NUCLEAR INSTRUMENTS AND METHODS 127 (1975) 253–260; © NORTH-HOLLAND PUBLISHING CO.

ANALYTIC AND GRAPHICAL METHODS FOR ASSIGNING ERRORS TO PARAMETERS IN NON-LINEAR LEAST SQUARES FITTING†

D. W. O. ROGERS

Physics Division, National Research Council, Ottawa K1A OR6, Canada
and
*Oxford Nuclear Physics Laboratory, Oxford, England**

Received 6 March 1975

Using a single notation, a summary is presented of the various analytic and graphical methods used for assigning errors to individual parameters determined in linear and non-linear least squares fitting procedures, with emphasis on cases with few degrees of freedom (e.g. the multipole mixing ratio problem in gamma ray spectroscopy). The ranges of applicability of each method and the relationships between the methods are discussed with particular emphasis placed on distinguishing between the cases for which the magnitudes of variances on the input data are known and those for which they are not known. It is shown directly that the graphical $\chi^2_{\min}+1$ rule is equivalent to the analytic internal error estimates and the corresponding graphical limit for the external error estimates is given by $\chi^2_{\min}[1+(1/(n-p))\,F_\alpha(1, n-p)]$. An explicit formula is given for transforming from one error estimate to another. A direct analytic procedure is presented for determining the diagonal elements of the error matrix without calculating or inverting the normal matrix. Tables of the statistical F and t distributions at the 68.3% and 95.5% confidence levels are also given. A simulation of a simple fitting experiment is discussed to clarify the concepts involved.

1. Introduction

This paper deals with the following rather specific question which has wide application in data fitting: "How do we assign confidence intervals to individual parameters determined in non-linear least squares fitting procedures where: i) the normal matrix and its inverse are not explicitly calculated; ii) there may be a very few degrees of freedom; and iii) the variances on the input data (which will be assumed to be from a normal distribution) may or may not be known?"

This work has been stimulated by the mixing ratio problem in gamma ray spectroscopy. However, the results are of general applicability and can be used in many areas (e.g. recoil distance lifetime measurements with only a few data points; DWBA analysis of reaction data; determining shell model matrix elements using least squares techniques).

This problem still needs to be discussed for the following reasons. Firstly, the non-linear aspects are not dealt with in standard texts and they tend to obscure the underlying statistical problems. More seriously, there are many statistical factors which are frequently ignored because they tend to unity in cases with more than a few degrees of freedom (e.g. the students t distribution at the 68.3% limit). However in cases with only a few degrees of freedom these factors are important. Finally, there has been a great deal of confusion since the proper distinctions have not always been made between the statistical methods which apply when the magnitudes of the variances on the input data are known and those which apply when the variances are not known. As a result of these factors there are many apparently contradictory methods and some erroneous methods found in the literature. For example, in the graphical method discussed below, the following limit points have all been suggested:

$$\chi^2_{\text{limit}} = \frac{\chi^2_{10\%}}{\chi^2_{50\%}} \times \chi^2_{\min} \qquad \text{ref. 13}$$

$$\chi^2_{\text{limit}} = \chi^2_{\min}+1 \qquad \text{ref. 3}$$

$$\chi^2_{\text{limit}} = \chi^2_{\min}\left(1+\frac{1}{n-p}F_\alpha(1, n-p)\right) \qquad \text{refs. 6, 9}$$

$$\chi^2_{\text{limit}} = \chi^2_{\min}\left(1+\frac{p}{n-p}F_\alpha(n, n-p)\right) \qquad \text{ref. 4}$$

The implications of the use of one or other of these formulae is often misunderstood and their relationships to the standard analytic techniques are often obscure.

Using a unified notation, a brief review of the analytic and graphical methods which can be used to assign error estimates will be given. Special care will be taken to define the ranges of applicability of each method. Then, based on the derivation of the graphical method, an analytic method for indirectly calculating the inverse of the normal matrix (i.e. the error matrix)

† NRCC 14188.
* NRC postdoctoral fellow until November 1973.

and hence the desired confidence intervals will be presented. Once the various methods have been delineated, a general method for transforming from one error estimate to another will be given.

The non-linear aspect of the problem is not solved. It must be assumed that the fitting function is nearly linear in the region of the fitted parameters before the confidence regions determined have a proper statistical meaning. However, this is frequently the case and methods for directly checking the linearity will be given.

2. Statement of the problem

Consider a series of n data points y_i, measured as a function of some variable x_i. We assume:

$$y_i = f(\boldsymbol{\beta}, x_i) + e_i, \tag{1}$$

where $\boldsymbol{\beta}$ is a series of p unknown but fixed parameters we wish to estimate, $f(\boldsymbol{\beta}, x_i)$ is a homogeneous fitting function (i.e. $f(\mathbf{0}, x_i) = 0$) and the e_i come from independent, normal distributions each with mean 0 and variance σ_i^2*. Under these assumptions, there are two physically distinct cases and a subcase to be considered.

