Citation Count Modelling: The Role of Statistical Distributions

Dr Gautam Mukhopadhyay

Chandrapur College P.O. & Vill. Chandrapur, Dist. Purba Bardhaman, Pin. 713145, West Bengal, India.

Abstract

The statistical distributions develop more accurate models of research impact, such as citation counts and h-index. The statistical models play a crucial role in scientometric analysis and citation mapping. The distributions help identify typical citation pattern, outliers, anomalies.These enable comparison across fields, authors, and papers. The probability distributions aid in predicting future citation counts and emerging research trends as well as areas of research. These indicate papers with immoderately high impact and metrics of research quality, such as citation-based metrics. By applying statistical distributions, scientometric analysis can improve accuracy, reveal underlying citation behaviour and how citations accumulate over time, trends, predict research phenomena. Advance theory development helps contribute to progress of theories of research behaviour and impact. This paper intends to gain insights into research impact, quality and citation flow that support the advancement of scientific research. Main objectives of this study are to understand and describe the patterns of citation counts and to predict the number of citations a paper or author will receive.

Keyword: Citation counts, Statistical distributions, Skewness, Variability, Mapping

INTRODUCTION

Statistical distributions play a crucial role in scientometric analysis and models. Statistical distributions help identify typical citation pattern, outliers, anomalies. The distributions develop more accurate models of research impact, such as citation counts and h-index. These enable comparison across fields, authors, and papers. The probability distributions aid in predicting future citation counts and emerging research trends as well as areas of research. These indicate papers with immoderately high impact and metrics of research quality, such as citation-based metrics. Analyzing the collaboration networks distributions model coauthorship and collaboration patterns. By applying statistical distributions, scientometric analysis can improve accuracy; reveal underlying citation behaviour and how citations accumulate over time, trends, predict research phenomena. Advance theory development helps contribute to progress of theories of research behaviour and impact. Detecting citation biases, numerical distributions show biases (e.g. geographic or linguistic) in citation behaviour. This paper intends to gain insights into research impact, quality and citation flow that support the advancement of scientific research.

Objectives

Main objectives of this study are:

1. To understand and describe the patterns of citation counts;

2. To predict the number of citations a paper or author will receive;
3. To detect papers with immoderately high impact;
4. To compare citation counts across papers;
5. To develop metrics model of research quality based on citation distributions;
6. To understand citation flow i.e. how citations accumulate over time;
7. To develop more accurate and robust metrics of research impact and to measure the influence and dissemination of research; and
8. To identify emerging topics and trends of research.

Review of literature

Some studies have been used statistical methods to analyze citation patterns and detect anomalies. Radicchi et al. (2008) [1] studied in their paper "Universality of citation distributions" that citation counts follow a universal log-normal distribution across disciplines. Another study conducted by Redner (1998) [2] was on analysis of citation patterns in Physical Review journals, revealing a power-law distribution. A paper by Seglen (1992) [3] discussed the skewed distribution of citations and its implications for research evaluation. A book by Egghe (2005) [4] provides a comprehensive overview of scientometrics, including the application of statistical distributions.

Role of statistical distributions in citation count

Statistical distributions play a significant role in understanding and analyzing citation counts, which are key research metrics in evaluating research impact, academic performance, and scientific advancement. Usage of statistical distributions in citation counts is crucial in scientometrics and research evaluation. Distributions like the Poisson, Negative Binomial, or Power-law capture the dispersion of citations, which inform about the uneven distribution of citations across papers and researchers. Statistical distributions enable the detection of outliers, such as highly cited papers or authors, which can provide insights into exceptional research impact. Citation distributions indicate how papers are being cited, which papers are highly cited, and how citation patterns change over time. Statistical distributions are the important tools and techniques that develop models concerning the variability in citation counts, allowing researchers to identify patterns and trends. Use of statistical models in citation distributions helps to analyse research impact, comparing citation patterns across fields, R&D institutions, and researchers. Citation distributions help develop more accurate research evaluation metrics, such as citation-based indices (e.g., h-index) and percentile-based rankings. Understanding citation patterns reveals the potential research collaborators, patterns in coauthorship networks and invisible colleges of research communities. Analyzing citation distributions can depict unique citation patterns, indicating potential manipulation or gaming of the citation system. Usage of statistical distributions can suggest research policy decisions, such as funding allocations and evaluation measures.

