Information Processing & Management Vol. 28, No. 1, pp. 5-17, 1992
Printed in Great Britain.
0306-4573/92 $5.00 + .00

ANATOMY OF THE GENERALIZED INVERSE GAUSSIAN-POISSON DISTRIBUTION WITH SPECIAL APPLICATIONS TO BIBLIOMETRIC STUDIES

H.S. SICHEL
Department of Statistics and Actuarial Science, University of the Witwatersrand,
Johannesburg, South Africa

Abstract — The vast number of observed bibliometric and scientometric datasets display a definite downward deviation from a straight line in the upper tail, when plotted in a double logarithmic coordinate grid. For this reason customary theoretical distribution laws are very poor representations of the observed phenomena. This disadvantage also extends to recently suggested models such as the Yule, the two- and the three-parameter Waring distributions. The main types of the GIGP distribution are described and two important limiting cases are discussed. The constrained minimum χ^2 method is developed for the estimation of the three parameters α, b, and γ. Finally it is argued that the Kolmogorov-Smirnov goodness-of-fit test is not applicable in the field of bibliometrics.

1. INTRODUCTION

With a few exceptions, observed bibliometric size-frequency distributions are zero-truncated and reverse J-shaped, with extremely long upper tails. Their random variables are discrete and are advancing in steps of one unit. A good graphical picture is obtained by plotting their observed frequencies against the associated number of event counts in a double logarithmic coordinate grid. Indeed, this is what Zipf (1949) did originally with amazing tenacity and great enthusiasm. As he perceived all such plots as more or less linear, he formulated his inverse power law for size-frequencies. Earlier, Lotka (1926) had applied the same procedures to author productivity and postulated that the inverse power should be 2.

What both these authors failed to see was the significant systematic departures of observations from the straight line. All bibliometric studies are based on relatively large sample sizes, and even the crudest of significance tests would have shown that the linear hypothesis should have been rejected. Unfortunately, neither Zipf nor Lotka applied any statistical tests to their data fits. This omission led to the false belief that the observed deviations from the line were random and not systematic and significant.

The largest deviations are known to occur at the "head" (low event counts) and in the "tail" (high event counts) of observed bibliometric distributions. In particular, Vlachý (1980) has given many useful examples to show that the plotted points in the upper tail deviate strongly downwards from the straight line drawn through the main body of observations.

It may be appropriate here to quote Kunz (1988):

> Vlachý pointed out that practically all compiled Lotka distributions deviated at some distance from their heads from linearity, but theoreticians did not recognize the importance of this fact and further extrapolated initial results till infinity.
>
> The success breeds success phenomenon has its limits. A saturation takes place and instead of accelerating of a production rate, prolific authors are satisfied with their positions and produce less than could be expected from the Lotka law.

2. THE MAIN TYPES OF SIZE FREQUENCY DISTRIBUTIONS

In Fig. 1, five types of reverse J-shaped, observed scientometric size-frequency distributions are drawn schematically in a double logarithmic grid. Type (a) refers to a downward sloping line as envisaged by Zipf and Lotka.

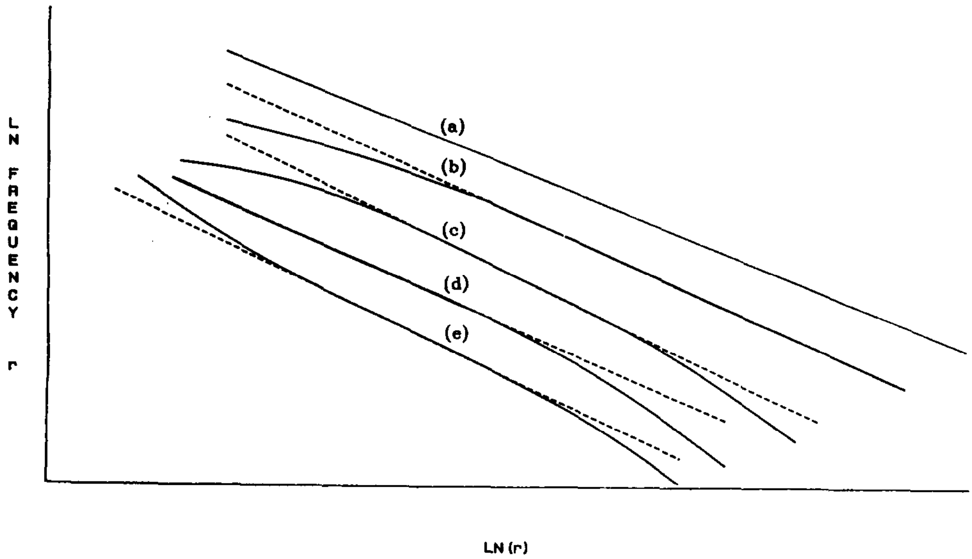

Fig. 1. Five types of reverse J-shaped scientometric frequency distributions in double logarithmic grid.

Very rarely do observations show such an oversimplified trend. Type (b) is convex at the head of the distribution and asymptotically approaches a downward sloping line at the tail. Data sets corresponding to this particular trend are also extremely rare. Mathematically this type (b) is represented by the Yule and Waring distributions. Type (c) has convexity at both ends of the distribution and is frequently encountered in reality. Type (d) is linear at the head and in the middle and displays convexity in the tail. This trend is very common. Finally, type (e) shows concavity at the head, a linear portion in the middle, and convexity at the upper tail. Such a pattern also occurs very frequently in observed distributions.

