Lotka' s Law, Co-authorship and Interdisciplinary Publishing

Peder Olesen Larsen* Markus von Ins†

May 30, 2008

Abstract

The robustness or breakdown of Lotka's law about the frequency distribution of scientific productivity depends on scientific cooperation, counting methods, interdisciplinary publishing and selection methods for sample collections. We have chosen to analyse the relationship using Mandelbrot's equivalent distribution model because this model is sensitive and uses the original data (scores). Five sets of authors and publications, the two sets used by Lotka, a set from *High Energy Physics*, a set from *Microbiology* and a set based on applicants to a research programme promoting young researchers have been used. It is shown that even for a sample of authors in *High-Energy Physics* with extremely strong co-authorship, Mandelbrot's distribution law is robust when complete-normalized (fractional) counting is used whereas complete counting results in a breakdown. In the field of *Microbiology* with much weaker cooperation, both counting methods result in a breakdown of Mandelbrot's law. Today a field like *Microbiology* with the corresponding set of journals, probably has a large content of interdisciplinary publishing and therefore no more fulfills the precondition of Lotka's law, that the total production of the authors (sources) is considered. For a set of applicants for the *Emmy Noether Programme* of the *German Research Foundation.* Mandelbrot's law breaks down despite the fact that all publications co-authored by the applicants are taken into account. In agreement with Bayes' theorem of conditional probabilities these results lead to the conjecture that any selection process of authors and/or publications causes a breakdown of Mandelbrot's law and, as a consequence Lotka's law.

1 Introduction

Lotka's Law from 1926 introduced a new kind of hyperbolic distributions in statistics and a new kind of principles, leading to such distributions. Among these principles are the "Cumulative Advantage Processes" introduced by Price (1976) (if you have written a lot of papers, the possibility increases that you will write even more). Additional importance has been ascribed to Lotka's law because of its equivalence with for example Zipf-Mandelbrot's law and Bradford's law (Rousseau 1990). It is however interesting and amazing that Lotka's law is less robust than Zipf-Mandelbrot's law on the frequency of words in texts and Bradford's law about journals and publications. There are several examples of the breakdown of Lotka's law about the frequency distribution of scientific productivity in today's cooperative research (Kretschmer and Rousseau 2001; Rao et al. 2003)(with references). However, (Bino et al. 2005) also found a breakdown of Lotka's law in the field of economic research with a well known moderate extent of co-authorship. Therefore, in current research, effects other than co-authorship can result in a breakdown of Lotka's law. Bino et al. conjectured that young researchers may prefer other journals than do senior researchers and that this could explain the breakdown in the case of economics journals. However, this can hardly explain the breakdown of Lotka's law in fields like *Physics* and *Microbiology.* Therefore, there is a need to identify effects other than co-authorship

*Marievej 10A,2, DK-2900 Hellerup, Denmark, pol at webspeed dot dk

†Institute for Research Information and Quality Assurance, iFQ, Godesberger Allee 90, D-53175 Bonn, Germany, ins at forschungsinfo dot de

H. Kretschmer & F. Havemann (Eds.): *Proceedings of WIS 2008*, Berlin
Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting
Humboldt-Universität zu Berlin, Institute for Library and Information Science (IBI)

causing a breakdown of Lotka's law and increasing since the time of the data used by Lotka (1916). Here we propose that to an increasing degree authors in many fields use journals from various fields for publication and that this may lead to a breakdown of Lotka's law. We call this effect "interdisciplinary publishing". We have studied the robustness and breakdown of Lotka's law using several databases[1]. In all studies, the database is a set of publications (called items in Egghe and Rousseau (1990)) and a set of authors (called *sources* in (Egghe and Rousseau 1990) and *objects of study* in (Gauffriau et al. 2007)). With the counting method chosen, a total score (a positive real number) will be assigned to each author. In the context of Lotka's law two counting methods are of interest: complete counting (counting the number of occurrences of the author's name in the database; this number is equal to the length of the author's personal publication list) and complete-normalized counting (fractional or adjusted counting; giving an author a score of $1/m$ for a publication with m co-authors). The total scores for all authors are sorted in descending order and used for ranking the authors[2]. If the scores are integer numbers (e.g. for complete counting), the frequency $f(n)$ of authors with a score n also follow a power law distribution. The frequency was shown by Lotka (1926) to follow an inverse square law

