An iterative technique for the rectification of observed distributions

L. B. Lucy*

Departments of Physics and Astronomy, The University of Pittsburgh, Pittsburgh, Pennsylvania 15213

(Received 15 January 1974; revised 26 March 1974)

An iterative technique is described for generating estimates to the solutions of rectification and deconvolution problems in statistical astronomy. The technique, which derives from Bayes' theorem on conditional probabilities, conserves the constraints on frequency distributions (i.e., normalization and non-negativeness) and, at each iteration, increases the likelihood of the observed sample. The behavior of the technique is explored by applying it to problems whose solutions are known in the limit of infinite sample size, and excellent results are obtained after a few iterations. The astronomical use of the technique is illustrated by applying it to the problem of rectifying distributions of $v \sin i$ for aspect effect; calculations are also reported illustrating the technique's possible use for correcting radio–astronomical observations for beam-smoothing. Application to the problem of obtaining unbiased, smoothed histograms is also suggested.

INTRODUCTION

A fundamental problem in statistical astronomy is that of *estimating* the frequency distribution $\psi(\xi')$ of a quantity ξ' when the available measures $x_1', x_2', \ldots, x_N'$ are a *finite* sample drawn from an infinite population characterized *not* by $\psi(\xi')$ but by

$$\phi(x)=\int \psi(\xi)P(x\,|\,\xi)d\xi, \tag{1}$$

where $P(x|\xi)dx$ is the probability (presumed known) that x' will fall in the interval $(x, x+dx)$ when it is known that $\xi'=\xi$. Equation (1) is an integral equation of the first kind with the conditional probability density $P(x|\xi)$ as kernel.

The classic example of this problem is that of correcting an observed distribution $\phi(x)$ for the effect of observational errors (Eddington 1913). If these errors follow a normal (Gaussian) distribution with variance σ^2, then

$$P(x\,|\,\xi)=\frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\exp\left\{-\frac{(x-\xi)^2}{2\sigma^2}\right\}, \tag{2}$$

and the right-hand side of Eq. (1) becomes a convolution integral.

Another example is the problem of correcting the distribution of stellar rotation velocities $x=v \sin i$ for aspect effect in order to derive the distribution of equatorial velocities $\xi=v$. If we assume that the stars' rotation axes are orientated randomly, then

$$P(x\,|\,\xi)=\frac{x}{\xi}(\xi^2-x^2)^{-\frac{1}{2}}H(\xi-x), \tag{3}$$

where H is Heaviside's unit function [$H(y)=1$ for $y\geq 0$ and $=0$ for $y<0$]. In this case, Eq. (1) reduces to Abel's integral equation (see e.g., Chandrasekhar and Münch 1950).

Among the many other problems of essentially the same character are the following: (1) the determination of the variation of star density along a line-of-sight from the distribution of proper motions in that direction; (2) the determination of the space distribution of radio sources from number counts; (3) the determination of the radial variation of star density in a globular cluster from star counts; (4) the correction of radio–astronomical and spectrographic observations for the effect of the instrumental profile; and (5) the determination of the temperature stratification in the solar atmosphere from limb-darkening data. These examples suffice to show that the problem under consideration arises in many branches of astronomy and that its solution is vital to the process of extracting useful information from observations.

* On leave of absence from the Department of Astronomy, Columbia Univerity.

I. STATISTICAL METHODS

If nothing is known about $\psi(\xi)$, other than that it obeys the constraints

$$\int \psi(\xi)d\xi=1 \quad \text{and} \quad \psi(\xi)\geq 0 \tag{4}$$

that apply to all frequency distributions, then the most obvious method for estimating ψ is by direct numerical solution of the integral Eq. (1). To do this, we first approximate the integral in Eq. (1) by a summation to obtain

$$\phi_i=\sum_{j=1}^{J} P_{ij}\psi_j, \tag{5}$$

where $\phi_i=\phi(x_i)\Delta x$, $\psi_j=\psi(\xi_j)\Delta\xi$, and $P_{ij}=P(x_i|\xi_j)\Delta x$. We then demand that $\phi_i=\tilde{\phi}_i\equiv n(x_i)/N$, the fraction of the observed sample that lies in the interval $(x_i-\frac{1}{2}\Delta x, x_i+\frac{1}{2}\Delta x)$, and thereby obtain the system of linear equations

$$\sum_{j=1}^{J} P_{ij}\psi_j=\tilde{\phi}_i, \quad i=1, 2, \ldots, I, \tag{6}$$

which will, in general, have a unique solution if we set $J=I$.

Unfortunately, unless the sample size N is large, this direct approach yields poor results, with the estimates ψ_j showing little evidence of continuity and often containing violations of the constraint $\psi_j \geq 0$ (cf. Trumpler and Weaver 1953, p. 112–114). We can understand this behavior if we realize that ϕ, being the result of a folding of $\psi(\xi)$ with $P(x|\xi)$, is a smoother function than is ψ; consequently, when we solve for ψ, short-wavelength errors in $\tilde{\phi}$ will be greatly magnified. This effect, applied to the statistical fluctuations in $\tilde{\phi}$, explains the poor results obtained by the direct method.

A way of perhaps avoiding this difficulty would appear to be first to smooth the $\tilde{\phi}_i$ and then to solve the linear Eqs. (6). Any smoothing operation applied to the $\tilde{\phi}_i$ does, however, correspond to a convolution of $\tilde{\phi}_i$ with some appropriate function, and this additional convolution must be allowed for in the rectification procedure if we are to avoid a biased answer. If we do this, however, the original difficulty simply reappears. We must conclude, therefore, that smoothing is not an acceptable solution.

A more promising way of possibly avoiding the difficulty follows from recognizing that an exact solution of Eq. (6) is not required since the $\tilde{\phi}_i$ themselves are not exact. In fact, any vector ψ_j, for which the corresponding ϕ_i is close enough to $\tilde{\phi}_i$ for the differences to be ascribable to sampling errors, is a possible solution (cf. Trumpler and Weaver 1953, p. 114). This remark then suggests that we should recognize that the problem under consideration is basically one of statistical estimation rather than an exercise in solving integral equations.

