THE ASTROPHYSICAL JOURNAL, 183: 1–13, 1973 July 1


# MAXIMUM-LIKELIHOOD ESTIMATION OF THE NUMBER–FLUX-DENSITY DISTRIBUTION OF RADIO SOURCES IN THE PRESENCE OF NOISE AND CONFUSION

HUGH S. MURDOCH* AND DAVID F. CRAWFORD*
Cornell-Sydney University Astronomy Center

AND

DAVID L. JAUNCEY†
National Astronomy and Ionosphere Center, and
Cornell-Sydney University Astronomy Center
*Received 1972 June 29; revised 1973 February 12*

## ABSTRACT

The application of the method of maximum likelihood (ML) to the determination of the slope of the number–flux-density relationship is extended to include the presence of experimental errors in the flux-density measurements. It is shown that these experimental errors have a significant effect on the number counts at higher ratios of flux density to error than is often recognized.

The case of noise-limited flux-density measurements is treated in some detail, and it is found that, provided the lower limit of a survey is chosen to be at least five times the rms noise, the enhancement in the source density as a function of flux density can be readily calculated. For the case of significant confusion errors in the flux-density measurements the importance of a Monte Carlo approach is emphasized. Several methods that have been used previously are discussed and a number of shortcomings noted.
*Subject heading:* radio sources

## I. INTRODUCTION

In a previous paper (Crawford *et al.* 1970, hereafter referred to as Paper I), we discussed the application of the maximum-likelihood (ML) method to the determination of the exponent of the radio-source-count distribution. There we used the simplifying assumption that the observed flux density $F$ is identical with the true flux density $S$ for all sources, an approximation that is only valid when the flux density of the source is much greater than the experimental errors. In this paper we give a theoretical treatment showing how the ML method can be used in the presence of experimental errors. We show that the effects of the errors are significant at rather higher ratios of flux density to error than is often recognized.

The basic problem is to determine, over some range of flux density, the best estimate of the true flux-density distribution $P(S)dS$ from the observed distribution $P(F)dF$, or rather from a limited though hopefully large number of observations. Here $P(S)$ represents the probability density at $S$ and $P(S)dS$ gives the probability of any given source having *true* flux density in the range $dS$ at $S$. The function $P(F)$ has similar meaning in relation to the *observed* flux density $F$.

In the presence of measurement errors of any type, the distributions $P(S)$ and $P(F)$ are not identical. The observation of an increased number of sources serves to better define $P(F)$ but does not eliminate the difference between $P(F)$ and $P(S)$.

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

† Center for Radiophysics and Space Research, Cornell University, Ithaca, New York. The National Astronomy and Ionosphere Center is operated by Cornell University under contract to the National Science Foundation.

In § II we discuss the error distribution which links the source distributions $P(S)$ and $P(F)$. In § III we discuss the use of the ML method to obtain the best estimate of $P(S)$ from the individual measurements $F_i$ and a knowledge of the error distribution.

The ML method is ideally suited to a model-fitting approach. The most frequently used model is that, over a limited range of flux density, $P(S)$ is given by a power law,

$$P(S)dS = kS^{-(\alpha+1)}dS\,, \tag{1}$$

where $\alpha$ is the usual integral slope. However, any model which can be put in parametric form may readily be used.

In § IV we examine whether it is possible to adopt an approach which is the antithesis of model fitting, namely, a direct correction to the experimental counts. We then give some examples of various approaches which have been made to the problem, most of them falling between the two extremes of model fitting and direct correction.

## II. THE ERROR DISTRIBUTION

The distributions $P(S)dS$ and $P(F)dF$ are linked by the error distribution $P(F|S)dF$ which is the probability that a source of flux density $S$ will be observed with a flux density between $F$ and $F + dF$.

From the nature of probability it follows that

$$P(F) = \int P(F|S)P(S)dS\,. \tag{2}$$

Where the dominant source of error is additive random noise, the error distribution expressed in the form $P(F - S|S)$ is independent of $S$ and may be assumed to be Gaussian. In this case $P(F)$ is the convolution of the error distribution with the flux-density distribution. Noise is the dominant source of error in certain unfilled aperture telescopes such as the 1-mile (1.6 km) Mills Cross of the Molonglo Observatory, or of the Northern Cross at Bologna. One should never neglect, however, a possible contribution including a mean bias from the fitting procedure used to deduce $F$ from the record. This is discussed further at the end of this section.

For most radio telescopes at low frequencies, the dominant source of error in the estimation of the flux density of an individual source is the presence in the beam of other, generally weaker sources, normally referred to as confusion. A typical example of an error distribution which includes significant confusion is given by Pilkington and Scott (1965) for the 4C survey. Here $P(F - S|S)$ is probably a slowly varying function of $S$ and is asymmetric with positive errors predominating. Such an instrument is often said to be *confusion limited*, and measurements made on a single source are subject to a confusion error due to the presence, in the same beam area, of other weaker sources not physically associated.

Ideally one would like a model of the source distribution, with suitable adjustable parameters, which would predict the distribution of total observed flux density per beam area, including not only a dominant source but any weaker sources as well. We discuss in § IV the limited use which has been made of such a method developed by Scheuer (1957). The difficulty in applying such a method is that one needs a model of the source distribution over a wide range of flux densities. We restrict our attention here to the common situation where one wishes to deduce the form of the distribution in the range of measured flux densities, without the need to know anything about the distribution at very low flux densities.

