

# Revisiting Lotka's law: patterns of author productivity in sustainability science

Marco Schirone\* and A. I. M. Jakaria Rahman\*\*

\*marco.schirone@chalmers.se

ORCID: 0000-0002-4166-153X

Department of Communication and Learning in Science, Chalmers University of Technology, Sweden  
and the Swedish School of Library and Information Science, University of Borås, Sweden

\*\*jakaria.rahman@chalmers.se

ORCID: 0000-0001-7876-4631

Department of Communication and Learning in Science, Chalmers University of Technology, Sweden

## Abstract

On the centenary of Alfred J. Lotka's seminal paper on the frequency distribution of scientific productivity, we revisit the classical inverse-square law using a ten-year (2016 – 2025) Scopus dataset of sustainability science journals. The dataset comprises 143,183 publications, yielding 707,094 authorship records from 398,085 unique authors, forming a large-scale, co-authored corpus well suited to examining productivity patterns through the lens of Lotka's law. We analyze author productivity distributions for heavy-tailed data, including maximum-likelihood estimation, goodness-of-fit testing, and model comparison. The estimated scaling exponent is substantially higher than Lotka's classical value, and statistical tests reject a pure power-law specification in favor of a lognormal model. The findings reveal strong concentration of productivity among a small group of authors while showing departure from the classical inverse-square exponent in the upper tail. A better-fitting lognormal model does not invalidate Lotka's law; rather, it refines its application to modern, multi-author research corpora.

## 1. Background

Alfred J. Lotka's seminal paper on the frequency distribution of scientific productivity remains widely recognized as a classic contribution to bibliometrics and continues to receive substantial attention in contemporary bibliometric research (Li et al., 2024; Lotka, 1926). Lotka's law states that the number of authors publishing  $n$  papers is approximately proportional to  $1/n^2$  of those publishing a single paper, implying that a small number of authors account for a large share of publications. This 'square law' became one of the foundational regularities of bibliometrics, shaping how researchers conceptualize the concentration of scientific output. Several extensions of Lotka's original formulation have been proposed, including the shifted Lotka function, which allows the distribution to include sources with zero items (Egghe & Rousseau, 2011). Subsequent empirical research, however, has shown that this regularity may break down under certain conditions. Studies of scientific collaboration have demonstrated that datasets containing publications with very large numbers of coauthors can deviate substantially from the classical Lotka distribution (Kretschmer & Rousseau, 2001). Statistical methods based on maximum-likelihood estimation and goodness-of-fit (GOF) testing (Clauset et al., 2009) have demonstrated that many empirical distributions previously labelled as power laws are better described by alternative heavy-tailed models such as the lognormal. Marking the centenary of Lotka's seminal contribution, a rigorous re-examination of author productivity distributions using contemporary bibliometric data and modern statistical methods is therefore both timely and necessary.

Sustainability science, as a rapidly expanding and highly interdisciplinary field, provides a useful context for testing whether classical bibliometric regularities are held in emerging research domains. This field has grown rapidly over the past two decades, attracting researchers from diverse disciplinary backgrounds (Schirone, 2024). Whether author productivity in such a young, interdisciplinary field follows the same distributional patterns observed in more established disciplines remains an open question. On the centenary of Lotka's (1926) paper, we forwarded the following two research questions (RQ) to examine whether classical bibliometric regularities are held in contemporary sustainability science:

RQ1: Does the productivity distribution across sustainability science journals conform to a power-law model consistent with the classical inverse-square law?

RQ2: How concentrated is scientific output, and what does the degree of inequality imply for the structure of the field?

## 2. Data and methods

Publication data covering 2016-2025 were retrieved via the Scopus API for 29 journals in sustainability science. We found 143,183 publications (articles and reviews only), and 707,094 authored publications counts (mean productivity = 1.78). Using Scopus Author ID, 398,085 unique authors were found.

To address RQ1, which tests whether the productivity distribution follows Lotka's law, a discrete power-law model was fitted using maximum-likelihood estimation implemented in the '*powerLaw*' R package (Gillespie, 2015). The lower cutoff parameter was estimated by minimizing the Kolmogorov–Smirnov statistic between the empirical distribution and the fitted model, and model adequacy was evaluated using a parametric bootstrap GOF test (500 simulations). The power-law model was compared with a lognormal alternative using the Vuong likelihood ratio test to determine which model provides the better statistical description of the data (Vuong, 1989).