3. THE TAIL BEHAVIOUR OF BIBLIOMETRIC DISTRIBUTIONS

Sichel (1986) used Stirling's theorem to show that the upper tail of the Generalized Inverse Gaussian-Poisson (GIGP) distribution becomes

$$\phi(r) \sim \frac{C\theta^r}{r^{1-\gamma}} \tag{1}$$

where the discrete random variable r is large, $-\infty < \gamma < \infty$, $0 < \theta \leq 1$, and C is a normalizing constant. Taking logarithms on both sides of eqn (1) gives

$$\ln \phi(r) = \ln C - (1 - \gamma)\ln r + (\ln \theta)e^{\ln r}.$$

Now write

$$Y = \ln \phi(r),$$

$$X = \ln r,$$

$$A = \ln C,$$

$$-B = \ln \theta.$$

This leads to

$$Y = A - (1 - \gamma)X - Be^{X}. \tag{2}$$

From (2) we infer that in a double logarithmic grid the GIGP tail is first linear with a negative slope as long as $\gamma < 1$. For very large r, the tail curve will break away downwards from the line (i.e., it becomes convex). Hence it follows that the GIGP distribution conforms to the most often observed tail behaviour of bibliometric or scientometric size-frequency distribution, as depicted in Fig. 1 for the types (c), (d), and (e).

From eqn (2) one can see that for $\theta = 1$ exactly, $B = 0$ exactly

$$Y = A - (1 - \gamma)X. \tag{3}$$

This means that the GIGP tail could also be linear throughout, corresponding to the Zipf and Lotka distributions shown as types (a) and (b) in Fig. 1.

Some authors have suggested the Waring distribution (Burrell, 1988) in its 2- or 3-parameter form as a model for bibliometric data. Others (Kochen *et al.* 1982) prefer the Yule distribution, which is a special case of the Waring. With the help of Stirling's theorem it is easy to show that the Waring and Yule distributions have linear tails in a double logarithmic grid, and hence they are unsuitable for representing the upper tails of most observed bibliometric size-frequency data, as indicated for types (c), (d), and (e) in Fig. 1.

There is another advantage to the GIGP model. As long as $\theta < 1$, all population moments do exist, even if $\theta = 0.99999$. This cannot be said of the Waring or Yule distributions, where some of the population moments are infinitely large. In the worst scenario, the Waring and Yule distributions have no population arithmetic means at all!

4. THE ENTIRE SWEEP OF BIBLIOMETRIC DISTRIBUTIONS IN A DOUBLE LOGARITHMIC GRID

The Yule distribution can only take the form of type (b) in Fig. 1, whereas the Waring distribution is capable of representing types (a) and (b) only. In contrast, the GIGP distribution easily models all five cases given as types (a), (b), (c), (d), and (e) in Fig. 1.

The zero-truncated GIGP distribution may be written as

$$\phi(r) = [(\alpha/b)^{\gamma}K_{\gamma}(b) - K_{\gamma}(\alpha)]^{-1}\,\frac{[(\alpha^2 - b^2)/2\alpha]^r}{r!}\,K_{r+\gamma}(\alpha), \qquad \text{for } r = 1,2,3\ldots\infty, \tag{4}$$

where $\alpha \geq 0$, $b \geq 0$, $-\infty < \gamma < \infty$ and $K_{\nu}(z)$ is the modified Bessel function of the second kind of order ν and argument z. Readers of a previous paper by Sichel (1985) may notice that in the new parameterization

$$b = \alpha\sqrt{1 - \theta}, \tag{5}$$

with $0 < \theta \leq 1$. If one is interested in parameter θ, one simply obtains from (5)

$$\theta = 1 - (b/\alpha)^2. \tag{6}$$

As previously mentioned by Sichel (1982), parameter θ is largely responsible for the tail of the distribution, in contrast to parameter α, which determines the shape of the head. Parameter γ is important for the entire sweep of the GIGP.

Based mainly on experience, some useful rules of thumb have now emerged.

- If the head of the plotted observations in the double logarithmic grid is linear, α will be small, and if it is concave, α will be very small and may be set as $\alpha = 0$ a priori.

- If the head is convex, α will be medium to large and must be estimated from the data.
- If the downward break-away of the tail observations from the straight line drawn through the main body of the data in a double logarithmic grid is substantial, parameter θ will be relatively small, say 0.6–0.8. This is usually, but not always, associated with a low arithmetic mean $\bar{r}$.
- If the break-away of plotted tail observations from the straight line is minimal, we have $\theta \to 1$. This is usually, but not always, accompanied by a high arithmetic mean $\bar{r}$. It must, however, be pointed out most emphatically, that $\theta = 0.999$ still displays a very perceptible downward deviation from the line in the upper tail of the distribution.
- The slope of a line drawn through the middle portion of the observed frequency plot in a double logarithmic grid gives some indication of the magnitude of parameter γ. The steeper this line, the greater parameter γ (with a negative sign). This rule applies if θ is fairly large, say $\theta > 0.96$, with a relatively high arithmetic mean $\bar{r}$. An exception may arise if the arithmetic mean is low and the reducing action of a small θ bends the line prematurely downwards.

To facilitate the graphical interpretation of the slope constant $-\gamma$, it may be useful to plot "guidelines" in addition to the observed frequencies in the double logarithmic grid. We stipulate that all these guidelines should go through the point defined by $r = 1$ and $\phi(1) = 100$. Hence in logarithmic units

$$X = \ln 1 = 0,$$

$$Y = \ln 100.$$

From eqn (3) we find for these two values $A = \ln 100$.

