1976ApJ...208..177L

THE ASTROPHYSICAL JOURNAL, 208:177–190, 1976 August 15
© 1976. The American Astronomical Society. All rights reserved. Printed in U.S.A.

PARAMETER ESTIMATION IN X-RAY ASTRONOMY

MICHAEL LAMPTON, BRUCE MARGON, AND STUART BOWYER
Space Sciences Laboratory, University of California, Berkeley
Received 1975 August 25; revised 1976 January 19

ABSTRACT

The problems of model classification and parameter estimation are examined, with the objective of establishing the statistical reliability of inferences drawn from X-ray observations. For testing the validities of classes of models, the procedure based on minimizing the χ^2 statistic is recommended; it provides a rejection criterion at any desired significance level. Once a class of models has been accepted, a related procedure based on the increase of χ^2 gives a confidence region for the values of the model's adjustable parameters. The procedure allows the confidence level to be chosen exactly, even for highly nonlinear models. Numerical experiments confirm the validity of the prescribed technique.

The $\chi^2_{\min} + 1$ error estimation method is evaluated and found unsuitable when several parameter ranges are to be derived, because it substantially underestimates their joint errors. The ratio of variances method, while formally correct, gives parameter confidence regions which are more variable than necessary.

Subject headings: X-rays: sources — X-rays: spectra

I. INTRODUCTION

X-ray astronomy observations have reached a level of sophistication at which systematic errors are usually insignificant in comparison with photon counting statistics, particularly in studies of faint extragalactic objects. Inferences drawn from such data, which might concern the spatial structure, spectral characteristics, or temporal variability of a source, are necessarily subject to random errors originating in this photon noise. It is consequently important to adopt data analysis procedures whose statistical reliability can be established.

In this paper, we adress two important problems often faced by X-ray astronomers: the assessment of the validity of models, and the determination of values of adjustable parameters which valid models possess. A feature common to these problems is that the worth of an experimental measurement can be judged by what it rules out. In the first case, demonstrating an incompatibility between some supposed emission mechanism and one's data suffices to rule out that mechanism; yet demonstrating agreement does not prove that the supposed explanation is correct. In the second case, one may constrain model parameters by demonstrating that some large region of parameter space would lead to incompatibilities with the observations; yet, with any finite stretch of data, the exact parameter values cannot be determined. In both cases, the statistical reliability of these determinations may be established through the use of techniques of testing hypotheses concerning classes of models, and subsets of classes of models.

In parameter estimation, it is the *range* of parameter values to which a theory is restricted that is the useful result of an experiment. The discrete best-fitting values of the parameters are essentially statistical artifacts subject to a variety of correlated random errors originating in the counting statistics of the original data. If an experiment could be repeated without systematic changes, best-fitting parameter values would differ, while their properly derived allowed ranges will overlap. In this paper, we derive a prescription for obtaining joint parameter ranges which can be constructed at any desired confidence level, and present the results of extensive numerical simulations which verify this prescription.

We stress that we are not discussing a specialized problem in error analysis which is subordinate to the primary data reduction effort; rather, this issue involves the techniques used to calculate the major experimental results. Preliminary discussions of this approach have been given elsewhere (Margon 1974; Margon, Bowyer, and Lampton 1975*a*; Margon *et al.* 1975*b*).

II. HYPOTHESIS TESTING

a) Simple Hypotheses

Numerous hypotheses can, in principle, be formulated about astrophysical processes. A hypothesis can be quantitatively tested if it provides a complete prediction of the expected data and a statistical description of the expected deviations or errors. In X-ray astronomy, hypotheses must include models for detector performance, background, dead time losses, etc., as well as the source spectrum and interstellar absorption. Deviations in count-rate accumulations are described by Poisson's law, which under the usual circumstance of many counts per bin is accurately represented by a normal distribution whose variance is equal to the mean.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1976ApJ...208..177L

A hypothesis is said to be composite if it contains adjustable parameters. We shall first discuss the testing of simple hypotheses, i.e., those having a fixed predicted count distribution. Testing such a hypothesis is done with a three-step procedure (see, for example, Brownlee 1965, p. 97). First, a test statistic S is devised, so constructed as to be sensitive to significant discrepancies between the data and the model. Second, the statistic is evaluated for the data on hand. Third, this value is compared with the theoretical distribution of S-values expected if the hypothesis were correct. When S is reasonably ranked among the values expected, the hypothesis is not rejected by the test: the model may be wrong, but no significant discrepancy has been seen. If, however, S lies in an improbable "critical region," the hypothesis is rejected; it is wrong, or else an improbable data set was obtained. The policy of rejecting a simple hypothesis whenever S lies in the critical region has an easily computable reliability. Suppose the hypothesis is correct. Then the probability α of wrongly rejecting it is just the probability that S lies in the critical region:

$$\alpha = \int f(S)dS\,, \tag{1}$$

where $f(S)$ is the probability density of S, given the hypothesis, and the integration is performed over the critical region. The quantity α is usually termed "the level of significance" of the test, i.e., of the critical region. Because ruling out a correct model is a fairly serious error, it is important to minimize the chance of this happening, i.e., to avoid claiming incompatibilities with models unless the test statistic is quite improbably situated. Above all it is important to state the significance level at which rejection of a model is claimed, to allow readers to understand the reliability of the claim.

An appropriate and widely used test is based on Pearson's χ^2 statistic. Let

$$S \equiv \sum_{i=1}^{N} \frac{(D_i - F_i)^2}{\sigma_i^2}\,, \tag{2}$$

where D_i is the (integer) number of observed counts in bin i and the F_i are the accumulations predicted by the model. Each term of this sum is of the form of a deviation squared divided by the expected variance $\sigma_i^2 = F_i$. The deviations are approximately Gaussian; and if the hypothesis is correct, they have zero mean and are independent. Therefore, under the hypothesis, S will be χ^2 distributed with N degrees of freedom.[1] If the hypothesis is wrong, there will be additional contributions to S from the systematic errors in the fit, and S will on the average be larger than expected. The hypothesis is to be rejected if S exceeds some critical decision threshold, which can be chosen at any desired significance level α from tables of the χ^2 probability function. We feel that significance levels of greater than 10 percent are not secure grounds for rejecting a hypothesis: there would be an appreciable chance of falsely ruling out a correct model, and we view this as a serious error. A 10 percent significance level appears to us to offer a reasonable compromise between security of conclusions and sensitivity toward detecting false models.

In this connection, a variant of the procedure outlined above is often used. From the observed value of S, one calculates the observed significance of the fit (sometimes called the confidence of the fit), $\alpha_{\rm obs}$, from

$$\alpha_{\rm obs} = \int_S^\infty f(\chi^2)d\chi^2\,, \tag{3}$$

where f is the density of the χ^2_N distribution. If $\alpha_{\rm obs}$ is found to lie in the range 10–100 percent, no significant discrepancy exists between the data and the model. If instead $\alpha_{\rm obs} < 10$ percent, the model can be rejected with the knowledge that this policy would fail only once in $1/\alpha_{\rm obs}$ applications.

b) Composite Hypotheses

In X-ray astronomy we may wish to determine whether, in an absolute sense, any given class of emission model is compatible with the data at hand. This question may be answered in a quantitative fashion independent of the existence of other similar or dissimilar models, through methods of testing composite (adjustable parameter) hypotheses.

A composite hypothesis can be rejected only when its best-fitting adjustment can be rejected. Thus in attempting to rule out all (say) power-law emission models for a source, one must explore all power-law indices, absorption column densities, background-level uncertainties, etc., to identify the best-fitting example and test it.