One estimation procedure readily applied to this problem is that of minimizing χ^2. To do this, we overdetermine the Eqs. (6) by taking $J<I$ and then make the problem determinate by asking for the solution vector ψ_j that minimizes

$$\chi^2 = N \sum_{i=1}^{I} \frac{(\tilde{\phi}_i - \phi_i)^2}{\phi_i}, \tag{7}$$

subject, of course, to the constraints

$$\sum_{j=1}^{J} \psi_j = 1 \quad \text{and} \quad \psi_j \geq 0. \tag{8}$$

Another possibility (Lucy and Ricco 1974) is to determine ψ_j by maximizing the likelihood of the observed sample,

$$L = \phi(x_1')\phi(x_2') \cdots \phi(x_N') \tag{9}$$

with respect to the ψ_j, subject again to the constraints (8). (The quantities $\phi(x_N')$ are, of course, evaluated with the quadrature formula (5), so that L is a function of the ψ_j.)

By not insisting that the estimate ψ_j exactly reproduce the statistical fluctuations in $\tilde{\phi}_i$, these statistical methods are clearly preferable to direct solution of Eqs. (6). However, in minimizing χ^2 or maximizing L, the ψ_j are treated as independent parameters [to the extent allowed by the constraints (8)] with no regard for any expected smoothness and continuity of $\psi(\xi)$. Because of this, the estimate ψ_j may still fail to show the degree of smoothness that we might expect. To overcome this, Lucy and Ricco (1974) added a constraint on the length of the curve defined by the ψ_j and showed that, with a suitable choice of the constrained length, the independence of the ψ_j could be curbed sufficiently for a smooth curve to be obtained. The imposition of this extra constraint does, of course, diminish the likelihood of the final solution; thus, in effect, we accept a smaller likelihood for the benefit of a smoother, and therefore simpler solution. This trade-off is analogous to that commonly made in curve-fitting, where one chooses a curve with few parameters giving a reasonable fit to the data in preference to a many-parameter curve giving a close, or exact fit. A smoothing constraint is also used by Phillips (1962) and Twomey (1963) in their method for solving integral equations in the presence of noise, a method used recently for astronomical problems by Carswell (1973) and Kunasz, Jefferies, and White (1973).

The above statistical methods, when used with a smoothing constraint on the ψ_j, have considerable merit and could be usefully applied to any of the problems mentioned in the Introduction. Their wide application is, however, restricted by their demands on computer time and programming skill. Fortunately, the rather simple iterative technique described below achieves essentially the same ends, provided that no more than a few iterations are made starting from a suitable first guess.

II. ITERATIVE TECHNIQUE

Let $Q(\xi|x)d\xi$ be the ('inverse') probability that ξ' comes from the interval $(\xi, \xi+d\xi)$ when it is known that the measured quantity $x'=x$. The probability that $x' \in (x, x+dx)$ and $\xi' \in (\xi, \xi+d\xi)$ is then $\phi(x)dx \times Q(\xi|x)d\xi$. This, however, is identical to the probability that $\xi' \in (\xi, \xi+d\xi)$ and $x' \in (x, x+dx)$, which is $\psi(\xi)d\xi \times P(x|\xi)dx$. Equating these two expressions and substituting for $\phi(x)$ from Eq. (1), we find that

$$Q(\xi|x) = \frac{\psi(\xi)P(x|\xi)}{\int \psi(\xi)P(x|\xi)d\xi}, \tag{10}$$

which is Bayes' theorem for conditional probabilities. From this theorem and the normalization of the probability $P(x|\xi)dx$, it follows that

$$\psi(\xi) \equiv \int \phi(x)Q(\xi|x)dx, \tag{11}$$

which is an identity having the appearance of being the inverse of the integral Eq. (1) with $Q(\xi|x)$ as the reciprocal kernel. Equation (11) cannot be used to calculate ψ, however, since $Q(\xi|x)$ depends on ψ—the true reciprocal kernel is, of course, a functional of $P(x|\xi)$ only.

Although the identity (11) does not solve the integral Eq. (1), it does suggest the following iterative procedure for generating estimates to ψ: From a guess to ψ and the known function $P(x|\xi)$, we use Eq. (10) to calculate an estimate for $Q(\xi|x)$. Then, taking the hint provided by the identity (11), we integrate this estimate over $\tilde{\phi}(x)$, the approximation to $\phi(x)$ obtained from the observed sample, and thereby generate an 'improved' estimate for $\psi(\xi)$. The procedure is then repeated as often as is necessary or wise. Thus, if ψ^r is the r-th in the sequence of estimates, the $(r+1)$-th estimate is

$$\psi^{r+1}(\xi)=\int \tilde{\phi}(x)Q^r(\xi|x)dx, \tag{12}$$

where

$$Q^r(\xi|x)=\frac{\psi^r(\xi)P(x|\xi)}{\phi^r(x)} \tag{13}$$

with

$$\phi^r(x)=\int \psi^r(\xi)P(x|\xi)d\xi. \tag{14}$$

We may readily show that this iterative scheme conserves the constraints (4). Eliminating $Q^r(\xi|x)$ from Eq. (12), we obtain

$$\psi^{r+1}(\xi)=\psi^r(\xi)\int \frac{\tilde{\phi}(x)}{\phi^r(x)}P(x|\xi)dx, \tag{15}$$

from which it follows that $\psi^{r+1}\geq 0$ if $\psi^0\geq 0$. Proof of the normalization constraint follows from integrating Eq. (12) with respect to ξ and then using the normalizations of the probabilities $Q^r(\xi|x)d\xi$ and $\tilde{\phi}(x)dx$.

Equation (15) shows that the iterative scheme converges if $\phi^r=\tilde{\phi}$. However, if the ψ corresponding to $\phi=\tilde{\phi}$ violates the constraint $\psi\geq 0$, then convergence to this solution is impossible and is, in any case, undesirable. From Eqs. (14) and (15), we see also that deviations of $\tilde{\phi}/\phi^r$ from unity on a length scale large compared to that of $P(x|\xi)$ will be removed in essentially one iteration. On the other hand, deviations on a small length scale will, to a large extent, be averaged out when folded with $P(x|\xi)$ and will result, therefore, in only small corrections to ψ^r. Thus, the scheme is responsive to long wavelength 'errors' in ϕ^r, but unresponsive to those of short wavelength. This is clearly a desirable characteristic of the scheme since the shorter the wavelength of the 'errors' in ϕ^r the more likely it is that they are due to statistical fluctuations in $\tilde{\phi}$. On the basis of these remarks, we may therefore anticipate that, after a few iterations, the scheme will usually have taken account of all significant information in the sample, and that further iterations will result only in small corrections that slowly tend to match the successive estimates ϕ^r to the statistical fluctuations in $\tilde{\phi}$.