$$f(n) = A\, n^{-2}. \tag{1}$$

where A is the number of authors with score 1. A is 60.8 per cent[3] of the number of all the authors in the set (N). However, the condition of integer scores is too restrictive for many realistic examples (Rousseau 1988; Rousseau 1992). Therefore, alternative (equivalent) formulations of Lotka's law have been sought. As a result Rousseau (1990) could show that original sorted scores were distributed in agreement with a Mandelbrot distribution

$$y(r) = C_1/(1 + r \times C_2) \tag{2}$$

where $y(r)$ is the score of the author ranked r, $A = C_1/C_2 < N$, C_1 and C_2 constants. Because C_2 is usually small, C_1 is close to the highest score (rank 1). C_2 must not be too small because $C_1/C_2 < N$. We have studied sets with large values for N (about 10^5). In such cases plausible values for C_2 are $\geq 10^{-3}$. There is also a second plausibility argument for the size of C_2. Let us assume that we have kN authors from k different fields all following a Mandelbrot distribution of scores with the same parameters and the same N. If we superpose all these fields we obtain an aggregate with k subfields and a distribution y_k following

$$y_k(kr) = y(r) \text{ or} \tag{3}$$

$$y_k(r) = y(r/k) = C_1/(1 + r \times (C_2/k)). \tag{4}$$

According to the last equation, the distribution of the authors' scores in the aggregate field is a Mandelbrot distribution with a k times smaller value of C_2 (cf. (Price 1963) for Lotka's law and subfields). Therefore a value of C_2 much smaller than 10^{-2} is hard to justify in a subfield or even a journal. This consideration indicates that in cases, with no subfields C_2 must be close to 1. In such cases the Pareto (inverse linear) distribution can hardly be distinguished from the Mandelbrot distribution. This is relevant for the following empirical samples. All the above considerations hold for continuous ranks and scores and therefore must be shown to be valid in empirical, discrete cases (Rousseau 1990). We therefore simulated the discrete case with an exact Lotka-law with $A = 1000$ authors, all with exactly 1 publication. The total number of authors, N, is then 1625[4]. For their scores a good fit is given by a Mandelbrot distribution with C_1 about 47 and C_2 about 0.04[5]. The disadvantage of the Mandelbrot distribution is that it cannot be subjected to a statistical interpretation or test. However, the Mandelbrot distribution has also advantages found convenient in our treatments of empirical data:

[1] To include the studies of Lotka (1926) we also denote printed registries and bibliographies as databases.

[2] We do not apply statistical tests here, but argue with plausibility considerations. It is therefore irrelevant, that equal scores give different author ranks.

[3] We used $A/N \approx 1/\zeta(2) \approx 60.8$ per cent. This leads to a low estimate of A/N (cf. (Rousseau 1990) p201).

[4] We used $A/n^2 \geq 0.5$ or $n < 45$. This leads to a larger $A/N = 1000/1625 \approx 61.5$ per cent, than in the estimate of footnote 5.

[5] Caveat: This simulation is not a proof, but just a plausibility consideration. It is also somewhat careless in the treatment of 1 as a "unit" in the rounding to integers (cf.the considerations in Rousseau (1990) about the hidden parameter u in Lotka's law), because we do not know the information content (e.g. in bits) of a paper at all.

- for large values of the rank, $r > 30$, and logarithmic scales the distribution gives virtually a straight line. It is therefore easy to observe (small) deviations even for large values of r,
- the distribution can be applied to all positive real valued scores,
- distributions with values for C_2 near 1 can easily be checked with linear regressions on logarithmic scales,
- it is not necessary to transform empirical data to frequencies,
- all the parameters – including the hidden parameter (Rousseau 1990) – in Lotka's law are explicit and enter the curve fitting processes.