For the general eqn (3) it follows that

$$\ln \phi(r) = \ln 100 - (1 - \gamma) \ln r. \tag{7}$$

Now set $\phi(r_s) = 1$, that is, $\ln \phi(r_s) = \ln 1 = 0$, which makes eqn (7).

$$\ln 100 - (1 - \gamma) \ln r_s = 0. \tag{8}$$

Finally we solve equation (8) for r_s:

$$r_s = (100)^{1/1-\gamma}. \tag{9}$$

All guidelines go through the logarithms of $r = 1$ and $\phi(1) = 100$. The second points on these lines are the logarithms of r_s, as defined by eqn (9), and $\phi(r_s) = 1$.

Below is given a table for the r_s values of the second points on the guidelines, for various parameters γ.

γ	$r_s = (100)^{1/1-\gamma}$
−2	4.642
−3/2	6.310
−1	10.000
−1/2	21.544
0	100.000
1/2	10 000.000

In general practice a single guideline for $\gamma = -\frac{1}{2}$ is quite sufficient.

5. TWO LIMITING DISTRIBUTIONS FOR THE GIGP

If in eqn (4) parameter $b \to 0$, two limiting probability distributions arise: The first case occurs if $\alpha \to 0$ and $0 < \theta < 1$, because $\lim_{\alpha\to 0} b = \lim_{\alpha\to 0} \alpha\sqrt{1-\theta} = 0$, from eqn (5). We make use of the Bessel function approximation

$$K_\nu(z) \sim \left(\frac{2}{z}\right)^\nu \frac{\Gamma(\nu)}{2} \tag{10}$$

which is valid if $z \ll \nu$.

Substitution of

$$K_\gamma(b) \sim \left(\frac{2}{b}\right)^\gamma \frac{\Gamma(\gamma)}{2},$$

$$K_\gamma(\alpha) \sim \left(\frac{2}{\alpha}\right)^\gamma \frac{\Gamma(\gamma)}{2},$$

and

$$K_{r+\gamma}(\alpha) \sim \left(\frac{2}{\alpha}\right)^{r+\gamma} \frac{\Gamma(r+\gamma)}{2}$$

into eqn (4) leads to

$$\phi(r) = \frac{[(\alpha/b)^{2\gamma} - 1]^{-1}}{\Gamma(\gamma)} \frac{\Gamma(r+\gamma)}{r!} [1 - (b/\alpha)^2]^r, \qquad \text{for } r = 1,2,3\ldots\infty. \tag{11}$$

Now from eqn (5) we have

$$(b/\alpha)^2 = 1 - \theta$$

and hence eqn (11) becomes

$$\phi(r) = \frac{[(1-\theta)^{-\gamma} - 1]^{-1}}{\Gamma(\gamma)} \frac{\Gamma(r+\gamma)}{r!} \theta^r, \qquad \text{for } r = 1,2,3\ldots\infty. \tag{12}$$

Equation (12) is the traditional zero-truncated negative binomial distribution with $\gamma > 0$ and $0 < \theta < 1$.

If in (12) we make $-1 < \gamma < 0$ for $0 < \theta < 1$, we obtain Engen's (1974) extended negative binomial distribution:

$$\phi(r) = \frac{\gamma[1 - (1-\theta)^\gamma]^{-1}}{\Gamma(1-\gamma)} \frac{\Gamma(r-\gamma)}{r!} \theta^r, \qquad \text{for } r = 1,2,3\ldots\infty. \tag{13}$$

Finally, if we set in (12) $\gamma = 0$ exactly, we obtain Fisher's logarithmic series distribution

$$\phi(r) = [-\ln(1-\theta)]^{-1} \frac{\theta^r}{r} \qquad \text{for } r = 1,2,3\ldots\infty. \tag{14}$$

To summarize: The first limiting distribution is obtained if $\alpha \to 0$ (and hence $b \to 0$). The zero-truncated GIGP becomes a two-parameter distribution with γ either positive (negative binomial), or γ negative (extended negative binomial), or $\gamma = 0$ (Fisher's LSD). Parameter θ is defined in the interval $0 < \theta < 1$.

The second limiting distribution is reached if $\theta \to 1$ and, concomitantly, $b \to 0$ once again. Further $\gamma < 0$.

Substitution of $K_\gamma(b) \sim (2/b)^\gamma[\Gamma(\gamma)/2]$ into (4) gives

$$\phi(r) = \left[(2/\alpha)^{-\gamma}\frac{\Gamma(-\gamma)}{2} - K_{-\gamma}(\alpha)\right]^{-1}\frac{(\alpha/2)^r}{r!}K_{r+\gamma}(\alpha), \qquad \text{for } r = 1,2,3\ldots\infty. \quad (15)$$

As γ is negative, we may reverse the sign of γ in (15) and then write

$$\phi(r) = \left[(2/\alpha)^{\gamma}\frac{\Gamma(\gamma)}{2} - K_{\gamma}(\alpha)\right]^{-1}\frac{(\alpha/2)^r}{r!}K_{r-\gamma}(\alpha), \quad (16)$$

where now γ is taken as positive.

This limiting distribution has two parameters: $\alpha > 0$ and $\gamma < 0$ if eqn (15) is used. Its higher population moments are infinitely large and in extreme cases the mean and variance may not exist. Clearly then, if $\theta \to 1$ and $b \to 0$, the GIGP becomes a stable Paretian distribution like those of Zipf, Lotka, and Yule. Fortunately, most real scientometric or bibliometric data are fitted well by the GIGP long before this limiting condition is reached.