The test is again accomplished by comparing the value of a test statistic with the population of expected values of the test statistic, under the hypothesis. The χ^2 test again can be used. The S statistic, given by equation (2), is minimized by varying p nonequivalent adjustable parameters. This may conveniently be done using a grid search method. The minimum value found, $S_{\rm min}$, is, even for nonlinear models, theoretically distributed as χ^2 with $N - p$ degrees of freedom (Fisher 1924; Cramer 1946, p. 424; Cochran 1952). The composite hypothesis is rejected at significance level α if $S_{\rm min}$ is found to exceed the α-point of the χ^2_{N-p} distribution, defined by

$$\alpha \equiv \int_\Upsilon^\infty f(\chi^2)d\chi^2 \tag{4}$$

where $\Upsilon \equiv \chi^2_{N-p}(\alpha)$ and f is the density of the distribution. Again, only low significance fits constitute evidence against the class of models being tested. We recommend that a model be regarded as incorrect only if it can be rejected with $\alpha \leq 10$ percent; that is, only if $S_{\rm min} > \chi^2_{N-p}(0.10)$.

[1] Pearson's (1900) original formulation tested only the shape, and not the size, of the histogram. Thus the F_i were normalized to the D_i sum. With that constraint, the S statistic has only $N - 1$ degrees of freedom.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1976ApJ...208..177L

III. PARAMETER ESTIMATION

a) General Considerations

The foregoing discussion involves a best-fitting *point estimate* of the adjustable parameter set. A point estimate is, however, worthless for parameter estimation unless some quantitative statement can be supplied as to its accuracy. Methods for generating point estimates can, for linear models, be extended to error estimation through covariance matrix methods or variational procedures. Thus, least-squares estimators, maximum-likelihood estimators, etc., can be perturbed to find the degree to which the fit is sensitive to these misadjustments.

If one parameter of a linear model is being estimated from the data, and if all others are perfectly known from other data, then the standard error in the estimate may be found from the parameter range required to increase the χ^2 statistic S by one (e.g., Bevington 1969, p. 243). Early efforts at multiparameter estimation in X-ray astronomy (Gorenstein, Giacconi, and Gursky 1967; Gorenstein, Gursky, and Garmire 1968) attempted to extend this by successively holding all parameters but one fixed, and noting the range of the remaining parameter for which $\Delta S = 1$. However, many parameter estimators of interest in X-ray astronomy are highly correlated. Perturbations in one parameter can be to some extent compensated by readjustments in other parameters. An error estimate which holds the other parameters fixed neglects this correlation and underestimates the error (Eadie *et al.* 1971, pp. 197 ff.).

An illustration of the interdependence of parameter estimates is provided by plots of constant χ^2 in parameter space. This technique was first applied to X-ray astronomy spectra by Lampton *et al.* (1971), and has been adopted by other groups (Bunner *et al.* 1972; Lea *et al.* 1973; Davison *et al.* 1975; Wolff, Helava, and Weisskopf 1975; Kellogg, Baldwin, and Koch 1975). It is important to establish the reliability of such constructions. Equivalently, we wish to construct a region that encloses the true value of the parameters, known to nature but not to the experimenter, a quantitatively predictable fraction of all experiment repetitions. We refer to this fraction as the "confidence level" of the construction method. Because the experiment is conducted only once, we are forced to focus our attention on analyzing one discrete section of data. Hence we will usually have available as our experimental result only one such contour. However, if this contour is constructed according to a prescription which yields a calculable confidence level, we regard this confidence as equivalent to the probability that the one available contour encloses the true parameter set. We emphasize, however, that the region is a statistic constructed from the data; its properties, including its reliability, can be computed from probability theory.

Confidence regions have received careful treatment in the literature (Cramer 1946, p. 511; Brownlee 1965, p. 121; Eadie *et al.* 1971, p. 200). There are in principle infinitely many ways to construct regions at a given confidence level. These methods give regions which can differ substantially in size, shape, and location. It is of course desirable that the region be as small as possible while achieving the required confidence level. In nonlinear estimation there is no satisfactory general way to construct minimum-volume confidence regions. Covariance matrices, for example, do not adequately describe errors in nonlinear parameter estimation. However, with the prescription given below, regions can be constructed which do have usefully small parameter-space volumes and which can be chosen to correspond to a given confidence level.

Before considering methods for constructing general confidence regions for parameters in nonlinear models, we shall discuss the simple example of a linear, independent, two-dimensional confidence region. Suppose some quantity X is measured by one sample x known to have normal errors and variance σ^2. Then with 68 percent confidence, one could claim

$$x - \sigma < X < x + \sigma .$$

We emphasize that the true value X is an unknown number and not a statistic; it is the interval $(x - \sigma, x + \sigma)$ which is the statistic, and the confidence level is the probability of the "plus and minus one sigma" construction procedure being successful. We could also have claimed

$$x - 0.47\sigma < X < \infty$$

with 68 percent confidence. There are infinitely many procedures which give a 68 percent interval for X. In this linear normal model, the symmetric interval is shortest (Brownlee 1965, p. 126).

Now if a second quantity Y is to be estimated from a sample y having normal errors independent of x and variance σ^2, one could claim

$$y - \sigma < Y < y + \sigma$$

with 68 percent confidence. The truth of this statement is independent of the truth of any claims about X. The following statements can be made about X and Y jointly: (1) The square $x \pm \sigma$, $y \pm \sigma$ contains X, Y with 46 percent confidence, (2) The rectangle $x \pm \sigma$, $y \pm \infty$ contains X, Y with 68 percent confidence, (3) The half-plane $x(+\infty, -0.47\sigma)$, $y \pm \infty$ contains X, Y with 68 percent confidence, (4) The square $x \pm 1.36\sigma$, $y \pm 1.36\sigma$ contains X, Y with 68 percent confidence, (5) The circle centered on x, y having radius 1.52σ contains X, Y with 68 percent confidence.

The above statements illustrate the fact that, even in this linear independent-error example, joint confidence regions are larger than single-parameter intervals having equal confidences. The reason for this growth of interval lengths is simply that, in estimating many parameters simultaneously, there are many possible coordinates in which the region may fail to enclose the true value (Eadie *et al.* 1971, p. 197).

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1976ApJ...208..177L

These statements also illustrate the variety of equal-confidence regions which may be constructed, and show that if some parameter is not of interest, it may be possible to weaken its constraint and correspondingly tighten the constraint on the remaining interesting parameter(s). However, when the errors are correlated, this separation is not trivial, because uncertainties in the uninteresting parameters contribute to the uncertainties in the parameters sought.

Confidence regions can also be constructed in such a way as to include irrelevant statistical variables. By an irrelevant variable we mean one which affects the contour but which is not affected by, nor can be used to infer, any of the desired parameters. For example, suppose that a procedure constructs a 10 percent confidence region from a data set, and that another procedure constructs a 90 percent confidence region from the same data. Suppose that then a fair coin flip is used to select one of these regions. The overall procedure yields a 50 percent confidence region. Such constructions are unsatisfactory because the inclusion of the irrelevant variable makes the inferred parameter ranges more variable and hence less trustworthy than they might be in any one experiment. It is obvious that irrelevant variables should be avoided where possible.

b) A Correct Procedure for X-Ray Astronomy

We have noted in § II*a* that for a correct model with the true (but usually unknown) parameter values, the S statistic S_{true} will be χ^2 distributed with N degrees of freedom. We have noted in § II*b* that the best-fitting adjustment of the p parameters yields an $S_{\min}$ statistic which will be χ^2 distributed with $N - p$ degrees of freedom. The argument given in the Appendix establishes the difference of these, $\Delta S \equiv S_{\text{true}} - S_{\min}$, as being distributed as χ^2 with p degrees of freedom, and proves that the ΔS variate is independent of $S_{\min}$. This derivation is closely related to the addition theorem for independent χ^2 variates (Cramer 1946, p. 234),

$$\chi^2_{N-p} + \chi^2_p \sim \chi^2_N ,$$

where $\sim$ denotes "is distributed as."

Our remarks provide a prescription for finding a confidence region for the parameters, whenever the form of the model is correct. Imaging repeating an experiment whose data analysis procedure consists of exhibiting the p-dimensional parameter space region R for which $S - S_{\min}$ is less than some fixed limit T. In some iterations, R will include the true parameter point; others, not. But because the true parameter $\Delta S \sim \chi^2_p$, the confidence is

$$C = \int_0^T f(\chi^2) d\chi^2 ,$$

where f is the density of the χ^2_p distribution. This establishes R as a confidence region. Its shape cannot be arbitrarily chosen as in the preceding simple example because its extent is determined strictly by the (one-dimensional) ΔS distribution. However, this construction method has the advantage of nowhere relying of the linearity, separability, or independence of the model parameters. It therefore provides exact confidence regions even in nonlinear X-ray astronomy applications.