2 Data and Method

Five empirical datasets have been studied: 1 and 2. A reproduction of the scores used by Lotka (1926) and obtained from Auerbach (1910) (all author names used) and from *Chemical Abstracts* 1907-1916 (only authors with surnames beginning with A and B). Only complete counting scores can be reproduced from these two data sets, but co-authorship is assumed to be moderate before 1916. 3. The publications of the 350 co-authors in Achard et al. (2006), participants in the *LEP(L3)* collaboration at *CERN*, were found in a search on the *Web of Science*® (*WoS*) for the years 1985-2001 (the *LEP*-experiments were performed from 1990-2000. An extra year, 2001, was added to adjust for publication time, and the five years 1985-1989 were included for publications preparing the *LEP*-experiments). Publications due to authorship of homonyms of the 350 physicists were eliminated from the resulting set of publications by using information on subject categories, journals and titles of the publications. This resulted in a set of 4,586 publications with about 22,000 author names. The set was used to calculate complete and complete-normalized (fractional) counting scores for 339 physicists (11 physicists were not found in the years observed due to changes in the collaboration). 4. A set constructed from the articles, notes, reviews, and letters in the journals of *Current Contents*® Subject Category *Microbiology* downloaded from the *SCI CD ROM Editions* 1998-2002. The set contained 60,110 records with about 146,000 different author names. The set was used to calculate complete and complete-normalized (fractional) counting scores for the 146,000 co-authors. 5. The publication data (in the four years before application) for the 148 applicants from medical sciences to the *Emmy Noether Programme* of the *German Research Foundation (DFG)* (Böhmer and Hornbostel 2008; Hornbostel et al. 2008) were used for the calculation of complete and complete-normalized (fractional) counting scores. When the datasets did not show an obvious breakdown of the Mandelbrot-law, the two parameters of the Mandelbrot "best fit" distribution were approximated using the least-square method. The iterative Gauss-Newton method was applied to find an approximate value for the parameters. In cases where a fit was possible, three to four iterative steps – 20 steps were programmed – were enough for convergence.

3 Results and Discussion

The data derived from Auerbach (1910) are presented in Figure 1. With $C_1 = 41.9$ and a reasonable $C_2 = 0.06$ the figure shows a good fit. The value for C_2 indicates that there were about 16 subfields in Physics before 1900. For the *Chemical Abstracts* data Figure 1 shows a good fit for $C_1 = 121.7$ and a reasonable $C_2 = 0.035$.

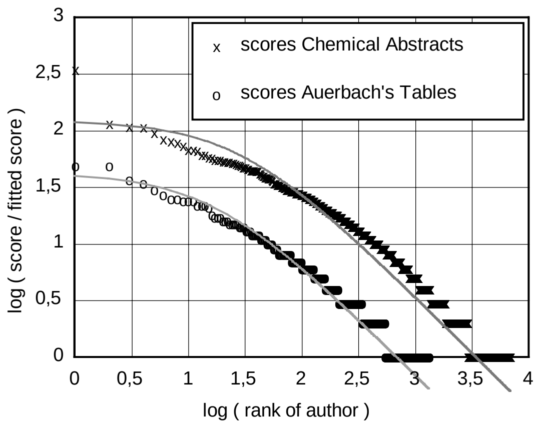

Figure 1: Scores of authors calculated from Lotka's data for *Chemical Abstracts* 1907-16 and *Auerbach's Historical Tables of Physics* until 1900. Mandelbrot fits are indicated with fine lines.