Case A1. The e_i are related to the measuring process or procedure involved. The magnitude of the σ_i^2 are known and refer to the experimental accuracy of the measurement of the y_i. In the absence of these measurement uncertainties, the model used would predict some quantity exactly (e.g. the length of a rod).

Case A2. The same as case A1 except that the magnitude of the σ_i^2 are unknown.

Case B. The e_i and hence the σ_i^2 are attributes of what has been measured and *not* of the measuring process. In this case the σ_i^2 are parameters which must be determined from the data along with the parameter $\boldsymbol{\beta}$. In this case the model does not predict the value of a single observation, but only the distribution of a series of observations. For example, the heights of boys whose fathers have the same height will be distributed about some mean value. The width of this distribution is a physical attribute of what is being measured and is not related to measurement "errors".

If the same data is analysed under any of the above assumptions, the *same* least squares estimators of $\boldsymbol{\beta}$ will be obtained. However, *different* confidence intervals will be calculated for the parameters β_j. Some confusion arises because the confidence regions obtained for case A2 and case B are the same. Furthermore, books on statistics and computer program packages often deal exclusively with case B whereas case A occurs frequently in physics. Lastly, there is confusion since there should be very little difference between the results for case A and case B when there are many degrees of freedom. The fact that significant differences are frequently expected when there are only a few degrees of freedom is often forgotten.

In the remainder of this paper, case A will be of prime interest. Traditionally the case of nuclear counting is considered in this light (and hence the χ^2 test can be used) although, strictly speaking, one could also interpret it as case B in some respects.

To allow a unified notation, the variances on the input data y_i will be written as $\theta\sigma_i^2$ where θ represents a possibly unknown scale factor. For $\theta = 1$, case A1 is obtained. In cases A2 and B, θ is unknown and the σ_i^2 become just the relative weightings of the data points and not the input variances (they may be unity).

The solution of the least squares problem leads to estimators $\hat{\boldsymbol{\beta}}$ for the parameters $\boldsymbol{\beta}$ given by (in the linear case):

$$\hat{\boldsymbol{\beta}} = \mathbf{S}^{-1}\mathbf{X}\tilde{\mathbf{y}}, \tag{2}$$

where

$$X'_{jk} = \frac{1}{\sigma_j}\frac{\partial f(\boldsymbol{\beta}, x_j)}{\partial \beta_k}, \qquad \begin{matrix} k = 1, \ldots p, \\ j = 1, \ldots n, \end{matrix} \tag{3}$$

$$\mathbf{S} = \mathbf{X}\mathbf{X}' \text{ or } S_{jk} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}\frac{\partial f(\boldsymbol{\beta}, x_i)}{\partial \beta_j}\frac{\partial f(\boldsymbol{\beta}, x_i)}{\partial \beta_k}, \quad j = 1, \ldots p, \quad k = 1, \ldots p, \tag{4}$$

$$\tilde{y}_j = y_j/\sigma_j\dagger, \qquad j = 1, \ldots n. \tag{5}$$

$\mathbf{S}$ is the so called "normal matrix". In a linear model (where "linear" refers to the parameters $\boldsymbol{\beta}$ and not to the x_i), the matrices $\mathbf{X}$ and $\mathbf{S}$ are functions of x_j only.

Most statistical results are derived for this linear model case. However, it is justifiably argued that as long as the fitting function $f(\boldsymbol{\beta}, x_j)$ is approximately linear near $\hat{\boldsymbol{\beta}}$, then the results obtained for the linear model are valid. Approximate linearity near $\boldsymbol{\beta}$ implies that the second and higher order terms in a Taylor expansion are negligible, i.e. for any k,

$$f(\boldsymbol{\beta} + \Delta\beta_k, x_i) \approx f(\boldsymbol{\beta}, x_i) + \Delta\beta_k \frac{\partial f(\boldsymbol{\beta}, x_i)}{\partial \beta_k}. \tag{6}$$

* The assumption of independence can be removed (see sect. 1.5 of ref. 10), but the assumption of normality is critical to the development. Normality can often be justified in nuclear counting experiments but not necessarily in other data.

† If $f(\boldsymbol{\beta}, x_i)$ is inhomogeneous, then set $\tilde{y}_j = (y_j - f(\mathbf{0}, x_j))/\sigma_j$.

If condition (6) is valid when the $\Delta\beta_k$ are the "errors" determined by the procedures outlined below, then the confidence specified will be approximately correct. In practice approximation (6) can be verified explicitly or, if graphical methods are used, other checks can be made (see sect. 6). In the remainder of the paper, exact results will be derived for the linear case only.

