Science Communication: Research Productivity and Data Analytics using Open Source Software

Vijay Kumar Verma

Central Library

IIT Delhi

Week: 05

Lecture 18: Lotka's Law

Dear Learners, Welcome to our NPTEL course on Science Communication, Research Productivity and Data Analytics using Open Source Software.

Now we have covered science, science communication, its importance, why we need different metrics, and bibliometrics and its applications. Now we are moving to Scientometrics. Scientometrics is basically a subset of bibliometrics. Scientometrics is a branch of bibliometrics that focuses especially on quantitative analysis of science and scientific research. As Pritchard said in bibliometrics, it is the metrics of books and other forms of scientific communication.

Here in Scientometrics basically it's focused on quantitative analysis of science and scientific research. Hence when we are discussing Scientometrics, we are basically dealing with science and scientific research. Scientometrics involved different bibliometric methods to study the various aspects of scientific interaction to various aspects of scientific literature, including its publications, citations, authors, journals and institutions. That means when we are discussing Scientometrics, Scientometrics is basically applying bibliometric methods in a study of the various domains of scientific literature. And those domains may include the different science publications or citations or the author profiles, the journal profiles or the institution profiles.

The goal of Scientometrics is to measure and evaluate the productivity, impact and dynamics of scientific research. Why do we need Scientometrics? Scientometrics is needed to evaluate the impact of science. It's needed to measure the dynamics of science. It's needed to measure the productivity of science. Hence when we talk about Scientometrics, Scientometrics is basically measuring the science productivity, it measures the impact and it measures the dynamics of science.

Scientometrics provide quantitative insight into the structure, evolution and impact of scientific research. It provides how science has evolved. when you say that Scientometrics is a measure of science, how science is evolved, what is the structure of science and what's the impact of scientific research on society. These are basically the areas which we cover when we use in Scientometrics. Scientometrics is very much required to analyze the research productivity quantitatively, and is very much required for evaluating and influencing scientific work. What scientific work is influencing our society, what's the influence of that work on other domains of science. This is the goal of Scientometrics.

That means when we are dealing with Scientometrics we are trying to find out what's the influence of research A on research B and vice versa. It supports evidence based decision making in academia, funding agencies and government institutions. As I discussed in bibliometrics, bibliometrics is very much essential for funders, is very much essential for the policymakers, for the planners. Similarly Scientometrics is very much required for evidence based decision making by the government, by the funding agencies, by the research organizations.

I am going to cover Lotka's inverse law of scientific productivity. When we are discussing scientific laws there are three prominent laws of Scientometrics. The first is Lotka inverse square law of productivity, second is Bradford's law of scattering of scientific papers, and the third one is Zipf’s law of word occurrence.

Let us discuss Lotka's inverse square law of scientific productivity. Lotka’s inverse square law of scientific productivity is named after Alfred J. Lotka. And this is perhaps the first Scientometrics or informetrics law. Lotka proposed the inverse square law in 1926 which correlates contributors of scientific papers with their contribution. Hence the basic concept of Lotka’s law which was produced by Lotka in 1926 is the number of contributions made by contributors. Thatș means the number of contributions made by an author.

Lotka observed that the number of authors producing a certain number of publications follows a mathematical pattern. While proposing this law Lotka came to know that the number of contributions made by a particular author has some mathematical pattern. And what he did for obtaining the data Lotka used he counted the number of personal names in the 1907 to 1916 Decennial Index of the Chemical Abstracts. What Lotka did? Lotka took the decennial that means ten year index of the Chemical Abstracts of 1907 to 1916. He took the alphabets A and B of that particular index, and he counted the number of publications. He listed the number of publications against the alphabets A and B 1, 2, 3, 4 like that.