When the number of observations N is not large, we might well be concerned at the loss of information involved in forming the data into a histogram in order to obtain $\tilde{\phi}$. This loss of information can be avoided by taking

$$\tilde{\phi}(x)=\frac{1}{N}\sum_{n=1}^{N}\delta(x-x_n), \tag{16}$$

where the x_n are the individual measures and $\delta(x)$ is Dirac's delta function. Substitution of this expression into Eq. (12) yields

$$\psi^{r+1}(\xi)=\frac{1}{N}\sum_{n=1}^{N}Q^r(\xi|x_n), \tag{17}$$

a result with considerable intuitive appeal. If we do, in fact, treat the data this way, the smoothness of the estimates $\psi^1, \psi^2, \ldots$ depends not only on the smoothness of ψ^0, but also on there being a large enough sample for the overlapping of the functions $Q^r(\xi|x_n)$ to produce a smooth function. When the sample is too small for this to happen, it is intuitively clear that the statistical uncertainty in any estimate of ϕ is greater than the rectification corrections. In such cases, there is no point in attempting the rectification. [The approach fails when $P(x|\xi)$ contains singularities, since the estimates ψ^{r+1} will then also contain singularities. In such cases we must make a smoother choice for $\tilde{\phi}$ than that given by Eq. (16).]

A further matter of practical concern, especially when a distribution is being corrected for errors, is the data's frequent lack of homogeneity. If the fraction ν_k of the sample corresponds to the conditional probability density function $P_k(x|\xi)$, and if we define $\phi_k(x)$ and $Q_k(\xi|x)$ accordingly, then the identity (11) becomes

$$\psi(\xi)\equiv\sum \nu_k\int \phi_k(x)Q_k(\xi|x)dx. \tag{18}$$

The corresponding form of Eqs. (12)–(14) is then obvious. If we choose not to group the data, then Eq. (17) applies with $Q^r(\xi|x_n)$ replaced by $Q_n{}^r(\xi|x_n)$, which is obtained from Eqs. (13) and (14) with $P(\xi|x_n)$ replaced by $P_n(\xi|x_n)$.

Above we have developed the iterative technique for one-dimensional distributions; the generalization to multidimensional distributions, $\phi(x,y,\ldots)$ and $\psi(\xi,\eta,\ldots)$ is immediate, however. To achieve this generalization, we interpret x and ξ as representing the vectors $(x,y,\ldots)$ and $(\xi,\eta,\ldots)$ and the integrals and summations as representing the appropriate multiple integrals and summations. For example, if $\tilde{\phi}(x,y)$ is the intensity distribution recorded by an image-forming device and if $A(x-\xi,\, y-\eta)$ is the point spread function of that device, then image restoration may be achieved with the following iterative scheme:

$$\psi^{r+1}(\xi,\eta)=\psi^r(\xi,\eta)\int\int\frac{\tilde{\phi}(x,y)}{\phi^r(x,y)}A(x-\xi,y-\eta)dxdy \quad (19)$$

where

$$\phi^r(x,y)=\int\int\psi^r(\xi,\eta)A(x-\xi,y-\eta)d\xi d\eta. \quad (20)$$

The starting approximation $\psi^0(\xi,\eta)$ should be a smooth, non-negative function having the same integrated intensity as the observed image. (N.B. If off-axis aberrations are significant, the point spread function is no longer simply a function of $x-\xi$ and $y-\eta$. This presents no difficulty for the present scheme, but does prevent the use of Fourier techniques.)

III. INCREASING LIKELIHOOD

Although the qualitative discussion in Sec. II suggests the usefulness of the iterative scheme, its relationship to the methods discussed in Sec. I is not obvious. We shall now show, however, that, when the integrals are approximated by sums and when only the fraction ϵ of the correction to ψ_j^r is actually applied, the scheme converges as $r\rightarrow\infty$ to a solution of the corresponding maximum likelihood (ML) problem, provided that ϵ is sufficiently small. Moreover, we demonstrate that the ML-solution is, in general, unique when $J\leq I$ and that the multiple solutions when $J>I$ all have equal likelihood. Finally, we prove that the direct solution is identical to the corresponding ML-solution, provided that the direct solution does not violate the constraint $\psi_j\geq 0$.

It is convenient to consider the latter statement first. Adopting the notation of Sec. I and noting that $\phi_i/\Delta x$ is the probability density assigned to each of the $N\tilde{\phi}_i$ observations in the interval $(x_i-\frac{1}{2}\Delta x, x_i+\frac{1}{2}\Delta x)$, we find that the likelihood is given by

$$\ln L=N\cdot H(\psi_j)+\text{constant} \quad (21)$$

with

$$H(\psi_j)=\sum_{i=1}^{I}\tilde{\phi}_i ln\phi_i, \quad (22)$$

and where ϕ_i is given by Eq. (5). Maximizing H is clearly equivalent to maximizing L.

In order to obtain the ML-solution subject to the constraint

$$\sum_{j=1}^{J}\psi_j=1,$$

we look for the stationary points of the function

$$G(\psi_j;\mu)=H(\psi_j)+\mu(\sum_{j=1}^{J}\psi_j-1). \quad (23)$$

(Notice that the constraint $\psi_j\geq 0$ is not being imposed.) Setting the derivatives of G with respect to ψ_j and μ equal to zero, we obtain

$$\sum_{i=1}^{I}\frac{\tilde{\phi}_i}{\phi_i}P_{ij}=-\mu, \quad (24)$$

and

$$\sum_{j=1}^{J}\psi_j=1. \quad (25)$$

Multiplying Eq. (24) by ψ_j and summing over j, we find that the Lagrangian multiplier $\mu=-1$. Then, since

$$\sum_{i=1}^{I}P_{ij}=1,$$

we see that Eq. (24) is satisfied if ψ_j is such that $\phi_i=\tilde{\phi}_i$. Moreover, since

$$\sum_{i=1}^{I}\tilde{\phi}_i=1,$$

this ψ_j also satisfies Eq. (25). Now, because the condition $\phi_i=\tilde{\phi}_i$ yields the linear equations of the direct method (Sec. I), we have shown that the direct solution is also a solution of the corresponding ML-problem when the constraint that $\psi_j\geq 0$ is relaxed. Obviously then, if the direct solution happens to be such that $\psi_j\geq 0$, it is also a solution of the corresponding fully-constrained ML-problem. (That this latter solution is in general unique is demonstrated below.)