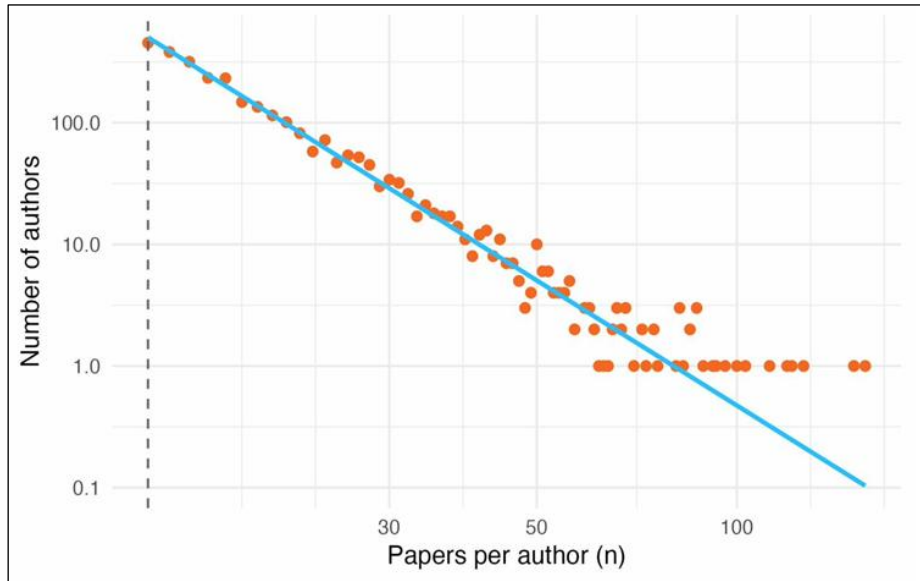
To address RQ2, which examines the concentration of scientific output, inequality was measured using the Gini coefficient, a standard summary indicator of inequality ranging from 0 (perfect equality) to 1 (maximum concentration) (Gini, 1997). The distribution of inequality was also visualized using the Lorenz curve, which plots the cumulative share of publications against the cumulative share of authors ordered from least to most productive.

## 3. Results and discussion

### 3.1. Power-law fits with Lotka's law

Figure 1 shows that the productivity distribution is strongly right skewed, with most authors publishing one or two papers and a long but sparse upper tail reaching a maximum of 156 publications. Fitting a discrete power-law model to the upper tail yields  $x_{min} = 13$ , the lower threshold above which the power-law behaviour is assumed to hold, and  $\alpha = 3.42$ , the scaling exponent that determines the rate at which the frequency of highly productive authors declines. These estimates indicate that potential scaling behavior applies only to highly productive authors with 13 or more publications. Bootstrap goodness-of-fit testing rejects the pure power-law specification ( $p = 0.002$ ). A Vuong likelihood ratio test comparing the power law with a lognormal alternative yields a test statistic of  $-2.45$  ( $p = 0.014$ ), indicating that the lognormal distribution provides a statistically better fit to the data.

**Figure 1: Observed productivity distribution vs. power-law fit on log-log scales ( $x_{min} = 13$ ,  $\alpha = 3.42$ ; GOF  $p = 0.002$ ; Vuong  $p = 0.014$ ). Dashed line marks  $x_{min}$ .**

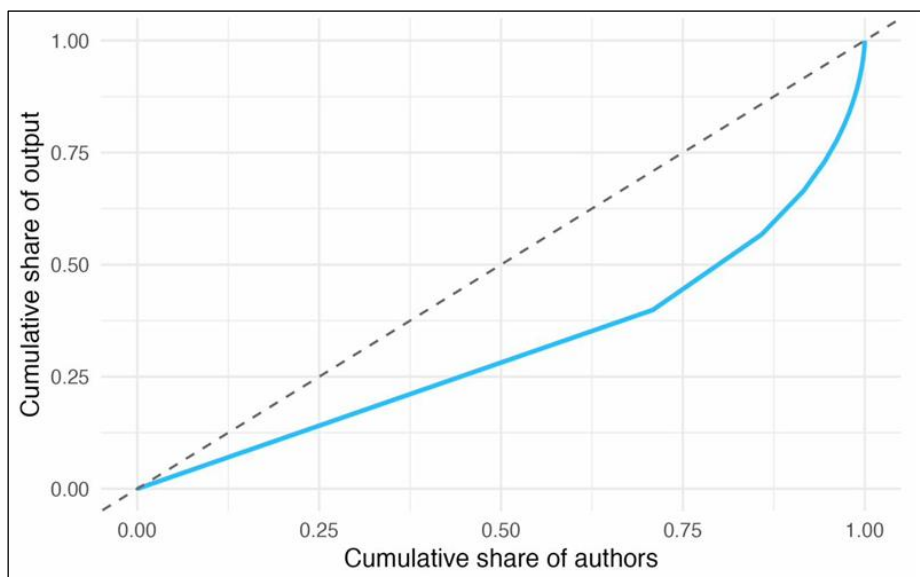


Taken together, the goodness-of-fit test rejects the pure power-law specification in the upper tail, and the estimated exponent ( $\alpha = 3.42$ ) is far steeper than the classical inverse-square value. Where the two models can be compared, the lognormal fits better. We note, however, that a better-fitting lognormal does not by itself refute a power law, as the two are difficult to distinguish over a limited range, and that this comparison concerns only authors with 13 or more publications rather than the full distribution to which Lotka's formulation refers.