He treated letter A and letter B both separately and in aggregate. And while formulating this law Lotka omitted the name of firms. That means he took the Chemical Abstracts of ten years from 1907 to 1916, took the alphabets A and B and counted down the listed publications under the alphabets A and B. This was one source. Another source he took the index in was Felix Auerbach of Physics. And the basic reason for taking this what Lotka explained is that it covers the entire range of history of physics through 1900. And it is a quality source which only takes into account not only the volume of production, but also quality. Because, it listed only the outstanding contribution in physics. That means while in the first source the Chemical Abstracts he took just a ten year index and took the alphabets A and B, but in the second source that is Auerbach Lotka took the complete source starting from the 1900. The reason behind taking this is. Because, it was published as an authentic source for the history of physics and listed only the quality contribution.

And while making this count while counting Lotka only credited the senior authors in joint publications. If a publication having two or more than two authors Lotka only took the senior author while using his law while counting his number of publications against the author. On a logarithmic scale he plotted the number of authors against their contribution. Because he has taken the author and the contribution he already counted the author and contribution using Chemical Abstracts and Auerbach physics. And now he plotted the number of authors against their contribution. He found that in each case the points are rather closely scattered about, essentially a straight line having a slope of 2 to 1. He drew on a logarithmic scale. He drew a straight line. He drew a straight line like that in logarithmic scale and he found that the point is very close to this straight line which has the slope of 2 to 1.

The approach of this ratio is particularly close in case of data taken from the Auerbach Table. The Auerbach Table is quite close to the straight line as compared to Chemical Abstracts data. Now the general formula Lotka found that there is a relationship exists between the frequency y of a person making x contribution. There is a relationship between the person and the contribution, and that relationship is $x^n y$ is equal to constant. Let us see this constant is C. That means when Lotka found out analyzing the result found from the Chemical Abstracts and Auerbach physics he came to know that there is a relationship exists between frequency y of a person, making x contribution and the formula for that relationship is $x^n y$ is equal to constant.

On the basis of this data Lotka derived what is a law and this is termed as the inverse square law of scientific productivity. Inverse square law correlates contributors of scientific papers with their contribution. It says what is the relationship between the number of papers produced by a contributor and that relationship is governed by $x^n y$ is equal to constant. For the special case n is equal to 2, that is the inverse square law of scientific productivity. the value we can find the value of constant as you see. We have already seen that here $x^n y = C$, $x^n y$ is equal to C.

If you are taking this one as y1, then y1 is equal to $C/1^2$. Similarly y2 is equal to $C/2^2$, y3 is equal to $C/3^2$ and So on. Now if you are adding all this that means, a $\Sigma 1$ to ∞y is equal to you are taking C outside its $1/1^2 + 1/2^2 + 1/3^2 + - - -$ just like that. Now this one is equal to C and this is $\Sigma 1/1^2 + 1/2^2 + 1/3^2$ is $\Sigma 1$ to ∞ $1/x^2$. That is equal to C $\pi^2/6$. And π^2C, $\pi^2/6$, and for C it would be for C it would be $6/\pi^2$. $\Sigma 1 = \infty y$. This here though y is frequency and the summation of this frequency gives unity hence then we get $C = 6/\pi^2$. We know the value of $\pi = 3.14$. If we put the value of π here. Then we get the value of C as 0.60 approximately. So, that means, we are getting the constant value of constant here that is 0.60. So, that is approximately 60 percent. That means what Lotka proposed here that the proportion of all contributors who contribute a single item should be just over 60 percent. That's what Lotka wants to say. Lotka wants to say that if we have a group of authors then in that group 60 percent of the authors will contribute a single contribution. 60 percent of authors will have only one article in their account that Lotka proposed here.

In the case he examined for the Chemical Abstracts and the Auerbach physics the actual proportion of this class was 59.2 percent in Auerbach data and 57.7 percent in Chemical Abstracts data under initial letter A and 57.9 percent in initial letter B and 57.9 percent under A and B jointly. That means what we are observing here is approximately 60 percent what we have found here the value of constant. That means approximately 60 percent of authors will have a single contribution. Now he found that if you take the total number of authors who have published only one article then the number of authors who have published k articles would approximately be $1/k^2$ times the number of times the number of authors who have published only one article.

