Information Processing & Management Vol. 30, No. 2, pp. 267-277, 1994
Printed in Great Britain.

0306-4573/94 $6.00 + .00

BRADFORD CURVES

RONALD ROUSSEAU
University of Antwerp (UIA), Speciale Licentie Informatie-en Bibliotheekwetenschap,
Universiteitsplein 1, B-2610, Wilrijk, Belgium and KIHWV,
Zeedijk 101, B-8400 Oostende, Belgium

(*Received* 22 *December* 1992; *accepted in final form* 12 *March* 1993)

Abstract – It is shown that generalized Leimkuhler functions give proper fits to a large variety of Bradford curves, including those exhibiting a so-called Groos droop or a rising tail.

1. INTRODUCTION

This article is written as a sequel to "An empirical examination of the existing models for Bradford's law" (Qiu, 1990). In that article the author investigated 22 models related to Bradford's law of scatter in classical bibliographies. In that investigation she made a distinction between size-frequency forms, such as the classical Lotka distribution (this is the case $\alpha = 2$ in formula (1) below), rank-frequency forms, such as Zipf's law, and cumulative rank-frequency forms such as Leimkuhler's function (a precise definition of this function is given below as formula (3)).

In Rousseau (1988) we demonstrated that the general Lotka function, that is,

$$f(y) = \frac{C}{y^{\alpha}} \quad y = 1, \ldots, y_{\max} \tag{1}$$

where $f(y)$ denotes the number of sources with production y, is, under certain hypotheses, mathematically equivalent to the following cumulative rank-frequency function:

$$R(r) = \frac{K}{\delta}\left(M^{\delta} - \left(M^{\delta-1} - \frac{\delta-1}{K}r\right)^{\delta/\delta-1}\right) \tag{2}$$

where $R(r)$ denotes the cumulative number of items produced by the r most productive sources. Parameters K, δ, and M correspond, respectively, to the parameters C, $2 - \alpha$ (if $\alpha \neq 2$), and $y_{\max}$ of Lotka's distribution (1). For $\alpha = 2$ we obtained the well known Leimkuhler function (eqn (3)) as the cumulative rank-frequency distribution corresponding to Lotka's size-frequency distribution:

$$R(r) = a\ln(1 + br). \tag{3}$$

Formula (2) was obtained by Egghe (1989, 1990b) as well, as one element in his general duality theory for Information Production Processes (IPPs). Equation (2) will be referred to as the generalized Leimkuhler function or representation.

One of Qiu's conclusions was that none of the models studied in her article could fit cumulative rank distributions, which showed an inflection point in the semi-logarithmic representation (a so-called Groos droop). Since we showed in Rousseau (1988) that for δ positive – this is the case corresponding to $\alpha < 2$ – Bradford curves always have an inflection point when represented on a semi-logarithmic scale, it seems natural to investigate whether eqn (2) can adequately fit data presented in a cumulative rank-frequency form. We will indeed show that generalized Leimkuhler functions usually give excellent fits to a wide variety of data. For other theoretical aspects of the generalized Leimkuhler curve, the reader may consult Rousseau (1988).

2. PRACTICAL PROBLEMS AND CHOICES TO BE MADE WHEN FITTING INFORMETRIC DISTRIBUTIONS

There exist different methods of fitting curves to data and of estimating parameters for the functional description of these curves. As in Qiu (1990), we have used a nonlinear, least squares fit to estimate parameters in eqn (2). Informetric data, however, present some specific problems; they usually are heavily skewed and we do not consider a source producing a singleton as important as a source with a much higher production. Consequently a weighting scheme had to be adopted. Where possible, we have opted for the simple weighting scheme in which from all sources with the same production we have used only one, namely the last one (i.e., that source that happened to have the highest rank among all sources with an equal production).

To test goodness-of-fit we have used the Kolmogorov-Smirnov test, as in Qiu's article (1990). Due to our weighting scheme we sometimes had problems at the end of the curve. Indeed, we are actually trying to fit a curved function to a straight line (a long tail of sources with a singleton production). This is illustrated in Fig. 1, curve a. In those cases where our fit was rejected by a Smirnov-Kolmogorov test at a 10% level of significance, we added a number of sources to the set of data used in determining parameters of a best fitting curve. This almost invariably led to an acceptable fit. See Fig. 1, curve b.

3. PRESENTATION OF DATA AND RESULTS

Next we will present the data set used to test the generalized Leimkuhler function (Table 1). Table 2 shows parameters of a best fitting, nonlinear, least squares function, the residual sum of squares and R^2, the coefficient of determination. Note that both the residual sum of squares and the coefficient of determination only give information on how well the resulting curve fits the data points used to calculate parameters of the generalized Leimkuhler function. Adding extra data points will invariably lead to a larger residual sum of squares and a lower coefficient of determination. Data sets marked * are those that were augmented by adding sources to the original weighting scheme. The exact number of sources used in finding best fitting parameters can be read from Table 3. We note that fits are excellent. Only two data sets out of 30 are not accepted by a Kolmogorov-Smirnov test at a 10% level (note that acceptance at a 10% level automatically implies acceptance at a 5% and a 1% level; moreover, to test goodness-of-fit we have used all data points, not just those used in the weighting scheme): namely Zipf's data on Chinese word frequencies (CHINESE) and the occurrences of Hebrew words in Numbers, Leviticus, and Deuteronomy (LND).