H. Kretschmer & F. Havemann (Eds.): *Proceedings of WIS 2008*, Berlin
Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting
Humboldt-Universität zu Berlin, Institute for Library and Information Science (IBI)

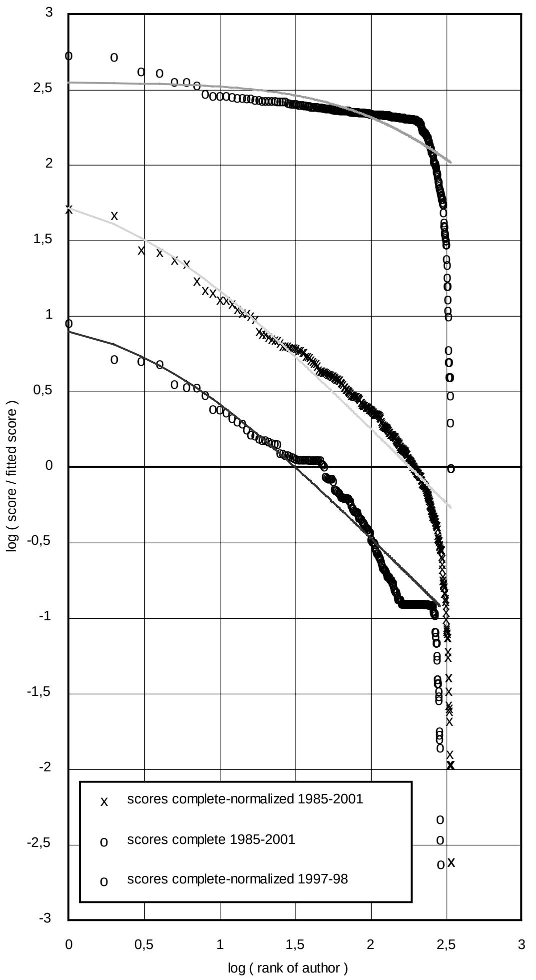

Figure 2: Complete and complete-normalized scores of 339 *High Energy physicists* 1985-2001. In the lowest graph only the publications in the years 1997/1998 are covered and complete-normalized scores are used. Mandelbrot fits are indicated with fine lines.

However, if we consider the majority of authors between ranks 100 and 5000, we see that the empirical data are slightly curved, whereas the Mandelbrot fit is nearly a straight line. Therefore, the fit for the *Chemical Abstracts* data is not so good as that found for the *Auerbach Tables.* Why is the fit in the case of *Auerbach Tables* better than in the case of *Chemical Abstracts*? A possible explanation is that *Chemical Abstracts* include not only publications from the journals of *Chemistry* but also a part of the publications in journals in *Life Sciences, Physics, Engineering* and other neighbouring fields. The work of authors in these neighbouring fields is therefore only partially/selectively covered by the *Chemical Abstracts.* Auerbach collected publications before 1900 in *Physics* and did not include partially covered neighbour fields.

Figure 2 presents results derived from the third dataset, publications co-authored by 339 *High Energy physicists.* The example was chosen due to the strong co-authorship in this field. For the 4,586 publications, the average number of authors was about 140, and one publication had 743 authors. Fitting the scores obtained by complete counting leads to $C_1 = 352.4$ and a seemingly reasonable $C_2 = 0.05$ (the top graph in Figure 2). However, C_1/C_2 should be smaller than the total number of authors, N, but is more than 20 times larger. Just two authors have scores of 1 or 2. Therefore, Lotka's law breaks down for *High Energy physicists* and complete counting. This is caused by many large scores. On the other hand, for scores obtained by complete-normalized counting, we find $C_1 = 71.9$ and a rather large $C_2 = 0.39$. $C_1/C_2 = 184.4$ is smaller than 339 and we see to the right of the graph (the middle graph in Figure 2) a couple of authors with very low scores. About 22 per cent of the authors account for much less than "predicted" by the Mandelbrot distribution. A simple explanation may be the long time period of 17 years in the study. Many young scientists stay in research only as long as is necessary for completing a master's thesis (1-2 years) or a doctoral thesis (3-4 years) and ergo much less than 17 years. This hypothesis has been checked using a shorter time period of two years 1997/98. The result is shown in the lowest graph in Figure 2. Just 287 of the 339 author names occur in this sub-period and just about 7 per cent[6] of these have smaller scores than expected. We find $C_1 = 10.2$ and a rather large $C_2 = 0.29$. $C_1/C_2 = 35.2$ is much smaller than 287. The two obvious "steps" in the figure, reducing the quality of the fit, stem from physicists only contributing to the multi-authored publications (0.12 in normalized scores, close to -1 in the logarithmic scale) and from physicists with 1 single-

[6]This value is smaller than the expected turnover-rates in scientific institutions.