We can then adopt the usual point of view that a strong source is *confused* by any weaker sources present in the beam, and regard the incremental flux density as a confusion error. Weak sources which cannot be individually detected may be said to constitute confusion noise as distinct from the blending of two sources, each of which would be separately detectable if they occurred far enough apart. The total confusion error may be much greater than the confusion noise. Where two sources, each of which would be separately detectable, are blended due to random association, one observes the presence only of a single source; in this case it seems desirable to preserve the distinction that the stronger source is confused by the weaker source, and that the weaker source is obscured by the stronger source. In any multiple blend, the strongest source should be regarded as confused by the weaker ones which are in turn obscured by the strongest source. With the above definition of confusion error, the probability $P(F|S)dF$ refers to the probability that a source $S$ will be observed in the range $F$ to $F + dF$, *and* that it is the strongest source contributing to the observed flux density $F$. If the latter condition is not fulfilled, then $P(F|S) = 0$, since the source $S$ is obscured by one or more stronger sources.

The fitting procedure for determining $F$ will normally assess the local background or base-line level as the mean of receiver noise plus confusion noise. The combined error due to noise, confusion, gain fluctuations, and any error due to the fitting process, may be either positive or negative. The preponderance of positive errors in the case of confusion comes from sources which are strong enough to be detected if occurring singly, but which happen in fact to be obscured by stronger sources.

For a digitally recorded survey it is very convenient to use the Monte Carlo method to obtain directly the error distribution $P(F|S)$. Simulated sources can be added to the record at random positions and then analyzed as though they were real sources. Any error due to the fitting procedure, including any mean over- or underestimation is therefore included as well as noise and confusion errors, but errors due to calibration and to gain fluctuations are not included. The method also gives the probability of a source being obscured by a stronger source. This situation is recognized where a randomly inserted Monte Carlo source happens to occur at such a position on the record as to be blended with a source stronger than itself which already exists on the record. Whatever the source of error, it is necessary to determine the error distribution for positive errors out to quite low levels of probability. The reason for this will become clear in the next section.

## III. THE ML METHOD

### *a) Outline of the Method*

The ML method for determining $P(S)$ was first introduced by Jauncey (1967), and its use in both grouped and ungrouped form was further discussed in Paper I. Since nothing is gained by grouping the data, we will discuss here only the ungrouped form although the method can readily be adapted for grouped data.

The method consists of maximizing the probability of occurrence of the observed values of $F_i$, given any assumed form of the distribution. Following Paper I, this reduces to maximizing

$$L = \prod_{i=1}^{M} P(F_i) . \tag{3}$$

It is usually more convenient to maximize $\mathscr{L} = \ln L$ given here by

$$\mathscr{L} = \sum_{i=1}^{M} \ln P(F_i) , \tag{4}$$

where the summation is over all the observed flux densities. The treatment in Paper I covers the case $F_i \equiv S_i$ for all $S_i$.

Here $P(F_i)$ is given from equation (2) by

$$P(F_i) = N\int_0^\infty P(F|S)P(S)dS\,, \tag{5}$$

where $N$ is a normalizing factor given by

$$N^{-1} = \int_{F_0}^{F_u}\int_0^\infty P(F|S)P(S)dSdF\,, \tag{6}$$

and where $F_0$ and $F_u$ are the *predetermined*[1] lower and upper limits of observed flux density. Where no upper limit is set, $F_u = \infty$. For a source distribution of the form of equation (1),

$$P(F_i) = N\int_0^\infty P(F|S)S^{-(\alpha+1)}dS\,. \tag{7}$$

In practice the lower limit of integration of $S$ will be not zero but some chosen value $S_0$ for $S$, for which the integrand is close to zero. This is discussed in § III.

In the error-free case discussed in Paper I it was possible to obtain an analytic solution for the ML estimate $a$ of $\alpha$, and also obtain the exact sampling distribution and hence the standard deviation $\sigma_a$ of $a$. In general, where $P(F|S)$ is not known analytically, it is not possible to obtain an exact solution either for $a$ or for its sampling distribution. However, it is possible in such cases to form the likelihood distribution given by $L$ as a function of $\alpha$. This is not itself a probability distribution but has the well-known property that for large $M$, it approaches the sampling distribution which is asymptotically Gaussian. From the properties of the Gaussian distribution it follows that, to obtain the standard deviation of $a$, it is merely necessary to note the values $a_1$ and $a_2$ of $\alpha$ for which $\mathscr{L}$ is 0.5 less than its maximum value. The standard deviation of $a$ is then given by $\sigma_a = a_1 - a \simeq a - a_2$. The concept of the likelihood distribution asymptotically approaching the sampling distribution can readily be extended to cases where more than one parameter is fitted.

Having obtained the ML estimation $a$ of $\alpha$ from equation (4) it is necessary to test the goodness of fit to equation (1). To do this the transformation

$$Y_i = \int_{F_0}^{F_i} P(F)dF \tag{8}$$

may be used and, as in Paper I, the $Y_i$ tested for uniformity in the range 0 to 1 by the Kolmogorov test (see, e.g., Kendall and Stuart 1961). This does not require grouping of the data as does the $\chi^2$ test.