We have already mentioned that the classical Leimkuhler function (3) corresponds to a Lotka function with a parameter $\alpha = 2$. However, looking at this from a stochastic point of view, the probability that α is exactly equal to two, is zero, Thus, we have calculated best fitting classical Leimkuhler curves for all 30 data sets. See Table 4. As a rule, we see that when $0.1 > \delta > -0.4$ (i.e., the case corresponding to $1.9 < \alpha < 2.4$), acceptable curves

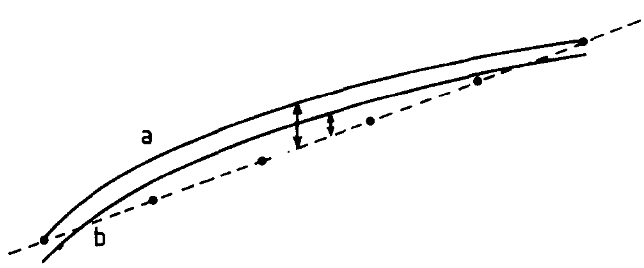

Fig. 1. Parts of best fitting curves based on endpoints (a) and based on both endpoints and added data points (b). It is shown that the latter curve has a smaller maximum distance to the data points than the former.

Table 1. Presentation of data sets

Symbol	Description	First published in	As studied, e.g., in
AG	Applied geophysics	(Bradford, 1934)	(Egghe, 1990a) (Qiu, 1990)
AUERBACH	Publications of physicists	(Auerbach, 1910)	(Lotka, 1926)
BIOL	Publications of biologists	(Williams, 1944)	(Rao, 1980)
CANADA	Canadian authors in information science	(Chu & Wolfram, 1991) p. 22	
CHINESE	Chinese word frequencies	(Zipf, 1932)	(Rousseau & Zhang, 1992)
COMPMUS	Computational musicology; authors	(Pao, 1979)	(Nicholls, 1986)
COMPSC	Computer science	(Radhakrishnan & Kerdizan, 1979)	(Pao, 1986) (Egghe, 1990b)
DOLBY	Words in song texts	(Dolby, 1992)	(Rousseau & Rousseau, in press)
DRESDEN	American mathematicians	(Dresden, 1922)	(Potter, 1981) (Nicholls, 1988)
ECON	Economics	(Rao, 1990)	
FAUNA	Mammalian fauna in Miocene	(De Bonis *et al.*, 1992)	
HUMAN	Humanities	(Murphy, 1973)	(Rao, 1980) (Egghe, 1990b)
INFO	Information science	(Pope, 1975)	(Egghe, 1990a) (Qiu, 1990)
INFORMET	Informetrics	(Egghe & Rousseau, 1990)	(Rousseau, 1992a)
LND	Hebrew words in Leviticus, Numbers, and Deuteronomy	(Morris & James, 1975)	(Brookes, 1982)
LUBR	Lubrication	(Bradford, 1934)	(Egghe, 1990a) (Qiu, 1990)
MAP	Map librarianship	(Schorr, 1975)	(Rao, 1980)
MAST CELLS	Mast cells	(Seley, 1968)	(Goffman & Warren, 1969) (Egghe, 1990a) (Qiu, 1990)
MICRO	Text information management software for microcomputers	(Nieuwenhuysen, 1988)	(Rousseau, 1990)
ORSA	Operational research	(Kendall, 1960)	(Egghe, 1990a) (Qiu, 1990)
PAGEREF	Page references	(Gellner, 1974)	(Brookes, 1984)
PERFECT	Artificial example	(Egghe, 1990a)	
PROGRAM	Small Pascal program	(Prather, 1988)	
RICEa	Rice literature	(Zhang, 1986)	See appendix for these data
RICEb	Rice literature	(Zhang, 1986)	See appendix for these data
RUDMANN	Language studies	(Rao, 1980)	
SACHS	Statistical methods	(Sachs, 1986)	(Egghe, 1990a) (Qiu, 1990)
SCHISTO	Schistosomiasis	(Warren & Newill, 1967)	(Goffman & Warren, 1969) (Egghe, 1990a) (Qiu, 1990)
STAT	Statistical methods	(Kinnucan *et al.*, 1987)	(Rousseau, 1992a)
TOP 40	Billboard's Top 40	(Cook, 1989)	

are found. Note, however, that generalized Leimkuhler functions always give a better fit. This is an obvious result, as the generalized Leimkuhler function contains one parameter more. Normally, the simpler curve should be preferred; see Kinnucan and Wolfram (1990), where this issue was studied.