6. PARAMETER ESTIMATION FOR THE GIGP DISTRIBUTION

Stein *et al.* (1987) described a Maximum Likelihood method for the joint estimation of the three parameters γ, b, and α if the observed distributions include the zero-events. As most scientometric distributions are zero-truncated, the MLE method becomes far too complicated. Some authors try to overcome this difficulty by shifting the random variable r by one unit (i.e., they write $x = r - 1$ where now $x = 0,1,2,\ldots\infty$). This approach is not to be recommended, as it changes the whole GIGP distribution shape and as the credibility into the theoretical aspects of the GIGP model is seriously undermined.

Instead, we equate the *observed* proportion of singletons to the corresponding population proportion of the zero-truncated GIGP. Further, we equate the *observed* sample mean to the population mean of the zero-truncated GIGP. For a given initial γ we solve the two simultaneous equations and, after determining the expected frequencies, we calculate the χ^2 statistic. We now change the γ parameter up and down and re-estimate parameters α and b until the χ^2 statistic is minimized. This method may be called a "constrained minimum χ^2 estimation."

This technique has several important theoretical underpinnings:

1. Anscombe (1950) and Sichel (1982) have shown that, if the first cell proportion in a discrete distribution is exceeding 30% of all frequencies, we obtain very efficient parameter estimates by making use of this first cell proportion. Virtually all observed scientometric distributions have more than 30% of frequencies in the first cell.
2. For the GIGP, the sample mean is an MLE estimator.
3. For large sample sizes, as found in observed bibliometric distributions, minimum χ^2 and MLEs give almost identical results, as may be theoretically proved.
4. Even if a traditional MLE procedure is used, we still would have to determine a χ^2 test for goodness-of-fit subsequently, to show that the chosen theoretical distribution (in our case the GIGP) is a reasonable model for the observed frequencies.

The whole constrained minimum χ^2 estimation method has been programmed for the GIGP model and no difficulties have been encountered.

7. WHICH GOODNESS-OF-FIT TEST?

Many authors working in the bibliometric field use the Kolmogorov-Smirnov one-sample test for describing the goodness of fit of their data to a chosen theoretical distribution model. Some of the authors pay lip service to the theoretical condition that the K-S test is strictly applicable only to continuous data and distributions. They then proceed to quote one or other authority who said that the K-S test is "conservative" if applied to discrete data.

None of the users of the K-S test mention that it is incorrect to first estimate parameters from the data and then apply existing tables to establish significant or nonsignificant departures from the chosen distribution model. Being a nonparametric test, this wrong procedure leads invariably to the acceptance of the hypothesis. Furthermore, the K-S test is only superior in power to the χ^2 test if the sample size is smaller or equal to 30 observations. Bibliometric and scientometric data sets are measured in hundreds or thousands of observations. In such cases, the χ^2 test has a far higher power of discrimination as it compares a good number of cells, whereas the K-S test depends entirely on a *single* maximum deviation of the cumulative distributions, without taking into account the other deviations before or after the maximum difference.

The present investigator has applied the conventional χ^2 test to the many published bibliometric data sets where other authors used the K-S test to justify the choice of a particular distribution model. Almost invariably the χ^2 test rejected these models, whereas the K-S test accepted the given hypotheses. The reader is referred to the excellent discussion on the merits and demerits of the Kolmogorov-Smirnov test by Stephens (1983). The χ^2 test of goodness of fit is applicable for both continuous and discrete data sets and distributions.

As bibliometric laws have large frequencies for the random variable at $r = 1,2,3\ldots$, no difficulties arise with respect to grouping except for the very high tail. To quote Nelson (1989) "Another disadvantage of the chi-square statistic is that in the tail of the distribution frequencies must be grouped, so it is not always sensitive to differences in this area." However, it should be pointed out that usually we have quite a few individual cell groupings in the upper tail due to the enormous scattering of individual events for the larger rs. Under such conditions it is very likely that the χ^2 test will monitor a significant departure for these high-frequency terms if the chosen distribution model is inappropriate.

Another worthwhile method of testing the fit or nonfit in the upper tail is to plot individual (ungrouped) observed frequencies in the double logarithmic coordinate grid. The superimposition of the theoretically expected frequencies—also ungrouped—will reveal to the eye whether the fit of the chosen distribution law is acceptable (or not) for the larger event numbers. This particular method has been used in the application section of this study.

Nelson (1989) also correctly mentions that traditional hypothesis testing using chi-square is not strictly applicable in the context of bibliometric data sets "since these are very large samples which are not simple random samples, but chi-square can be used as a comparison statistic." One should add to these remarks that probability levels of significance should be looked at as goodness-of-fit criteria, not to be interpreted in the customary sense of probability theory.

To sum up: The Kolmogorov-Smirnov test should never be used for bibliometric frequency distributions, whereas the χ^2 test of goodness-of-fit statistic is very useful to establish whether a proposed theoretical model is feasible and credible.

8. APPLICATIONS

The first example refers to the number of journals having $r = 1,2,3\ldots$ articles on schistosomiasis during the period 1852–1962, as listed by Goffman and Warren (1969). Altogether there were $n = 1738$ journals carrying $N = 9914$ articles. The observed frequency distribution is given in the first two columns of Table 2 and the plot of ln(frequency) against $\ln r$ is shown in Fig. 3. From the latter we see that the start of the graph is almost linear, which indicates that parameter α will be relatively small. Further, the general sweep of the observations in Fig. 3 displays a greater negative slope than the dotted guideline referring to $\gamma = -\frac{1}{2}$. This means that we should expect a parameter $\gamma < -\frac{1}{2}$. Finally, because of the almost linear trend even in the upper tail of the graph, we should expect a θ parameter slightly smaller than 1.