Statistical distributions used in citation counts

Here are some statistical distributions used in citation counts:

1. Binomial and Negative Binomial Distribution

Binomial Distribution is the discrete probability distribution of the numbers of successes in a sequence of 'n' independent trials, each of which yields success with probability 'p'. Citation counts can be calculated using the binomial distribution.

The probability mass function (PMF) of the Binomial Distribution is:
$P(X=k)= nCk * p^{k} * (1-p)^{(n-k)}$
Where, P(X=k) is the probability of k citations; n (number of trials) is the total number of publications; p (probability of success) is the probability of receiving a citation; k (number of success) is the actual citation count; nCk is the number of combinations of 'n' items taken 'k' at a time (binomial coefficient). Hence, nCk = n! / k!(n – k)!
Example:
Suppose, a scientist published 10 scientific research papers (n = 10); each paper has a 20% chance (the probability of receiving a citation i.e. p = 0.2) and the actual citation count (k = 3). Then, the probability of 3 citations is
$P(X=3) = 10C3 * (0.2)^{3} * (1-0.2)^{(10-3)} \approx 0.2013$
The value of P(X=3) indicates that the probability of the researcher receiving exactly 3 citations is approximately 20.13%.
Consider two scientists, S1 and S2, with 5 publications (n) each. Scientist S1 has a citation probability (p) of 0.2, while another scientist S2 has a citation probability (p) of 0.1.
$P(X=2) = 5C2 * (0.2)^{2} * (1-0.2)^{(5-2)} \approx 0.2048$
$P(X=2) = 5C2 * (0.1)^{2} * (1-0.1)^{(5-2)} \approx 0.0729$
Scientist S1 has a higher probability (20.48%) of receiving exactly 2 citations compared to Scientist (7.29%).
Negative Binomial Distribution is a discrete probability distribution of the number of successes in sequence of independent and identically distributed Bernoulli trials before a specified (non-random) number of failures (denoted r) occur. This distribution is normally used to model citation counts in academic research, as it effectively indicates the skewness (most papers receive few citations, while a few papers receive many) and over dispersion (variance greater than mean).
The Probability Mass Function (PMF) of the Negative Binomial Distribution is:
$P(k) = (k + r - 1)! / (k! * (r - 1)!) * p^{r} * (1 – p)^{k}$
Where, k is the number of citations; r is the number of failures until the experiment stops (e.g. maximum number of citations) and p is the probability of success (e.g. probability of receiving a citation).
Example:
Suppose, number of citations (k) = 5, the number of failures (r) = 2 and the probability of receiving a citation (p) = 0.2. Now, it can be possible to model the citation counts of a set of research papers:
$P(5) = (5+2-1)! / (5! * (2-1)!) * 0.2^{2} * (1-0.2)^{5}$
$= (6! / (120 * 1)) * 0.04 * 0.32768$
$= 6 * 0.04* 0.32768 \approx 0.0786$
Therefore, the probability of receiving citations of the given set of research papers is 7.86%.

2. Gaussian Distribution (Normal Distribution) [5]

Any of a family of continuous probability distributions such that the probability densities function is the Gaussian function. It is also known as the normal distribution (or bell curve) which can be used to model citation counts, assuming mean (μ) i.e. the average citation count and standard deviation (σ) i.e. the spread or dispersion of citation counts.
The Probability Density Function (PDF) of the Normal Distribution is:
$f(x) = (1/\sqrt{}(2\pi\sigma^{2})) * e\hat{}(-((x-\mu)^{2})/(2\sigma^{2}))$
Where, x is the citation counts.

Example:

Suppose, one wants to find out probability of research paper receiving exactly 15 citations. The mean citation count (μ) is 10 and the dispersion of citation count i.e. standard deviation (σ) is 5. Now, using the Gaussian function one can calculate the probability of research papers receiving 15 citations:

$f(15) = (1/\sqrt{}(2\pi(5)^2)) * e\text{^}(-((15-10)^2)/(2(5)^2))$

$=(1/\sqrt{}(50\pi)) * e\text{^}(-((15-10)^2)/50)$

Approximately equal value is

$\approx 0.0797 * e\text{^}(-((5)^2)/50) \approx 0.0483$

Therefore, the probability of receiving exactly 15 citations of a research papers is 4.83%.