Table 2. Parameters of a best fitting, nonlinear, least squares curve; residual sum of squares; coefficient of determination

A	B	C	D	E	F
AG	−0.0664	320.6	130.31	4888	0.99796
AUERBACH*	−0.1606	1091	40.12	11899	0.99959
BIOL*	−0.9457	1857	7.55	7195.8	0.99901
CANADA	−0.0868	205.8	22.226	238.77	0.99942
CHINESE*	−0.1083	2778	595.5	717945	0.99824
COMPMUS	−0.3199	442.4	37.75	791.4	0.99936
COMPSC	−1.263	345.4	8.065	0.021	1.00000
DOLBY	0.0027	413.3	145.53	7355	0.99902
DRESDEN	0.0391	253.41	60.57	3317	0.99833
ECON	0.1939	698.66	57.37	47801	0.99903
FAUNA	0.7719	4.25	60.19	24.6	0.99546
HUMAN	−0.4078	144.98	5.321	0.5117	0.99998
INFO	0.2642	528.90	313.72	47500	0.99972
INFORMET*	−0.6131	424.38	55.16	4327.1	0.99698
LND*	0.1466	3439	8075	45.7×10^6	0.99719
LUBR	0.0324	104.04	23.55	14.98	0.99989
MAP	−0.8120	340.3	11.18	11.135	0.99994
MAST CELLS	0.2253	318.96	67.058	3031.5	0.99975
MICRO	0.4035	59.22	38.34	843.2	0.99771
ORSA	−0.0217	303.69	416.41	6134.8	0.99822
PAGEREF	−0.0210	144.9	42.804	1080.1	0.99584
PERFECT*	−0.0278	935.34	30.775	11351	0.99971
PROGRAM	0.0583	19.006	13.144	1.086	0.99928
RICEa	−0.0319	87.15	899.1	146.4	0.99871
RICEb	0.1088	67.56	86.31	147.6	0.99871
RUDMANN*	−0.1182	844.8	14.539	7898.7	0.99932
SACHS	0.6103	39.50	67.04	2677	0.99715
SCHISTO*	0.1078	1181.5	336.47	66624	0.99983
STAT*	−2.1454	276.90	12.847	185.55	0.99634
TOP 40*	0.0800	1578.0	43.588	112195	0.99942

*Set augmented by adding sources to original weighting scheme.
A: Data set symbol.
B: Parameter δ.
C: Parameter K.
D: Parameter M.
E: Residual sum of squares.
F: Coefficient of determination: R^2.

4. BRADFORD CURVES AND THE GROOS DROOP

In a sense, the Groos droop can be considered an artefact of using semi-logarithmic scales. Indeed, using normal scales, all Leimkuhler curves are concave, and do not show any trace of an inflection point (cf. Fig. 2, showing Sachs' data and a best fitting Bradford curve). Moreover, we note that in this case, Leimkuhler curves are actually equivalent to Lorenz curves, well known from concentration theory (cf. Burrell, 1991a; Egghe & Rousseau, 1990a; Rousseau, 1992b).

On the other hand, we demonstrated in Rousseau (1988) that when $\delta > 0$ (corresponding to $\alpha < 2$), the semi-logarithmic representation of Bradford curves always shows an inflection point. Not surprisingly, we see that generalized Leimkuhler functions can fit data sets showing a so-called Groos droop in semi-logarithmic representation. See Fig. 4 (Sachs' data on semi-logarithmic scales).

5. A DISCUSSION OF RESULTS

More than half of the data sets used may be fitted by a classical Leimkuhler curve: AG, CANADA, COMPMUS, DOLBY, DRESDEN, FAUNA, HUMAN, LUBR, ORSA,

Table 3. Data related to the volume of the data sets used in this investigation

A	B	C	D	E
AG	326	1332	4.09	24
AUERBACH*	1325	3398	2.56	30
BIOL*	1527	2229	1.46	12
CANADA	268	629	2.35	15
CHINESE*	3342	13252	3.97	66
COMPMUS	501	1088	2.17	17
COMPSC	301	381	1.26	7
DOLBY	611	2233	3.65	33
DRESDEN	278	1124	4.04	26
ECON	744	4130	5.54	41
FAUNA	19	130	6.84	7
HUMAN	170	238	1.40	5
INFO	1011	7368	7.29	65
INFORMET*	554	890	1.61	20
LND*	2372	60615	25.55	102
LUBR	164	395	2.41	14
MAP	326	479	1.47	10
MAST CELLS	587	2378	4.05	39
MICRO	157	545	3.47	24
ORSA	370	1763	4.77	26
PAGEREF	123	519	4.22	18
PERFECT*	1449	3567	2.46	34
PROGRAM	21	55	2.62	7
RICEa	91	535	5.88	14
RICEb	91	402	4.42	14
RUDMANN*	1111	2291	2.06	19
SACHS	143	828	5.79	22
SCHISTO*	1738	9914	5.70	75
STAT*	204	236	1.16	9
TOP 40*	2311	7526	3.26	43

*Set augmented by adding sources to original weighting scheme.
A: Data set symbol.
B: Number of sources.
C: Number of items.
D: Average production.
E: Number of ranks used in least squares solution.