Table 1. Showing the relationship between the parameter estimates and the minimization of χ^2, for *schistosomiasis* literature

$\hat{\gamma}$	$\hat{\alpha}$	$\hat{b}$	$\hat{\theta}$	χ^2	*d.f.*	$p(\chi^2 \mid d.f.)$
−0.4594828	0	0	0.9879963	31.098	30	0.411
−0.500	0.116942	0.0118967	0.9896506	26.289	30	0.660
−0.600	0.339000	0.0281169	0.9931209	19.806	30	0.921
−0.700	0.529827	0.0341804	0.9958382	21.315	30	0.878
−0.800	0.704827	0.0327262	0.9978441	30.146	30	0.458
−0.900	0.869578	0.0252034	0.9991600	46.976	30	0.025
−1.000	1.026694	0.0134436	0.9998285	73.999	30	0.000
−1.100	1.176631	0.0019895	0.9999971	115.402	30	0.000
−1.1537	1.253617	0	1	145.443	30	0.000
			Minimum χ^2			
−0.628	0.394539	0.0306740	0.9939555	19.578	30	0.927

As mentioned before, we estimate the parameters via the constrained minimum χ^2 method. Hence the mean $\bar{r} = N/n = 5.704258$ and the proportion of singletons is $\hat{\phi}(1) = \hat{\text{freq}}(1)/n = 0.5224396$. These two observed statistics are kept the same right through the estimation process.

We start with an initial $\hat{\gamma} = -\frac{1}{2}$ parameter and solve the two simultaneous equations for $\hat{\alpha}$ and $\hat{b}$. We find $\hat{\alpha} = 0.116942$ and $\hat{b} = 0.0118967$, from which $\hat{\theta} = 1 - (\hat{b}/\hat{\alpha})^2 = 0.9896506$. We now calculate the χ^2 statistic from the first two columns in Table 2 using the above parameter estimates for the expected frequencies. This gives $\chi^2 = 26.289$. Next we repeat the whole process with $\hat{\gamma} = -0.6$, giving us different estimates for $\hat{\alpha}$, $\hat{b}$, and $\hat{\theta}$, and also a different χ^2.

Table 1 shows the solutions for the estimates of the parameters for 10 different $\hat{\gamma}$s. Included in this table are the two limiting distributions for the GIGP:

- $\hat{\gamma} = -0.4594828$, $\hat{\alpha} = 0$, $\hat{b} = 0$, and $\hat{\theta} = 0.9879963$. This is an "extended negative binomial distribution," first described by Engen (1974).

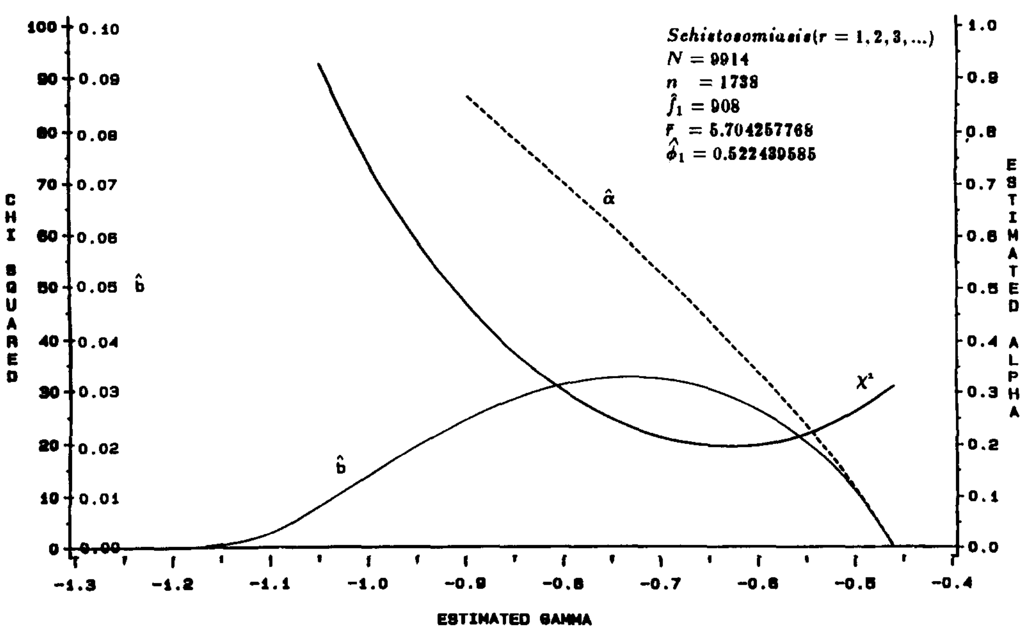

Fig. 2. Goodness-of-fit test statistic, χ^2, and parameter estimates $\hat{\alpha}$ and $\hat{b}$ as a function of parameter $\hat{\gamma}$, compiled from Table 1.