If the data are not a good fit to equation (1), then a more complicated form of distribution must be used. It is often very useful, as a guide to the correct form, to obtain estimates over several intermediate ranges of flux density. In using these as a guide for a more complex expression, it is necessary not only to take account of the separate values of $a$, but also to ensure the correct normalization where various ranges of flux density join.

Sometimes observations are made over a wide area of sky for strong sources and over a restricted region for weaker sources. In some such cases it may be desirable to fit a value of $\alpha$ separately for each region in the first instance. It is very easy, however,

[1] Our use of the symbol $S_m$ for the predetermined upper limit in Paper I was confusing since in equation (4) of that paper it could have been interpreted as the flux density of the strongest source.

to test the general applicability of any assumed form of distribution over the whole range of $S$.

If $F_{0j}$ and $F_{uj}$ are the designated upper and lower limits of flux density for a region of sky whose area (in steradians) is $A_j$, then the probability density $P(F_{ij})$ of observing a source with flux density $F_i$ in area $j$ is given by

$$P(F_{ij}) = NA_j \int_0^\infty P(F|S)P(S)dS \tag{9}$$

with

$$N^{-1} = \sum_j \int_{F_{0j}}^{F_{uj}} A_j \int_0^\infty P(F|S)P(S)dSdF\,. \tag{10}$$

The best fit maximizes $\sum_i \sum_j \ln P(F_{ij})$. In applying the goodness of fit test, equation (8) becomes

$$Y_i = \sum_j \int_{F_{0j}}^{F_i} P(F_{ij})dF_i\,. \tag{11}$$

In this case the use of the relative area factors $A_j$ automatically ensure the correct normalization of the combined result for the whole area. The derivation of the ML estimate for the error-free case is given in the Appendix.

### *b*) *Application to Gaussian Noise*

We will now discuss in more detail the application of the method in the simple situation where the error distribution, $P(F|S)$, is Gaussian of standard deviation $\sigma$. This should be approximately the case in a noise-dominated situation. Writing $\psi(S, F)$ for the product $P(F|S)P(S)$ in equation (5), we have

$$\psi(S, F) = \frac{1}{(2\pi)^{1/2}\sigma} S^{-(\alpha+1)} \exp\left(\frac{-(F-S)^2}{2\sigma^2}\right)\,. \tag{12}$$

For convenience we shall express $S$ and $F$ in the dimensionless ratios $s = S/\sigma$, and $r = F/\sigma$, and in order to indicate the form of this distribution we have plotted in figure 1 for $\alpha = 1.5$,

$$\phi(s, r) = (s/r)^{-(\alpha+1)} \exp\left[-\tfrac{1}{2}(s-r)^2\right] \tag{13}$$

as a function of $s/r(= S/F)$. Note that large positive errors are expressed by small values of $s/r$. For example, the $s/r = 0.5$ point on the $r = 5$ curve represents an error of 2.5 $\sigma$, and in this case $\phi$ is by no means negligible. For comparison the Gaussian curve for $r = 5$ is also drawn.

The $\phi$ curves represent the relative probability of a given ratio of true-to-observed flux density for a given signal-to-noise ratio $r$ (Jauncey 1968; Jauncey *et al.*, in preparation). For $r < 5$ it is readily apparent that there is appreciable probability of a given observed flux density being chiefly due to noise. The accuracy of the curves at low $s/r$ depends on the validity of the assumption that the noise error is Gaussian out to many standard deviations. It also depends on the assumed form of the source distribution $P(S)$ being valid for $S \ll F$.

For $r < 5$ there is an appreciable contribution to the integral $P(F_i)$ from the uncertain region of low $s/F$. It is in fact necessary to assume some modification to $P(F_i)$ at small $s/F$, for example a cutoff at $t_0 = S_0/F$, in order to get a finite answer. If the