3. Poisson Distribution [6]

Any of a class of discrete probability distributions that express the probability of a given number of event occurring in a fixed time interval, where the events occur independently and at a constant average rate; describable as a limit case of either binomial or negative binomial distributions. Citation counts can be found out using the Poisson distribution, which is commonly used in scientometrics to analyze the distribution of citations scientific research publications.

The Probability Mass Function (PMF) of the Poisson distribution is:

$P(k) = (e^{(-\lambda)} * (\lambda^k) / k!$

Where, k is the number of events (citations); λ (lambda) is the average rate of events (citations) and e is the base of the natural logarithm (almost value 2.71828).

Example:

Suppose, a researcher wants to calculate the probability of a research paper receiving a specific number of citations.

Say, the average number of citations per paper is 5 ($\lambda = 5$).

$P(0) = (e^{(-5)} * (5^0) / 0! \approx 0.0067$ (0.67% probability of 0 citation)

$P(1) = (e^{(-5)} * (5^1) / 1! \approx 0.0337$ (3.37% probability of 1 citation)

Citation Count (k)	Probability P(k; λ = 5)
0	0.0067
1	0.0337
2	0.0842
3	0.1404
4	0.1755
5	0.1755
6	0.1462
7	0.1044
8	0.0653
9	0.0634
10	0.0181

Table 1: Data showing specific citation counts using Poisson Distribution

Table 1 shows the probability of a paper receiving a specific number of citations using Poisson Distribution. For example, the probability of a paper receiving exactly 10 citations is 0.0181 (or 1.81%).

4. Galton or Log-normal Distribution [7]

The Galton Distribution, popularly known as Log-normal Distribution, is widely used in numerous fields including scientometrics, to model citation counts. It assumes that the logarithm of the citation count follows a normal distribution. A logarithmic function that has a normal distribution. Log-normal Distributions are skewed to the right, reflecting the majority of papers with few citations and a long tail of highly cited papers. The distribution has a heavy tail that means extreme values (very highly cited papers) are more likely than in a normal distribution. It is continuous probability distribution of which parameters are mean (μ) and standard deviation (σ).

The Probability Density Function of Log-normal Distribution formula is:

$f(x; \mu, \sigma) = (1 / (x*\sigma\sqrt{(2*\pi)})) * e(-((In(x) – \mu)^2) / (2*\sigma^2))$

Where, x is the citation count, μ is the mean of the logarithm of the citation counts, σ is the standard deviation of the logarithm citation counts, In(x) is the natural logarithm of x, and π is the mathematical constant (approximately 3.14159).

Example:

Suppose, a researcher wants to model the citation counts of research papers in a particular field, with $\mu = 2$ and $\sigma = 1$.

Using the formula,

$f(x) = (1 / (x*\sigma\sqrt{(2*\pi)})) * e(-((In(x) – \mu)^2) / (2*\sigma^2))$

$f(10) = (1 / (10 * 1 * \sqrt{(2 * \pi)})) * exp(-((ln(10) - 2)^2) / (2 * 1^2)$

≈ 0.053

Citation Count (x)	Probability (f(x))
10	0.053
50	0.133
100	0.053
500	0.004

Table 2: Data showing specific citation counts using log-normal distribution

Table 2 shows the probability of a paper receiving a specific number of citations using Galton or Log-normal Distribution. For example, the probability of a paper receiving exactly 10 citations is 0.053.

5. Pareto Distribution [8]

Pareto Distribution is a probability distribution such that for a random variable X with that distribution holds that the probability that X is greater than some number x is given by $Pr(X>x) = (x/x_m)^{-k}$ for all $x>=x_m$, where x_m is the (necessarily positive) minimum possible value of X, and k is a positive parameter. In the context of citation counts, the Pareto distribution is often used to describe the distribution of citations among academic papers, where a small number of papers receive a large number of citations.

The Probability Density Function of the Pareto Distribution is given by:

$f(x; x_m, \alpha) = (\alpha * x_m) / x^{(\alpha+1)}$

Where, x is the number of citations or citation count, x_m is the scale parameter which indicates the minimum citation count, α is the shape parameter or Pareto index that shows the proportion of papers with a small number of citations and f(x) is the probability density function.

Example:

Suppose, an author wants to model the citation counts of academic papers using the Pareto Distribution with

$\alpha = 2$ and $x_m = 1$.