Table 2. Observed and expected frequency distributions for *schistosomiasis* literature during the period 1852–1962. Parameters estimated via the constrained minimum χ^2 method

Number of articles r	Observed number of journals $freq(r)$	Expected number of journals $\hat{freq}(r)$
1	908	908.000
2	266	274.251
3	137	130.484
4	76	77.788
5	57	52.394
6	44	38.047
7	27	29.070
8	29	23.040
9	19	18.774
10	14	15.632
11	11	13.245
12	6	11.383
13	10	9.901
14	10	8.699
15	9	7.710
16	10	6.885
17	8	6.189
18	10	5.596
19	4	5.086
20–21	7	8.904
22–23	6	7.542
24–25	5	6.474
26–27	2	5.620
28–30	7	7.166
31–33	7	5.983
34–36	4	5.070
37–40	6	5.665
41–44	4	4.700
45–50	3	5.709
51–60	5	7.053
61–70	5	5.053
71–90	5	6.655
91–120	7	5.556
121 and over	10	8.676
Total	1738	1738.000

Number of journals: $n = 1738$
Number of articles: $N = 9914$
Number of journals with only one article: $\hat{freq}(1) = 908$
Average number of articles per journal: $\bar{r} = 5.704257768$
Observed proportion of journals with only one article: $\hat{\phi}(1) = 0.522439585$

Parameter Estimates
$\hat{\gamma} = -0.628$
$\hat{\alpha} = 0.394539$
$\hat{b} = 0.0306740$
$\hat{\theta} = 1 - (\hat{b}/\hat{\alpha})^2 = 0.9939555$
Number of cells: 34
Degrees of freedom: 30
$p(\chi^2/d.f.) = 0.927$
$\chi^2 = 19.578$

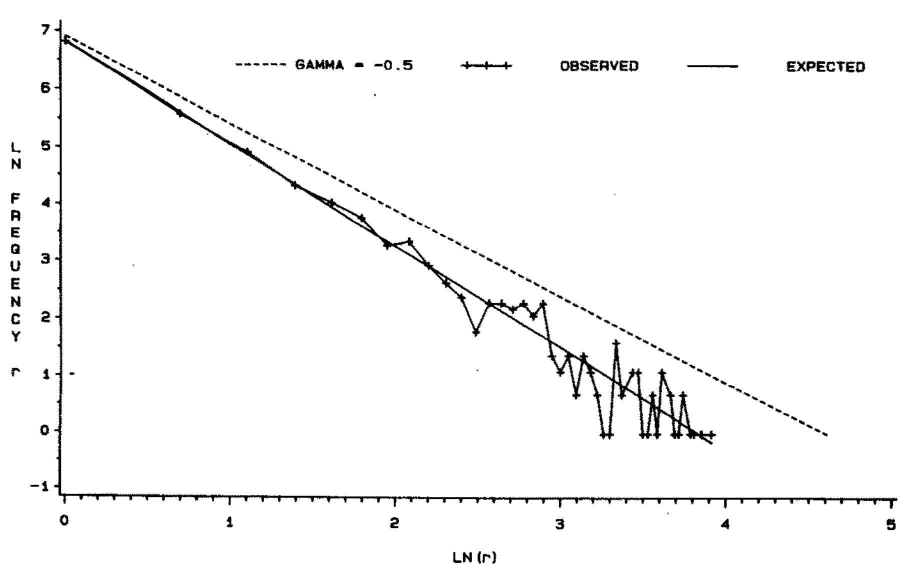

Fig. 3. Observed and expected frequencies: Schistosomiasis literature plotted on double logarithmic grid. $N = 9914$, $n = 1738$, $\hat{freq}(1) = 908$, $\hat{\gamma} = -0.628$, $\hat{\alpha} = 0.3945$, $\hat{\theta} = 0.99396$.

- $\hat{\gamma} = -1.1537$, $\hat{\alpha} = 1.253617$, $\hat{b} = 0$, and $\hat{\theta} = 1$. This is a stable Pareto distribution for which some of the lower population moments exist, but higher moments tend to infinity.

The "best" parameter estimates are those that minimize χ^2, given at the bottom of Table 1, that is, $\hat{\gamma} = -0.628$, $\hat{\alpha} = 0.394539$, $\hat{b} = 0.0306740$, and $\hat{\theta} = 0.9939555$, with a $\chi^2 = 19.578$ and an associated probability of $p = 0.927$ for 30 degrees of freedom. The full comparison of observed and expected frequencies is shown in Table 2. It will be seen that there are 15 cell groupings in the tail of the distribution. The lowest *expected* frequency is 4.7. Another comparison of goodness-of-fit is given in Fig. 3, where the expected frequencies (ungrouped) are shown as the solid curve. It will be noticed that the fit in the upper tail is very plausible.

The contents of Table 1 are drawn as curves in Fig. 2. In this diagram the functional relationship between $\hat{\gamma}$, $\hat{b}$, $\hat{\alpha}$, and χ^2 is clearly demonstrated. With increasing $\hat{\gamma}$:

- $\hat{b}$ starts at zero, rises to a maximum, and comes back to zero. The two zeros indicate the two limiting distributions of the GIGP.
- $\hat{\alpha}$ decreases monotonically and reaches zero at the traditional NBD or at the extended NBD limiting distribution.
- χ^2 follows a parabolic trend, where its minimum leads to the "best" parameter estimates for $\hat{\alpha}$, $\hat{b}$, and $\hat{\gamma}$.