3. Analytic methods

Define:

$$L(\mathbf{y}, \boldsymbol{\beta}) = \sum_i \frac{1}{\sigma_i^2}(y_i - f(\beta, x_i))^2 \tag{7}$$

and $\mathbf{C}$ = the variance (covariance) matrix for the estimators $\hat{\boldsymbol{\beta}}$ (the "error matrix")

$$= E[(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'], \tag{8}$$

where $E(x)$ is the expectation value of x. $L(\mathbf{y}, \boldsymbol{\beta})$ is the function minimized to give $\hat{\boldsymbol{\beta}}$. The following statements are standard results in statistics (see e.g. ref. 10). In eqs. (9), (12) and (14), $\boldsymbol{\beta}$ refers to the true values of the parameters.

For cases A1, A2 and B:

a) $\hat{\boldsymbol{\beta}}$ as given by eq. (2) is an unbiased estimator of $\boldsymbol{\beta}$ (i.e. the expectation value of $\hat{\boldsymbol{\beta}}$ is $\boldsymbol{\beta}$). (9)

b) $\mathbf{C} = \theta\mathbf{S}^{-1}$. (10)

c) $L(\mathbf{y}, \hat{\boldsymbol{\beta}})/\theta$ has a chi-squared distribution with $n-p$ degrees of freedom. (11)

d) $\hat{\boldsymbol{\beta}}$, considered as a random-variable, has a multivariate normal distribution with mean value $\boldsymbol{\beta}$, and variance $\theta\mathbf{S}^{-1}$, or equivalently, the function

$$N(0, 1) = \frac{\beta_i - \hat{\beta}_i}{\sqrt{\theta(\mathbf{S}^{-1})_{ii}}} \tag{12}$$

has a normal distribution with mean 0 and variance 1.

For Cases A2 and B:

e) $$\hat{\theta} = L(y, \hat{\boldsymbol{\beta}})/(n-p) \tag{13}$$

is an unbiased estimator of the unknown θ.

f) The function

$$t(n-p) = \frac{\beta_i - \hat{\beta}_i}{\sqrt{\hat{\theta}(\mathbf{S}^{-1})_{ii}}} \tag{14}$$

has a students t distribution with $n-p$ degrees of freedom. Note that $L(\mathbf{y}, \hat{\boldsymbol{\beta}})$ is the usual χ^2 statistic but it has a chi-squared distribution *only* if $\theta \equiv 1$ (i.e. in case A1).

Statements d) and f) lead directly to the following results regarding confidence intervals.

Case A1: Internal errors.

In the case in which the variances on the y_i are known ($\theta \equiv 1$), then result d implies that the following interval contains the true value of the parameter β_i, α percent of the time:

$$(\hat{\beta}_i - N_\alpha\sqrt{(\mathbf{S}^{-1})_{ii}},\ \hat{\beta}_i + N_\alpha\sqrt{(\mathbf{S}^{-1})_{ii}}), \tag{15}$$

where N_α is the normal distribution at the α% point (e.g. $N=1.0$ for $\alpha=68.27\%$, $N_\alpha=2.0$ for $\alpha=95.45\%$). This has been referred to as the internal error by Ferguson[8]). For later purposes it is useful to think of this result as follows. Rewrite eq. (12) to give:

$$\beta_i = \hat{\beta}_i + N_\alpha(0, 1)\sqrt{(\mathbf{S}^{-1})_{ii}}. \tag{16}$$

Conceptually the meaning of this equation is vague although operationally it can be well defined. β_i should no longer be thought of as the true value of the parameter but as a function ($F(\beta_i) = \beta_i$). The end points of the confidence interval in eq. (15) are just those values of β_i which produce the limiting values of the function (i.e. β_i) given by eq. (16).

Cases A2 and B: External errors.

In the cases in which the variances on the y_i are unknown, result f implies that the following interval contains the true value of the parameter β_i, α percent of the time:

$$(\hat{\beta}_i - t_\alpha\sqrt{\hat{\theta}(\mathbf{S}^{-1})_{ii}},\ \hat{\beta}_i + t_\alpha\sqrt{\hat{\theta}(\mathbf{S}^{-1})_{ii}}), \tag{17}$$

where t_α is the α% point of the t distribution with $n-p$ degrees of freedom (see appendix A) and $\hat{\theta}$ is given by eq. (13). Ferguson has referred to these as the external errors[8]).

All of the above results are well known and are included here to provide a complete summary in a single notation and to allow comparison to the graphical methods.

4. Graphical solutions

The intervals given in eq. (15) and eq. (17) can be used to obtain confidence regions for any parameter determined in a least square fitting procedure as long as $\mathbf{S}^{-1}$ is known. However, in practice they may be useless since there are many ways of determining $\hat{\boldsymbol{\beta}}$ which do not require finding $\mathbf{S}$ or $\mathbf{S}^{-1}$ (e.g. grid search methods or methods which avoid direct calculation of the matrix $\mathbf{S}^{-1}$, e.g. ref. 11).

