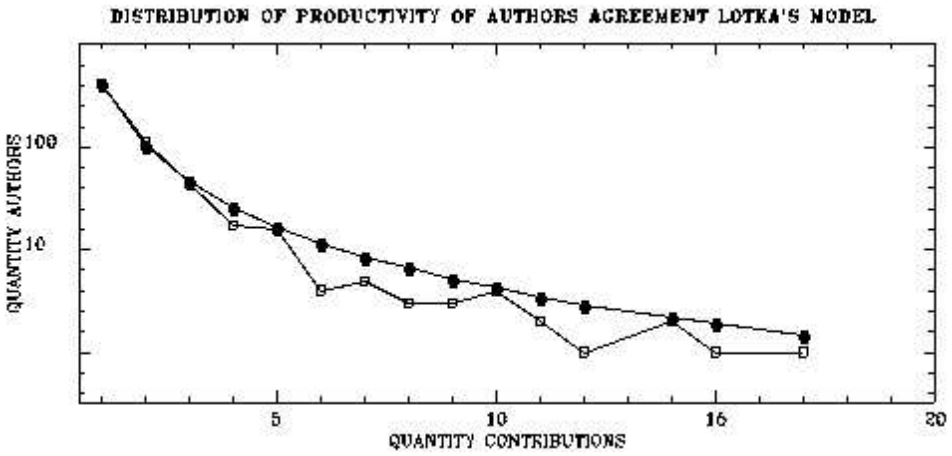

Figure 1. The figure show the observed authors (□) and calculated authors (•)

A = the quantity of observed contributions.
B = the quantity of observed authors.
C = the quantity of observed and accumulated authors.
D = the rate of observed accumulated authors $s(x)$.
E = the calculation of the Lotka model (calculated authors).
F = the theoretical accumulated quantity of authors.
G = the rate of calculated accumulated authors $f(x)$.
H = the difference between $f(x)$ and $s(x)$.
I = the absolute value of the data obtained with the H column.

The calculation of the E column was carried out according to the Lotka formula. In order to verify the statistical equality between the observed sample and the calculated one, the Kolmogrov-Smirnov statistical graph was used:

$$D = \text{maximum}[f(x) - s(x)]$$

where $f(x) =$ is the accumulated theoretical distribution. $s(x) =$ is the observed accumulated distribution. $D =$ maximum $= 0$ as indicated in the table. $N = 640$ main authors of the journal.

The calculation was as follows:

$$\begin{aligned} K - S &= 1.63/\sqrt{n} \\ K - S &= 1.63/\sqrt{640} \\ K - S &= 1.63/25.29 = 0.064452352 \\ 0.064452352 &> 0 \end{aligned}$$

This statistical graph establishes that when the resulting or calculated value is greater than or equal to the maximum identified distance between the observed and the calculated sample it can be affirmed that statistically such behaviors are regular.

Therefore, the data of the sample do not fit the Alfred Lotka model's postulates on authors' productivity. We can verify this in the Figure 1, which according to the Lotka's propositions, results in a logarithmic scale for both axes. The correlation between the curves representing the observed authors and the one representing the observed authors is shown. These curves do not show the characteristics of a typical inverse square curve of the data they represent.

This same exercise was also applied on the data obtained at the library, and gave us the following results : (Table 2)

Table 2.

A CONTRIB.	B AUT. OBSER.	C ACU. OBSER.	D $s(x)$	E $1/(n)^2$	F TEOR. ACU.	G $f(x)$	H $f(x) - s(x)$	I $\lvert f(x) - s(x) \rvert$
1	481	481	0.611959288	481.00	481.00	0.62883157	0.01687228	0.01687228
2	128	609	0.77480916	120.25	601.25	0.78603946	0.01123030	0.01123030
3	71	680	0.865139949	53.44	654.69	0.85590964	-0.00923031	0.00923031
4	34	714	0.908396947	30.06	684.76	0.89521161	-0.01318533	0.01318533
5	23	737	0.937659033	19.24	704.00	0.92036488	-0.01729416	0.01729416
6	10	747	0.95038168	13.36	717.36	0.93783242	-0.01254926	0.01254926
7	12	759	0.965648855	9.82	727.17	0.95066572	-0.01498314	0.01498314
8	5	764	0.972010178	7.52	734.69	0.96049121	-0.01151897	0.01151897
9	5	769	0.978371501	5.94	740.63	0.96825456	-0.01011694	0.01011694
10	4	773	0.98346056	4.81	745.44	0.97454288	-0.00891768	0.00891768
11	2	775	0.98600509	3.98	749.41	0.97973983	-0.00626526	0.00626526
12	3	778	0.989821883	3.34	752.75	0.98410672	-0.00571516	0.00571516
13	2	780	0.992366412	2.85	755.60	0.98782762	-0.00453880	0.00453880
14	1	781	0.993638677	2.45	758.05	0.99103594	-0.00260274	0.00260274
16	1	782	0.994910941	1.88	759.93	0.99349231	-0.00141863	0.00141863
17	1	783	0.996183206	1.66	761.60	0.99566820	-0.00051500	0.00051500
18	1	784	0.997455471	1.48	763.08	0.99760904	0.00015357	0.00015357
22	1	785	0.998727735	0.99	764.08	0.99890828	0.00018054	0.00018054
24	1	786	1	0.84	764.91	1.00000000	0.00000000	0.00000000
TOTALES	786							