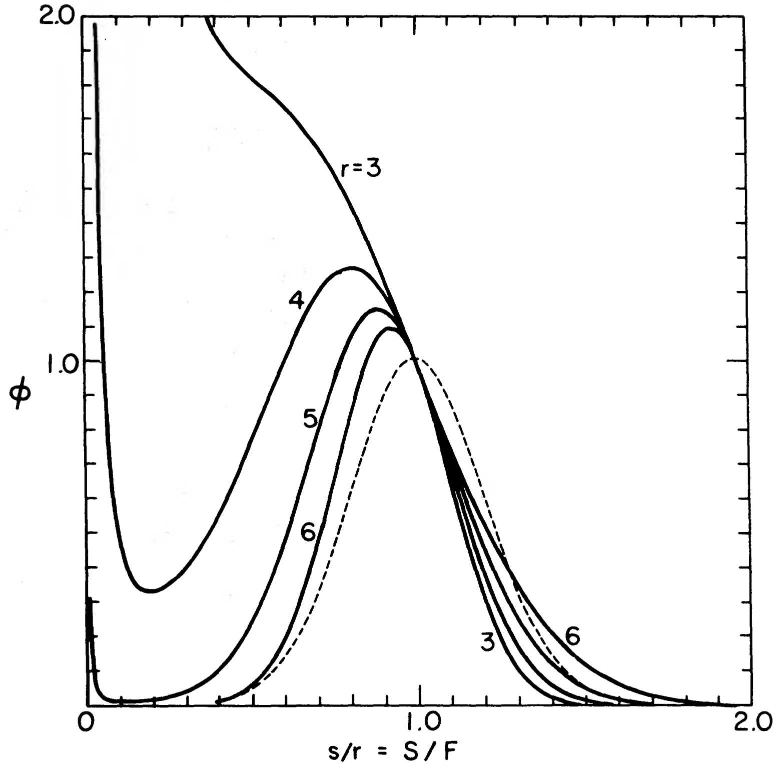


FIG. 1.—The function $\phi(s, r)$ (eq. [13]) is plotted as a function of the ratio of true to observed flux density for several values of the ratio $r$ of observed flux density to rms error. Dashed curve is the corresponding Gaussian error distribution for $r = 5$.

deduced value of $\alpha$ is to be essentially independent of $P(S)$ at very small $s$, then the assumed cutoff must have a negligible effect on the integrals $P(F_i)$.

In table 1 we present, for $\alpha = 1.5$, the value of the ratio $R$ of the number of sources observed with flux density $F$ to the number of sources which in fact have this value of flux density. $R$ is given by

$$R = F^{(\alpha+1)} \int_{S_0}^{\infty} \psi(S, F) dS . \tag{14}$$

For $r$ less than 5, the value of $R$ is very sensitive to $t_0$. On the other hand, for $r$ greater than 5, it is virtually independent of $t_0$ within fairly wide limits, provided one

TABLE 1

$R$ AS A FUNCTION OF $r$ AND $t_0$ FOR $\alpha = 1.5$

| | $t_0$ | | |
|---|---|---|---|
| $r$ | 0.05 | 0.1 | 0.2 |
| 3........ | 7.200 | 4.445 | 3.038 |
| 3.5...... | 3.500 | 2.523 | 2.087 |
| 4........ | 1.901 | 1.684 | 1.584 |
| 5........ | 1.253 | 1.249 | 1.247 |
| 6........ | 1.149 | 1.149 | 1.149 |

TABLE 2

$R$ AS A FUNCTION OF $r$ AND $\alpha$ (for $t_0 = 0.1$)

| $r$ | $\alpha$ = 1.0 | 1.2 | 1.4 | 1.5 | 1.6 | 1.8 | 2.0 |
|---|---|---|---|---|---|---|---|
| 5. . . . . . . . . . . . . | 1.157 | 1.189 | 1.228 | 1.249 | 1.272 | 1.325 | 1.390 |
| 6. . . . . . . . . . . . . | 1.099 | 1.117 | 1.138 | 1.149 | 1.161 | 1.187 | 1.215 |
| 7. . . . . . . . . . . . . | 1.068 | 1.081 | 1.095 | 1.103 | 1.111 | 1.128 | 1.146 |
| 8. . . . . . . . . . . . . | 1.047 | 1.060 | 1.071 | 1.076 | 1.082 | 1.094 | 1.107 |
| 10. . . . . . . . . . . . | 1.032 | 1.037 | 1.043 | 1.047 | 1.050 | 1.057 | 1.065 |
| 12. . . . . . . . . . . . | 1.022 | 1.025 | 1.030 | 1.032 | 1.034 | 1.039 | 1.044 |
| 15. . . . . . . . . . . . | 1.014 | 1.016 | 1.019 | 1.020 | 1.021 | 1.024 | 1.028 |
| 20. . . . . . . . . . . . | 1.008 | 1.009 | 1.010 | 1.011 | 1.012 | 1.014 | 1.015 |
| 30. . . . . . . . . . . . | 1.003 | 1.004 | 1.005 | 1.005 | 1.005 | 1.006 | 1.007 |

ignores the apparent divergence as $s$ approaches zero. The justification for ignoring this region is given at the end of § III*c*. In table 2 we present values of the correction factor $R$ as a function of signal-to-noise $r$ for $r \geq 5$ and for various values of $\alpha$. In this region $R$ is virtually independent of $t_0$ and the figures may be of practical utility in making corrections. In table 3, we present corresponding ratios of the total count above a flux-density limit $F_0$ to the corresponding count above the same limiting flux density in the absence of errors. Here $r_0$ is the signal-to-noise ratio at the lower limit; $F_0 = r_0\sigma$.

From the values of $R$ as a function of $r$ in tables 1 and 2 we may calculate the effect on the estimation of $\alpha$ of completely ignoring the errors in the estimation of the individual flux densities $F_i$. To do this for any given $\alpha$, we apply the ML method of Paper I to $P(S)$ as defined by equation (5) with Gaussian $P(F|S)$:

$$\frac{1}{a'} = \int_{F_0}^{\infty} \ln (F/F_0) P(F) dF . \tag{15}$$

The results are presented in tables 4 and 5 where the deduced value $a'$ is plotted as a function of $r_0$ and $\alpha$. The deduced estimate $a'$ of $\alpha$ depends significantly on $t_0$ for $r_0$ less than 5, while for $r_0$ greater than 5 it is essentially independent of $t_0$. In table 5 we present values of $a'$ as a function of $\alpha$ for $r_0 \geq 5$.