Because Eq. (24) is not linear in ψ_j, the surface defined by $H(\psi_j)$ is not quadratic; consequently, we must examine the possibility that this surface has local maxima at which the likelihood is less than the greatest likelihood attained in the permitted region. This possibility does not arise, however, because $H(\psi_j)$ is a concave function of position in D, the convex region of J-dimensional vector space defined by the constraints (8). To demonstrate this, let $\psi_j{}^a$ and $\psi_j{}^b$ be *distinct* vectors belonging to D and let $\phi_i{}^a$ and $\phi_i{}^b$ be the corresponding ϕ_i-vectors. If we now define $\psi_j=\lambda\psi_j{}^a+(1-\lambda)\psi_j{}^b$ with $0<\lambda<1$, we see at once that ψ_j also belongs to D, which is, therefore, a convex region of vector space. Because ψ_j is indeed in the permitted region, it is of interest to calculate $H(\psi_j)$ and to compare it with $H(\psi_j{}^a)$ and $H(\psi_j{}^b)$. We have

$$H(\psi_j)=\sum_{i=1}^{I}\tilde{\phi}_i ln[\lambda\phi_i{}^a+(1-\lambda)\phi_i{}^b]$$

which is

$$\geqq\sum_{i=1}^{I}\phi_i[\lambda ln\phi_i{}^a+(1-\lambda)ln\phi_i{}^b],$$

since $\ln(x)$ is a concave function. We have, therefore, established that

$$H(\psi_j)\geq\lambda H(\psi_j{}^a)+(1-\lambda)H(\psi_j{}^b), \quad (26)$$

a result that eliminates the possibility of multiple maxima, since it implies that the chord joining any two points on the surface $H(\psi_j)$ always lies in or below that surface.

Because $\ln(x)$ is a strictly concave function, the inequality (26) is strict provided that ψ_j^a and ψ_j^b being distinct implies that ϕ_i^a and ϕ_i^b are distinct. The linear mapping defined by Eq. (5) will, in general, have this property when $J \leq I$, but not otherwise. Accordingly, when $J \leq I$, the inequality (26) is in general strict, which implies that $H(\psi_j)$ attains its greatest value in D at a point; the solution of the ML-problem is then unique. On the other hand, when $J > I$, there is not a unique vector ψ_j corresponding to the ϕ_i that maximizes H; the greatest likelihood is attained, therefore, at points on a 'plateau' of the $H(\psi_j)$ surface. This behavior is, of course, a consequence of asking for more information than exists in the grouped data.

Let us now examine the behavior of the iterative scheme. With the integral approximated by a sum, Eq. (15) becomes

$$\psi_j^{r+1} = \psi_j^r \sum_{i=1}^{I} \frac{\tilde{\phi}_i}{\phi_i} P_{ij}, \tag{27}$$

so that the correction to ψ_j^r is

$$\delta\psi_j^r = \psi_j^r \left(\sum_{i=1}^{I} \frac{\tilde{\phi}_i}{\phi_i^r} P_{ij} - 1 \right). \tag{28}$$

Remembering that $\mu = -1$, we have, from Eqs. (23) and (24),

$$\delta\psi_j^r = \psi_j^r \frac{\partial G^r}{\partial \psi_j^r} \tag{29}$$

and

$$\frac{\partial H^r}{\partial \psi_j^r} = \frac{\partial G^r}{\partial \psi_j^r} + 1; \tag{30}$$

consequently, when we apply the correction $\epsilon\delta\psi_j^r$ to ψ_j^r, the change in H is

$$dH_\epsilon = \epsilon \sum_{j=1}^{J} \left(\frac{\partial G^r}{\partial \psi_j^r} + 1 \right) \delta\psi_j^r + 0(\epsilon^2). \tag{31}$$

Because the scheme conserves the constraints (8), we have

$$\sum_{j=1}^{J} \delta\psi_j^r = 0$$

so that

$$dH_\epsilon = \epsilon \sum_{j=1}^{J} \psi_j^r \left(\frac{\partial G^r}{\partial \psi_j^r} \right)^2 + 0(\epsilon^2). \tag{32}$$

From Eq. (29), we see that the iterative scheme converges (i.e., $\delta\psi_j = 0$ for all j) at the point ψ_j^* if and only if

$$\frac{\partial G}{\partial \psi_j} = 0 \quad \text{when} \quad \psi_j^* \neq 0. \tag{33}$$

Accordingly, if ψ_j^* is entirely within D, convergence occurs at a point of maximum likelihood. On the other hand, if ψ_j^* is not entirely within D, convergence occurs at a point of greatest likelihood on the boundary of D. (From earlier results, we know that these points are in general unique when $J \geq I$.)

When ψ_j^* satisfies the condition (33), the first-order correction to H, given by Eq. (32), is of course zero. When condition (33) is not satisfied, this first-order correction is strictly positive, so that, for sufficiently small ϵ, the iteration will result in an increase in the likelihood. Thus, we see that the iterative scheme can be used to generate a sequence of estimates to $\psi(\xi)$, each member of which assigns a greater likelihood to the sample than did its predecessor.

Although these analytical results demonstrate increasing likelihood only for sufficiently small ϵ, numerical calculations reveal increasing likelihood even with $\epsilon = 1$, and it seems likely that increasing likelihood will turn out to be rigorously true for $\epsilon = 1$. Accordingly, in applying this technique, it is recommended that the full correction be applied.

IV. NUMERICAL EXPERIMENTS

The results of Sec. III indicate that the iterative scheme converges monotonically as $r \to \infty$ to the ML-solution. No attempt should be made to achieve convergence, however, because, after a few iterations, further gains in likelihood are achieved in general only by fitting $\phi^r(x)$ to the statistical fluctuations in $\tilde{\phi}(x)$ at the price of an increasingly complicated $\psi^r(\xi)$ (cf. Sec. II). We now illustrate this point by reporting the results of numerical experiments for a case whose exact solution is known as $N \to \infty$.