Now, the probability of a paper receiving 10 citations is:

$f(10) = (2 * 1) / 10^{(2+1)} = 0.002$

This means that approximately 0.2% of papers are expected to receive 10 citations.

6. Weibull Distribution [9]

This distribution is any of continuous probability distributions applied to measure the amount of time for which something can be used until it ceases to be operable. The Weibull Distribution is a probability distribution used in various fields, including social sciences. In the measure of citation count, the Weibull Distribution can be applied to model the distribution of citations received by research articles. Citation distributions are often skewed, with many papers receiving few citations and a few papers receiving many citations. Actually, Weibull Distribution analyzes the characteristics of the citation distribution (e.g., skewness, heavy-tailedness) as well as compares the citation patterns across different research domains.

The two-parameter Weibull Distribution is given below:

$f(x; \alpha, \beta) = (\beta/\alpha) * (x/\alpha)^{(\beta - 1)} \exp(-(x/\alpha)^{\beta}) \qquad \text{if } x \geq 0 ; \alpha , \beta > 0$

Where, x is the citation count, α is the scale parameter and β is the shape parameter which controls the shape of the distribution.

Example:

Suppose, a researcher wants to model the citation count distribution of his research papers in a particular field. Say, the parameters α and β using maximum likelihood estimation.

Let's say the estimated parameters are:

$\alpha = 10$ (scale parameter)

$\beta = 1.5$(shape parameter)

Using the Weibull Distribution, we can calculate the probability of a paper receiving exactly x citations.

Now, the probability of a paper receiving 20 citations is:

$f(x; \alpha, \beta) = (\beta/\alpha) *(x/\alpha)^{(\beta - 1)} \exp(-(x/\alpha)^{\beta})$

$f(20; 10, 1.5) = (1.5/10 * (20/10)^{(1.5-1)} \exp(-(20/10)^{1.5}) \approx 0.0025$

6. Zero-Inflated Poisson (ZIP) Distribution [10]

The Zero-Inflated Poisson Distribution is a statistical model that accounts for excess zeros in citation count data. Citation count shows excess zeros i.e. many papers receive no citations and skewness that is a few papers receive a large number of citations.

Let Y be the citation count. The ZIP distribution is defined as:

$f(Y = y; \pi, \lambda) =\{ \pi + (1 - \pi) * e^{(-\lambda)} \qquad \text{if } y = 0$

$((1 - \pi) * (e^{(-\lambda)} * \lambda^{\wedge y}) / y! \qquad \text{if } y > 0$

Where, π (pi) is the probability of zero-inflation (probability of excess zeros or zero citations), λ (lambda) is the mean citation rate or mean of the Poisson distribution (average citations). Actually,

π represents the probability of a paper receiving no citations (zero-inflation) and λ represents the average number of citations for papers that do receive citations.

Example:

Suppose, a researcher has citation data for 100 papers, with 60 papers receiving no citations. Using ZIP model,

$\pi = 0.6$ (60% chance of zero citations)

$\lambda = 2.5$ (average 2.5 citations for papers with citations)

Using the formula,

$f(Y = 0; \pi = 0.6, \lambda = 2.5) = 0.6 + (1 - 0.6) * e^{(-2.5)} \approx 0.745$

$f(Y = 1; \pi = 0.6, \lambda = 2.5) = (1 - 0.6) * (e^{(-2.5)} * 2.5^{1}) / 1! \approx 0.186$

$f(Y = 2; \pi = 0.6, \lambda = 2.5) = (1 - 0.6) * (e^{(-2.5)} * 2.5^{2}) / 2! \approx 0.058$

7. Power Law Distribution [11]

The Power Law Distribution is a statistical model where the frequency or size of events follows a power-law relationship. This model is used to describe the distribution of citation counts. In the analysis of citation counts, it means that a small number of papers receive a disproportionately large number of citations, while most papers receive very few citations.

Discrete Power Law Distribution formula:

$P(c) = K * c^{(-\alpha)}$

Where, P(c) is the probability of a paper receiving c citations; K is the normalization constant; c is the variable i.e. the number of citations and α is the power law exponent (typically positive).

Example:

Suppose, a research worker wants to model the citation count of a set of research papers in a specific subject area. It is seen that the citation counts follow a power law distribution with $\alpha = 2.5$ and K 0.087. Table 3 shows the probability of a paper receiving 1, 2, 3, 4 and 5 citations respectively applying the normalization constant formula of Power Law Distribution.