TABLE 3

RATIO OF TOTAL COUNT TO ERROR-FREE TOTAL COUNT

| $r_0$ | $\alpha$ = 1.0 | 1.2 | 1.4 | 1.5 | 1.6 | 1.8 | 2.0 |
|---|---|---|---|---|---|---|---|
| 5. . . . . . . . . . . . . | 1.043 | 1.062 | 1.081 | 1.092 | 1.103 | 1.128 | 1.157 |
| 6. . . . . . . . . . . . . | 1.031 | 1.041 | 1.053 | 1.060 | 1.066 | 1.081 | 1.098 |
| 7. . . . . . . . . . . . . | 1.022 | 1.029 | 1.037 | 1.042 | 1.047 | 1.057 | 1.069 |
| 8. . . . . . . . . . . . . | 1.016 | 1.022 | 1.028 | 1.031 | 1.035 | 1.043 | 1.051 |
| 10. . . . . . . . . . . . | 1.010 | 1.014 | 1.017 | 1.020 | 1.022 | 1.026 | 1.032 |
| 12. . . . . . . . . . . . | 1.007 | 1.009 | 1.012 | 1.013 | 1.015 | 1.018 | 1.022 |
| 15. . . . . . . . . . . . | 1.005 | 1.006 | 1.008 | 1.008 | 1.009 | 1.011 | 1.014 |
| 20. . . . . . . . . . . . | 1.003 | 1.003 | 1.004 | 1.005 | 1.005 | 1.006 | 1.008 |
| 30. . . . . . . . . . . . | 1.001 | 1.001 | 1.002 | 1.002 | 1.002 | 1.003 | 1.003 |

TABLE 4

FITTED VALUE $a'$ FOR $\alpha = 1.5$ (ignoring errors)

| $r_0$ | $t_0$ 0.05 | 0.1 | 0.2 |
|---|---|---|---|
| 3........ | 2.734 | 2.242 | 2.010 |
| 3.5...... | 1.974 | 1.838 | 1.775 |
| 4........ | 1.697 | 1.672 | 1.661 |
| 5........ | 1.580 | 1.580 | 1.580 |

The above results show that provided the lower limit of a survey is chosen to be at least five times the rms noise, one may readily calculate for any given $\alpha$: (*a*) the enhancement in the observed source density as a function of flux density, and (*b*) the error to be expected in the deduced estimate of $\alpha$ if noise errors in the measurement are ignored.

One may also use the ML method to obtain the best estimate of $\alpha$. On the other hand, if one attempts to carry the counts from a survey to lower values of flux density, the calculations are meaningless in the absence of precise knowledge of *both* the error distribution *and* the form of the source distribution at very much lower flux densities than those observed.

The results of tables 1 and 2 also show that if the lower limit of the survey is set at five times the rms error, then the enhancement in rate due to noise errors at the lower limit, and the effect of the latter on the deduced estimate for $\alpha$, are both quite substantial. Only at much higher flux densities is it valid to ignore the errors altogether. An $r_0 \sim 5$ represents a fairly sharp transition from a region where the errors still have a significant effect but may be readily taken into account, to a region where it is pointless to try.

Extending the counts from a given survey below this limit greatly increases the source numbers and hence reduces the sampling errors. This reduction, however, is offset by the increased uncertainty in accounting for the measurement errors, and by the need to know the form of $P(S)$ at flux densities much below the survey limits. This results in the curious situation whereby the addition of sources of low $r$ may serve to increase rather than decrease the uncertainty in determining $P(S)$.

### *c) The Case Where Confusion Error Is Significant*

In the case of confusion, the error distribution cannot be treated analytically as readily as can Gaussian noise. The error distribution $P(F|S)$ should be obtained by

TABLE 5

VALUES OF $a'$ AS A FUNCTION OF $\alpha$ AND $r_0$ ($t_0 = 0.1$)

| $r_0$ | $\alpha$ 1.0 | 1.2 | 1.4 | 1.5 | 1.6 | 1.8 | 2.0 |
|---|---|---|---|---|---|---|---|
| 5.............. | 1.030 | 1.248 | 1.468 | 1.580 | 1.693 | 1.924 | 2.162 |
| 6.............. | 1.021 | 1.231 | 1.444 | 1.551 | 1.659 | 1.878 | 2.099 |
| 7.............. | 1.015 | 1.222 | 1.431 | 1.536 | 1.642 | 1.854 | 2.069 |
| 8.............. | 1.011 | 1.216 | 1.423 | 1.527 | 1.631 | 1.840 | 2.051 |
| 10............. | 1.007 | 1.210 | 1.414 | 1.517 | 1.619 | 1.825 | 2.032 |
| 20............. | 1.002 | 1.202 | 1.403 | 1.504 | 1.605 | 1.806 | 2.008 |
| 30............. | 1.001 | 1.201 | 1.402 | 1.502 | 1.602 | 1.803 | 2.003 |

the Monte Carlo method discussed in § II. The error $F - S$ is not necessarily independent of $S$ but may be a slowly varying function of $S$, as can be seen from the error histograms of Pilkington and Scott (1965). Since a range of values of $S$ corresponds to any given observed $F$, the points on each $\psi(S, F)$ curve will come from different values of $S$ and hence from different $P(F|S)$ curves. Provided $P(F-S|S)$ varies slowly with $S$, it is sufficient in practice to establish the function at several values of $S$ by the use of Monte Carlo sources of those particular values of $S$. The curve for intermediate values can be obtained by interpolation.