What Lotka observed? Lotka observed that if you take the total number of authors who have published only one article then the number of authors who have published k articles would be $1/k^2$ times the number of authors who have published one article. For example let us see if an author published two articles. then the number of authors published two articles would be $1/2^2$ times. That means 1/4 times.

The number of authors who publish two articles would be 1/4 times the number of authors who published one article. This is basically what Lotka proposed. In simpler terms we can say that the number of authors who have published two articles is approximately 1/4th of the number of authors published in one article. The number of authors published three articles would be 1/9th, $1/3^2$, that is one ninth of the authors who have published one article. The number of authors published four articles would be 1/16 and So on. Hence this number of authors making n contribution would be $1/n^2$ times the number of authors who have published one article.

It is very clear that the number of authors publishing an article would be $1/n^2$ times the number of authors published in one article. For example, as I said in my previous slide, the number of authors publishing two articles would be 1/4 the number of authors publishing one article, three articles would be 1/9th and So on. The proportion of all contributors who make a single article is about 60% and this law is known as the Lotka inverse law of scientific productivity. That means when we talk about the Lotka inverse law of scientific productivity we see that most of the authors published one article in their lifetime and very few authors contribute the most number of articles. That means very few authors are prominent in that particular field of study who are contributing to the development of that field.

This is Lotka's inverse law of scientific productivity. This is the graph which Lotka originally published when he published his article in the Journal of Washington Academy of Sciences in June 1926. You see here this is the fully drawn line and this is the dashed line. This fully drawn line basically indicates the point given by inverse square law. Now as I mentioned that Lotka evaluated two sources one is Auerbach and the other was Chemical Abstracts for letter A and B.

In Auerbach what you see is basically this cross. This cross is for Auerbach physics and this particular circle is for Chemical Abstracts. What you are finding here means that in the case of Auerbach the cross is very much near to this fully drawn line. While this one in the case of Chemical Abstracts these circles are very near to the dashed line. That means when we are drawing a graph on a logarithmic table using the data derived by Lotka, then we come to know they are quite close to the exponential line having slope of 2/1.

This is basically Lotka's inverse law of scientific productivity which is very much essential for growth and development of science, and which predicts how science is moving. Who is the most prolific author, who is contributing in science? Now if you see the various applications of Lotka's Law. Lotka's Law can be used for research evaluation. Lotka's Law can be used for evaluating research productivity and identifying the highly productive authors. Because, we have already seen that some of the authors are contributing the maximum number of articles in a particular domain of science. With the help of Lotka's Law we come to know who the authors are and how they are contributing to science productivity.

Lotka's Law can be used for authorship analysis. We can analyze, we can understand the authorship pattern in that particular branch of science using Lotka's Law in scientific publications. How that particular scientific publication is happening, who are the most eminent authors, who are the scientists, who are the researchers, who contributed for the development of the domain can be found out using Lotka's Law. Lotka's Law can be used for resource allocation. It helps in taking the logical decision for the allocation of the resources. Because, based on the Lotka's Law we can find out the author's productivity and by that way it is possible to take the decision on allocation of resources.

Lotka's Law gives the research trends and the research pattern. With the help of Lotka's Law we can find out the most prolific author, the most contributing author and by that way we can find out in which particular domain that particular author is working. By that way we can find the research trend and pattern in that particular domain. Lotka's Law can be used for career development. It is used for mentoring. Because, Lotka's Law finds out the productivity of the author. And based on that we can find who are the most eminent authors to whom we need to pay attention. It provides the insight of the career development and the mentorship program. Hence Lotka's Law can be used for career development and mentorship.

Lotka's Law can also be used for the publication strategy. Lotka's Law can guide publication strategy for the researcher. They understand the distribution of the author's productivity and can set realistic goals. You see, every discipline is unique. it is not necessary that discipline A is contributing more. So, that discipline B in science. They have their unique characteristics, they have their unique features.