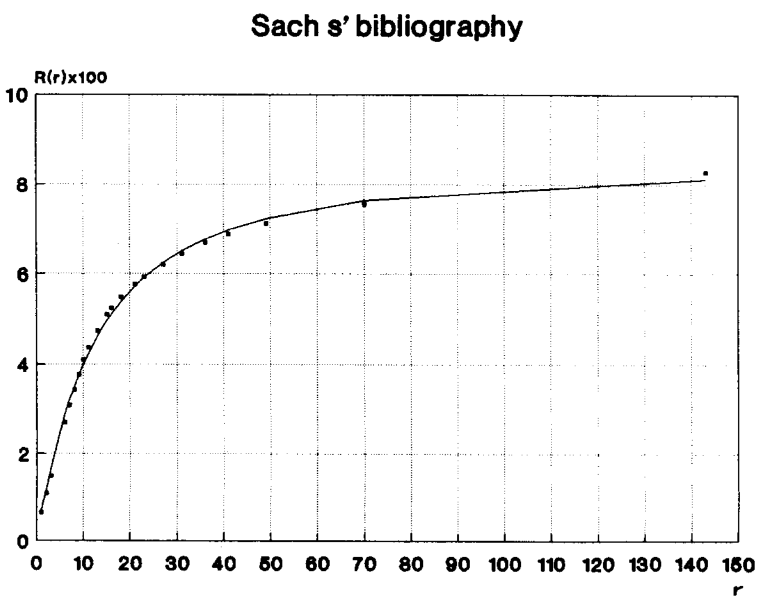

Fig. 2. Sachs' data and a best fitting generalized Leimkuhler curve. The curve is concave because logarithmic scales are not used.

Table 4. Results of fitting a classical Leimkuhler curve ($R(r) = a\ln(1 + br)$) to the data

A	B	C	D	E	F
AG	271.43	0.4211	6181.6	0.9974	10%
AUERBACH*	831.27	0.0400	35916	0.9988	1%
BIOL*	1080.14	0.0041	23633	0.9968	NOT
CANADA	181.72	0.1132	327.66	0.9992	10%
CHINESE*	1952	0.2116	2079083	0.9949	NOT
COMPMUS	276.66	0.0887	6297	0.9949	10%
COMPSC	188.86	0.0215	254.65	0.9976	NOT
DOLBY	416.32	0.3515	7364	0.9990	10%
DRESDEN	277.08	0.2287	3499	0.9982	10%
ECON	1138.96	0.0593	108867	0.9978	NOT
FAUNA	29.69	5.53	377	0.9300	10%
HUMAN	115.91	0.0398	10.328	0.9997	10%
INFO	1397.78	0.3664	2222924	0.9868	NOT
INFORMET*	284.99	0.0353	43438	0.9697	NOT
LND*	7344.39	3.1115	3.18×10^8	0.9805	NOT
LUBR	109.21	0.2223	21.117	0.9998	10%
MAP	191.81	0.0331	782.99	0.9959	NOT
MAST CELLS	563.26	0.1505	52479	0.9956	NOT
MICRO	139.35	0.4015	5651.2	0.9847	1%
ORSA	284.51	1.3543	6616.7	0.9981	10%
PAGEREF	138.75	0.3021	1085.95	0.9958	10%
PERFECT*	893.6	0.0335	12184	0.9997	5%
PROGRAM	20.465	0.6655	1.149	0.9992	10%
RICEa	79.813	9.039	186.35	0.9984	10%
RICEb	86.68	1.2259	308.95	0.9973	10%
RUDMANN*	737.32	0.0181	9301.5	0.9992	10%
SACHS	207.12	0.6013	31877	0.9660	NOT
SCHISTO*	1670.23	0.2673	112661	0.9971	NOT
STAT*	192.49	0.0111	622.67	0.9877	10%
TOP 40*	1855.2	0.0254	157936	0.9992	5%

*Set augmented by adding sources to original weighting scheme.
A: Data set symbol.
B: Parameter a.
C: Parameter b.
D: Residual sum of squares.
E: Coefficient of determination R^2.
F: Result of a Kolmogorov-Smirnov test.
10%: accepted at a 10% level.
5%: accepted at a 5% level.
1%: accepted at a 1% level.
NOT: not even accepted at a 1% level of significance.

PAGEREF, PROGRAM, RICEa, RICEb, RUDMANN, and STAT (at a 10% level); further, PERFECT and TOP 40 (at a 5% level).

Out of these data sets, FAUNA, RICEb, and STAT are only accepted because they are small sets for which statistically acceptable fits can easily be found. We note that ORSA, RICEa, RICEb, and FAUNA have b-values greater than 1, which is considered exceptional. See Fig. 3 for a typical example of this family of traditional Leimkuhler curves (DOLBY).

Curves with a clear Groos droop are ECON, FAUNA, INFO, MAST CELLS, MICRO, RICEb, SACHS, SCHISTO. Fits are excellent (accepted at a 10% level), which contrasts with Qiu's observations. We note that, with the exception of FAUNA, all these data sets have journals as sources. See Fig. 4 for a representative (SACHS) of this group of curves.

