THE ASTROPHYSICAL JOURNAL, 162:405–410, November 1970

MAXIMUM-LIKELIHOOD ESTIMATION OF THE SLOPE FROM NUMBER–FLUX-DENSITY COUNTS OF RADIO SOURCES

DAVID F. CRAWFORD,* DAVID L. JAUNCEY,† AND HUGH S. MURDOCH*
Cornell–Sydney University Astronomy Center
Received 1969 September 29; revised 1970 June 19

ABSTRACT

The application of the method of maximum likelihood (ML) to the determination of the slope of the number–flux-density distribution of radio sources is discussed. It is shown that the ML estimate of the slope, which uses the maximum available information, may be obtained by a simple calculation from the individual flux densities. The efficiency and bias of other methods are compared with the ML method by means of numerical calculations on simulated data.

The error of using integral counts is stressed and an example is given from the literature to demonstrate that incorrect conclusions may be drawn through the use of statistically invalid procedures. The need to test the adequacy of the power-law model is stressed.

I. INTRODUCTION

Radio-source counts have been extensively investigated as a possible tool for cosmological investigations. The most important parameter to be determined is the exponent of the number–flux-density distribution. Clearly, when one is estimating the exponent from experimental data, it is necessary to have an objective estimate of that exponent and its standard error so that any comparison with theory will be valid and meaningful.

There are many methods currently in use for estimating the exponent from radio-source counts. It is the maximum-likelihood (ML) method alone, however, that makes full use of the data.[1] The purpose of the present paper is to examine the application of the ML method to radio-source counts, and to compare it with other currently used methods.

II. THE MAXIMUM-LIKELIHOOD METHOD

We shall assume here that the measurements of flux density are error free—a common, though not always justifiable, assumption. The more complex problem of treating data which are subject to measurement error will be dealt with in a later paper.

We shall start with the assumption that the flux-density distribution may be approximated by the expression

$$N(S) = kS^{-\alpha} \tag{1}$$

over a given range of flux densities, $S_0 \leq S \leq S_m$, where $N(S)$ is the number of sources with flux density greater than S. Since this assumption is not necessarily justified, we shall draw attention in § III to the need for a "goodness of fit" test to check its validity.

The likelihood L as a function of α is defined as the probability, given α, of obtaining the observed set of results. The ML estimate a of α is that value of α for which L is a maximum.

Before deriving the general results for ungrouped data, we will consider first the more familiar form for grouped data. Consider a total of M sources in the flux-density in-

* School of Physics, University of Sydney, Sydney, N.S.W., Australia.

† Center for Radiophysics and Space Research, Cornell University, Ithaca, New York.

[1] For a full discussion of the properties of the ML method, see, for example, Kendall and Stuart (1961), chapters 17 and 18.

terval S_0–S_m, grouped into m ranges of flux density S_{i-1}–S_i, with i ranging from 1 to m. If p_i is the probability that any source will occur in the ith interval, then L is given by the multinomial distribution

$$L\prod_{i=1}^{m} n_i! = M!\prod_{i=1}^{m} p_i^{n_i}, \tag{2}$$

where n_i is the number of sources in the ith range and

$$\sum_{i=1}^{m} n_i = M .$$

It is usually more convenient to maximize the logarithm of L. Since the factorial term is independent of α for a given set of observations, the ML method reduces to maximizing the expression

$$\mathfrak{L} = \sum_{i=1}^{m} n_i \ln p_i \tag{3}$$

as a function of α. Jauncey (1967) has previously used this method, and it may readily be shown that maximizing equation (3) is equivalent to maximizing his expression (6).[2]

From equation (1) it follows that

$$p_i = \frac{S_{i-1}^{-\alpha} - S_i^{-\alpha}}{S_0^{-\alpha} - S_m^{-\alpha}} . \tag{4}$$

Although we have defined the *integral* distribution in equation (1), it is clear from equation (4) that the method operates on the *differential* distribution, i.e., the number n_i of sources in each range of flux density. The n_i are thus independent. On the other hand, the integral counts N_i are not independent, and any estimation method which assumes that the N_i are independent is statistically incorrect.