With the help of Lotka's Law, a budding author can find out, can derive their, derive his or her strategy and work on that. He or she can fix a realistic goal based on the development of that particular domain which can be achieved by him or her in the next 15 or 20 years of time. Lotka's Law is used for the bibliometric studies like we have already covered, this is the bibliometric study-like authorship pattern with the help of Lotka's Law. We can come to know about the authorship pattern, we can come to know about the production which is the most productive field. we can come to know about the productivity distribution. We come to know about the collaborative network. All these are possible with the help of Lotka's Law and Lotka's Law can be used in this category. Lotka's Law is very much essential for the management of libraries and information centers.

Library and information management. Lotka's Law helps in the management of library and information services. This can be applied to analyze the productivity of authors in academic libraries and information management and this can be useful for the librarians and information specialists in research allocation and the collection management you see. We know that with the help of Lotka's Law, we can easily find out the author's profile. We can easily find out how the authors are performing in particular domains of research, in particular domains of science. Hence, the library information science professionals use Lotka's Law in their collection development, use Lotka's Law in the resource allocation, use Lotka's Law in guiding the institution to adopt different development fields where the further study can be done. Lotka's Law can also be used for making a good information storage and retrieval system. Because, a good information storage and retrieval system is very much required to retrieve the stored information precisely and is very much required for getting the required information at the right time in the right way. Lotka's Law can be used for the search engine by understanding the distribution of authorship and improving the ranking of search results.

You see a good storage and information system that gives you the maximum precise result with minimum time. And for that, Lotka's Law can be used for improving the search result. Lotka's Law can be used for ranking the search result by that way it would be helpful in making or in designing a good information storage and retrieval system. There are some limitations of Lotka's Law also. Because, you have already seen that Lotka consulted two sources and these two sources having one was Chemical Abstracts, other was qualitative based physics literature. In first there was a 10 year index covering only later A and B, in the second he covered the history of physics without any restriction on the later. Hence for Lotka's Law a single source in a limited time frame may not produce favorable results. It's important. Because, I have already told you in the beginning of my lecture that science is ever growing, science is dynamic, science is expanding.

Hence if we are taking any source for a limited period of time for one year, two years then the result which we obtain won't favor Lotka's Law. If you are taking a single publication for a given journal then the Lotka's Law it might be possible that Lotka's Law won't be true or correct it won't follow the Lotka's Law. Data compiled from a comprehensive source in order to capture true representation of author publications using either quality or quantity as a selection criteria. Lotka's Law is valid when we are considering various comprehensive data for a longer period of time. then the Lotka's Law can be true or it can give true results.

Another limitation of Lotka's Law is the assumption of stability. Lotka's Law assumes a stable system over time. You see when Lotka proposed his law in 1926, basically he took Chemical Abstracts 10 year data with another physics source having the data from 1900 onwards. But science is dynamic, hence he assumed the stability of science basically. Lotka assumes a stable system over time. But that system is not stable. If you are taking the data of 1907 to 1916 from the Chemical Abstracts today it is not necessary that whatever produced 100 years back would be true this time also. That means we can't assume stability in the case of science. Because, it is a developing area.

Hence without taking a stability without assuming a stability we can go for the Lotka's Law. But Lotka's Law when he proposed, he assumed stability. Because he took the data for a limited period. Lotka's Law is basically a linear model based on the power law relationship assuming a linear model for the distribution of scientific productivity. This may over simplify the complex nature of scientific productivity. You see, Lotka only counted the number of publications against the contributor. How many articles are published by a particular contributor. But science, the nature of science, is something complex. It is not only limited to the number of publications by a particular author A or

B. Hence whatever the model proposed by Lotka is something which is very much linear in nature and this was one of the limitations. He made it. So, he oversimplified that he didn't consider the complex nature of science while proposing this model.

This was one of the limitations of Lotka's Law. There is another limitation of Lotka's Law and that is the dependence on data quality. The law highly depends on the accuracy and the quality of data. So, obtained. Incomplete or inaccurate data gives the false result of Lotka's Law. Hence this is quite dependent on data quality and basically it's the static analysis. Because, the law is more suited for static analysis instead of dynamic analysis. Because, it doesn't capture all the changing modes of science communication. Thanks for your patience. Thank you.