We have noted that the ΔS statistic is independent of the $S_{\min}$ statistic. Indeed, ΔS has only as many degrees of freedom as the parameter vector has dimensions. This situation assures us that irrelevant variables (e.g., components of $S_{\min}$) have been excluded from the confidence region construction procedure. The actual value of $S_{\min}$ is irrelevant to the determination of the parameters' values because, at $S_{\min}$, all partial derivatives of S with respect to the parameters vanish. Indeed, the $S_{\min}$ distribution is independent of the parameter vector (Fisher 1925, p. 99). Information as to the size, location, and shape of the parameter confidence region must come from the variations of S in parameter space rather than from the value of $S_{\min}$.

In terms of the S statistic (eq. [2]), the recommended procedure can be established by the following argument. By

$$\Delta S \sim \chi^2_p$$

we mean for any number T

$$\text{Prob}\,(\Delta S > T) = \text{Prob}\,(\chi^2_p > T) .$$

With the limiting contour value S_L defined as $S_{\min} + T$,

$$\text{Prob}\,(\Delta S > S_L - S_{\min}) = \text{Prob}\,(\chi^2_p > S_L - S_{\min}) ,$$

or, from the definition of ΔS,

$$\text{Prob}\,(S_{\text{true}} > S_L) = \text{Prob}\,(\chi^2_p > S_L - S_{\min}) .$$

The left-hand side is just the probability α of the contour failing to enclose the true value, hence

$$\alpha = \text{Prob}\,(\chi^2_p > S_L - S_{\min}) .$$

But, the α-point of the χ^2 distribution is defined by

$$\alpha \equiv \text{Prob}\,[\chi^2_p > \chi^2_p(\alpha)] ,$$

so that $S_L - S_{\min}$ is exactly the α-point of χ^2_p, and the required contour for significance α is exactly

$$S_L = S_{\min} + \chi^2_p(\alpha) . \qquad (5)$$

In this expression, $\chi^2_p(\alpha)$ is the tabulated value of the χ^2 distribution for p degrees of freedom and significance α . We reiterate that the value of p to be used in this expression is the number of parameters adjusted in minimizing S. The p-dimensional parameter region for which S is less than S_L would enclose the true parameter vector in $1 - \alpha$ of all experiments. Equivalently, any one observation's contour has a confidence $C = 1 - \alpha$ of enclosing the true parameter vector.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1976ApJ...208..177L

TABLE 1

VALUES OF THE TERM $\chi^2_p(\alpha)$

SIGNIFICANCE α	CONFIDENCE C	χ^2 $p = 1$	$p = 2$	$p = 3$	$p = 4$
0.80.....	0.20 "0.25σ"	0.06	0.45	1.00	1.65
0.32.....	0.68 "1σ"	1.00	2.3	3.5	4.7
0.10.....	0.90 "1.6σ"	2.71	4.61	6.25	7.78
0.01.....	0.99 "2.6σ"	6.63	9.21	11.3	13.3

In Table 1 we list the value of the term $\chi^2_p(\alpha)$ for several cases of interest. Also shown in the table, for the purpose of comparison with commonly used nomenclature (Eadie *et al.* 1971, p. 190), is the number of standard deviations of the normal distribution equivalent to each of the values of α. These values are given in quotation marks to emphasize that this problem does *not* involve normally distributed parameter estimators.

We note several interesting features of the table. For theories with one free parameter, a 68 percent confidence interval may be generated at

$$S_L = S_{\min} + 1.00 .$$

Therefore equation (5) reduces in the limit $p = 1$ to the commonly used formula. As pointed out by Margon *et al.* (1975*b*), however, for the typical X-ray astronomy analysis one has $p = 3$, and therefore to jointly estimate these three parameters with 68 percent confidence,

$$S_L = S_{\min} + 3.5 ;$$

or to the more realistic confidence level of 90 percent,

$$S_L = S_{\min} + 6.25 .$$

Kellogg, Baldwin, and Koch (1975; hereafter KBK) have suggested that $S_{\min} + 1$ is appropriate for the joint estimation of several parameters. Here we note that our analytic derivation does not support the use of $S_{\min} + 1$ for problems in which several parameters are simultaneously estimated. Indeed, for $p = 3$, Table 1 shows that $S_{\min} = 1$ regions have only 20 percent confidence; only one in five such regions encloses the true value. However, equation (5) provides an equally simple prescription that can be correctly applied for any desired values of p and α.

It is important to bear in mind that statistical parameter estimation methods cannot be expected to function reliably in the presence of large systematic errors. One should proceed to estimate parameters of a model only after the model has been found to give acceptable fits. In the present context, this is equivalent to requiring that $S_{\min}$ be reasonably ranked in the χ^2_{N-p} distribution, i.e., that $S_{\min}$ not greatly exceed $N - p$.

We should explicitly point out that the formalism developed here has a much wider range of application than just the fitting of energy spectra. It can be applied to any binned Poisson accumulations for which the model prescribes the F_i, as long as all $F_i \gg 1$ so that deviations are closely Gaussian. There is no restriction on N, except that it must exceed p. One application to X-ray astronomy is the testing of models of spatial source structure using binned angular scan data (Lea *et al.* 1973). Another would be a test for the presence of a source with background in an on-source off-source measurement. A joint hypothesis about source and background spectra can be tested with a χ^2 statistic similar to equation (2) but containing not only background-plus-source terms but also background-alone terms. In this fashion, the problem of uncertain constants may be dealt with.

Tests for statistically significant differences in parameter values obtained from two or more data sets may be carried out with the formalism presented here. If the confidence regions overlap, there is of course no reason to reject the hypothesis of parameter constancy. Indeed, their overlap is a set of parameter values which reconcile all the data. Suppose, however, that two confidence regions having significances α are disjoint. Under the best-fitting constancy hypothesis, this circumstance could occur with probability α. The policy of rejecting the constancy hypothesis whenever the regions are disjoint would thus fail in a fraction α of all experiment repetitions. So, we regard the policy of rejecting parameter constancy when regions are disjoint, as having the same confidence as the regions, i.e., $1 - \alpha$.

Recently, Wolff, Helava, and Weisskopf (1975) have obtained X-ray spectra of the Perseus cluster of galaxies, and argue that the spectral parameters of NGC 1275 differ significantly from those of the remainder of the cluster, because the best-fit points differ. We caution that best-fit points are not suited to the task of hypothesis testing; in fact, a best-fitting point is a zero-confidence region whose location is a statistic dependent upon fluctuations in the data. Since the contours shown by Wolff *et al.* overlap even at the $S_{\min} + 1$ level, there is no need to postulate differing spectra based on their parameter contours.

c) *Likelihood Ratio Methods*

A related method of parameter estimation is based on the extent of the likelihood function in parameter space. In X-ray astronomy, likelihood functions have been recommended (e.g., Hearn 1969) and used (e.g., Giacconi *et al.* 1972, 1974) for parameter error estimation. We briefly sketch the likelihood ratio

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1976ApJ...208..177L

method here. For Gaussian deviations, the likelihood function

$$L = \prod_{i=1}^{N} \exp\left[-\frac{1}{2}\frac{(D_i - F_i)^2}{\sigma_i^2}\right]$$

is proportional to the probability that a model with arbitrary parameter values gives the observed data. To approximate a Poisson process, we set the variances $\sigma_i^2 = F_i$. For a given set of data, L can be considered to be strictly a function of the adjustable parameters of the model, through the F_i parameter dependences. By comparison with equation (2), we note that $L = \exp(-\frac{1}{2}S)$; hence the parameter set which minimizes S, maximizes L. This maximum-likelihood estimator is, of course, unlikely to be the true parameter value. Its error may be expressed by constructing the region in parameter space for which the likelihood ratio $\lambda \equiv L/L_{\max}$ exceeds some cutoff value λ_{crit} (Brownlee 1960, p. 113). This region will enclose the true parameter value in a predictable fraction of all experimental repetitions, and indeed its boundary coincides with the surface of constant S, with $S = S_{\min} - 2\ln\lambda_{\text{crit}}$. The likelihood ratio method is thus equivalent to the χ^2 procedure summarized in equation (5), as noted by Cramer (1946, p. 426), and can be employed to construct parameter regions with any desired confidence level by choosing $\lambda_{\text{crit}}(\alpha) = \exp[-\frac{1}{2}\chi^2{}_p(\alpha)]$, as recommended by Eadie *et al.* (1971, p. 207).

d) Reduced Numbers of Parameters

The foregoing formalism provides a straightforward method to jointly estimate values of all parameters in fitted functions. For some purposes, however, it is desirable to obtain constraints on some subset of those parameters, without regard to the remaining ones. Examples of this situation are found wherever relevant but uninteresting variables (e.g., background levels) affect results, or where for display purposes the parameter region must be reduced to a two-dimensional contour.