We now proceed to show that grouping the data is unnecessary. Expressing all flux densities relative to S_0, we write s_i for the ratio S_i/S_0 and b for its maximum value S_m/S_0. Equation (4) may then be written

$$p_i = \frac{s_{i-1}^{-\alpha} - s_i^{-\alpha}}{1 - b^{-\alpha}} . \tag{5}$$

From equation (3) it is clear that it is only those flux-density ranges which actually contain sources that contribute to $\mathfrak{L}$. We may, if we wish, reduce the flux-density intervals to the point where we have M intervals each containing one source, with the remaining intervals devoid of sources. In the limit equation (5) becomes

$$dp_i = \frac{\alpha s_i^{-(\alpha+1)} ds_i}{1 - b^{-\alpha}} . \tag{6}$$

If terms independent of α are neglected, equation (3) becomes

$$\mathfrak{L} = M \ln \alpha - \alpha \Sigma_i \ln s_i - M \ln (1 - b^{-\alpha}) . \tag{7}$$

Equating $d\mathfrak{L}/d\alpha$ to zero to obtain α, we have

$$0 = \frac{M}{\alpha} - \Sigma_i \ln s_i - \frac{M \ln b}{b^{\alpha} - 1} . \tag{8}$$

[2] Due to a typographical error, pointed out by Dr. L. Golden, a factor M_i has been omitted from Jauncey's (1967) paper. His equation (6) should read

$$\log k^* \Sigma_i M_i + \Sigma_i M_i \log \left[\int_i S^{-\alpha} dS \right] .$$

Where no upper limit is set to the flux density (i.e., b infinite), we have the extremely simple analytic solution

$$\frac{1}{a} = \frac{1}{M}\Sigma_i \ln s_i \,. \tag{9}$$

The inverse of a is simply the mean of the natural logarithms of the normalized flux densities. For finite b equation (9) may be used as an initial solution to obtain the solution to equation (8) by iteration. For any reasonable range b, this process converges rapidly.

The simple form of equation (9) is not surprising since the substitution $\ln s_i = t_i$ converts the power-law distribution (1) into an exponential distribution $\exp(-\alpha t_i)$. There is thus a close analogy to the statistics of radioactive decay, with $1/\alpha$ corresponding to the mean lifetime τ.

The asymptotic variance of a for large M is given by

$$[\mathrm{var}(a)]^{-1} = -\left\langle \frac{d^2L}{da^2} \right\rangle \approx -\left(\frac{d^2L}{da^2}\right)_{a=\alpha} . \tag{10}$$

For $b^\alpha \gg 1$, a condition which usually holds, we obtain from equations (8) and (10) a good approximation,

$$\sigma_a = \frac{\alpha}{\sqrt{M}}\left[1 - \frac{\alpha^2(\ln b)^2}{b^\alpha}\right]^{-1/2} . \tag{11}$$

For $b = 10$ and $\alpha = 1.5$, the standard deviation is about 30 percent higher than for $b = \infty$, for the same number of sources.

For $b = \infty$, the exact sampling distribution of a given a for finite M is readily obtained from that given by Annis, Cheston, and Primakoff (1953) for the corresponding exponential distribution, as

$$f(a) = \frac{\alpha^M}{M!}\left(\frac{M}{a}\right)^{M+1} e^{-\alpha M/a} . \tag{12}$$

This is the well-known gamma distribution which is closely related to the χ^2 distribution. For large M it can be closely approximated by a Gaussian distribution with a standard deviation of

$$\sigma_a = \frac{M\alpha}{(M-1)(M-2)^{1/2}} \simeq \frac{\alpha}{\sqrt{M}} . \tag{13}$$

This is the same approximation to the standard deviation, for large M, as Refsdal's equation (13) (Refsdal 1969).

An estimate is unbiased if its expectation value is equal to the true value of the parameter being estimated. From equations (6) and (9) it follows that, for infinite range, the expected value of a^{-1} is α^{-1}. Although this shows that a^{-1} is an unbiased estimate of α^{-1}, a is not an unbiased estimate of α. From the sampling distribution (12) it follows that for infinite range, an unbiased estimate of α is $(M-1)a/M$.