The basic idea behind these graphical methods is to calculate $L_j(\mathbf{y}, \boldsymbol{\beta})$ as a function of each β_j near the value $\hat{\boldsymbol{\beta}}_j$ found in minimizing $L(\mathbf{y}, \boldsymbol{\beta})$. $L_j(\mathbf{y}, \boldsymbol{\beta})$ is the value of

$L(\mathbf{y}, \boldsymbol{\beta})$ which, for a specified value of β_j, is minimized with respect to all other $\beta_i (i \neq j)$. Limiting values of $L_j(\mathbf{y}, \boldsymbol{\beta})$ are specified (differently for each method) and the values of β_j which produce these limits are defined as the end points of the confidence interval for the parameter β_j. Thus in fig. 1, the interval $(\hat{\beta}_j - \Delta, \hat{\beta}_j + \Delta)$ is the confidence interval defined by the limiting value of $L_j(\mathbf{y}, \boldsymbol{\beta})$.

In order to see the relationships between the graphical methods and the analytic methods of sect. 3, it is instructive to start with the following result, although other derivations have used different approaches[6,3,9]). In the present notation, eq. (12) of Archer et al.[2]) becomes:

$$L_j(\mathbf{y}, \boldsymbol{\beta}) - L(\mathbf{y}, \hat{\boldsymbol{\beta}}) = \frac{(\beta_j - \hat{\beta}_j)^2}{(\mathbf{S}^{-1})_{jj}} \; * \qquad (18)$$

This equation tells us that a plot of L_j versus β_j must be parabolic about the minimum value $\hat{\beta}_j$.

To obtain limiting values of $L_j(\mathbf{y}, \boldsymbol{\beta})$ for use in the graphical methods outlined above, one substitutes for the right hand side of eq. (18) using eqs. (12) or (14) and considers the resulting equation in exactly the same manner as eq. (16).

Association of the right hand side of eq. (18) with the square of the normal distribution gives (using eq. (12) for case A1 with $\theta = 1$):

$$L_j(\mathbf{y}, \boldsymbol{\beta})_{\text{lim}} = L(\mathbf{y}, \hat{\boldsymbol{\beta}}) + (N_\alpha(0, 1))^2. \qquad (19)$$

Similarly for case A2, substitution of eq. (14) into eq. (18) gives:

$$L_j(\mathbf{y}, \boldsymbol{\beta})_{\text{lim}} = L(\mathbf{y}, \hat{\boldsymbol{\beta}}) + \hat{\theta} t_\alpha^2 (n - p) \qquad (20)$$

or, using eq. (13) and the general relationship $t_\alpha^2(n-p) = F(1, n-p)$ where $F(1, n-p)$ is the statistical F distribution with 1 and $n-p$ degrees of freedom:

$$L_j(\mathbf{y}, \boldsymbol{\beta})_{\text{lim}} = L(\mathbf{y}, \boldsymbol{\beta}) \left[1 + \frac{1}{n-p} F_\alpha(1, n-p) \right]. \qquad (21)$$

Since $N_\alpha(0, 1) = 1$ for $\alpha = 68.3\%$, eq. (19) gives the standard "$\chi^2 + 1$" rule for the limiting value of $L_j(\mathbf{y}, \boldsymbol{\beta})$[3]). Eq. (19) also gives the more general rule for any confidence level. Similarly eq. (20) is the prescription given by Davies[6]) for the limiting value (see also refs. 5, 9). From the derivations it is clear that application of eq. (19) is equivalent to the analytic solution for the internal errors given by the interval in eq. (15) and applies when the variances on the input data are known (case A1). On the other hand, application of eq. (21) is equivalent to the external errors interval given in eq. (17) and applies when the scale factor for the input data's variance is unknown (case A2).

A third graphical method is available, but it treats a slightly different problem (see e.g. refs. 7, 10). The two methods presented above deal with confidence intervals for individual parameters, considered independently of the values of all other fitted parameters. However, one may be interested in a confidence region (ellipsoid) for all parameters considered simultaneously. In this case, the following equation defines a p dimensional volume (in the $\boldsymbol{\beta}$ parameter space) which includes the true value of $\boldsymbol{\beta}$, $\alpha\%$ of the time:

$$L(\mathbf{y}, \boldsymbol{\beta}) = L(\mathbf{y}, \hat{\boldsymbol{\beta}}) \left(1 + \frac{p}{n-p} F_\alpha(p, n-p) \right). \qquad (22)$$

(Note that $L_j(\mathbf{y}, \boldsymbol{\beta})$ does *not* occur as suggested in ref. 4). This method is not normally of interest in the nuclear mixing ratio problem but it does define a valid confidence region when the variances on the input data are not known (i.e. in case A2 or B).