A rigorously valid constraint on the parameter subset values can be obtained by projecting the original p-dimensional confidence region onto the desired subspace of, say, q parameters. The confidence of the projected region is at least as high as that of the original region, because it contains the original region as a subset. Consequently the exactly known confidence of the original region is a definite lower limit to the confidence of the projected region.

The projection method may be implemented in the following way. First, a grid of parameter points is established in the desired subspace. Then, at each point in the grid, all $p - q$ remaining parameters are adjusted to minimize S there. On this grid, the set of parameter values having $S < S_L$ will contain the true value with confidence $C \geq 1 - \alpha$, where S_L is taken to be $S_{\min} + \chi^2{}_p(\alpha)$ in accord with equation (5).

A second method can be used to construct exact confidence regions for q of the p parameters when the model is linear, i.e., when the F_i are linear functions of the parameters. In this case, the p-dimensional covariance matrix V_{pp} provides a complete description of the fitting errors. The covariance matrix for the desired parameters, V_{qq}, is obtained by simply deleting the unwanted rows and columns from V_{pp} (Eadie *et al.* 1971, p. 199). The contribution to ΔS from random errors in the q parameters α is just the quadratic form $\alpha^T V_{qq}{}^{-1}\alpha$, which is distributed as χ^2 with q degrees of freedom (Eadie *et al.* 1971, p. 63; Brownlee 1965, p. 277). Consequently a joint confidence region for the desired q parameters, independent of the values of the remaining $p - q$ parameters, is the set of all points for which $S < S_L$ with $S_L = S_{\min} + \chi^2{}_q(\alpha)$. The construction of this region can proceed as in the first method, with a q-dimensional grid of trial parameter values and with S minimized at each point with respect to the remaining $p - q$ variables. Nothing can, of course, be said about these remaining parameters; thus this is a specialized technique which has not commonly been invoked in X-ray astronomy analyses.

We note that this method must be used with great caution because its derivation explicitly depends upon model linearity. In X-ray astronomy, spectral parameters fix predicted count rates through highly nonlinear functions, and we should not in general expect a linear matrix decomposition to satisfactorily remove unwanted parameter constraints. Specifically, when parameter-space constant-S contours deviate appreciably from ellipsoids, substantial model nonlinearity is indicated. In such cases, projection of the full $\chi^2{}_p$ region gives a secure upper limit to the joint errors in the q parameters. Nonetheless, the nonlinear numerical simulations conducted by Avni (1976) show that the confidence level achieved by projection onto the subspace of q parameters is accurately given by the $\chi^2{}_q$ function.

A third method for reducing the number of constrained parameters is possible whenever parameter intervals can be independently constructed. Specifically, if the parameters to be fitted can be chosen in such a manner as to permit factoring the likelihood function, e.g.,

$$L(a, b, c) = L_a(a)L_{bc}(b, c)\,,$$

or (equivalently) decomposing the χ^2 statistic,

$$S = S_a(a) + S_{bc}(b, c)\,,$$

then the separated parameters may be assigned confidence intervals independently of constraints on the remaining parameters. In this example, an interval for the a parameter would be the interval for which $S < S_L$ with $S_L = S_{\min} + \chi^2{}_1(\alpha)$. Similarly, a joint confidence region for the parameters b and c is the $S < S_L$ region with $S_L = S_{\min} + \chi^2{}_2(\alpha)$.

We remark that this separation method is usually inapplicable to X-ray astronomy analyses, due to the complicated interdependences of scientifically useful parameters. In some cases parameters may be concocted which are reasonably independent of the

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1976ApJ...208..177L

others; examples are the use of total count rate (Giacconi *et al.* 1972) or 1–10 keV fluxes (Bowyer *et al.* 1970) in specifying X-ray intensities. However, the intensity coefficient of a spectrum is not generally separable from the other parameters in spite of the fact that the fitting function is linearly dependent on it. This problem is due to the coefficient estimator being highly correlated with the other parameters. If its value is needed, it must be estimated jointly with the other parameters.

IV. NUMERICAL SIMULATIONS

In Margon *et al.* (1975*a*, *b*) we described several numerical simulations designed to measure the confidences of regions constructed according to various $S_{\min} + T$ prescriptions. The purpose of conducting those simulations was to verify that equation (5) is applicable to the specific nonlinear models used in X-ray astronomy spectrum analysis. They involved picking models by assigning parameter values to an equation of the form

$$I(E) = aE^{-c} \exp(-bE^{-3}) \tag{6}$$

where a, b, and c set the intensity, absorption, and index of the hypothetical source. The simulations established that with T chosen to be 1.00, 3.50, and 6.25, the reliability of the contours was 17 ± 10 percent, 64 ± 7 percent, and 84 ± 4 percent, in excellent agreement with the χ^2_3 distribution appropriate to this example (eq. 5).

A more realistic simulation could be carried out by including the effects of the detector and pulse height analyzer on the observed count distribution. Although we have no reason to believe that the theory of § II should not be valid in both cases, the latter clearly offers a more unambiguous test of the technique.

We have therefore generated data for this simulation by calculating a nominal response curve for a gas proportional counter of a type employed by numerous experimenters. We have assumed a gas mixture of argon and methane at surface densities of 4.3 and 0.2 mg cm^{-2}, respectively, and a $\frac{1}{8}$ mil Mylar window with a conductive Nichrome coating.

One of the more mathematically interesting properties of a proportional counter is its finite energy resolution. An incoming monochromatic photon flux is detected as a series of ion pulses having a distribution of different pulse heights. This convolution could conceivably have an effect on the error estimation process, so we have included it in the simulated observations. We have used a Poisson energy resolution kernel of the form

$$R(E, Y) = \frac{\exp(-E/Q)}{Q} \frac{(E/Q)^{Y/Q}}{\Gamma(Y/Q + 1)}, \tag{7}$$

where E is the incident photon energy, Y the resulting pulse height, and Q the mean ionization energy per primary electron. This function has been shown experimentally to be a good approximation to actual proportional counter pulse-height distributions (Campbell and Ledingham 1966). For most of the simulations, unless otherwise stated, we have set $Q = 0.043$ keV, which yields a resolution of 20 percent FWHM at the ^{55}Fe line at 5.9 keV.

The simulations proceeded as follows. An incident spectral function such as equation (6) is chosen, and realistic values are selected for the free parameters. This then completely defines the "true" spectrum. This spectrum is then convolved with the counter response and resolution functions to yield an energy-dependent pulse-height spectrum of the form

$$F_i = \int_0^\infty \int_{Y_i}^{Y_{i+1}} I(E)\eta(E)R(E, Y)dEdY, \tag{8}$$

where $\eta(E)$ is the counter response function defined above, Y_i and Y_{i+1} are the lower and upper energy limits of each pulse-height interval, and the other symbols have been previously defined. The pulse-height intervals have been selected to be 1 keV wide, centered on energies from 1 to 7 keV. The normalization coefficient in the function $I(E)$ was normally selected to yield hundreds or thousands of counts per pulse-height channel, i.e., similar to that obtained from a typical satellite or long rocket exposure.