The next example is from Rao (1989) giving the number of journals with $r = 1,2,3...$ articles in the International Bibliography of Economics. The observed frequency distribution is listed in the first two columns of Table 3, and the observations are plotted in a double logarithmic grid in Fig. 4. According to Rao, the negative binomial distribution, which he tried as a model, did not fit these data at all. From Fig. 4 we clearly see that

Table 3. Observed and expected frequency distributions for articles in the *International Bibliography of Economics*. Parameters estimated from mean and first observed proportion

Number of articles r	Observed number of journals $\widehat{freq}(r)$	Expected number of journals $\widehat{\widehat{freq}}(r)$
1	229	229.000
2	138	135.768
3	88	84.462
4	61	56.621
5	40	40.405
6	29	30.231
7	20	23.441
8	14	18.682
9	10	15.213
10	12	12.605
11	7	10.594
12	9	9.010
13	11	7.740
14	6	6.706
15–16	10	10.997
17–18	8	8.581
19–21	12	9.728
22–24	8	7.134
25–27	8	5.367
28–37	14	10.601
38 and over	10	11.114
Total	744	744.000

Number of journals: $n = 744$
Number of articles: $N = 4130$
Number of journals with only one article: $\widehat{freq}(1) = 229$
Average number of articles per journal: $\bar{r} = 5.55107527$
Observed proportion of journals with only one article: $\hat{\phi}(1) = 0.3077957$

Parameter Estimates
$\gamma = -0.5$ a priori
$\hat{\alpha} = 1.452712$
$\hat{b} = 0.264343$
$\hat{\theta} = 1 - (\hat{b}/\hat{\alpha})^2 = 0.9668887$
Number of cells: 21
Degrees of freedom: 18
$p(\chi^2/d.f.) = 0.932$
$\chi^2 = 9.995$

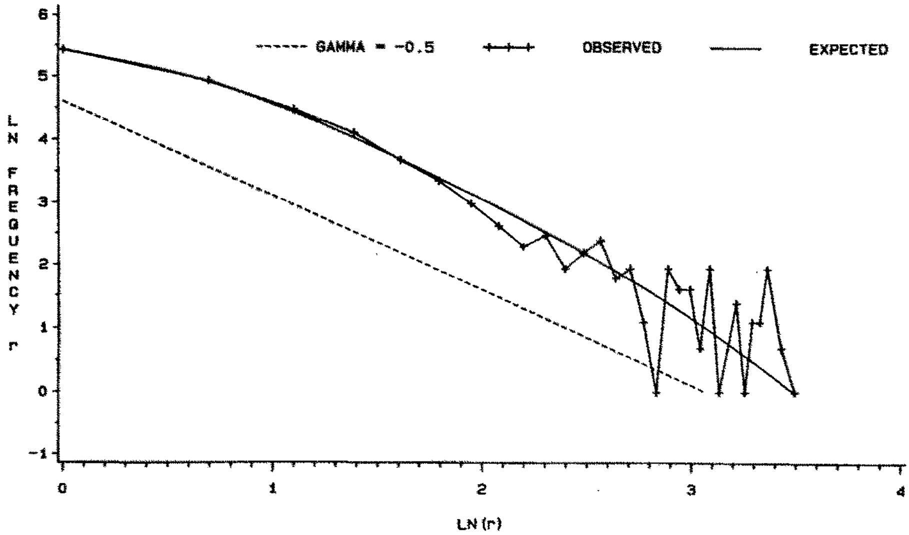

Fig. 4. Observed and expected frequencies for articles in the international bibliography of economics, plotted on double logarithmic grid. $N = 4130$, $n = 744$, $\hat{\text{freq}}(1) = 229$, $\gamma = -0.5$ a priori, $\hat{\alpha} = 1.4527$, $\hat{\theta} = 0.96689$.

- The start of the observations is convex, which indicates a relatively large α parameter.
- The middle portion of the curve is almost parallel to the $\gamma = -\frac{1}{2}$ dotted guide line. This suggests that we could use $\gamma = -\frac{1}{2}$ as given a priori.
- The upper tail deviates strongly downwards from the line projected through the middle portion of the observations. In consequence thereof, we should expect a θ parameter less than 0.98.

Table 4. Observed and expected frequency distributions for index terms of the MEDLARS database. Parameters estimated from mean and first observed proportion

Number of postings r	Observed number of terms $freq(r)$	Expected number of terms $\hat{freq}(r)$
1	2598	2598.000
2	640	689.354
3	362	345.296
4	245	213.627
5	151	147.262
6	90	108.452
7	85	83.520
8	70	66.427
9	62	54.136
10–11	96	82.904
12–13	52	60.378
14–16	69	64.658
17–20	52	57.433
21–25	44	45.932
26–33	45	42.807
34–48	41	36.719
49–52	4	5.275
53–57	3	5.128
58–64	4	5.220
65–76	5	5.588
77 and over	8	7.884
Total	4726	4726.000

Number of journals: $n = 4726$
Number of postings: $N = 18304$
Number of terms with only one posting: $\hat{freq}(1) = 2598$
Average number of postings per term: $\bar{r} = 3.873042742$
Observed proportion of terms with only one posting: $\hat{\phi}(1) = 0.549724925$

Parameter Estimates
$\hat{\gamma} = -0.45404$
$\alpha = 0$ a priori
$b = 0$ a priori
$\hat{\theta} = 0.9720139$
Number of cells: 21
Degrees of freedom: 18
$\chi^2 = 19.811$
$p(\chi^2/d.f.) = 0.344$