Let $\Phi(x;\mu,\sigma)$ denote the normal distribution with mean μ and variance σ^2. Then, if we take $\psi(\xi) = \Phi(\xi; 0, 1/\sqrt{2})$ and $P(x|\xi) = \Phi(x; \xi, 1/\sqrt{2})$, we may readily show that $\phi(x) = \Phi(x; 0,1)$. Accordingly, our experiment is to apply the scheme to a random sample of N numbers drawn from $\Phi(x; 0,1)$ and to compare the resulting sequence of estimates ψ^r with the exact ψ.

A criterion measuring the goodness-of-fit of ψ^r to ψ is obtained by calculating a pseudo-χ^2 in the following way: we determine the ten-percentage points ξ_k of the ψ-distribution and then calculate n_k^r, the expected number of observations in the k-th interval (ξ_k, ξ_{k+1}) when N observations are distributed according to ψ^r. Then, since $n_k^r = N/10$ when $\psi^r \equiv \psi$, we define our pseudo-χ^2 to be

$$\chi^2\{\psi^r\} = \sum_{k=1}^{10} \frac{(n_k^r - N/10)^2}{N/10}. \tag{34}$$

It is also of interest, of course, to examine the goodness-of-fit of both ϕ and ϕ^r to the sample. We do this by calculating $\chi^2\{\phi\}$ and $\chi^2\{\phi^r\}$, the conventional χ^2s,

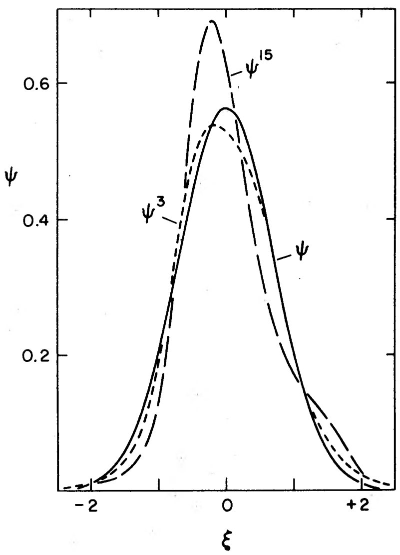

FIG. 1. Estimates obtained after the third and fifteenth iterations for the experiment described in the text. The solid curve is the known answer.

with the data binned, in both cases, into the ten intervals defined by the ten-percentage points of ϕ.

The results from one such experiment are shown in Figs. 1 and 2. We take $N=150$ and, because of the modest sample-size, use Eq. (16) for $\tilde{\phi}(x)$; Eq. (17) then yields the estimates ψ^r starting with $\psi^o=(\sqrt{2}/\pi)/(1+\xi^4)$. In Fig. 1, the exact ψ as well as ψ^r for $r=3$ and 15 are plotted, and we see that the result of many iterations is indeed inferior to the excellent fit obtained in a few iterations. Further illustration of this effect is found in Fig. 2, where $\chi^2\{\psi^r\}$, $\chi^2\{\phi^r\}$, and $N^{-1}\ln L$ are plotted against r. This plot shows that each ψ^r is an "improvement" over its predecessor if we judge success by the likelihood, which increases monotonically with r, or by $\chi^2\{\phi^r\}$, which decreases monotonically. However, if we judge success by $\chi^2\{\psi^r\}$, our measure of the goodness-of-fit of ψ^r to ψ, we find that the iterations beyond $r=3$ are not, in this case, improvements. It is notable also that the minimum in $\chi^2\{\psi^r\}$ coincides with $\chi^2\{\phi^r\}$ dropping below $\chi^2\{\phi\}$—that is, with ϕ^r becoming a closer fit to the data than is the actual sampling distribution, ϕ. The subsequent iterations, therefore, can indeed be described as merely fitting the statistical fluctuations in $\tilde{\phi}$; and the penalty is seen to be worsening estimates ψ^r.

To investigate the degree of success that might typically be obtained with this scheme, the above experiment has been repeated for forty independent samples, again with $N=150$. In Fig. 3, the results obtained after the third iteration are presented as histograms of $\chi^2\{\phi\}$, $\chi^2\{\phi^3\}$, and $\chi^2\{\psi^3\}$. Comparison of the histograms of $\chi^2\{\phi\}$ and $\chi^2\{\phi^3\}$ shows that by stopping at $r=3$ we have trespassed only slightly into the domain where the data are being fitted too closely. Comparison of the histograms of $\chi^2\{\psi^3\}$ and $\chi^2\{\phi\}$ shows that the former is strongly biased to small values, indicating that good fits are indeed generally obtained. In fact, this degree of success might seem surprising, since one might have anticipated intuitively that the most we could hope for is $\chi^2\{\psi^r\}\simeq E(\chi^2)$, the

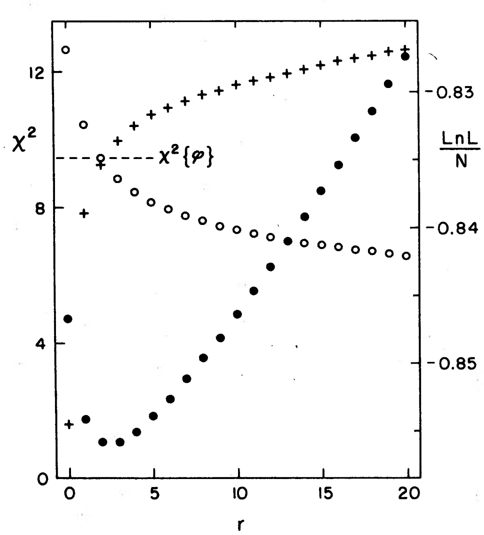

FIG. 2. Success criteria plotted against iteration number. Crosses denote $N^{-1}\ln L$; filled circles denote $\chi^2\{\psi^r\}$; open circles denote $\chi^2\{\phi^r\}$.