* The proof of this equation is messy but depends solely on algebraic techniques. The proof is available on request.

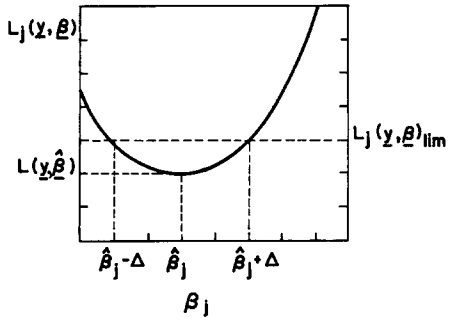

Fig. 1. In the graphical method of assigning confidence intervals, a graph of $L_j(\mathbf{y}, \boldsymbol{\beta})$ versus β_j is constructed where $L_j(\mathbf{y}, \boldsymbol{\beta})$ is the function $L(\mathbf{y}, \boldsymbol{\beta})$ at β_j, minimized with respect to all $\beta_i (i \neq j)$. The confidence interval is defined by the limiting values of $L_j(\mathbf{y}, \boldsymbol{\beta})$ given by the various methods discussed in the text. In a linear model, the plot is given by:

$$L_j(\mathbf{y}, \boldsymbol{\beta}) = L(\mathbf{y}, \hat{\boldsymbol{\beta}}) + (\beta_j - \hat{\beta}_j)^2 / (\mathbf{S}^{-1})_{jj},$$

where $\mathbf{S}^{-1}$ is the "error matrix".

5. An "analytical-graphical" method

The prime purpose of the graphical methods presented in section 4 is to determine $(\mathbf{S})_{jj}^{-1}$ without explicitly calculating $\mathbf{S}$. The methods outlined were "graphical" in the sense that in practice they have been applied by considering a plot of $L_j(\mathbf{y}, \boldsymbol{\beta})$ versus β_j. However, from a computational point of view, there is a much more direct method. Rather than calculating a complete table of $L_j(\mathbf{y}, \boldsymbol{\beta})$ values, one

merely needs $\hat{\boldsymbol{\beta}}$, $L(\mathbf{y}, \hat{\boldsymbol{\beta}})$ and $L_j(\mathbf{y}, \boldsymbol{\beta})$ at a single known value of β_j. Then, using eq. (18), $(\mathbf{S}^{-1})_{jj}$ can be calculated and the internal or external confidence intervals calculated using eqs. (15) or (17) respectively. This saves computing time compared to the presently used methods and allows a computer program to explicitly output the desired confidence intervals. Thus, in general, finding the confidence interval in this manner for a single parameter in a problem with p fitted parameters is equivalent to solving the fitting problem with $p-1$ parameters. Although there are good initial estimates of the $p-1$ parameters, this procedure can be prohibitive in terms of computing time unless p is small or there are only a few non-linear parameters.

One practical point is worth noting. This method is subject to numerical accuracy problems unless $\hat{\boldsymbol{\beta}}$ has been determined to a tolerance which is much smaller than the value of $\sqrt{(\mathbf{S}^{-1})_{jj}}$ (e.g. if a grid search has been used to find $\hat{\boldsymbol{\beta}}$, then the mesh size must be much smaller than the confidence interval found). Some of the methods outlined in the following section can be used to verify the numerical accuracy of a particular solution.

6. Non-linear functions

As pointed out in section 2 the above results apply exactly only if $f(\boldsymbol{\beta}, x_i)$ is linear in $\boldsymbol{\beta}$ (i.e. if eq. (6) holds exactly). However they will be approximately valid as long as eq. (6) is approximately valid near $\boldsymbol{\beta}$. There are several ways to check on this assumption. The most direct method is to substitute the exact values into both sides of eq. (6). This of course implies that the derivatives needed to calculate $\mathbf{S}$ are available.

If methods have been used which avoid calculating derivatives then eq. (18) can be used to verify certain conditions which are necessary consequences of approximate linearity. This equation states that a plot of $L_j(\mathbf{y}, \boldsymbol{\beta})$ versus β_j must be approximately parabolic near β_j if $f(\boldsymbol{\beta}, x_i)$ is approximately linear there. This can be verified visually from a plot or by explicitly calculating $L_j(\mathbf{y}, \boldsymbol{\beta})$ with $\beta_j = \hat{\beta}_j \pm N_\alpha \sqrt{(\mathbf{S}^{-1})_{jj}}$ (or $\hat{\beta}_j \pm t_\alpha \sqrt{\hat{\theta}(\mathbf{S}^{-1})_{jj}}$) and comparing these values of $L_j(\mathbf{y}, \boldsymbol{\beta})$ to those predicted by eq. (19) or (21)). The sources of any discrepancies must be investigated further since they may be due to non-linearities in $f(\boldsymbol{\beta}, x_i)$ or they may be due to the numerical accuracy problems mentioned in section 5.