The iterative simulation is now ready to procede. Each iteration consists of applying a Poisson count rate fluctuation to each of the seven F_i. This is done by calling a random number generator that supplies a normally distributed random number with mean and variance both equal to F_i. Define the resulting perturbed count rate spectrum as D_i; for each iteration it represents the results of one "experimental measurement" of the true pulse-height spectrum. For this iteration we may now immediately calculate S_{true}, the value of the statistic S in equation (2) for the true free parameter values, from the expression

$$S_{\text{true}} = \sum_{i=1}^{N} \frac{(D_i - F_i)^2}{F_i}. \tag{9}$$

Note that the σ_i^2's are known precisely since we specified them to the random number generator.

We now determine the best-fit free-parameter value for this iteration by varying the free parameter values about the known true values, minimizing equation (2) with a steepest-descent algorithm. Minimization continued until successive values of S differed by less than one part in 10^4. The result is a value for $S_{\min}$ and the associated values of the free parameters. It is now possible to calculate the quantity $S_{\text{true}} - S_{\min}$ needed to estimate the ΔS distribution. This completes one iteration of the simulator; the values of S_{true}, $S_{\min}$, and the best-fit parameters are stored. For each choice of an initial function $I(E)$, 100 to 150 such iterations were performed.

At the conclusion of the simulation, certain summary statistics are computed. The mean and sample standard deviation of the stored values of S_{true}, $S_{\min}$, and $S_{\text{true}} - S_{\min}$, are derived, and compared against the known mean and standard deviation for the χ^2 distribution with N, $N - p$ degrees of freedom,

© American Astronomical Society • Provided by the NASA Astrophysics Data System

Table 2

Results of Numerical Simulations

Model	A		B		C		D		E	
No. of Iterations	113		152		111		98		134	
	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed
Mean S_{true}	7.00	6.80±0.35	7.00	6.78±0.31	7.00	6.02±0.30	7.00	6.66±0.32	7.00	7.25±0.35
S.D. S_{true}	3.74	3.70±0.25	3.74	3.83±0.22	3.74	3.17±0.21	3.74	3.13±0.23	3.74	4.08±0.25
Mean S_{min}	4.00	3.83±0.24	4.00	3.89±0.23	4.00	3.49±0.21	4.00	3.97±0.25	4.00	4.21±0.28
S.D. S_{min}	2.83	2.60±0.17	2.83	2.80±0.16	2.83	2.21±0.16	2.83	2.43±0.18	2.83	3.19±0.20
Mean $(S_{true}-S_{min})$	3.00	2.97±0.26	3.00	2.89±0.22	3.00	2.52±0.19	3.00	2.69±0.24	3.00	3.05±0.20
S.D. $(S_{true}-S_{min})$	2.45	2.80±0.19	2.45	2.72±0.16	2.45	1.96±0.13	2.45	2.40±0.17	2.45	2.37±0.15
Mean a_{min}	50,000	50179±105	50,000	50110±95	50,000	50086±54	3000	3012±16	50,000	49956±171
Mean b_{min}	4.60	4.61±0.01	4.60	4.60±0.006	4.60	4.61±0.01	4.60	4.62±0.03	4.60	4.60±0.01
Mean c_{min}	1.50	1.50±0.001	3.00	3.00±0.004	400	1150±134	100	460±65	3.0	3.008±0.008
Fraction of $(a,b,c)_{true}$ within $S_{min}+6.25$	0.90	0.90±0.03	0.90	0.91±0.02	0.90	0.94±0.02	0.90	0.92±0.03	0.90	0.87±0.03
Fraction of $(a,b,c)_{true}$ within $S_{min}+3.5$	0.68	0.73±0.04	0.68	0.74±0.04	0.68	0.78±0.04	0.68	0.73±0.04	0.68	0.64±0.04
Fraction of $(a,b,c)_{true}$ within $S_{min}+1$	0.20	0.21±0.04	0.20	0.23±0.03	0.20	0.20±0.04	0.20	0.23±0.04	0.20	0.15±0.03

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1976ApJ...208..177L

respectively. A histogram of the number of occurrences of each value range for each of the S-statistics is also plotted, and compared with the appropriate χ^2 distribution function, to ensure that the higher moments of the distributions are similar. The means and standard deviations of the free parameter values are computed, and compared with the known true parameter values. Finally, the fraction of $S_{\rm true} - S_{\rm min}$ which are less than 6.25, 3.5, and 1 were calculated to compare with Table 1. Note that if $S_{\rm true} - S_{\rm min} \leq T$ for any value of T, then a contour in S-space drawn about $S_{\rm min}$ of size $S_{\rm min} + T$ will encompass $S_{\rm true}$. Thus this calculation is the simplest way to verify that a test contour does or does not encompass the true values, without the necessity of graphically constructing the entire contour in parameter space. Note that the procedure of verifying the contour volume in S-space, i.e., asking whether $S_{\rm true}$ lies in the interval $[S_{\rm min}, S_{\rm min} + T]$, is identical to performing this construction in the fully three-dimensional parameter space because of the correspondence given in equation (2).

The results for each phase of the simulation [each choice of an $I(E)$] are given in Table 2. We have taken some care to compute errors on the values of statistics and quantities derived from the simulation, so that they may be properly compared with expected values. These errors were computed as follows:

a) *Mean* $S_{\rm true}$, *Mean* $S_{\rm min}$, *Mean* $S_{\rm true} - S_{\rm min}$.—Error quoted is the standard deviation of the mean of the list of statistics generated in the simulation.

b) *S.D.* $S_{\rm true}$, *S.D.* $S_{\rm min}$, *S.D.* $S_{\rm true} - S_{\rm min}$.—Error quoted by calculating the normal error in n independent samples of the variance of a Gaussian random variable. The fractional error in such a variance estimate is $(2/n - 1)^{1/2}$.

c) *Mean* $a_{\rm min}$, *Mean* $b_{\rm min}$, *Mean* $c_{\rm min}$.—Error quoted is the standard deviation of the mean of the list of parameter values that minimized S for each iteration.

d) *Fraction of* $(a, b, c)_{\rm true}$ *within* $S_{\rm min} + T$, *for* $T = 6.25, 3.5, 1$.—Error quoted is the Bernoulli error for the known number of iterations.

We now discuss each phase of the simulation separately.

a) *Power-Law Model with Attenuation*

The function given in equation (6) was used with $a = 50{,}000$, $b = 4.6$, $c = 1.5$. The value of b was chosen to give significant attenuation in the low-energy channels, and the value of the slope c is typical of that found in many celestial X-ray sources. A total of 113 iterations yielded the results in Table 2. The values of the statistics and free parameters are precisely as predicted to within the errors; a contour of size $S_{\rm min} + 6.25$ enclosed 90 percent of the true values, as predicted. In Figure 1 we present histograms of the values of the three S-statistics obtained in the simulation, and appropriate χ^2 distribution functions, to show that the higher order moments of the distributions are also in agreement with theory.