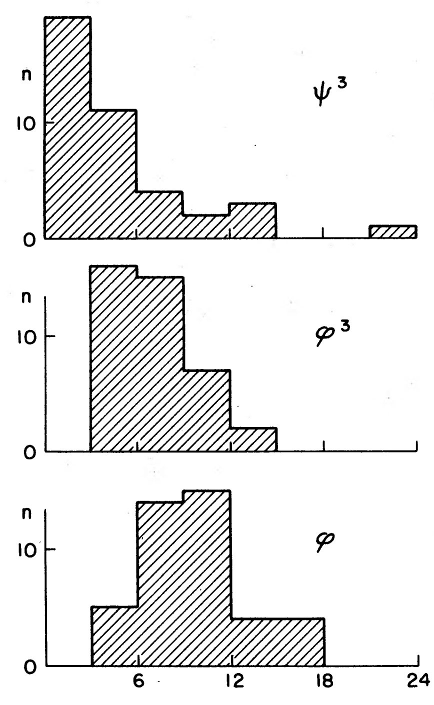

FIG. 3. Results obtained after three iterations in forty independent experiments. The upper, middle, and lower plots are histograms of $\chi^2\{\psi^3\}$, $\chi^2\{\phi^3\}$, and $\chi^2\{\phi\}$, respectively.

expected value of χ^2 for samples drawn directly from ψ. That the scheme does better than this is a consequence of its built-in smoothing operation [see the remarks following Eq. (17)]. From the histograms, we see that $\chi^2\{\psi^3\}$ is typically one half of $\chi^2\{\phi\}$—that is, half the χ^2 obtained by random sampling from a given distribution function. Therefore, since for fixed n_k^r/N, $\chi^2 \propto N$, the smallness of $\chi^2\{\psi^3\}$ is due to n_k^r matching $N/10$ with the closeness expected in random sampling from ψ when the sample size is $\simeq 2N$.

In the above experiments excellent results are obtained in a few iterations. There are two circumstances, however, in which the convergence to a good fit will be slow. The first arises when our choice for ψ^o is such that the corresponding ϕ^o introduces short-wavelength deviations from unity in $\tilde{\phi}/\phi^o$ additional to those caused by statistical fluctuations in $\tilde{\phi}$ [cf. the remarks following Eq. (15)]. Such would be the case, for example, had we taken ψ^o in the above experiments to be $\Phi(\xi;0,\sigma)$ with $\sigma^2 \ll \frac{1}{2}$, the variance of the exact ψ. This circumstance may be avoided if we resist the temptation of anticipating the outcome of the rectification and take for ψ^o a simple, smooth fit to $\tilde{\phi}$.

The second circumstance leading to many iterations arises when the exact ψ has significant short-wavelength structure and the sample size is sufficiently enormous that $\tilde{\phi}$ still contains information from which this structure in ψ can be discovered. In this case, if we start from a smooth ψ^o, many iterations are necessary before the estimates ϕ^r acquire the short-wavelength structure required if ϕ^r is to be a good fit to $\tilde{\phi}$.

It is important to note that, in both the above circumstances, we need not know ψ in order to recognize the need for many iterations. This need will be apparent from large values of $\chi^2\{\phi^r\}$, provided that the χ^2-test is carried out with optimum resolution—that is, with bin sizes whose expected contents are all $\sim$5–10 observations. Since in practical applications of this scheme ψ is not known, the basis for stopping the iterations has to be the value of $\chi^2\{\phi^r\}$, or of some other measure of the goodness-of-fit of ϕ^r to $\tilde{\phi}$. If the χ^2-test is used and if $P(\chi^2 > \chi_o^2)$ denotes the probability that random sampling will give χ^2 greater than a given value χ_o^2, then the iterations should be stopped as soon as $P(\chi^2 > \chi^2\{\phi^r\})$ is not small (e.g., is not <0.05). This recommendation implies, of course, that the iterations not be continued to the point that $1-P$ is small. (In calculating these probabilities ν, the number of degrees of freedom, should be taken as the number of bins minus one. By not reducing ν further, to allow for our use of the data in estimating ϕ^r, we are treating ϕ^r exactly as if it were the true ϕ.)

In many applications, the errors in $\tilde{\phi}_i$ will be measurement errors rather than sampling errors. In such cases, we may still use the χ^2-test in deciding when to stop the iterations by comparing the actual and the expected distributions of residuals $\tilde{\phi}_i - \tilde{\phi}_i^r$. (If necessary, parameters in the expected distribution function may be estimated from the residuals themselves.)

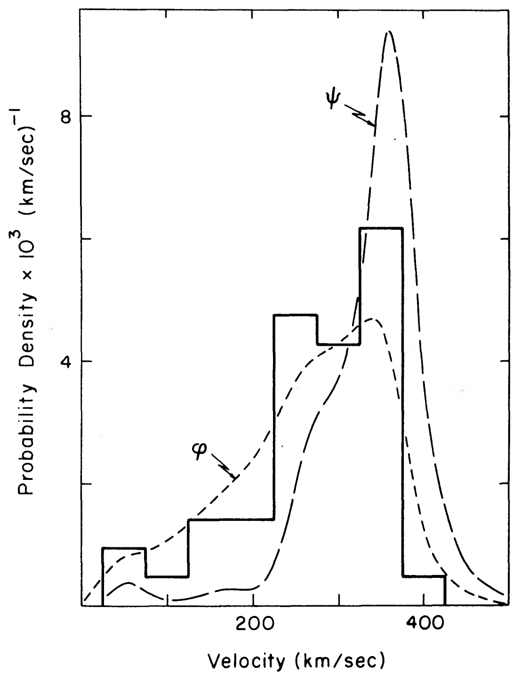

FIG. 4. Rectification of $v \sin i$ measures for 42 B6-B9e stars. The final estimates ψ and ϕ are shown together with the histogram of the raw data.

When the iterations have been stopped, we may wish to estimate the uncertainty of our estimate $\psi^{r+1}(\xi)$. This is readily done if we regard $Q^r(\xi|x)$ as exactly known, for then Eq. (12) becomes a linear mapping of $\tilde{\phi}(x)$ into $\psi^{r+1}(\xi)$ and as such may be used to study the propagation of errors.

V. APPLICATIONS

Having established the usefulness of the technique, we now illustrate and discuss its application to astronomical problems.

(*i*) *Stellar Rotation:* As mentioned in the Introduction, the rectification of distributions of $v \sin i$ is an example of the class of problems under consideration. We shall rediscuss two distributions from which significant conclusions have previously been drawn.