One final consequence of the approximate linearity of $f(\boldsymbol{\beta}, x_i)$ is that all confidence intervals deduced *must* be symmetric due to the parabolic shape of the L_j vs β_j plot.

7. Transformation properties

With the variety of methods used to assign confidence regions to fitted parameters, it is sometimes necessary to be able to transform results obtained using one method into those that would be obtained using another. Using the results in eqs. (13), (15) and (17) one obtains the following relationship between external and internal "errors" (determined analytically or graphically) at various confidence levels:

$$\frac{(\text{Internal at } \alpha\%)}{N_\alpha} = \frac{(\text{External at } \beta\%)}{t_\beta(n-p)\sqrt{L(\mathbf{y}, \hat{\boldsymbol{\beta}})/(n-p)}}. \tag{23}$$

It is also possible to transform intervals obtained graphically into internal or external confidence intervals. Consider the interval $(\hat{\beta}_j - \Delta', \hat{\beta}_j + \Delta')$ determined graphically using limiting values of

$$L_j(\mathbf{y}, \boldsymbol{\beta})_{\text{lim}} = L(\mathbf{y}, \hat{\boldsymbol{\beta}}) \times R, \tag{24}$$

where R is any specified factor (e.g. $R = 1 + (p/(n-p)) \times \times F_\alpha(p, n-p)$ as in ref. 4 or $R = \chi^2_{10\%}/\chi^2_{50\%}$ as in ref. 13). Denote the 68.3% internal "error" by Δ. One obtains the following general relationship from eq. (18):

$$\Delta = \Delta' / \sqrt{L(\mathbf{y}, \hat{\boldsymbol{\beta}}) \times (R-1)}. \tag{25}$$

Eqs. (23) and (25) emphasize the fact that $L(\mathbf{y}, \hat{\boldsymbol{\beta}})$ should *always* be reported, both to allow for an evaluation of the quality of the data and to allow others using the data to transform it to another estimate for comparison.

8. A simulation

To demonstrate explicitly the meanings of the various prescriptions presented for assigning confidence intervals, a simple least squares fitting experiment has been simulated. The simulation calculated estimators for the parameters a and b in the expression $y = a + bx$, given a series of pseudo-random data points $y_i \pm \sigma_i$ at known values of x_i.

The points y_i were generated as $y_i = a + bx_i + e_i$ where e_i was a normally distributed pseudo-random number, mean 0, variance $\sigma_i^2 = a + bx_i$. These data points simulated a nuclear counting experiment. The least squares estimators of a and b were calculated for each experiment making use of the true variances on each data point and confidence intervals were constructed using the various methods mentioned in the previous sections. The methods of section 5 were employed to obtain the "graphical results" and checks were made to ensure that these were equivalent to "drawing a graph". A record was kept of how often

TABLE 1

Percentage of intervals including the "true" value of the parameter "b" in 300 000 simulated experiments fitting $y = a + bx$ with 5 data points and using the correct variances on the input data for case 1 and incorrect variances for case 2.

Method	Text reference	Predicted % for case 1	% of intervals including "b" Case 1[d]	Case 2[e]
Internal				
Analytic	12, 15	68.3	68.2	99.9
Graphical[f]	19, §5	68.3	68.2	99.9
External				
Analytic	13, 14, 17	68.3	68.5	68.5
Graphical[g]	21, §5	68.3	68.5	68.5
Cline and Lesser				
Graphical	[a]	90	95.4	95.4
Analytic	[b]	68.3	61.0	61.0
$\chi^2_{10\%}/\chi^2_{50\%}$	[c]	68.3	88.7	88.7
Ellipsoid[h]	22	90.0	90.0	90.0

([a]) ref. 4, our eq. (22) using $L_j(\mathbf{y}, \boldsymbol{\beta})$ instead of $L(\mathbf{y}, \boldsymbol{\beta})$.
([b]) ref. 4, eq. (3), i.e. our eq. (17) with *no t* distribution factor.
([c]) ref. 13 suggests $L_j(y, \beta)_{\text{lim}} = L(\mathbf{y}, \boldsymbol{\beta}) \times (\chi^2_{10\%}/\chi^2_{50\%})$ gives a standard deviation error on the mixing ratio.
([d]) using "true" variances on input data.
([e]) using variances scaled up by a factor of $\sqrt{10}$.
([f]) equivalent to $\chi^2_{\text{min}} + 1$ rule (ref. 3).
([g]) equivalent to Davies method, ref. 6.
([h]) This gives an estimate of how often both a and b were within the ellipsoid defined by eq. (22).

these various confidence intervals actually included the "true" values of a and b. The results are presented in table 1.