It follows from the form of equation (7) that a is a *sufficient*[3] statistic for α and is, therefore, a minimum-variance estimate. This sufficiency implies that a contains all the information about the true exponent α that is available from the data.

III. "GOODNESS OF FIT" TESTS

The ML method outlined above allows full use to be made of the data in obtaining an estimate a of the exponent α, given that the distribution is of the form (1). It is, however, still necessary to test whether the form of equation (1) is an adequate representation of the data. To do this, some type of "goodness of fit" test must be used.

[3] For a formal definition of sufficiency see, for example, Kendall and Stuart (1961), § 17.31.

For grouped data the χ^2 test has found wide acceptance, but the grouping carries with it a loss in efficiency. The result of the test is also dependent on the actual grouping intervals chosen. Other tests exist which can be applied to ungrouped data. These tests make use of the fact that any continuous distribution may be transformed into a distribution which is uniform over the range (0, 1). For the assumed distribution (1) the required transformation is

$$y_i = \frac{1 - s_i^{-a}}{1 - b^{-a}} . \tag{14}$$

The various tests examine the observed y_i for departure from the expected uniform distribution.

Of the wide variety of possible tests the Smirnov test and the Kolmogorov test have achieved wide acceptance (Kendall and Stuart 1961). Both of these tests are easily used with digital computers.

IV. COMPARISON WITH OTHER METHODS

In order to compare the relative efficiencies of various methods of estimation we have used a Monte Carlo technique to produce simulated data from a power-law distribution over a given range of flux densities. With $M = 150$, $b = 1000$, and $a = 1.800$, 400 sets

TABLE 1

Method (1)	$\langle a \rangle$ (2)	$(\langle a \rangle - a)/(\text{SD of } \langle a \rangle)$ (3)	σ (4)	σ^2_{ML}/σ^2 (5)
ML.....	1.801	+ 0.1	0.142	1.00
GML.....	1.809	+ 1.2	0.151	0.88
LS.....	1.810	+ 1.2	0.154	0.84
χ^2.....	1.816	+ 2.2	0.149	0.91
Mχ^2.....	1.795	− 0.6	0.164	0.75
LSI.....	1.817	+ 1.8	0.184	0.59
R_1.....	1.704	−10.5	0.183	0.60
R_2.....	1.798	− 0.2	0.161	0.77

of data were produced. Similar calculations were done for $a = 1.500$ and 2.000 and also for various values of b and M, with results essentially the same as for the above.

For each set of data the slope a was estimated by each of the methods mentioned below. Except for ML and R_2, each set of 150 simulated sources was divided into ten groups with equal expected numbers. The group size was chosen to give a negligible probability of obtaining fewer than five sources in any one group. Different forms of grouping would produce slightly different results in each case. This is one of the disadvantages of grouping the data.

The methods used were the following:

ML.—Ungrouped maximum likelihood in which equation (8) is used.

GML.—Grouped maximum likelihood in which equations (3) and (5) are used.

LS.—Least squares, i.e., maximizing $\Sigma_i(Mp_i - n_i)^2$.

χ^2.—Maximizing $\Sigma_i[(Mp_i - n_i)^2/Mp_i]$.

Mχ^2.—Modified chi squared, i.e., maximizing $\Sigma_i[(Mp_i - n_i)^2/n_i]$.

LSI.—Least-squares regression of the logarithm of the integral numbers versus the logarithm of the flux density.

R_1.—Refsdal's method which uses his equation (11) which incorporates his weights based on his equation (12).

R_2.—Refsdal's method with the same data arranged into three groups with expected differential numbers 90, 45, and 15, respectively. This grouping is similar to that used by Refsdal in applying his method.

The results of the simulation are presented in Table 1, in which the method is listed

in column (1), and the mean value $\langle a \rangle$ of the estimate in column (2). Column (3) contains the difference $\langle a \rangle - a$ divided by each observed standard deviation of $\langle a \rangle$. Column (4) gives the observed standard deviation σ of the estimate about the true value a, and column (5) gives the ratio of the ungrouped ML variance to the variance, σ^2, of each method. The ML method in column (2) has been corrected for the known $(M - 1)/M$ bias. The observed variance is 0.0201 ± 0.0018 compared with an expected value of 0.0216.