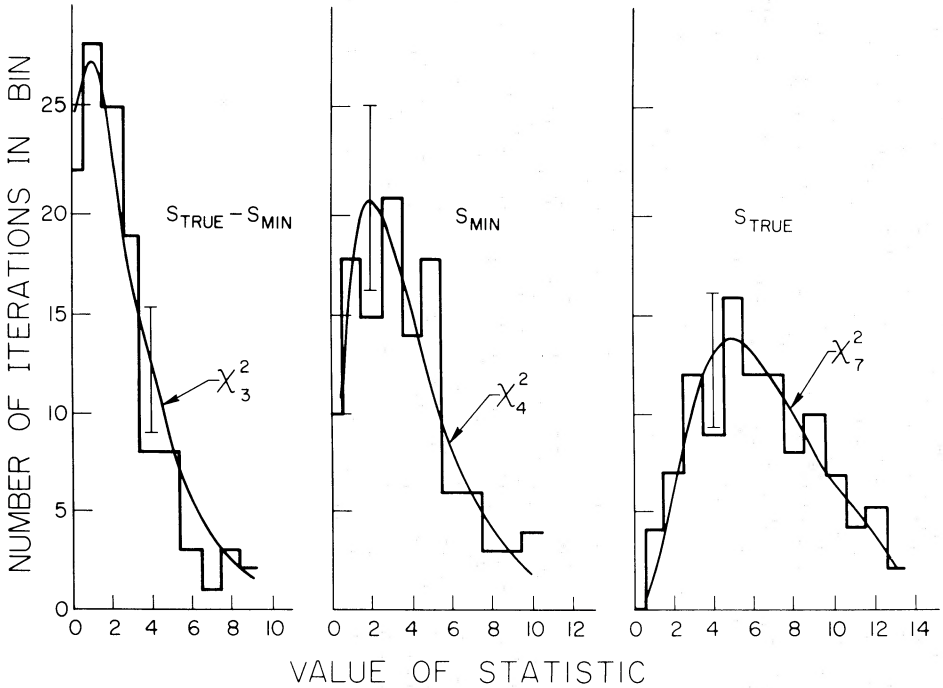

FIG. 1.—Histograms of the observed distributions of the statistics $S_{\rm true} - S_{\rm min}$, $S_{\rm min}$, and $S_{\rm true}$, for the 113 iterations of fits to power-law parent spectra described in § IV*a*. Also shown are the probability densities of the χ^2 distributions having 3, 4, and 7 degrees of freedom, normalized to 113 trials.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1976ApJ...208..177L

b) Exponential with Attenuation

To simulate thermal emission from a thin plasma, a function of the form

$$I(E) = aE^{-1} \exp(-E/c) \exp(-bE^{-3}) \quad (10)$$

was utilized in equation (8). The free parameters were set to resemble a 3 keV plasma, with moderate attenuation as in case A. After 152 iterations, Table 2 indicates that the distributions are again as predicted.

c) Highly Nonlinear Exponential with Attenuation

To create an extremely nonlinear case to test our technique, we repeated phase B with $c = 400$, i.e., the equivalent of a 400 keV plasma. When measured with an instrument with a bandpass of 1–7 keV, relatively large fluctuations of the value of c about c_{true} should have little effect on the derived spectrum. Thus the generated contours in parameter space should usually be open-ended in the direction of large c. The results in Table 2 show that the S-statistics still behave as we predict, after 111 iterations. The derived values of a and b also bracket the true values. The observed mean for c is 1150, substantially different from $c_{\text{true}} = 400$ and indicating that the desired nonlinearity has been achieved. As expected, continued iterations yield values for c in the range $\sim 10^2$ to infinity with relative uniformity. The important result to note (Table 2) is that even in this highly nonlinear situation, the contours drawn according to equation (5) still encompass the true parameter values the correct fraction of times, even though the average best-fit parameter might differ from its true value. This fact reinforces the utility of regarding the confidence region, and not the best-fit point, as the essential result of a data analysis procedure.

d) Low-Count-Rate Exponential

In a further effort to create a highly nonlinear problem, equation (10) was used again as the incident spectrum, but with $a = 3000$, $b = 4.6$, and $c = 100$. A 100 keV plasma with this value of normalizing coefficient a and the assumed instrument response function yields a very small total number of counts per channel, the least efficient channel containing only 16 detected counts. The results in Table 2 again show that the desired nonlinearity was achieved (with mean $c_{\text{min}} = 460$ while $c_{\text{true}} = 100$) and that the contour prescription is still an effective one, since the predicted fraction of free parameter values still fall within the calculated contours.

e) Instrumental Energy Resolution

One might ask whether the finite energy resolution of proportional counter instrumentation may invalidate or modify the formalism derived here, because the strong physical correlation between neighboring pulse-height channels may reduce the effective number of independent degrees of freedom in the problem. It certainly is true that an intense monochromatic flux in one channel will cause a correlated increase in counts in neighboring channels, independent of the true flux in those channels. However, this does not alter the number of statistical degrees of freedom in the problem. This quantity is determined solely by the number of independent samples of the randomized spectrum available to the observer, regardless of the physical process or algorithm which generated the data prior to its perturbation.

To numerically verify our conclusion, we have examined several phases of our simulation relevant to the energy resolution problem. First, phases (*a*)–(*d*) above, which use a finite resolution comparable to actual detectors, may be compared with the simulations reported by Margon *et al.* (1975*a*, *b*), which ignored resolution (and were thus equivalent to an experiment with perfect resolution). In all cases, equation (5) correctly sized the relevant contour. To further pursue this point, however, we conducted another simulation identical to phase (*b*), but with a value $Q = 0.3$ keV inserted in the energy resolution kernel, equation (7). This is equivalent to an energy resolution of roughly 53 percent at ^{55}Fe, far worse than would be tolerated in an actual experimental situation. The results in Table 2 indicate that the desired effect was achieved, because the errors on the derived values of a_{min}, b_{min}, c_{min} are roughly twice as large as for case (*b*), which is identical to the present case except for better resolution. This larger scatter in the derived values indicates that the degraded resolution evidences itself in a less precise determination of the free-parameter values, as we would expect. Thus the contours resulting from these experiments have more volume in *parameter* space than those for case (*b*). However, contours drawn in *S-space* by equation (5) still enclose the predicted fraction of the true free-parameter values. This difference is of course due to the much slower variation of S with parameter value in the present case. Thus we have demonstrated that our formalism is independent of the magnitude of energy resolution of the experiment, as we predicted on theoretical grounds.

V. THE SIMULATIONS OF KELLOGG *et al.* (*1975*)

Kellogg, Baldwin, and Koch (KBK) (1975) have reported results of numerical simulations which they feel support the use of $S_{\text{min}} + 1$ to provide an $\alpha = 0.32$ contour for the case of $q = 2$ and $p = 3$. In this section we describe their simulation and discuss their results.

The KBK simulation method is the following. One spectrum of actual experimental data is first selected, containing (unknown) Poisson fluctuations about the (unknown) true spectrum. A single best-fit parameter set is then calculated, using a least-squares grid technique. On this grid, a single $S_{\text{min}} + 1$ contour is drawn. The iteration scheme consists of randomizing the best-fit spectrum and locating a new best-fitting parameter set, to determine if it is enclosed by the

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1976ApJ...208..177L

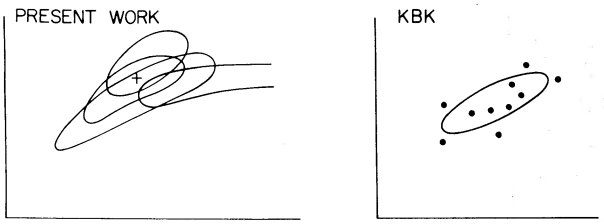

FIG. 2.—Schematic illustration of the simulations reported here, in which numerous confidence regions are tested to find their probability of containing the known true parameter value (*cross*). Also sketched is the method of Kellogg *et al.* (1975; KBK), where perturbed best-fit points are compared with a single parameter region. In both cases, the axes represent derived values of any two free parameters.

one original experimental contour. The fraction of these successes was taken to be the confidence of the contour construction method.

The conceptual difference between our simulations and those reported by KBK is shown schematically in Figure 2. We have generated a great many three-dimensional contours about one known "true"-value point in parameter space, to assess the probability that any one contour contains this known point. KBK have calculated a single two-dimensional contour about one value of $S_{\min}$; this contour may or may not contain S_{true}. They then calculate the fraction of further best-fit values falling in this one contour, a quantity which is unrelated to the probability that the contour drawing technique is yielding a fixed, desired confidence level. The initial contour is not tested, because the true (astronomical) parameter values and indeed the correct fitting functions are unknown. The subsequently randomized fits are of course not true parent spectra for the initial contour, so they do not test its confidence either.

Since we have provided an analytic derivation and extensive simulations whose results are in agreement, while KBK have no derivation and simulations which do not appear to establish the reliability of their method, we feel that we are justified in concluding that equation (5) gives the correct confidence region for the joint estimation of p parameters.