The first example is Slettebak's (1966) study of $v \sin i$ measures for 42 B6-B9e stars, the distribution of which he asserted to be consistent with the hypothesis that all such stars have equatorial velocities $v \simeq 350$ km/sec. The results obtained after four iterations with the iterative scheme are given in Fig. 4, which shows that the estimate for ψ has a sharp peak at $v=360$ km/sec; Slettebak's conclusion is, therefore, essentially confirmed. (The starting approximation was ψ^o=constant for $v<500$ km/sec, and $\tilde{\phi}$ was taken to be the

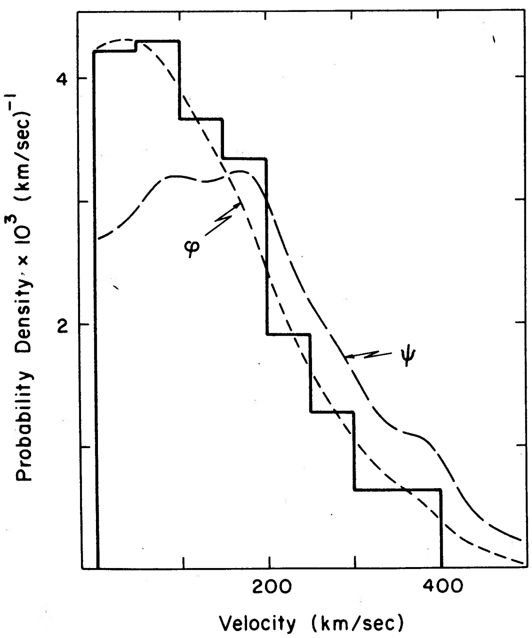

FIG. 5. Rectification of $v \sin i$ measures for 251 B5-B9 stars. The final estimates ψ and ϕ are shown together with the histogram of the raw data.

continuous function obtained by linear interpolation between the midpoints of each block of the histogram.)

The second example is Deutsch's (1968) study of $v \sin i$ measures for 251 B5–B9 stars in the main-sequence band in which he concluded that two distinct populations must be recognized, one comprising slow rotators with $v \simeq 15$ km/sec, the other rapid rotators with $v \simeq 150$ km/sec. After two iterations, we obtain the results given in Fig. 5. (The starting estimate and $\tilde{\phi}$ were obtained as in the previous example.) Comparison of the ϕ shown here with that obtained by Deutsch (1968, Fig. 2) reveals that we have failed to recover the sharp peak [$\phi \simeq 9 \times 10^{-3}$ (km/sec)$^{-1}$ at $v \simeq 10$ km/sec] that is the basis of his claimed discovery of a population of slow rotators.

The explanation of our disagreement with Deutsch is fairly obvious. In applying the iterative technique, we have made no assumptions about ψ other than that it satisfies the constraints (4); consequently, starting with a smooth ψ^o and working with data grouped in 50 km/sec bins, we have obtained an acceptable fit to $\tilde{\phi}$ without introducing variations in ψ or ϕ on the scale of 10 km/sec. Deutsch, on the other hand, does make additional assumptions about the form of ψ. Having found that a single Maxwellian distribution predicts too few slow rotators, he adopts a bimodal Maxwellian distribution for ψ and, not surprisingly, finds that the additional Maxwellian fits itself to the slow rotators. Since our results contain no hint of this bimodal character, we conclude that Deutsch's discovery is more a consequence of his assumptions concerning ψ than of the data. Deutsch's conclusions should be accepted only if there are independent arguments favoring the bimodal Maxwellian distribution.

In definitive studies of stellar rotation, one should first apply the iterative technique to correct the distributions of $v \sin i$ for observational errors and then apply the technique again to rectify the corrected distributions for aspect effect. It would also be of interest first to correct each star's $v \sin i$ for evolutionary expansion, so that the final result is the distribution of equatorial velocities at zero-age.

(*ii*) *Beam-smoothing:* As an example from the important area of deconvolving data for the effect of instrumental broadening, we consider the problem of correcting radio-astronomical observations for the smoothing due to the beam pattern of the antenna.

Following Bracewell and Roberts (1955), we take the beam pattern to be $P(x|\xi) = A(x-\xi)$ with $A(z) = (\sin\pi z/\pi z)^2$. Error-free observations of a point source of unit amplitude [i.e., $\psi = \delta(\xi)$] result, therefore, in the response $\tilde{\phi} = A(x)$. The problem, then, is to apply the iterative technique starting with $\psi^o = \tilde{\phi}(\xi)$, the estimate provided by the observations, and to examine the approach of the succeeding estimates to the known answer, $\psi = \delta(\xi)$. The results obtained after one and three iterations are given in the right-hand side of Fig. 6. With one iteration, the central intensity is

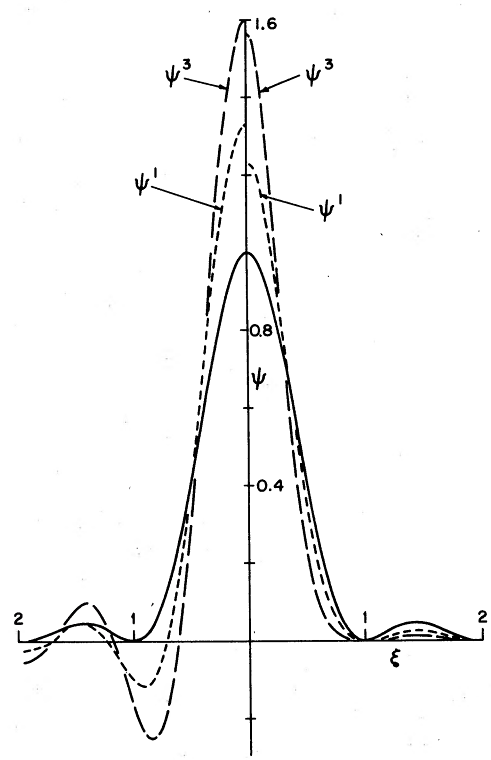

FIG. 6. Deconvolution of an antenna's response to a point source. The right-hand side shows results obtained after one and three iterations with iterative technique. The left-hand side shows the corresponding results with the Burger-van Cittert scheme.

increased from 1.00 to 1.22, and the intensity in the side lobes is halved.

The left-hand side of Fig. 6 shows the results obtained with the Burger-van Cittert iterative scheme (see Bracewell and Roberts 1955), which, in our notation, is

$$\psi^{r+1}(\xi)=\psi^r(\xi)+[\tilde{\phi}(x)-\phi^r(x)]_{x=\xi}. \tag{35}$$

Comparison of these results shows that the two schemes achieve comparable rates of increase in the central intensity; the Burger-van Cittert scheme is, however, clearly inferior because of the substantial negative intensities that appear in the region of the side lobes. This physically inadmissable behavior of the Burger-van Cittert scheme, which has been widely discussed in radio astronomy, emphasizes the great importance of the present scheme's built-in conservation of the constraint $\psi\geq 0$.