As the error distribution is asymmetric toward large errors, the $\phi$ curves will be relatively larger for $s$ less than 1 than for the case of symmetric Gaussian errors. The lower limit of flux density will need to be chosen by an actual plot of the $\phi$ curves. This will to some extent be influenced by the need to use a large number of Monte Carlo sources to determine with sufficient precision the upper tail of the error distribution.

There is a further problem with confusion which is best illustrated by means of an example. Suppose a 1.0 flux unit inserted Monte Carlo source combines with a genuine source of 1.3 f.u. to produce a flux density of 2.3 f.u. The straightforward interpretation is that the 1.0-f.u. source is obscured by the 1.3-f.u. source. Another possibility is that the 1.3-f.u. source is itself a blend of two or more sources, each less than 1.0 f.u. In this case the correct interpretation is that the 1.0-f.u. source is subject to a confusion error of 1.3 f.u. This is much less likely at flux levels where the mean number of beam areas per source is large, but the uncertainty can be a problem at low flux-density levels. In any case it can only arise for $s$ less than 0.5. Hence, an additional safety criterion is that the lower limit of flux density should be chosen so that $\phi$ is very small at $s = 0.5$. If this criterion were applied to Gaussian noise, it would lead to $r_0$ of a little greater than 5 which is very close to the limit suggested in § III*b* using quite a different criterion. Any analysis to lower levels than suggested here requires a detailed treatment of multiple blending.

It should be noted that at $S$ levels for which blending becomes important, $P(F|S)$ and hence $\psi(F, S)$ are sharply reduced since $P(F|S)$ is the probability that a source $S$ will appear as a source $F$ *without* any contribution to the blend by a source stronger than $S$. Hence for *very* small values of $s$, $P(F|S)$ and $\psi(F, S)$ approach zero. Even for a noise-limited instrument, blending will occur for sufficiently small values of $s$, since the total number of beam areas is finite. The blending of sources in a finite beam is ignored in the calculation of the curves of figure 1. This is valid in a noise-dominated situation except at very low $s/r$ where multiple blending occurs, and hence all the $\phi$ and $\psi$ curves approach zero as $s$ approaches zero. Thus irrespective of whether the dominant source of error is noise or confusion, the apparent divergence of $\phi$ as $s$ approaches zero may be ignored.

The above analysis shows that, for any reliable attempt at estimating the number–flux-density relationship from a radio survey, the lowest flux-density level considered should be no less than five times the rms flux-density error. Even then it is important to realize that the effects at the lowest flux-density levels of the survey are still quite significant. While the number–flux-density relationship may be deduced from the observations, the survey is *complete* at the lower limit only in the statistical sense. Near the limit of the catalog many of the *individual* sources will in fact be weaker than the survey limit, and similarly, others which are above the limit will have been missed. This should be particularly borne in mind if the sources are subdivided into different categories, as is done, for example, in a division on the basis of optical identifications.

## IV. OTHER METHODS

We have shown that the ML method is ideally suited to a model-fitting approach to the solution of the basic problem. In many circumstances one would like to be able to *correct* the measured distribution directly for the effect of measurement errors, in order to obtain the true distribution in an empirical form rather than as an analytic expression with one or more fitted parameters. This would have the great advantage that no assumptions would need to be made in advance about the true distribution.

Eddington (1913), in discussing the distribution of stellar magnitudes, considered the general question of whether the true distribution can be obtained analytically from the observed distribution and the error function. He obtained, for a Gaussian error distribution, an expression for the true distribution in the form of a series expansion involving the successive even-numbered derivatives of the observed distribution. This requires a high order of precision in smoothing the observed distribution, and hence it is impractical. For the form of $P(S)$ considered here, there is the additional problem that the series begins to diverge after the first few terms.

In a later paper (1940) Eddington gave an expression for the mean overestimation of individual sources which involves only the observed distribution and its first derivative. One might be tempted to think that this method could be used to correct the individual values for mean overestimation and then deduce the source distribution empirically from the *corrected* values. Eddington (1940) shows, with the help of an example, that this procedure cannot be expected to lead to the true source distribution.

To obtain the true source-count distribution from the observed results, it is necessary to assume something about that distribution. This assumption will either be about the form of distribution (but with one or more parameters to be fitted as in the ML method), or it may be an approximate assumption about the distribution itself in order to obtain a first-order correction to the observed distribution. To illustrate these approaches we will briefly discuss some methods which have been used in the past.

Mills and Slee (1957) calculated, for an assumed true distribution of constant $\alpha = 1.5$, the fraction of sources in any given flux-density interval which would be carried to an adjacent interval due to noise errors. They also calculated the effect of blending to several orders of multiple blending. This gave a percentage correction as a function of flux density to be applied to the true distribution in order to predict the observed distribution. Strictly, this procedure should be iterated until the predicted distribution leads to the observed distribution after allowing for errors. Theirs is a model-fitting approach similar to that discussed here, and remains one of the most thorough treatments of the subject.