We also have a number of curves with a rising tail: AUERBACH, BIOL, COMPSC, HUMAN, INFORMET, MAP, RUDMANN, and STAT. Of these, HUMAN, RUDMANN, and STAT have acceptable classical Leimkuhler curves; yet residual sums of squares show that generalized Leimkuhler functions fit considerably better. Figure 5 shows a representative curve of this family (STAT).

Finally, generalized Leimkuhler functions for CHINESE and LND cannot be accepted by a Kolmogorov-Smirnov test (not even at a 1% level). The first of these two has a much

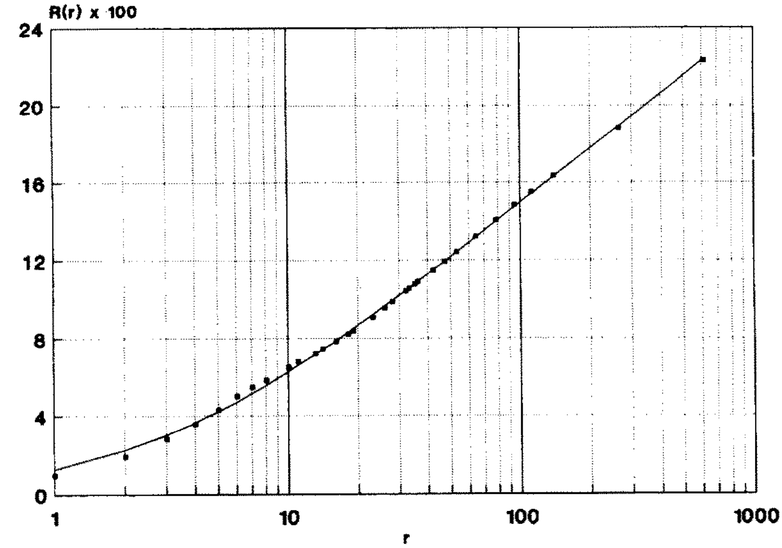

Fig. 3. An example (DOLBY) of a set of data points best represented by a classical Leimkuhler curve (logarithmic scale on the r-axis).

more productive rank one source than expected (cf. Rousseau & Zhang, 1992; note that in this article it has been shown that using a classical Leimkuhler curve with an additional additive term results in an acceptable fit). For the second one we have found a good-fitting generalized Leimkuhler function, which is only rejected because this data set has by far the largest number of items of our collection, making it very difficult to obtain a statistically acceptable result. Our best fit for the LND data is shown in Fig. 6. We further think it only accidental that the two collections that do not fit are data on word occurrences. The main reason for non-acceptance must be sought in the fact that both are large data sets (the largest in our collection). Moreover, DOLBY, which is a word occurrences set too, fits perfectly.

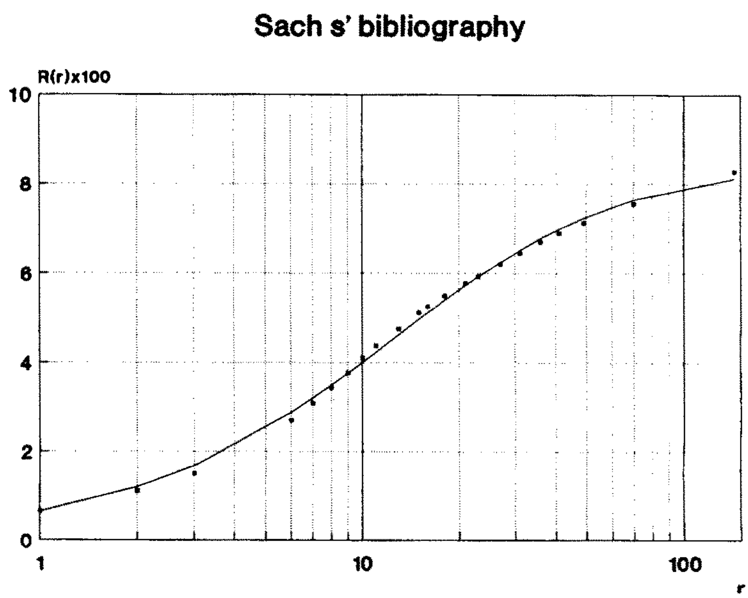

Fig. 4. An example (SACHS) of a best fitting generalized Leimkuhler curve with an inflection point (Groos droop).