VI. THE RATIO-OF-VARIANCES METHOD

Some recent X-ray astronomy papers (Charles, Culhane, and Zarnecki 1975; Fabian 1975) have employed the ratio-of-variances method (Draper and Smith 1966; Cline and Lesser 1970) to construct multiparameter confidence intervals. The method is appropriate in applications where a function F_i is to be fitted to data D_i subject to errors having variances which are neither known nor fixed by the model, but which must be estimated from the data. If the errors have a common variance σ^2 and are Gaussian, one may proceed in the following way. The unnormalized sum of squared errors[2] is formed from the N data:

$$\mathcal{S} = \sum_1^N (D_i - F_i)^2 .$$

Then, by adjusting the p parameters of the model, one finds the minimum value of $\mathcal{S}$. These situations now hold: (*a*) $\mathcal{S}_{\min}$ is in principle distributed as $\sigma^2\chi^2_{N-p}$; (*b*) $\Delta\mathcal{S} \equiv \mathcal{S}_{\text{true}} - \mathcal{S}_{\min}$ is in principle distributed as $\sigma^2\chi^2_p$, independently of $\mathcal{S}_{\min}$. It follows that the ratio of variances is independent of σ^2 and is

$$\frac{\mathcal{S}_{\text{true}}}{\mathcal{S}_{\min}} \sim 1 + \frac{\chi^2_p}{\chi^2_{N-p}} \sim 1 + \frac{p}{N-p} F(p, N-p)$$

where F is the Fisher-Snedecor distribution. A confidence region for the parameters is the parameter range having $\mathcal{S} < \mathcal{S}_L$ with

$$\mathcal{S}_L \equiv \mathcal{S}_{\min}\left[1 + \frac{p}{N-p} F(p, N-p, \alpha)\right] \quad (11)$$

in which F is chosen from the tables at the desired confidence level.

It is correct that equation (11) will generate a confidence contour at level $(1 - \alpha)$. In addition to the derivation provided above, we have numerically verified equation (11) using a simulation technique identical to that described earlier to check equation (5). For any one given iteration, equations (5) and (11) will yield different size contours; this is certainly not incompatible with their both providing the same confidence level. Over the ensemble of iterations, both will enclose the true parameter values the correct fraction of trials.

In the limit of infinite $N - p$ the residual variance $\mathcal{S}_{\min}/N - p$ approaches the true value σ^2. Also, $pF(p, N - p)$ approaches a χ^2_p distribution. It follows that in this limit a confidence region is defined by $\mathcal{S} < \mathcal{S}_L$ with

$$\mathcal{S}_L = \mathcal{S}_{\min} + \sigma^2\chi^2_p(\alpha) .$$

In this limit, the region is the same as is given by equation (5). Thus the confidence region construction methods approach equality as $N - p \to \infty$.

The ratio-of-variances method is self-consistent and has been recommended in a variety of applications where the variance is unknown (Beale 1960; Draper and Smith 1966, p. 292). Cline and Lesser (1970) have shown[3] that workers who choose to normalize $\mathcal{S}_{\min}$ to $N - p$ are actually using the ratio-of-variances method. That is, they point out that $(N - p)\Delta\mathcal{S}/\mathcal{S}_{\min}$ is not χ^2 distributed, but is instead distributed as $pF(p, N - p)$.

We wish to point out that the ratio-of-variances method is unnecessary and in fact undesirable in applications such as X-ray astronomy where term

[2] The formulation of Cline and Lesser (1970) allows for separate relative variances or weights for the N terms. We have adopted the simpler formulation of Draper and Smith (1966) to emphasize that the overall variance scale is a separately determined factor.

[3] The clarification noted by Lesser *et al.* (1972, p. 610) applies to the number of degrees of freedom in their particular problem, and not to the general formalism discussed here and in Draper and Smith (1966).

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1976ApJ...208..177L

variances are exactly predicted by the model. The ratio-of-variances presciption for the limiting contour contains the statistic $\mathscr{S}_{\min}$. This value is, however, irrelevant to the estimation of parameters for the same reason that $S_{\min}$ of § III*a* is irrelevant: it is unaffected by the parameters' values and so is of no help in estimating them. As a consequence, the contours produced are more variable than necessary and, as was discussed in § III, the best possible inferences about the parameters are not constructed. Further, there is another reason *not* to utilize equation (11). Because it normalizes the contour volume to the value $\mathscr{S}_{\min}$ for a given data sample, it will tend to provide larger contours for larger values of $\mathscr{S}_{\min}$, giving the erroneous impression that the technique is somehow compensating for the poor-quality fit of the best model. We stress again that if $S_{\min} \gg N - p$, *no* formalism which uses distributions describing random fluctuations can provide the proper error estimator. Thus this "normalizing" property of equation (11) is both illusory and misleading.

VII. SUMMARY

We summarize the chief conclusions of this paper as follows.

1. Equation (5) is the correct multivariate least-squares error estimator for problems of the type considered here, i.e., the joint estimation of p-parameter values with data variances given by the model.

2. Our numeric simulations indicate that on the ensemble average equation (5) yields the correct error estimator for the specific nonlinear equations of interest in X-ray astronomy. Regardless of the degree of nonlinearity introduced, there is no evidence of any departure of the actual uncertainties from those predicted. Thus we see no reason to avoid application of an analytic formalism to this problem on the grounds that it is too complex or generally insoluble.

3. We find no evidence, either analytic or numeric, that the $S_{\min} + 1$ technique is correct for problems in which two or more parameters are jointly estimated. Use of this technique in such cases will badly underestimate the errors. Specifically, in an analysis with three free parameters, $S_{\min} + 1$ regions will enclose the true values only one time in five.

4. The simulations reported by KBK in support of the use of $S_{\min} + 1$ are conceptually erroneous, in that they do not test the probability that any one such contour encompasses the true parameter values.

5. There is no evidence that the finite energy resolution of a gas proportional counter has any effect on the formalism derived here, in agreement with basic theoretical considerations.

6. The paper by Cline and Lesser (1970) was directed as a comment on a specific parameter estimation method which is unnecessary when the data variances are fixed by the model. The resulting equation (11) is formally correct but needlessly more complex and less reliable than equation (5). The normalization of equation (11) to the derived value of $\mathscr{S}_{\min}$ does not extend the validity of the analysis to cases where $\mathscr{S}_{\min} \gg N - p$. In such cases, the cause of the greater-than-Poisson fluctuations must be dealt with explicitly.

Financial support for this work was provided by the California Institute of Technology under a HEAO-A subcontract. We are pleased to acknowledge useful discussions with Drs. W. Baity, E. Beale, E. Boldt, A. Fabian, G. Garmire, E. Kellogg, D. Schwartz, and P. Serlemitsos. We thank Y. Avni for a preprint of his paper.

APPENDIX

THE $S_{\min}$ AND ΔS DISTRIBUTIONS

In this Appendix, we sketch the derivations of the $S_{\min}$ and ΔS distributions. The arguments are due to Fisher (1924) for the one-parameter case and to Cramer (1946) for the p-parameter case.

I. DEFINITIONS AND NOMENCLATURE

A function of the type

$$Q = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \sum_{k}^{N}\sum_{j}^{N} a_{jk}x_j x_k$$

is called a quadratic form (see Cramer, pp. 107–108). $\boldsymbol{A}$ is a symmetric $N \times N$ square matrix, and $\boldsymbol{x}$ is a vector of N components. If new variables $\boldsymbol{y}$ are introduced by a linear transformation $\boldsymbol{x} = \boldsymbol{C}\boldsymbol{y}$, the quadratic form above is equal to

$$Q = \boldsymbol{y}'\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C}\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{B}\boldsymbol{y}\,,$$

where $\boldsymbol{B}$ is the new matrix $\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C}$. The rank of a quadratic form is the smallest number of independent variables on which Q can be made to depend, via nonsingular transformations of $\boldsymbol{x}$ (equivalent to rotations in N-space; see Cramer, p. 110). Rank measures the minimum dimensionality of $\boldsymbol{A}$. Clearly, $0 \leq r \leq N$.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1976ApJ...208..177L

II. THE S_{true} STATISTIC

Consider the sum-of-squares function

$$S \equiv \sum_1^N (D_i - F_i)^2/\sigma_i^2 ,$$

where D_i are the data. If the true population mean and variance F_i^2 and σ_i^2 (known only to nature, unknown to the experimenter) were substituted, each term of the sum would be the square of an independent zero mean unit variance variable, x_i. In the limit of many events per bin these deviations are normally distributed, by the central limit theorem. With the true values of F_i and σ_i^2, S becomes

$$S_{\text{true}} = \sum^N x_i^2 = \boldsymbol{x}'\boldsymbol{x} ,$$

which is the quadratic form whose matrix is the $N \times N$ identity matrix I. Its rank is N. Its sampling distribution is the χ^2 distribution with N degrees of freedom (Cramer, p. 234 and p. 313).