Bennett (1962) calculated the correction to be applied to the observed rate as a function of $S$ and $\alpha$ by obtaining a series expansion in terms of the moments of the error distribution for an expression of the form of equation (2). His error distribution was for a phase-switched interferometer, but a similar expression can be derived for the general situation. Bennett considered only the first few terms of the expansion, but if more terms are calculated the series diverges, as does the Eddington series. The series-expansion method cannot circumvent the problem raised in § III*b* of rapid rise and uncertain value of the rate for $r$ less than 5. Indeed, it is only under the conditions for which the rate is essentially independent of $t_0$ that the series-expansion method is valid. The use of Bennett's method must therefore be treated with considerable caution, and it certainly cannot improve on the results obtained here. Harris and Kraus (1970) claim to have used a method similar to Bennett's to obtain corrected source counts at low flux levels for the Ohio survey. The corrections were carried to flux-density levels less than twice the rms confusion error and less than three times the confusion noise as defined in § II. No details of the calculation are given, but it is difficult to see how they could have failed to encounter the divergence problem discussed here. The effect of random errors appears to have been more than counteracted by the mean underestimation of weak sources pointed out by Jauncey and Niell (1971), and subsequently acknowledged for much of the survey (Harris and Kraus 1971). It is by no means clear how the uncorrected counts presented by Kraus (1972) are affected by random and systematic errors, but uncorrected counts at flux densities down to about 1 s.e. cannot be taken seriously.

It should also be noted that Bennett's correction is an approximation, because it is necessary to assume a value of $\alpha$ to obtain the correction. Gower (1966) used a Monte Carlo technique to obtain an approximate correction. The inserted sources were distributed at discrete flux densities but with a smoothed integral distribution that was similar to the observed integral distribution. The distribution of *observed* flux density for the inserted Monte Carlo sources (obtained by analyzing them as for real sources) was compared with their *true* distribution (i.e., the distribution of inserted values). A percentage correction as a function of flux density was therefore obtained. This too is an approximate *direct correction*, not an exact solution. Its weakness lies in starting with the assumption that the true distribution is like the observed distribution. Ideally, this process should be iterated until the assumed true distribution of Monte Carlo sources in fact leads to the actual observed distribution. There is unfortunately no guarantee that the process will converge in this way. There are other criticisms which can be made of Gower's calculations. For example, it would have been preferable to allow for multiple blending by adding Monte Carlo sources down to somewhat less than 0.625 times the minimum flux density of the survey and also at more closely spaced intervals of flux density. It should be noted that Gower's Monte Carlo treatment is quite different from the Monte Carlo treatment advocated here to obtain the error distribution. His method requires the insertion of a whole range of flux densities simulating the source distribution. In the present method a large number of sources of identical $S$ are inserted at each of several well-chosen values of $S$ in order to obtain the error distribution for those values of $S$.

The best overall compromise seems to us to be to obtain an approximate solution first, and then attempt to put the solution in a parametric form. Having done this, one can then take full advantage of the ML method to obtain the optimum value of the parameters. The chosen form should be no more complicated than necessary to give a good fit to the data.

The methods we have just been discussing were all designed to obtain an approximate correction in the region where the flux density is still significantly larger than the rms error. Scheuer (1957) adopted a somewhat different approach. He calculated, for an assumed distribution of constant $\alpha = 1.5$, the probability distribution $P(D)$ of deflections $D$ at arbitrary points on the record, a method that is particularly suitable for a phase-switched interferometer. He did not allow for noise, but the method was later extended by Hewish (1961) to include noise. The method was specifically designed to find out something about the form of the distribution at levels of flux density below which sources can be individually measured and counted. Hewish was able to set certain broad limits on the form of the distribution at low $S$. The best use of this method is also in conjunction with ML since

$$\mathscr{L} = \sum_i \ln P(D_i)\,. \tag{16}$$

The chief disadvantage of all the methods discussed here for calculating the correction to the observed distribution is that they take too simplistic a view of radio sources. It is usual to insert only Monte Carlo "point" sources. This will generally give an underestimate of confusion error. Ideally, one should insert a set of sources having the correct distribution in structure as well as in flux density. Unfortunately this is generally unknown for the weak sources contributing to confusion.

This also raises the question of randomness of sources. If there is any clustering of sources, insertion of Monte Carlo sources at random will also give an underestimate of confusion error. One must also choose some degree of resolution at which two sources are regarded as separate. It is not always easy to tell whether there is a genuine physical association of two close or partially resolved sources. These disadvantages are common to any method which to our knowledge has been used so far.

Ideally, one needs a more sophisticated model of radio sources, their distribution in space, luminosity, physical size, and structure, from which one can predict all the way to the observed flux densities or to the observed deflections on the record. This seems a fair way off yet, but may be necessary before radio-source counts can make any *definitive* contribution to cosmology. In the meantime it is important to make best use of the available data. We have shown that this requires a careful consideration of the interaction of the source-count distribution with the error distribution, especially near the lower limit of most surveys.