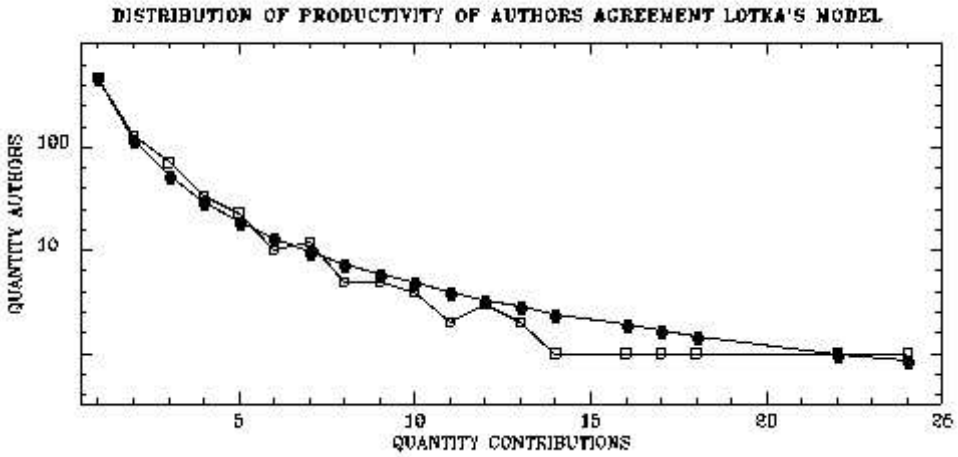

Figure 2. The figure show the observed authors (□) and calculated authors (•)

3. The Kolmogrov-Smirnov Graph

Applying the Kolmogrov-Smirnov statistical graph in the same way as in the preceding base, the following data were obtained:

$$D = \text{maximum}[f(x) - s(x)]$$

$F(x)$ =the theoretical accumulated distribution. $s(x)$ =the observed accumulated distribution. D = maximum = 0.01687228 as indicated in the table. $N = 786$ main authors of the journal.

The calculation remains as follows.

$$\begin{aligned} K - S &= 1.63/\sqrt{n} \\ K - S &= 1.63/\sqrt{786} \\ K - S &= 1.63/28.03 = 0.0581519 \\ 0.581519 &> 0.01687228 \end{aligned}$$

A graphic representation in a logarithmic scale for both axes, gives as a result the Figure 2. The correlation between the curve representing the observed authors and the one representing the calculated authors does fulfill the characteristics of a typical inverse square curve of the data it represents. In it, the results between the calculated and the observed sample can be seen showing that their behavior is statistically regular that is say, the model is observed. Therefore, the data in this sample do adjust to Alfred Lotka's postulates on the productivity of authors.

4. Conclusions

The behavior of the productivity of authors of the journal, based on the PERIODICA data base does not fulfill the Lotka postulated.

On the other hand the database carried out in the library does behaves according to that postulated by Lotka, that is to say, 60% of the authors of the journal appear with just one contribution, at the same time as a nucleus represented by 2.16% of the total of the authors produced 14.1% of the total of articles published by the journal.

The interesting of this type of works is that it shows us the most productive authors in a specific journal by identifying 60% of them. This, together with other bibliometric studies will give us the most thorough behavior of the journal.

Acknowledgements. Thanks to Juana Orta and Laura Parrao for all their help.

References

Lotka, A.J. (1926), "Statistics. The Frequency Distribution of Scientific Productivity", Journal of the Washington Academy of Sciences, Vol. 16(12), 317-323.