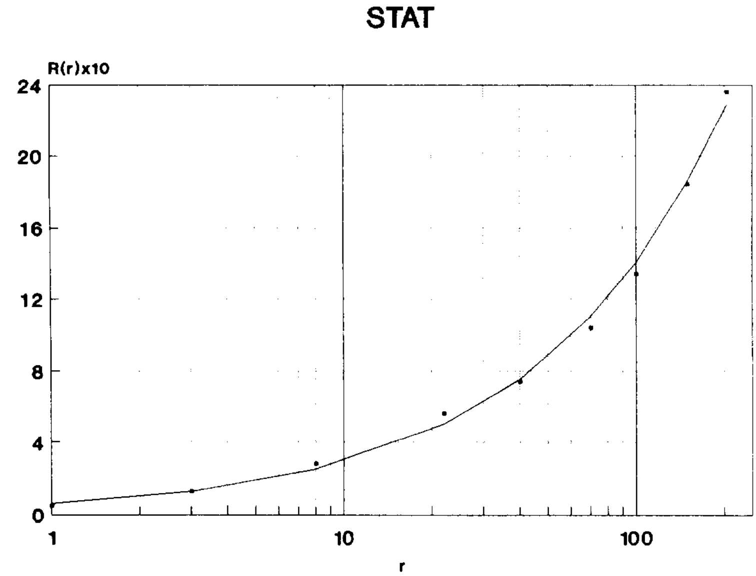

Fig. 5. An example (STAT) of a best fitting generalized Leimkuhler curve with a rising tail.

6. SOME FURTHER COMMENTS ON THE METHOD

When comparing our least squares fits with Egghe's method (Egghe, 1990a), we note two differences: one related to the underlying principles used to fit parameter values, and one related to the general form a cumulative data set must take. Egghe determines parameters for a best fitting Leimkuhler curve based on theoretical principles. By the term "theoretical principles" is meant that actual parameter values are determined based on values they must take according to formulae related to Bradford's law. We, on the other hand, have based our parameter estimations on standard statistical principles, adapted to the particular situation by using a special weighting scheme.

Secondly, we do not have an *a priori* idea of the form the curve must take: parameter values determine whether we find a classical Leimkuhler curve ($\delta = 0$, or $\alpha = 2$), a curve

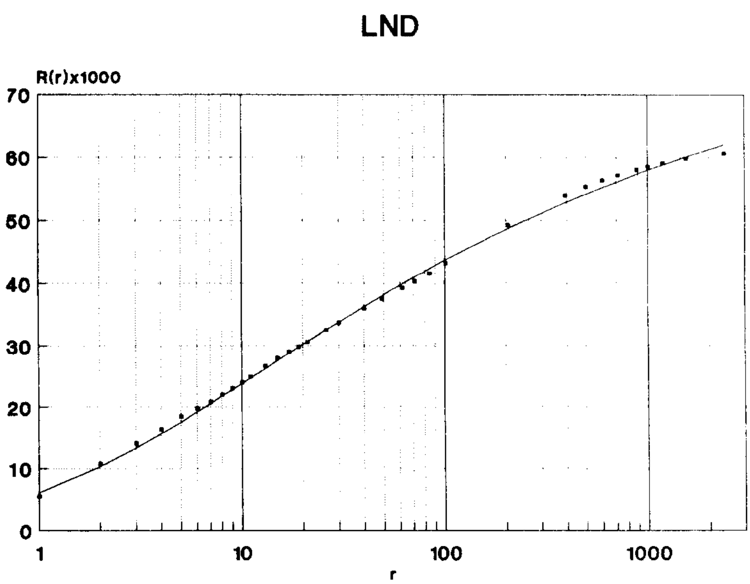

Fig. 6. An example (LND) of a generalized Leimkuhler curve that does not fit the data (in a statistical sense). Nevertheless, the curve yields a reasonable representation of the data points.

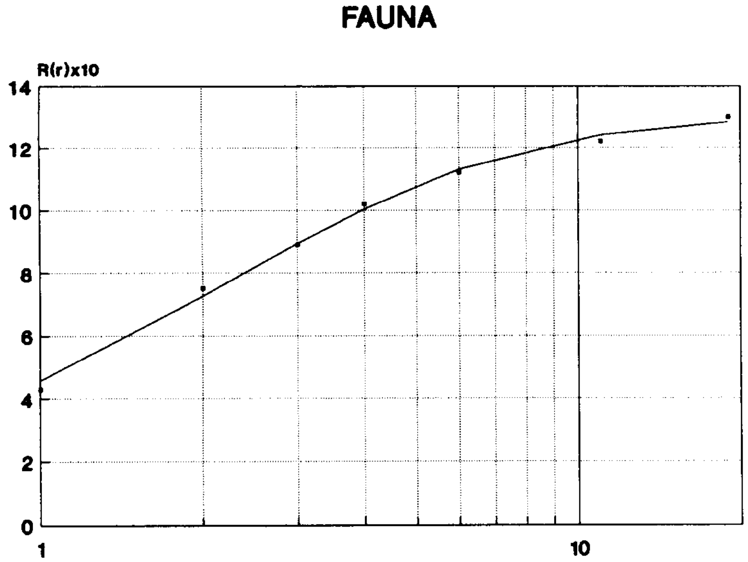

Fig. 7. An example (FAUNA) in which the inflection points is invisible.

with an inflection point ($\delta > 0$ or $\alpha < 2$), or a curve with a rising tail ($\delta < 0$ or $\alpha > 2$). Egghe's method, on the other hand, is based on the classical Leimkuhler curve. When a Groos droop occurs, it is considered an anomaly and is cut off. Curves with a rising tail are not considered in Egghe's model. Note, however, that in Egghe's general duality theory (Egghe, 1989), all curves can occur, and generalized Leimkuhler functions are part of this theory.