H. Kretschmer & F. Havemann (Eds.): *Proceedings of WIS 2008*, Berlin
Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting
Humboldt-Universität zu Berlin, Institute for Library and Information Science (IBI)

authored paper and a number of multi-authored papers (1.12 in normalized scores, close to 0 on a logarithmic scale). The rather large values for C_2 show that a linear regression fit for a limiting Pareto-distribution (Rousseau 1990) would also give a good result and that the number of subfields is small, if not 1. This may indicate that the 339 selected physicists themselves form a field or an *Information Production System (IPS)* (Egghe and Rousseau 1990). Summing up, we conclude that Mandelbrot's law is robust even in a field of extremely strong co-authorship, if complete-normalized counting is used. On the other hand, Mandelbrot's law visibly breaks down, when complete counting is used. Remembering that complete counting just provides the lengths (number of entries) of personal "publication lists", it must be commented that the multiple assignment of the same publications to hundreds of personal publication lists does not multiply the information contained in the original publications. This result conforms with Kretschmer and Rousseau (2001) who find a breakdown of Lotka's law – using complete counting – for Dutch Physics institutes working partly in groups with strong co-authorship.

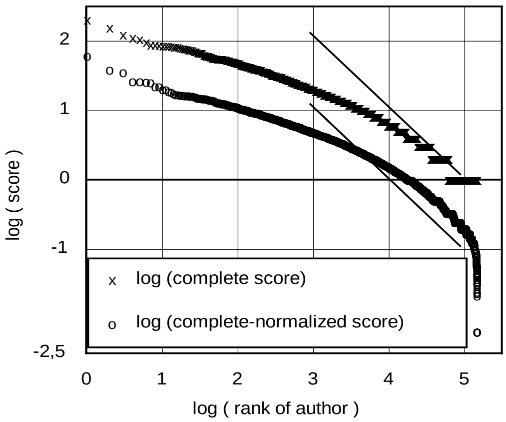

Figure 3: Complete and complete-normalized scores of about 146,000 authors found in journals of *Microbiology* in *SCI* 1998-2002.

In Figure 3 we display the data derived from the *Microbiology* set and find a breakdown of Mandelbrot's law both for complete and complete-normalized counting. Attempts to make a fit give rapidly divergent values of the two parameters C_1 and C_2 and a complete breakdown after 7-8 iterations, when the maximal allowable numbers[7] are reached. This is amazing, considering that co-authorship is moderate in *Microbiology* in comparison with *High Energy Physics*. In both cases (complete and complete-normalized counting) we recognize the reason for the breakdown by adding some straight lines to the graphs as shown in Figure 3. Comparing the data with the straight lines we recognize that the data form strongly curved graphs even in the region between ranks 103 and 105, comprising the majority of all authors. Even if Mandelbrot exponents different from 1 (Rousseau 1990) are considered, Mandelbrot's distribution becomes nearly a straight line for these high ranks, if C_2 is not chosen much smaller than 10^{-4}. On the other hand, in the left part of Figure 3, we see that the largest scores – and so also C_1 – are about 70-150. Therefore C_1/C_2 would be at least 7×10^5 and much larger than the total number (1.46×10^5) of authors found. With this contradiction, Mandelbrot's law is bound to fail for the set presented in Figure 3. In the right part of Figure 3 in the graph for the complete-normalized scores[8] we find, as in Figure 2, a low per- centage of authors with "too small scores". This may be caused by turnover effects as in the previous example. Again, these turnover effects relate to a small percentage of the data and cannot result in a complete breakdown of Mandelbrot's law.