The results given in column (3) give an indication of any possible bias. The only clearly significant bias is for Refsdal's method with ten groups (R_1). When his method is used with coarse grouping (R_2) as in his paper, there is no significant bias.

The sufficiency of the ML estimate ensures that it is a minimum-variance estimate and hence the most efficient. The variance ratio presented in column (5) is a measure of the relative efficiency of the various methods. The loss of efficiency is about 10 percent for grouped ML and χ^2, but is somewhat larger for the other methods.

When one is fitting to experimental data in grouped form, it is necessary to consider the statistical weight of each group. Murdoch (1958) has shown that, in a situation somewhat similar to source counts (the fitting of the slope of the density distribution of cosmic-ray showers), quite wrong conclusions can be drawn about the form of the distribution when an unweighted least-squares fit is carried out.

Jauncey (1967) has shown that the usual fitting procedures are not valid when applied to integral counts because of the lack of independence of the individual data points. The error in the exponent deduced from integral fitting will always be too small, and any differences in exponent will therefore be overemphasized. To illustrate this we have applied the usual formula for the standard deviation of the slope of a straight line regression equation for our Monte Carlo results for the LSI method. This yielded an average standard deviation of 0.043 compared with the observed value of 0.184.

A recent example from the literature is the claim, based on integral counts, of a significant difference between the slopes for radio sources with high and low brightness temperatures (Fomalont 1968). Application of equation (13) shows that his quoted standard errors are considerably less than the theoretical minimum value. Furthermore, in comparing two experimental distributions, the appropriate method is to compare the differential data directly instead of comparing exponents. Grouped data can be arranged in a simple $2 \times m$ contingency table and a χ^2 test applied (Kendall and Stuart § 33.28). Applying this test to Fomalont's data leads to a χ^2 of 5.4 for five degrees of freedom. For identical distributions the probability of getting this value of χ^2 or greater is 0.37. We conclude from this test that Fomalont's data show no significant difference between the high- and low-brightness source distributions.

It should be pointed out that, while Refsdal's method operates on integral numbers, he has shown that it is not subject to Jauncey's criticism mentioned above. We find that Refsdal's exact expression (Refsdal 1969, eq. [13]) for the standard deviation is a good estimate of the observed standard deviation about the mean. For R_1 the observed standard deviation (taken about a mean of 1.704) is 0.156 compared with the theoretical value of 0.152. For R_2 the observed standard deviation is 0.161 compared with the theoretical value of 0.155.

V. CONCLUSIONS

We can say the following in favor of maximum-likelihood estimation of the slope from the number–flux-density distribution of radio-source counts:

1. It is the most efficient method.
2. It is exceedingly simple to apply.
3. For any reasonable number of sources, the sampling distribution is a good approximation to a Gaussian.
4. An unbiased estimate of the slope may be obtained.
5. The use of ungrouped data gives a unique answer for a given set of data.

It is necessary to test the assumed form of the distribution. Adequate "goodness of fit" tests for ungrouped data are available.

We would like to thank the Science Foundation for Physics within the University of Sydney for computing facilities provided in the Basser Computing Department.

One of us (D. L. J.) would like to thank Professor B. Y. Mills for his hospitality during the writing of this paper.

The Arecibo Ionospheric Observatory is operated by Cornell University with the support of the Advanced Research Projects Agency and the National Science Foundation under contract to the Air Force Office of Scientific Research.

REFERENCES

Annis, M., Cheston, W., and Primakoff, H. 1953, *Rev. Mod. Phys.*, **25**, 818.
Fomalont, E. B. 1968, *Bull. Astr. Inst. Netherlands*, **20**, 69.
Jauncey, D. L. 1967, *Nature*, **216**, 877.
Kendall, M. G., and Stuart, A. 1961, *The Advanced Theory of Statistics*, Vol. **2** (London: Charles Griffin & Co.).
Murdoch, H. S. 1958, *Nucl. Phys.*, **8**, 157.
Refsdal, S. 1969, *Ap. J.*, **155**, 373.