### 3.2. Inequality in author productivity

Figure 2 shows that inequality in author productivity is moderate. The Gini coefficient equals 0.37. The top 10% of authors produce 33.5% of total output, while the top 1% account for 9.6%. The estimated scaling exponent ( $\alpha = 3.42$ ) is substantially larger than the classical Lotka value ( $\alpha \approx 2$ ), implying a thinner upper tail than predicted by Lotka's law.

**Figure 2: Lorenz curve of author productivity (Gini = 0.37), top 10 share = 33.5%, and top 1% share =9.6%.**



Consistent with this pattern, the lognormal distribution is favored over the power law in the model comparison regarding the upper tail, although this does not in itself refute a power-law form. The results indicate that scientific output is clearly concentrated among more productive authors, but the overall level of inequality is moderate (Gini = 0.37).

These results have several implications for bibliometric research and practice. First, the results suggest that Lotka's inverse-square law should not be interpreted as a universal regularity; instead, the scaling exponent appears to vary across fields and time periods, requiring empirical estimation rather than assumption (Kretschmer & Rousseau, 2001). Second, the finding that the lognormal distribution provides a significantly better fit than the power law suggests that researchers analyzing scientific productivity distributions should consider lognormal or other heavy-tailed models rather than assuming power-law scaling *a priori* (Clauset et al., 2009). Third, although scientific output is concentrated among more productive authors, inequality in sustainability science appears less extreme than in other fields. Prior research shows strong skewness in productivity across disciplines (Cho, 2021; Ruiz-Castillo & Costas, 2014), suggesting that the observed level (Gini = 0.37) is consistent with comparatively less concentrated authorship structures.

#### **4. Conclusion**

One hundred years after Lotka's seminal contribution, the productivity distribution in sustainability science remains heavy-tailed but does not conform to the classical inverse-square law. Modern maximum-likelihood methods indicate a steeper decay in the upper tail of the distribution and favor a lognormal specification over a power-law model. Productivity inequality, while clearly present, is moderate. These findings suggest that contemporary productivity patterns in sustainability science deviate from the classical Lotka formulation and highlight the importance of empirically testing bibliometric regularities rather than assuming universal scaling laws. By combining modern statistical tests for heavy-tailed distributions with inequality metrics, this study provides a systematic reassessment of Lotka's law in a rapidly growing interdisciplinary field.

#### **Open science practices**

Author identification relied on Scopus Author IDs, meaning the underlying records are proprietary and cannot be redistributed. To enable reproducibility, the R analysis code, derived aggregate data, figures, and publication DOI list (DOIs\_publications.txt) are openly available at <https://github.com/marcoschirone/lotka-sustainability-2016-2025>. These materials allow any researcher with Scopus access to reconstruct the corpus, reproduce the analysis, or pursue new research directions.

#### **Author contributions**

Both authors contributed equally to conceptualization, methodology, data collection & curation, data analysis, writing and editing. Both authors approved the final manuscript and share equal responsibility for its content.

#### **Competing interests**

The authors declare that they have no competing interests.

## References

- Cho, J. (2021). Analysis of global inequality in research outcome using the Gini coefficient. *Performance Measurement and Metrics*, 22(1), 25-37. <https://doi.org/10.1108/pmm-05-2020-0027>
- Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4), 661-703. <https://doi.org/10.1137/070710111>
- Egghe, L., & Rousseau, R. (2011). Theory and practice of the shifted Lotka function. *Scientometrics*, 91(1), 295-301. <https://doi.org/10.1007/s11192-011-0539-y>
- Gillespie, C. S. (2015). Fitting heavy tailed distributions: the powerLaw package. *Journal of Statistical Software*, 64, 1-16.
- Gini, C. (1997). Concentration and dependency ratios. *Rivista di politica economica*, 87, 769-792. (Original work published 1909)
- Kretschmer, H., & Rousseau, R. (2001). Author inflation leads to a breakdown of Lotka's law. *Journal of the American Society for Information Science and Technology*, 52(8), 610-614. <https://doi.org/10.1002/asi.1118>
- Li, L., Zhang, J., Yang, L., & Rousseau, R. (2024). Which older publications are still highly cited in the field of bibliometrics? Contemporary bibliometric citation classics. *Scientometrics*. <https://doi.org/10.1007/s11192-024-05102-3>
- Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12), 317-323. <http://www.jstor.org/stable/24529203>
- Ruiz-Castillo, J., & Costas, R. (2014). The skewness of scientific productivity. *Journal of Informetrics*, 8(4), 917-934. <https://doi.org/10.1016/j.joi.2014.09.006>
- Schirone, M. (2024). The formation of a field: sustainability science and its leading journals. *Scientometrics*, 129, 401–429. <https://doi.org/10.1007/S11192-023-04877-1>
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2), 307-333. <https://doi.org/10.2307/1912557>