Using the formula,

$P(c) = K * c^{(-\alpha)}$

$P(1) = 0.087 * 1^{(-2.5)} = 0.087$

Citation Count (c)	Probability P(c)
1	0.087
2	0.035
3	0.017
4	0.008
5	0.004

Table 3: Data showing specific citation counts using Power Law Distribution

8. Exponential Distribution

The Exponential Distribution calculates the time between citations, useful for analyzing citation rates over time. Consider a paper receiving citations at an average rate of 1 citation per month. Using the Exponential Distribution, we can calculate the probability of waiting 2, 5, or 10 months for the next citation.

9. Gamma Distribution

The Gamma Distribution models the distribution of citations, accounting for variability in citation rates. Suppose, a few papers receive citations at an average rate of 2 per year, with a shape parameter (α) of 1.5 and rate parameter (β) of 0.5. We can calculate the probability of receiving a certain number of citations within a specified time frame.

Citation metrics

Statistical models play a pivotal role in measuring citations metrics. Citation distributions are often skewed, making traditional metrics (e.g., mean, median) inadequate. Statistical models reveal underlying structures, such as citation rates, influential papers, and collaboration networks. Citation models enable fair comparisons between journals with different citation patterns. These models account for differences in citation rates across fields. Citation metrics are widely used quantitative measurements in evaluating the impact and relevance of research publications. Regression analysis examines relationships between citation counts and variables (journal, author, year). Network analysis studies citation networks using graph theory. Cluster analysis identifies groups of related papers or authors. Time-series analysis examines citation trends over time. Here are some common citation metrics using statistical distributions, along with examples:

i) H-Index: This is an athor-level metrics, measuring productivity and citation impact such as h-index 10 indicates papers with more than or equal to 10 citations each.

ii) G-Index: This is a generalization of h-index, accounting for citation distribution. The g-index measures the number of papers (g) that have at least g citations; for example, g-index 15 indicates 15 papers have at least 15 citations each. A higher g-index shows a stronger citation profile.

iii) Egghe Index: Similar to h-index, but weighted by citation counts.

iv) Citation rate or Citations per paper (CPP) : This is an average citations per paper per year. Citation rate 5 means 5 citations per paper per year.

v) Impact Factor: Average citations per paper in a journal over 2 years. As for example, a journal has an impact factor of 3.5.

vi) CiteScore: Similar to impact factor, but includes more citation types.

vii) SCImago Journal Rank (SJR): SJR is a prestige metric that evaluates scholarly journals based on their citation counts and the prestige of the citing journals. The SJR indicator is calculated by assigning different values to citations depending on the prestige of the journals where they come from. This signifies that citations from highly prestigious journals are given more weight than those from less prestigious ones.

viii) Eigenfactor: It is a prestige metric that evaluates scholarly journals based on their citation patterns, similar to SJR. It considers citations from the past five years. Eigenfactor analyses citation relationships between journals and scores reflect a journal's influence within its field. It is a network-based model. It uses data from the Web of Science database.

ix) Citation Count: Total number of citation received such as a paper with 30 citations.

Power Law, Log-normal models favour top-performing authors. Pareto, Eigenfactor models emphasize influential journals. Field normalization i.e. Gamma and Negative Binomial distributions account for field specific citation rates. Collaboration network analysis i.e. network-based models reveal complex relationships.

Comparative evaluation among the statistical distributions

Citation counts are an essential metric in bibliometrics and scientometrics. Different models of statistical distributions have been employed to measure and analyze them. Here's a comparative evaluation of some commonly used distributions. It is assumed that citations occur independently in Poisson Distribution and at a constant citation rate. Poisson Distribution is suitable for modeling low-to-moderate citation counts. It underestimates variability. Negative Binomial Distribution extends the Poisson model by allowing for over dispersion (variance greater than the mean). It is more complex and requires additional parameters. This is often used for analyzing citation counts with high variability. Zero-inflated Poisson distribution accounts for the presence of zero-citation papers and it is mainly used when modeling citation counts with a high proportion of zeros. Here it is assumed that two distinct processes for zero and non-zero citations. Log-normal distribution is suitable for measuring skewed citation count data. It can be sensitive to outliers, requires transformation. This is often used in bibliometric analysis to model citation mapping. Power Law Distribution is characterized by a long tail; this distribution is often used to model the distribution of most cited papers. It can be sensitive to parameter estimation and requires large datasets. Weibull Distribution is a flexible model that enables to measure various shapes including skewed and symmetric distributions. It can be complex, requires careful parameter estimation.