III. THE S_{min} STATISTIC

S_{min} is another function of the same data:

$$S_{\text{min}} \equiv \sum^N \frac{(D_i - F_i)^2}{\sigma_i^2} = \sum^N y_i^2 = \boldsymbol{y}'\boldsymbol{y} ,$$

where each y_i is the deviation of the datum D_i from the best fitting model's prediction (Cramer, p. 427). By best fitting, we mean that the model's parameters have been adjusted to minimize S. Cramer (pp. 427–431) shows that, under remarkably general circumstances, such a best adjustment exists. The model predictions' parameter dependences are not assumed linear, but must be twice differentiable.

The y_i are not independent of each other, because the best fitting predictions f_i introduce data-dependent correlations between them.

Cramer then shows (pp. 431–434) that there exists a linear nonorthogonal transformation $\boldsymbol{A}$ such that $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ where the $\boldsymbol{x}$ are the independent true deviations. The rank of the quadratic form $\boldsymbol{y}'\boldsymbol{y}$ is shown to be smaller than the rank of $\boldsymbol{x}'\boldsymbol{x}$ by an amount p, the number of nonequivalent adjustable parameters. (Note that the transformation applies to the data vector, which is normally distributed, not to the parameter vector. Thus we have incorporated no assumptions as to model linearity.) This result establishes S_{min} as being χ^2 distributed with $N - p$ degrees of freedom, by Cramer's argument of page 313. This result has been widely cited (e.g., Cochran 1952). The result is important in hypothesis testing.

IV. THE ΔS STATISTIC

The statistic

$$\Delta S \equiv S_{\text{true}} - S_{\text{min}}$$

is of importance in parameter estimation. From the preceeding discussion, ΔS is a quadratic form

$$\sum^N x_i^2 - \sum^N y_i^2 ,$$

where the y_i are functions of the x_i. One way to evaluate this expression is to use Cramer's transformation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ described above, followed by his $\boldsymbol{K}$ transformation (pp. 433–434) which explicitly diagonalizes the $\Sigma\, y_i^2$ quadratic form in terms of $\boldsymbol{x}$; it gives $N - p$ nonzero diagonal elements, each equal to 1. This can be trivially subtracted from $\boldsymbol{x}'\boldsymbol{x}$ to reveal that the ΔS statistic contains exactly p independent squares, and that none of these contribute to the S_{min} statistic. Hence, ΔS is independent of S_{min}, as stated in § III*b*. By Cramer's argument of page 313, ΔS is χ^2 distributed with p degrees of freedom and is independent of S_{min}. In Cramer's notation, ΔS is the quadratic form

$$\boldsymbol{x}'\boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{H}\boldsymbol{H}'\boldsymbol{x} .$$

An alternate derivation, which does not require explicit use of the $\boldsymbol{K}$ transformation, can be constructed from Fisher's lemma (Fisher 1925; Cramer, pp. 379–381). This important proposition establishes that the difference between a sum of N squares and S squares of orthogonal functions of the variables is a quadratic form of rank $N - S$, is χ^2 distributed, and is independent of the sum of S squares.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

1976ApJ...208..177L

REFERENCES

Avni, Y. 1976, preprint, submitted to *Ap. J.*
Beale, E. M. L. 1960, *J. Roy. Stat. Soc.*, Ser B., **22**, 41.
Bevington, P. R. 1969, *Data Reduction and Error Analysis for the Physical Sciences* (New York: McGraw-Hill).
Bowyer, C. S., Lampton, M., Mack, J., and de Mendonca, F. 1970, *Ap. J.* (*Letters*), **161**, L1.
Brownlee, K. A. 1965, *Statistical Theory and Methodology* (2d ed.; New York: Wiley).
Bunner, A. N., Coleman, P. L., Kraushaar, W. L., and McCammon, D. 1972, *Ap. Letters*, **12**, 165.
Campbell, J. L., and Ledingham, K. W. D. 1966, *British J. Appl. Phys.*, **17**, 769.
Charles, P. A., Culhane, J. L., and Zarnecki, J. C. 1975, *Ap. J.* (*Letters*), **196**, L19.
Cline, D., and Lesser, P. M. S. 1970, *Nucl. Instr. and Meth.*, **82**, 291.
Cochran, W. G. 1952, *Ann. Math. Stat.*, **23**, 315.
Cramer, H. 1946, *Mathematical Methods of Statistics* (Princeton: Princeton University Press).
Davison, P. J. N., Culhane, J. L., Mitchell, R. J., and Fabian, A. C. 1975, *Ap. J.* (*Letters*), **196**, L23.
Draper, N. R., and Smith, H. 1966, *Applied Regression Analysis* (New York: Wiley).
Eadie, W. T., Drijard, D., James, F. E., Roos, M., and Sadoulet, B. 1971, *Statistical Methods in Experimental Physics* (Amsterdam: North-Holland).
Fabian, A. 1975, *M.N.R.A.S.*, **172**, 149.
Fisher, R. A. 1924, *J. Roy. Stat. Soc.*, **87**, 442.
Fisher, R. A. 1925, *Metron*, **5**, 90.
Giacconi, R., Murray, S., Gursky, H., Kellogg, E., Schreier, E., Matilsky, T., Koch, D., and Tananbaum, H. 1974, *Ap. J. Suppl.*, No. 237, **27**, 37.
Giacconi, R., Murray, S., Gursky, H., Kellogg, E., Schreier, E., and Tananbaum, H. 1972, *Ap. J.*, **178**, 287.
Gorenstein, P., Giacconi, R., and Gursky, H. 1967, *Ap. J.* (*Letters*), **150**, L85.
Gorenstein, P., Gursky, H., and Garmire, G. 1968, *Ap. J.*, **153**, 885.
Hearn, D. 1969, *Nucl. Instr. and Meth.*, **70**, 200.
Kellogg, E., Baldwin, J. R., and Koch, D. 1975, *Ap. J.*, **199**, 299 (KBK).
Lampton, M., Bowyer, S., Mack, J. E., and Margon, B. 1971, *Ap. J.* (*Letters*), **168**, L1.
Lea, S., Silk, J., Kellogg, E., and Murray, S. 1973, *Ap. J.* (*Letters*), **184**, L105.
Lesser, P. M. S., Cline, D., Goode, P., and Horoshko, R. N. 1972, *Nucl. Phys.*, **A190**, 597.
Margon, B. 1974, Bull. *AAS*, **6**, 313.
Margon, B., Bowyer, S., and Lampton, M. 1975*a*, *Proc. International Conference on X-Rays in Space* (Calgary: University of Calgary), p. 1069.
Margon, B., Lampton, M., Bowyer, S., and Cruddace, R. 1975*b*, *Ap. J.*, **197**, 25.
Pearson, K. 1900, *Phil. Mag.*, **50**, 157.
Wolff, R. S., Helava, H., and Weisskopf, M. C. 1975, *Ap. J.* (*Letters*), **197**, L99.

Note added in proof.—The parameter-subset estimation procedure recommended by Avni, discussed here in § III, has been generally proven by W. Cash (preprint 1976).

STUART BOWYER, MICHAEL LAMPTON, and BRUCE MARGON: Space Sciences Laboratory, University of California, Berkeley, CA 94720

© American Astronomical Society • Provided by the NASA Astrophysics Data System