As expected the graphical and analytic methods gave identical results for both the internal and external errors. When the correct variances were used in fitting (i.e. case 1 in table 1), the internal and external confidence intervals included the "true" value of the fitted parameter the expected number of times. This does not mean that the confidence intervals were the same in each simulated experiment. The predicted internal confidence interval was the same in all experiments (since S is independent of the y_i) while the external confidence interval fluctuated and was less than one half as large as the internal confidence interval in nearly 20% of the cases.

As seen from the table, all other methods produce the wrong results, except for the ellipsoid technique which is concerned with values of both parameters and not just the value "b".

In case 2, the same simulated data points and variances were used as in case 1 but the variances associated with them in the fitting procedures were multiplied by a factor of $\sqrt{10}$. Here the internal error estimates are wrong since they depend on knowing the true variances on the input data. However, the external error estimates are not dependent on this knowledge and hence still provide "correct" estimates of the confidence intervals.

An analysis of the actual distribution of the estimators of "b" was made. The mean value was found to be the "true" value of "b" (i.e. the estimators were unbiased) and the variance and fourth moment were characteristic of a normal distribution with the variance predicted by the internal error method (see eq. (12)).

9. Applications

Graphical methods have been adopted in the nuclear mixing ratio problem primarily because there is frequently more than one local minimum. However the partial derivatives needed to calculated $\mathbf{S}^{-1}$ directly could be calculated quite simply and analytic techniques used to assign confidence intervals. However the use of plots of $L_j(\mathbf{y}, \boldsymbol{\beta})$ versus β_j is still a valuable aid since they display the quality of the data and the validity of the fitting procedure in a very graphic manner. In fig. 2 it is clear that in case A the linearity approximation is invalid, in case B the parameter is poorly defined and in case C the parameter is well defined.

In more complicated fitting procedures these graphical methods or the method of section 5 become essential (see e.g. ref. 3). However it is remarkable that in some areas of physics little attention is paid to extracting any type of error on various parameters. One example is in DWBA analysis of particle reaction data where seldom, if ever, are errors assigned to

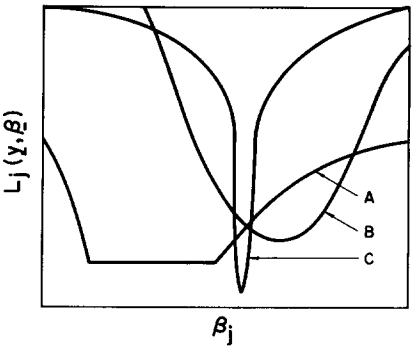

Fig. 2. Although the graphical method can be replaced by a computationally shorter analytic method (see section 5), a plot of $L_j(\mathbf{y}, \boldsymbol{\beta})$ versus β_j can still be very informative. From this plot it is immediately clear that: i) for case A the linearity approximation is invalid since the plot is not parabolic; ii) for case B the data is not very sensitive to β_j; and iii) for case C the data defines β_j well.

fitted spectroscopic factors or optical model parameters. Knowledge of such errors would clearly be useful.

One must be wary of interpreting the meanings of statistical errors in too direct a sense, both in complicated cases such as DWBA analysis or when fitting a simple straight line. One must assume the model fitting function is realistic. As has been shown for the case of nuclear charge densities deduced from electron scattering results[12]), the restrictions of the model may in fact obscure much of our real ignorance. Nonetheless, it is still important to determine whether, within a given model, our data gives a 1%, 10% or 100% statistical accuracy to fitted parameters and at least to determine which parameters in the model are sensitive to our data. The statistical errors and the value of $L(\mathbf{y}, \hat{\boldsymbol{\beta}})$ should always be reported, even if the model is known to be unrealistic.

10. Conclusions

The equivalence of several analytic and graphical methods of assigning confidence intervals to fitted parameters has been demonstrated and the relationship between internal and external errors has been explicitly presented. The central role of $L(\mathbf{y}, \hat{\boldsymbol{\beta}})$ in these relationships shows that it should always be reported to allow future comparisons to be made. It is also essential to explicitly state which method has been used in the error analysis.

The question of whether internal or external confidence intervals should be used hinges on whether or not the variances on the input data are known. If the input variances are known, and if use is being made of the standard chi-squared test, then logical consistency insists on use of the internal errors. This is normally the case in gamma-ray angular distribution studies. If input variances are unknown, or only relative weights are known, then external confidence intervals must be used. The conservative choice made by some physicists is to use the larger of the two confidence regions. This is acceptable but implies no confidence in any chi-square tests applied to the data.

The simple analytic method presented in section 5 for determining confidence regions using "the graphical method" should prove useful in programming although it is subject to possible numerical accuracy problems. The methods of section 6 should always be used to verify the linearity assumption inherent in this analysis. One immediate consequence of linearity is that the confidence intervals assigned using any of these methods must be symmetric.