When selecting a distribution for modeling citation counts, consider the factors like skewness, variability, presence of zeros and outliers. Whether a researcher is interested in modeling the overall citation distribution or focusing on highly cited papers should be an important research question. While considering sample size, the larger datasets may require more complex distributions whereas the smaller datasets may be adequately modelled by simpler distributions. One should a distribution model that provides meaningful insights into the citation patterns. By careful evaluating the above factors and taking into consideration the advantages and disadvantages of each distribution, one research worker can select the most suitable distribution for his/her citation count data.

Findings

Some salient findings in statistical distribution models in citation count analysis are: Poisson Distribution is often used as standard model, but is likely to underestimate citation counts. Negative Binomial Distribution provides a more appropriate than Poisson for a large volume of citation datasets, representing over dispersion. Zero-inflated Poisson Distribution adequately models datasets with a high rate of zero-citation papers. Log-normal Distribution most often used to model citation counts, as it can show skewness and variability. Power Law Distribution used in citation datasets, particularly for extensively cited papers. Weibull Distribution aptly fits a flexible model for citation counts capable of capturing various shapes. Citation counts often show skewness and high variability making distributions like Log-normal, Negative Binomial and Power Law suitable. Sometimes institutional and disciplinary differences matter while modeling citation counts. Citation patterns can vary remarkably across institutions and disciplines, emphasizing the need for context-specific modeling.

Conclusion

While the Gaussian Distribution can provide a rough estimate of citation counts, it has limitations: citation counts are often skewed, with many papers receiving few citations and a few papers receiving many citations, the distribution may have heavy tails, which can affect the accuracy of Gaussian

Distribution. There are alternative distributions, such as i) Log-normal Distribution ii) Power Law Distribution and iii) Negative Binomial Distribution may better capture the characteristics of citation counts. The mean and variance of the Poisson distribution is equal, which is useful for modeling citation counts, as it implies that papers with more citations are more likely to receive additional citations. The Poisson Distribution is skewed to the right, reflecting the fact that most papers receive few citations, while a few papers receive many citations. The Poisson Distribution assumes independence between citations, which may not always hold true, as citations can be influenced by various factors, such as author reputation or journal prestige.

References

1. Radichhi, F et al. (2008). Universality of citation distributions: toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences, USA* 105(43): 17268-72.
2. Redner, S. (1998). Citation statistics from 110 years of physical review. *Physics Today,* 58(6): 49-54.
3. Seglen, P.O. (1992). The skewness of science. *Journal of the American Society for Information Science,* 43(9): 628-638.
4. Egghe, L. (2005). *Scientometrics: an infometric approach to its literature*. Springer.
5. Gauss, C.F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Sumtibus F. Perthes et I.H. Besser, 269P.
6. Poisson, S.D. (1837). *Recherches sur la probabilite des jugements en maiere criminelle et en matiere civile*. Paris: Bachelier, Imprimeur-Libraire, 415P.
7. Galton, F. (1879). The geometric mean, in vital and social statistics. *Proceedings of the Royal Society*, 29: 365–367.
8. Pareto, V. (1896). *Cours d'economie politique*. v.1; Lausanne: F. Rouge, 430P.
9. Weibull, W. (1951). A statistical distribution function of wide applicability. *The ASME Journal of Applied Mechanics*, 18(3), 293-297.
10. Huang, M.H. (2012). Exploring citation count distributions using zero-inflated models.
11. Newman, M.E.J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics,* 46(5): 323-351.
12. Sinatra, R., et al. (2016). Quantifying the evolution of individual scientific impact. *Science*, 354(6312): 523-527.10
13. Egghe, Leo and Rousseau, R. (1990). *Introduction to informetrics: quantitative methods in library, documentation and information science.* Elsevier Science Publishers, 450P.
14. Price, D.J. de S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science,* 27(5): 292-306.
15. Lotka, A.J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences,* 16(12): 317-323.
16. Zipf, G.K. (1949). Human Behavior and the Principle of LeastEffort. Cambridge: Addison-Wesley.
17. Burrell, Q.L. (2008). Extending Lotkaian informetrics. *Information Processing & Management,* 44(5): 1794-1807.