## V. CONCLUSIONS

In Paper I we demonstrated that the ML method is superior to any other method of model fitting to obtain the slope of the source-count distribution. We restricted the treatment there to an idealized error-free situation. We also showed that, contrary to previous treatments, there is no need to group the data.

The main conclusions of the present paper are: (1) The ML method can readily be extended to take account of errors in the recorded flux densities due to noise and confusion. (2) It has long been recognized by cautious workers in this field that surveys and source counts should be restricted to flux densities *at least* five times the rms error. We have shown here that there is a sound theoretical basis for this restriction. Below this level it is virtually impossible to make valid corrections to the observed counts because of the high probability of an observed source being spurious (i.e., boosted by noise or confusion error from a very low value of flux density). (3) To obtain a reliable estimate of the true source-count distribution, a detailed knowledge of the error distribution is necessary, including any systematic bias and also the non-Gaussian tail in the confusion-error distribution. The latter is crucial in a confusion-limited survey and should be obtained by Monte Carlo techniques.

For a noise-limited survey free of systematic bias, a Gaussian error distribution is a good approximation. We have treated this case analytically and have given numerical tables for the correction factors. These show that the corrections can still be appreciable at flux density levels much greater than 5 $\sigma$. These results should be applicable to any noise-limited survey. It is not possible to give a general treatment for confusion-limited surveys, because of the specific nature of the confusion-error distribution for any given instrument.

We would like to thank the Science Foundation for Physics within the University of Sydney for facilities provided in the Basser Computing Department. H. S. M. would like to thank the National Astronomy and Ionosphere Center, Arecibo, and the Center for Radiophysics and Space Research, Cornell University, for hospitality.

## APPENDIX

We consider here the analytical treatment of the multiarea case in an error-free situation. Suppose one decides to accept sources in area $A_j$ with flux densities between $S_{0j}$ and $S_{uj}$, then the probability density $P(S_{ij})$ of a source of flux density $S_{ij}$ in area $j$ is given by

$$P(S_{ij}) = \alpha K^{-1} A_j S_{ij}^{-(\alpha+1)},$$

where

$$K = \sum_j A_j(S_{0j}^{-\alpha} - S_{uj}^{-\alpha}).$$

Hence

$$\mathscr{L} = M \ln \alpha - M \ln K + \sum_j m_j \ln A_j - (\alpha + 1) \sum_i \sum_j \ln S_{ij} ,$$

where $m_j$ = the number of sources observed in area $j$ and

$$M = \sum_j m_j .$$

We define the derivatives

$$\dot{K} = \frac{\partial K}{\partial \alpha} = -\sum_j A_j(S_{0j}{}^{-\alpha} \ln S_{0j} - S_{uj}{}^{-\alpha} \ln S_{uj}) ,$$

$$\ddot{K} = \frac{\partial^2 K}{\partial \alpha^2} = \sum_j A_j[S_{0j}{}^{-\alpha}(\ln S_{0j})^2 - S_{uj}{}^{-\alpha}(\ln S_{uj})^2] .$$

Then the maximum-likelihood value for $\alpha$ is obtained by solving

$$\frac{\partial \mathscr{L}}{\partial \alpha} = 0 = \frac{M}{\alpha} - \frac{M\dot{K}}{K} - \sum_i \sum_j \ln S_{ij} .$$

The standard deviation of this estimate is given, in the limit of large numbers, by the relation

$$\sigma^{-2} = -\langle \partial^2 \mathscr{L} / \partial \alpha^2 \rangle .$$

Or

$$\sigma^2 = \alpha^2 M^{-1}\left[1 - \frac{\alpha_2}{K^2}(\dot{K}^2 - K\ddot{K})\right]^{-1} ,$$

where, to sufficient approximation, we may replace $\alpha$ by its estimated value.

## REFERENCES

Bennett, J. G. 1962, *M.N.R.A.S.*, **125**, 75.
Crawford, D. F., Jauncey, D. L., and Murdoch, H. S. 1970, *Ap. J.*, **162**, 405.
Eddington, A. S. 1913, *M.N.R.A.S.*, **73**, 359.
———. 1940, *ibid.*, **100**, 354.
Gower, J. F. R. 1966, *M.N.R.A.S.*, **133**, 151.
Harris, B. J., and Kraus, J. D. 1970, *Nature*, **227**, 785.
———. 1971, *ibid.*, **230**, 140.
Hewish, A. 1961, *M.N.R.A.S.*, **123**, 167.
Jauncey, D. L. 1967, *Nature*, **216**, 877.
———. 1968, *Ap. J.*, **152**, 647.
Jauncey, D. L., Murdoch, H. S., and Crawford, D. F. 1973, in preparation.
Jauncey, D. L., and Niell, A. E. 1971, *Nature*, **229**, 223.
Kendall, M. G., and Stuart, A. 1961, *The Advanced Theory of Statistics*, Vol. **2** (London: Charles Griffin and Co.).
Kraus, J. D. 1972, *Nature*, **236**, 4.
Mills, B. Y., and Slee, O. B. 1957, *Australian J. Phys.*, **10**, 162.
Pilkington, J. D. H., and Scott, P. E. 1965, *M.N.R.A.S.*, **69**, 183.
Scheur, P. A. G. 1957, *Proc. Camb. Phil. Soc.*, **53**, 764.