It is a reasonable conjecture that many authors from neighbouring fields (for example *Life Sciences, Medical Sciences, Agricultural Sciences and Veterinary Sciences*) publish a part of their publications in *Microbiology* journals and that conversely, scientists in Microbiology Institutes publish also in journals from neighbouring fields. This effect, "Interdisciplinary Publishing", can result in a distorted Mandelbrot's distribution or a breakdown of Mandelbrot's law. To check this conjecture, we identified on the *WoS*® 506 of the publications from the *Department of Microbiology and Genetics* at *Har-*

[7] The programming language PERL was chosen.

[8] This effect cannot be made visible with complete scores, where the lowest score allowed is one.

H. Kretschmer & F. Havemann (Eds.): *Proceedings of WIS 2008*, Berlin
Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting
Humboldt-Universität zu Berlin, Institute for Library and Information Science (IBI)

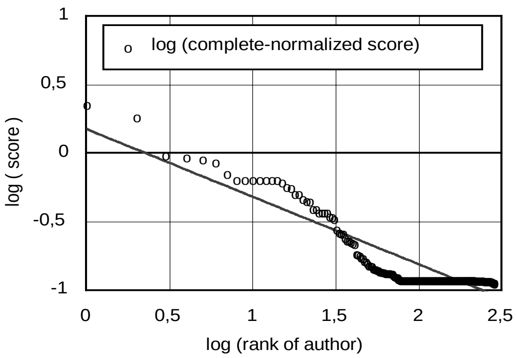

Figure 4: Complete-normalized (fractional) scores of 283 *High Energy physicists* 1997/98 in the journal *Physics Letters B*. A linear Pareto fit with slope -0.49 and correlation coefficient $R^2 = 0.82$ is indicated with a fine line.

vard Medical School in Boston, Massachusetts[9]. These publications occurred in about 120 journals in not less than 39 subject categories (subfields) of *Current Contents*® testifying to the inter- and multidisciplinary character of the research. As a second example, we used the publications in the frequently used journal *Physics Letters B* authored by the 287 *High Energy Physicists* behind the lowest graph in Figure 2. Complete-normalized scores were applied and a linear (Pareto) fit was attempted in agreement with the experience gained from Figure 2. The Pareto distribution has the disadvantage, that we cannot control the parameter C_2 of the Mandelbrot distribution but has the advantage of being simple and delivering an approximate value for the Mandelbrot/Pareto exponent (Rousseau 1990). The results are presented in Figure 4.

The long, nearly horizontal part of the data in the right part of the figure again stems from physicists contributing only to multi-authored publications. The linear regression fit is reasonable with a correlation coefficient $R^2 = 0.82$ and leads to a Pareto/Mandelbrot exponent close to 0.5. According to Rousseau (1990), a Pareto exponent close to 0.5 corresponds to a Lotka-exponent close to 3 and therefore indicates an "inverse cubic law" instead of the "inverse square law" found by Lotka. Therefore, this example, depending on publications in only one journal, shows also a breakdown of Lotka's law if we assume that the exponent 2 in Lotka's law is fundamental. In this example the reason for the breakdown is simple. Our data show that the 339 *High Energy physicists* published in about a dozen journals, "interjournal publishing". By the selection of only one journal, we obtain selected publications for many authors, resulting in deformed distributions. Figure 5 is based on all the publications (in the four years before application) co-authored by 148 applicants in *Medical Sciences* of the *Emmy Noether Programme* (Böhmer and Hornbostel 2008, Hornborstel et al. 2008) of the *German Research Foundation (DFG)*. Because the data included all publications covered by the definition, we might assume the resulting complete-normalized scores to be Mandelbrot distributed as in the case for the 339 *High Energy physicists*. However, a glance at Figure 5 shows that this is not the case. Both the graphs based on complete-normalized and complete counting are strongly curved in the region for author ranks between 30 and 120 (a majority of data). Trials to fit the Mandelbrot parameters end up with values of C_2 larger than 1. Therefore, a Pareto fit (linear regression on logarithmic scales) is adequate. This leads to a slope of -0.68 and a correlation coefficient $R^2 = 0.76$. The slope is not compatible with an inverse square law (slope -1), but in close agreement with an inverse cubic law (slope -0.5). Therefore, Lotka's law breaks down in this case, despite the fact, that we have adjusted for co-authorship by using complete-normalized counting, and that we considered all the publications and not only publications in specific fields.