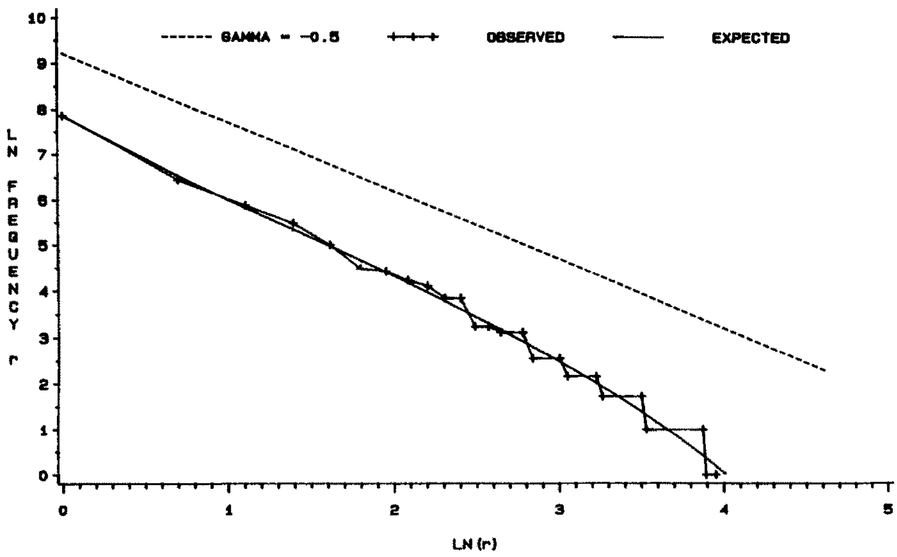

Fig. 5. Observed and expected frequencies for index terms in MEDLARS database, plotted on double logarithmic grid. $N = 18304$, $n = 4726$, $\hat{\text{freq}}(1) = 2598$, $\hat{\gamma} = -0.45404$, $\alpha = 0$ a priori, $\hat{\theta} = 0.972014$.

Parameters α, b, and θ were estimated from the observed $\bar{r}$ and $\hat{\phi}(1)$ with parameter $\gamma = -\frac{1}{2}$ given a priori. Observed and expected frequencies are shown in Table 3 and Fig. 4. The fit is extremely good, as $\chi^2 = 9.995$, for 18 degrees of freedom with an associated probability of $p = 0.932$. From the above it follows that the GIGP distribution, in contrast to the NBD, is a good model for the data on articles in economic journals.

The last observed bibliometric distribution is taken from Nelson and Tague (1985). The latter authors tried to model the index terms in the Medlars database by using a number frequency distribution for terms with low rs and a rank frequency distribution for high rs. The data are shown in the first two columns of Table 4, and they are plotted in a double logarithmic grid in Fig. 5. The frequency grouping is exactly the same as in Nelson and Tague's original paper. From Fig. 5 we see that

- The start of observations displays concavity, which means we could choose parameter $\alpha = 0$ as given a priori.
- The middle portion of the observations has a stronger negative slope than the $\gamma = -\frac{1}{2}$ dotted guide line. However, to expect a $\gamma < -\frac{1}{2}$ is fallacious in this case, as the downward break-away from the line through the centre of observations is substantial.
- Because of the downward deviation of observations from the line in the tail, we should obtain a θ parameter estimate, which is less than 0.98.

Parameter estimates were obtained from the observed $\bar{r}$ and $\hat{\phi}(1)$, with α taken as zero a priori. A comparison of observed and expected frequencies is given in Table 4 and Fig. 5. The fit is satisfactory for $\chi^2 = 19.811$ and 18 degrees of freedom. The associated probability is $p = 0.344$.

REFERENCES

Anscombe, F.J. (1950). Sampling theory of the negative binomial and logarithmic series distribution. *Biometrika*, *37*, 358–382.

Burrell, Q.L. (1988). Modelling the Bradford phenomenon. *Journal of Documentation*, 44, 1–18.

Engen, S. (1974). On species frequency models. *Biometrika*, *61*, 263–270.

Goffman, W.M., & Warren, K.S. (1969). Dispersion of papers among journals based on a mathematical analysis of two diverse medical literatures. *Nature*, *221*, 1205–1207.

Kochen, M., Crickman, R., & Blaivas, A. (1982). Distribution of scientific experts as recognized by peer consensus. *Scientometrics*, *4*, 45–56.

Kunz, M. (1988). Lotka and Zipf: Paper dragons with fuzzy tails. *Scientometrics*, *13*, 281–289.

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

Nelson, M.J. (1989). Stochastic models for the distribution of index terms. *Journal of Documentation*, *45*, 227–237.

Nelson, M.J., & Tague, J.M. (1985). Split size-rank models for the distribution of index terms. *Journal of the American Society for Information Science*, *36*, 283–296.

Rao, I.K.R. (1989). *Journal productitivy in economics.* Paper presented at the Second International Conference on Bibliometrics, Scientometrics and Informetrics. London, Ontario, Canada.

Sichel, H.S. (1986). The GIGP distribution model with applications to physics literature. *Czechoslovak Journal of Physics*, *B36*, 133–137.

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

Sichel, H.S. (1982). Asymptotic efficiencies of three methods of estimation for the inverse Gaussian-Poisson distribution. *Biometrika*, *69*, 467–472.

Stein, G.Z., Zucchini, W., & Juritz, J.M. (1987). Parameter estimation for the Sichel distribution and its multivariate extension. *Journal of the American Statistical Association*, *82*, 938–944.

Stephens, M.A. (1983). Kolmogorov-Smirnov type tests of fit. In Kotz, S., Johnson, N.L. & Read, C.B. (Eds.), *Encyclopedia of Statistical Sciences* (Vol. 4, pp. 398–402). New York: John Wiley & Sons.

Vlachý, J. (1980). Evaluating the distribution of individual performance. *Scientia Yugoslavica*, *6*, 155–189.

Zipf, G.K. (1949). *Human behavior and the principle of least effort.* Reading, MA: Addison-Wesley.