Although our description is static, change in the parameter δ (for the same data set, considered at different times) may well be related to evolution in time. These issues are studied in Burrell (1991b) and Oluić-Vuković (1992).

If δ is large (α small), the inflection point may be almost invisible, as is the case with FAUNA and LND. On a semi-logarithmic scale, the curve corresponding to a generalized Leimkuhler function seems convex. This also implies that the generalized Leimkuhler function incorporates Brookes' hybrid forms (Brookes, 1977) as well, making hybrid formulae unnecessary (cf. Fig. 7 (FAUNA)).

Traditional bibliometric distributions and their immediate generalizations have more positive points than sometimes thought. Thus, it is not always necessary to take recourse to more intricate descriptions such as Sichel's GIGP (Sichel, 1985), although, on the other hand, we admit that the GIGP is here (= in informetrics) to stay (Burrell & Fenton, 1993).

One last cautionary note: Best fitting parameter values for a cumulative distribution need not coincide with acceptable, let alone best fitting parameter values of the corresponding frequency distribution.

7. CONCLUSION

It has been shown that generalized Leimkuhler functions provide excellent fits to a large variety of data, including those exhibiting a so-called Groos droop or a rising tail.

Acknowledgement—We thank M. Dekeyser and J. Margodt (KIHWV) for their help during the preparation of this article.

REFERENCES

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

Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering*, *137*, 85–86. Reprinted in *Journal of Information Science*, *10* (1985) 176–180.

Brookes, B. C. (1977). Theory of the Bradford law. *Journal of Documentation*, *33*, 180–209.

Brookes, B. C. (1982). Quantitative analysis in the humanities: The advantage of ranking techniques. In H. Guiter & M. V. Arapov (Eds.), *Studies on Zipf's law* (pp. 65–115). Bochum: Brockmeyer.

Brookes, B. C. (1984). Towards informetrics: Haitun, Laplace, Zipf, Bradford and the Alvey programme. *Journal of Documentation, 40*, 120–143.

Burrell, Q. L. (1991a). The Bradford distribution and the Gini index. *Scientometrics, 21*, 181–194.

Burrell, Q. L. (1991b). The dynamic nature of bibliometric processes: A case study. Invited paper for presentation at the Third International Conference on Informetrics, Bangalore, India.

Burrell, Q. L., & Fenton, M. R. (1993). Yes, the GIGP really does work—and is workable! *Journal of the American Society for Information Science, 44*, 61–69.

Chu, C. M., & Wolfram, D. (1991). A survey of the growth of Canadian research in information science/Une enquête sur la croissance de la recherche canadienne en science de l'information. *The Canadian Journal of Information Science/Revue canadienne des sciences de l'information, 16*, 12–28.

Cook, K. L. (1989). Laws of scattering applied to popular music. *Journal of the American Society for Information Science, 40*, 277–283.

De Bonis, L., Bouvrain, G., Geraads, D., & Koufos, G. (1992). Diversity and paleoecology of Greek late Miocene mammalian faunas. *Palaeogeography, Palaeoclimatology, Palaeoecology, 91*, 99–121.

Dolby, T. (1992). *Astronauts and heretics (CD).* Virgin Records, CDV 2701.

Dresden, A. (1922). A report on the scientific work of the Chicago section, 1897–1922. *American Mathematical Society Bulletin, 28*, 303–307.

Egghe, L. (1989). The duality of informetric systems with applications to the empirical laws. Ph.D. thesis, The City University, London, U.K.

Egghe, L. (1990a). Applications of the theory of Bradford's law to the calculation of Leimkuhler's law and to the completion of bibliographies. *Journal of the American Society for Information Science, 41*, 469–492.

Egghe, L. (1990b). New Bradfordian laws equivalent with old Lotka laws, evolving from a source-item duality argument. In L. Egghe & R. Rousseau (Eds.), *Informetrics 89/90, Proceedings of the 2nd International Conference on Bibliometrics, Scientometrics and Informetrics* (pp. 79–96). Amsterdam: Elsevier.

Egghe, L., & Rousseau, R. (1990a). Elements of concentration theory. In L. Egghe & R. Rousseau (Eds.), *Informetrics 89/90, Proceedings of the 2nd International Conference on Bibliometrics, Scientometrics and Informetrics* (pp. 97–137). Amsterdam: Elsevier.

Egghe, L., & Rousseau, R. (1990b). *Introduction to informetrics.* Amsterdam: Elsevier.

Gellner, E. (1974). *Legitimation of belief.* Cambridge: Cambridge University Press.

Goffman, W., & Warren, K. S. (1969). Dispersion of papers among journals based on a mathematical analysis of two diverse medical literatures. *Nature, 261*, 1205–1207.

Kendall, M. G. (1960). The bibliography of operational research. *Operational Research Quarterly, 11*, 31–36.

Kinnucan, M., Nelson, M., & Allen, B. (1987). Statistical methods in information science research. *Annual Review of Information Science and Technology, 22*, 147–178.