In Figure 6 we display the complete scores[10] as in Figure 5 but in normal (not logarithmic) scales and try to find the reason for, and understand, this dilemma. In the left part of the figure the data seem to follow an inverse linear distribution with few large scores and many smaller scores. In the right part of the figure the data behave differently from the predictions of Lotka's law. According to the prediction, there should

[9]These are by far not all publications from the department, only a part being sufficient for our purpose.

[10]In Figure 6 we could have chosen to display the complete-normalized scores as basis for the argumentation. Complete scores were chosen, because they are more familiar.

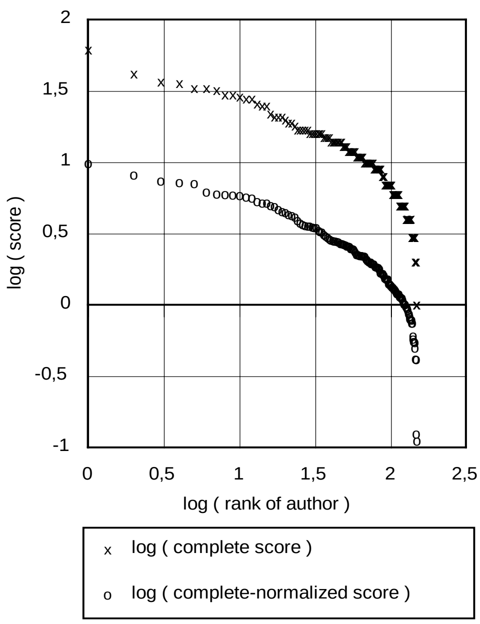

Figure 5: Complete and complete-normalized (fractional) scores of 148 applicants to the *Emmy Noether Programme* for young researchers in *Medical Sciences* in 4 years before application. Logarithmic scales are used for both ranks and scores.

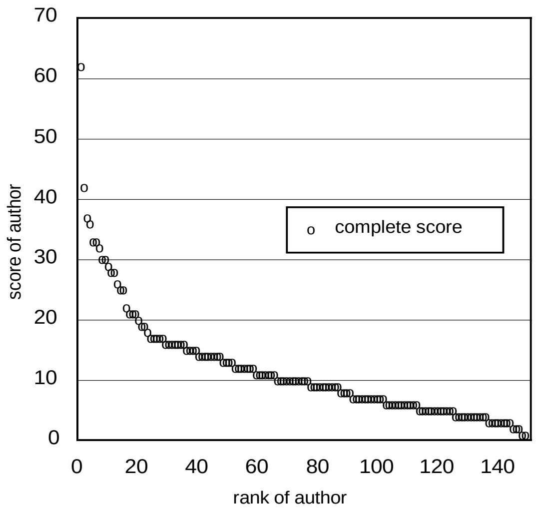

Figure 6: Complete scores of 148 applicants to the *Emmy Noether Programme* for young researchers in *Medical Sciences* in the 4 years before application.

be about 60.8 per cent (i.e. about 100) of the data in the lowest class with score 1, but there are only 2 with score 1. There are only 3 authors with score 2 whereas Lotka's law predicts a much larger number. A simple explanation is that postdoctoral researchers with low complete (or complete-normalized) scores generally did not apply for the *Emmy Noether Programme*. This effect can be described as a self-selection process and leads to a breakdown of Lotka's (Mandelbrot's) law, as recognized in Figure 5.