TABLE A1

Values of the F and t distributions at the one and two standard deviation limits for various numbers of degrees of freedom.

n	$F(1, n)$		$\left(1 + \frac{F(1, n)}{n}\right)^{c}$		$t(n)$	
	68.3[a]	95.5[b]	68.3	95.5	68.3	95.5
1	3.376	195.1	4.38	196.1	1.837	13.97
2	1.746	20.5	1.87	11.25	1.321	4.52
3	1.433	10.93	1.48	4.64	1.197	3.31
4	1.303	8.24	1.33	3.06	1.141	2.87
5	1.233	7.02	1.25	2.40	1.110	2.65
6	1.190	6.33	1.20	2.06	1.091	2.52
7	1.160	5.90	1.17	1.84	1.077	2.43
8	1.137	5.60	1.14	1.70	1.066	2.37
9	1.122	5.38	1.12	1.60	1.059	2.32
10	1.108	5.21	1.11	1.52	1.053	2.28
11	1.097	5.08	1.10	1.46	1.047	2.25
12	1.089	4.98	1.09	1.42	1.043	2.23
13	1.082	4.89	1.08	1.38	1.040	2.21
14	1.075	4.82	1.08	1.34	1.036	2.20
15	1.070	4.76	1.07	1.32	1.034	2.18
16	1.065	4.71	1.07	1.29	1.032	2.17
17	1.062	4.66	1.06	1.27	1.031	2.16
18	1.058	4.62	1.06	1.26	1.029	2.15
19	1.055	4.58	1.06	1.24	1.027	2.17
20	1.052	4.55	1.05	1.23	1.026	2.13
50	1.020	4.21	1.02	1.08	1.010	2.05
100	1.010	4.10	1.01	1.04	1.005	2.02

[a] ± 3 in last figure.
[b] ± 1 in last figure.
[c] Used in the graphical method for external errors.

I would like to thank Drs Brian Underwood and Bill Dixon for many stimulating discussions on this subject. I also want to thank Dr Walter Davies for making some of his unpublished notes available to me. They were directly responsible for stimulating much of this work. My thanks are also due to Dr John Becker for useful comments on the manuscript.

Appendix

Knowledge of the 68.27% and 95.45% limits* of the F and t distributions is necessary to calculate external confidence intervals at the one and two standard deviation limits. As these are not normally tabulated, they are presented in table A1. They have been cal-

* The present definitions of limit points are intuitively obvious but technically somewhat sloppy. In dealing with confidence regions with probability $(1-\alpha)\times 100\%$, we should strictly speaking deal with the following limiting points: $F_{1-\alpha}(n, m)$, $t_{1-\alpha/2}(n)$ and $N_{1-\alpha/2}(0, 1)$.

culated using the relationship:

$$P(F|n_1, n_2) = 1 - I_x(n_2/2, n_1/2),$$

$$x = n_2/(n_2 + n_1 F)$$

where $P(F|n_1,n_2)$ is the probability that an F distribution with n_1 and n_2 degrees of freedom is less than the value F and I_x is the incomplete beta function (eq. (26.6.2), ref. 1). An IBM 360 Scientific Subroutine Package routine was used to calculate the incomplete beta function. The t distribution was obtained using the relationship:

$$t_\alpha^2(n) = F_\alpha(1, n) .$$

References

1) M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
2) N. P. Archer, W. V. Prestwich and G. L. Keech, Nucl. Instr. and Meth. **44** (1966) 114.
3) R. A. Arndt and M. H. MacGregor, Meth. Comp. Phys. **6** (1966) 253, appendix 2; P. R. Bevington, *Data reduction and error analysis* (McGraw-Hill, 1969).
4) D. Cline and P. M. S. Lesser, Nucl. Instr. and Meth. **82** (1970) 291.
5) P. M. S. Lesser, D. Cline, P. Goode and R. N. Horoshko, Nucl. Phys. **A190** (1972) 597, footnote p. 610.
6) W. G. Davies, BAPS **16** (1971) 605, and private communication.
7) N. R. Draper and H. Smith, *Applied regression analysis* (Wiley, New York, 1966), p. 292.
8) A. J. Ferguson, *Angular correlation methods in gamma-ray spectroscopy* (North-Holland, Amsterdam, 1964).
9) A. N. James, P. J. Twin and P. A. Butler, Nucl. Instr. and Meth. **115** (1974) 105.
10) H. Scheffe, *The analysis of variance* (John Wiley and Sons, New York, 1959).
11) M. J. D. Powell, Computer J. **7** (1964) 155.
12) I. Sick, Nucl. Phys. **A218** (1974) 509.
13) M. F. Thomas, J. S. Lopes, R. Ollerhead, A. R. Poletti and E. K. Warburton, Nucl. Phys. **78** (1965) 298.