(*iii*) *Smoothing of Histograms:* An important role in observational investigations is often played by conclusions drawn from histograms of observations. Despite the added confidence that could be given to such conclusions by smoothing the histograms, this is rarely done. The reason for this, perhaps, is the realization that standard smoothing operations, in which the data are convolved with some appropriate function, result in smoothed distributions with biased variance. (Of course, even grouping data into a histogram leads to biased moments, the correction of which is the purpose of Sheppard's corrections.) This objection to smoothing may be overcome, however, with the following application of the iterative technique:

If the data sample is $\xi_1, \xi_2, \ldots, \xi_N$, then

$$\tilde{\psi}(\xi)=\frac{1}{N}\sum_{n=1}^{N}\delta(\xi-\xi_n) \tag{36}$$

contains all the information in the sample but is not a satisfactory estimate of $\psi(\xi)$, the true distribution, because the short-wavelength behavior of $\tilde{\psi}$ is a consequence of the discrete sampling. A smooth, but biased estimate, $\tilde{\phi}(\xi)$, for ψ may be obtained by convolving $\tilde{\psi}$ with a probability density function $P(x|\xi)$. Thus, we have

$$\tilde{\phi}(x)=\frac{1}{N}\sum_{n=1}^{N}P(x|\xi_n). \tag{37}$$

(We might, for example, take $P(x|\xi)=\Phi(x;\,\xi,\sigma)$, the normal distribution. We would then choose σ to be just large enough for $\tilde{\phi}$ to be a smooth function.) We now apply the iterative technique, starting with $\psi^o=\tilde{\phi}(\xi)$, in order to 'correct' for this convolution. From our earlier discussion (Secs. II and IV), we know that with a few iterations we can restore the long-wavelength behavior of $\tilde{\psi}$ but not the short-wavelength behavior. After a few iterations, therefore, we obtain an estimate ψ^r for ψ by, in effect, eliminating the short-wavelength behavior of $\tilde{\psi}$ without modifying its long-wavelength behavior. Accordingly, we may expect ψ^r to have unbiased low-order moments. (The success of the estimate ψ^r may, of course, be checked against the sample with the χ^2-test.)

The above procedure may be used to obtain smooth estimates, $\tilde{\phi}(x)$, for other rectification problems. Smoothed estimates are necessary when $P(x|\xi)$ contains a singularity, as in the stellar rotation problem. Another possible application is to the smoothing of data obtained in Monte-Carlo simulations in order to accelerate convergence.

(*iv*) *Model Testing:* In all of the earlier discussion we have regarded $P(x|\xi)$ as known. In some problems, however, the model leading to $P(x|\xi)$ will be in doubt. If the model is indeed wrong, then it may well be that no function ψ satisfying the constraints (4) will correspond to a function ϕ that provides an acceptable fit to the observational data. Therefore, since the iterative technique cannot violate the constraints, the model must be rejected if the technique fails to reduce the residuals $\tilde{\phi}_i-\phi_i{}^r$ to the point where they can be ascribed to observational errors.

VI. CONCLUSIONS

The main purpose of this paper has been to investigate a new approach to the wide, and astronomically important class of statistical estimation problems in which the desired distribution function, $\psi(\xi)$, is related to an observationally accessible function, $\phi(x)$, by the integral Eq. (1).

Because of sampling errors in $\tilde{\phi}(x)$, the observational estimate for $\phi(x)$, we have argued (Sec. I) that our aim should be to obtain an estimate $\hat{\psi}(\xi)$ for $\psi(\xi)$ that (a) obeys the constraints (4), (b) exhibits a degree of smoothness and continuity consistent with our expectations (often intuitive) for $\psi(\xi)$, and (c) corresponds to a $\hat{\phi}(x)$ that differs from $\tilde{\phi}(x)$ by amounts that can reasonably be ascribed to sampling errors in $\tilde{\phi}$, a condition that excludes improbably small as well as improbably large residuals.

Seen in the context of these conditions, the method of direct solution is unacceptable because it violates condition (c) by demanding zero residuals. Moreover, the penalty for this violation is usually an estimate $\hat{\psi}(\xi)$ that also fails to satisfy conditions (a) and (b). In marked contrast to this unfortunate behavior, the scheme proposed in Sec. II yields, after a few iterations, estimates $\psi^r(\xi)$ that satisfy all three conditions, provided only that the first estimate $\psi^o(\xi)$ is chosen sensibly and that the number of observations is less than enormous.

In addition to solving the problem, the technique has the further merits of being readily programmed and of not being extravagant in its demand for computer time. Because of this, it should be practicable to use the technique for two- and three-dimensional problems—the general beam-smoothing problem, for example.

ACKNOWLEDGMENTS

This work has been supported by the National Science Foundation under Grant GP-30927 and by a grant of computer time from the Goddard Institute for Space Studies, NASA. I am also grateful to K. H. Prendergast, E. Ricco, and L. Taff for useful conversations.

REFERENCES

Bracewell, R. N., and Roberts, J. A. (1954). Aust. J. Phys. **7**, 615.
Carswell, R. F. (1973). Mon. Not. R. Astron. Soc. **162**, 61.
Chandrasekhar, S., and Münch, G. (1950). Astrophys. J. **111**, 142.
Deutsch, A. J. (1970). *Stellar Rotation*, edited by A. Slettebak (D. Reidel, Dordrecht–Holland), p. 207.
Eddington, A. S. (1913). Mon. Not. R. Astron. Soc. **73**, 359.
Kunasz, C. V., Jefferies, J. T., and White, O. R. (1973). Astron. Astrophys. **28**, 15.
Lucy, L. B., and Ricco, E. (1974). In preparation.
Phillips, B. L. (1962). J. Assoc. Comp. Mach. **9**, 84.
Slettebak, A. (1966). Astrophys. J. **145**, 121.
Trumpler, R. J., and Weaver, H. F. (1953). *Statistical Astronomy* (U. Calif. Press, Berkeley).
Twomey, S. (1963). J. Assoc. Comp. Mach. **10**, 97.