Kinnucan, M., & Wolfram, D. (1990). Direct comparison of bibliometric models. *Information Processing & Management, 26*, 777–790.

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

Morris, P. M. K., & James, E. (1975). *A critical word-book of Leviticus, Numbers and Deuteronomy. The Computer Bible, vol. VIII.* Scholars Press, University of Montana.

Murphy, L. J. (1973). Lotka's law in the humanities? *Journal of the American Society for Information Science, 24*, 461–462.

Nicholls, P. T. (1986). Empirical validation of Lotka's law. *Information Processing & Management, 22*, 417–419.

Nicholls, P. T. (1988). Price's square root law: Empirical validity and relation to Lotka's law. *Information Processing & Management, 24*, 469–477.

Nieuwenhuysen, P. (1988). A bibliography of text information management software for IBM microcomputers and compatibles. *The Electronic Library, 6*, 264–320.

Oluić-Vuković, V. (1992). Journal productivity distribution: Quantitative study of dynamic behavior. *Journal of the American Society for Information Science, 43*, 412–421.

Pao, M. L. (1979). Bibliometrics and computational musicology. *Collection Management, 3*, 97–109.

Pao, M. L. (1986). An empirical examination of Lotka's law. *Journal of the American Society for Information Science, 37*, 26–33.

Pope, A. (1975). Bradford's law and the periodical literature of information science. *Journal of the American Society for Information Science, 26*, 207–213.

Potter, W. G. (1981). Lotka's law revisited. *Library Trends, 30*(1), 21–39.

Prather, R. E. (1988). Comparison and extension of theories of Zipf and Halstead. *The Computer Journal, 31*, 248–252.

Qiu, L. (1990). An empirical examination of the existing models for Bradford's law. *Information Processing & Management, 26*, 655–672.

Radhakrishnan, T., & Kerdizan, R. (1979). Lotka's law and computer science literature. *Journal of the American Society for Information Science, 30*, 51–54.

Rao, I. K. R. (1980). The distribution of scientific productivity and and social change. *Journal of the American Society for Information Science, 31*, 111–121.

Rao, I. K. R. (1990). Journal productivity in economics. In L. Egghe & R. Rousseau (Eds.), *Informetrics 89/90, Proceedings of the 2nd International Conference on Bibliometrics, Scientometrics and Informetrics* (pp. 251–280). Amsterdam: Elsevier.

Rousseau, R. (1988). Lotka's law and its Leimkuhler representation. *Library Science with a Slant to Documentation and Information Studies, 25*, 150–178.

Rousseau, R. (1990). A bibliometric study of Nieuwenhuysen's bibliography of microcomputer software for online information and documentation work. *Journal of Information Science, 16*, 45–50.

Rousseau, R. (1992a). 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, 43*, 645–647.

Rousseau, R. (1992b). Concentration and diversity in informetric research. Ph.D. thesis, UIA, Antwerp.

Rousseau, R., & Rousseau, S. (in press). Informetric distributions: A tutorial review/Les distributions informétriques: Un aperçu. *Canadian Journal of Information Science/Revue canadienne des sciences de l'information, 18.*

Rousseau, R., & Zhang, Q. (1992). Zipf's data on the frequency of Chinese words revisited. *Scientometrics, 24,* 201-220.

Sachs, L. (1986). *A guide to statistical methods and to the pertinent literature.* Heidelberg: Springer-Verlag.

Schorr, A. E. (1975). Lotka's law and map librarianship. *Journal of the American Society for Information Science, 26,* 189-190.

Seley, H. (1968). *The mast cells.* London: Butterworth.

Sichel, H. S. (1985). A bibliometric distribution which really works. *Journal of the American Society for Information Science, 36,* 314-321.

Warren, K. S., & Newill, V. A. (1967). *Schistosomiasis, a bibliography of the world's literature from 1852-1962.* Cleveland: Western Reserve University Press.

Williams, E. B. (1944). The number of publications written by biologists. *Annals of Eugenics, 12,* 143-146.

Zhang, Q. (1986). Obsolescence and Bradford distribution of rice literature. M.Sc. thesis, The City University, London, U.K.

Zipf, G. K. (1932). *Selected studies of the principle of relative frequency in language.* Cambridge, MA: Harvard University Press.

APPENDIX

As the following bibliography has not been published in the formal literature, data are presented here.

International bibliography of rice research (edited by IRRI); "Genetics, cytology and breeding of rice (1983-1984)," taken from Zhang (1986).

Frequency form

Number of items	Number of sources
1	37
2	22
3	11
4	3
5	3
6	2
8	3
10	2
12	2
13	1
14	1
15	1
41	2
190[a] 57[b]	1

[a]With supplement of *Japanese Journal of Breeding.*
[b]Without.

Cumulative rank-frequency form

	Cumulative number of items produced	
Rank	a	b
1	190	57
3	272	139
4	287	154
5	301	168
6	314	181
8	338	205
10	358	225
13	382	249
15	392	261
18	409	276
21	421	288
32	454	321
54	498	365
91	535	402