4 Conclusions

- Various presentations of the original (score) data can be used to study breakdown and robustness of Lotka's inverse square law. The Mandelbrot/Pareto distributions are very sensitive to deformations.
- Selection of publications, for example in sets of journals, lead to a breakdown of Mandelbrot distributions.
- (Self)selection by prolific scientists leads to a breakdown of Mandelbrot distributions.
- Complete scores lead to a breakdown of Mandelbrot distributions in fields with strong co-authorship. Therefore, it needs to be studied, whether complete scores causes a kind of selection.
- A necessary condition for a Mandelbrot (Lotka) distribution of productions is that a set of all productions from a set of sources is considered and that complete-normalized scores are used when collaboration is considerable.
- The breakdown is related to the more general Bayes Theorem of conditional probabilities and distributions.
- These results lead to a general conclusion: Lotka's law is valid if all the preconditions are fulfilled, if we have all items from a set of sources and therefore no selection, and

H. Kretschmer & F. Havemann (Eds.): *Proceedings of WIS 2008*, Berlin
Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting
Humboldt-Universität zu Berlin, Institute for Library and Information Science (IBI)

if we have no preselection of some publications by using non-normalized counting methods. Therefore, the change from single-authored to multiauthored research has not changed the fundamental production principles of cumulative advantage and the Matthew principle.

Acknowledgement

Support from the Carlsberg Foundation to POL is gratefully acknowledged. We are indebted to Wilfred T. Hastings for correction of the language.

References

Achard et al., P. (2006). Measurement of hadron and lepton-pair production in e(+)e(-) collisions at root s = 192-208 GeV at LEP. *European Physical Journal C 47*(1), 1–19.

Auerbach, F. (1910). *Geschichtstafeln der Physik*. Barth, Leipzig.

Bino, P., S. Subash, and A. Ramanathan (2005). Concentration in knowledge output: A case of economics journals. *European Journal of Comparative Economics 2*(2), 261–279.

Böhmer, S. and S. Hornbostel (2008). *Postdocs in Deutschland: Evaluation des Emmy Noether-Programms*, Volume No. 3 of *iFQ Working Paper*. Institut für Forschungsinformation und Qualitätssicherung (iFQ), Bonn. (to be published).

Egghe, L. and R. Rousseau (1990). *Introduction to Informetrics – Quantitative Methods in Library, Documentation and Information Science*. Elsevier Science Publishers, Amsterdam.

Gauffriau, M., P. Larsen, I. Maye, A. Roulin-Perriard, and M. von Ins (2007). Publication, cooperation and productivity measures in scientific research. *Scientometrics 73*(2), 175–214.

Hornbostel, S., S. Böhmer, B. Klingsporn, J. Neufeld, and M. von Ins (2008). Funding of young scientist and scientific excellence. *Special Issue of Scientometrics from the 11th International Conference of the International Society for Scientometrics and Informetrics Madrid 2007*. (to be published).

Kretschmer, H. and R. Rousseau (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.

Lotka, A. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences 16*(12), 317–323.

Price, D. d. S. (1963). *Little Science, Big Science*. Columbia University Press, New York.

Price, D. d. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science 27*(5), 292–306.

Rao, I., B. Sahoo, and L. Egghe (2003). A distribution of papers based on the fractional counting: an empirical study. In Guohua J. (Ed.), *Proceedings of the Ninth International Conference on Scientometrics and Informetrics*, pp. 241–250. Dalian University of Technology Press, Beijing.

Rousseau, R. (1988). Citation distribution of pure mathematics journals. In L. Egghe and R. Rousseau (Ed.), *Informetrics 87/88*, pp. 249–262. Elsevier, Amsterdam.

Rousseau, R. (1990). Relations between continuous versions of bibliometric laws. *Journal of the American Society for Information Science and Technology 41*, 197–203.

Rousseau, R. (1992). Breakdown of the robustness property of lotka's law: The case of adjusted counts for multiauthorship attribution. *Journal of the American Society for Information Science and Technology 42*, 645–647.