Statistical Error Propagation

Joel Tellinghuisen[†]

Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235

Received: September 26, 2000; In Final Form: February 1, 2001

The simple but often neglected equation for the propagation of statistical errors in functions of correlated variables is tested on a number of linear and nonlinear functions of parameters from linear and nonlinear least-squares (LS) fits, through Monte Carlo calculations on $10^4-4 \times 10^5$ equivalent data sets. The test examples include polynomial and exponential representations and a band analysis model. For linear functions of linear LS parameters, the error propagation equation is exact. Nonlinear parameters and functions yield nonnormal distributions, but their dispersion is still well predicted by the propagation-of-error equation. Often the error computation can be bypassed by a redefinition of the least-squares model to include the quantity of interest as an adjustable parameter, in which case its variance is returned directly in the variance-covariance matrix. This approach is shown formally to be equivalent to the error propagation method.

Introduction

Perhaps one of the "best-kept secrets" in experimental physical science is the simple matrix expression for error propagation[1−3]

$$\sigma_f^{\,2} = \mathbf{g}^{\mathrm{T}}\,\mathbf{V}\,\mathbf{g} \tag{1}$$

in which σ_f^2 represents the variance in some function f of a set of parameters $\boldsymbol{\beta}$, whose variance-covariance matrix is $\mathbf{V}$, with the ith element in the vector $\mathbf{g}$ being $\partial f/\partial\beta_i$. To be sure, undergraduate chemistry and physics students are drilled in the form of this equation that applies for *uncorrelated* variables:

$$\sigma_f^{\,2} = \sum\left(\frac{\partial f}{\partial\beta_i}\right)^2 \sigma_{\beta_{\mathrm{i}}}^{\ 2} \tag{2}$$

which is obtained from eq 1 by dropping the off-diagonal terms (the covariances) in $\mathbf{V}$. However, in very many applications of interest, the parameters $\boldsymbol{\beta}$ themselves result from a least-squares analysis, and in general their covariances are not negligible. Accordingly, estimates based on eq 2 can be grossly in error. While this inadequacy has been recognized, the correct eq 1 still appears to be underutilized. This may be because this equation does not feature very prominently (if at all) in the data analysis reference sources most physical scientists use.[4−6]

Another possible reason for the neglect of eq 1 is the sense that error propagation is "only approximate". The present work was undertaken to examine that notion. Monte Carlo calculations involving at least 10^4 equivalent data sets are used to compare "experiment" with the predictions of eq 1 in a number of problems of interest, involving linear and nonlinear functions of parameters that themselves result from both linear and nonlinear least-squares (LS) fits. The computations employ methods like those used recently to investigate the distributional properties of LS parameters from nonlinear fits[7] and from linear fits to transformed (nonnormal) data.[8] With the usual assumptions of normal, unbiased data having an a priori known error structure, eq 1 is rigorous in application to linear functions of linear LS parameters. Such parameters are themselves normally distributed, with variances (the diagonal elements of $\mathbf{V}$) known exactly at the outset:[1−3,7]

$$\mathbf{V} = \mathbf{A}^{-1} \tag{3}$$

where $\mathbf{A}$ is the matrix of the normal equations. Accordingly, linear functions of such parameters are unbiased and normal, with variances σ_f^2 known exactly. On the other hand, nonlinear parameters and nonlinear functions of linear parameters are not normally distributed and in fact are usually biased.[7] Nevertheless, for the cases examined here, this nonnormality seldom translates into a serious deficiency in the predictions of eq 1 and its "normal" interpretation for establishing confidence limits. Indeed, the 10% "rule of thumb" suggested for nonlinear LS parameters[7] seems also to apply to functions of such parameters: If the relative standard error σ_f/f is $<1/10$, confidence limits based on eq 1 should also be reliable to within 10%. However, in several of the cases examined here, asymmetry in the distributions is more severe than in the examples studied in ref 7.

It is not surprising that functions of LS parameters behave in a fashion similar to the parameters themselves, because often it is possible to bypass eq 1 in the calculation of the propagated error for a particular f, by redefining the fit to include f among the adjustable parameters. Then its variance is returned directly by the LS fit. As is shown below, this approach is formally equivalent to the use of eq 1, a point which has also been verified computationally.

For specific illustration of some of these points, suppose that data are fitted to a straight line, $y = a + bx$, and that the usual assumptions for the data apply, namely that the model is correct and the data have random, normally distributed error in y only. Then the LS estimates of a and b are unbiased and normally distributed about the true values, with standard errors that are exactly predictable if the error structure of the data is known: $\sigma_a^{\,2} = V_{11} = A_{11}^{\ -1}$ and $\sigma_b^{\,2} = V_{22}$. Now consider the three functions $f_1 = a + bx$ (the fit function itself), $f_2 = a \pm b$, and $f_3 = b/a$. The row matrices of eq 1 for these three cases are $\mathbf{g}^{\mathrm{T}} = (1, x)$, $(1, \pm 1)$, and $(-b/a^2, 1/a)$, respectively. The propagated variances in f_1 and f_2 are

[†] FAX: 615-343-1234. E-mail: tellinjb@ctrvax.vanderbilt.edu.

10.1021/jp003484u CCC: $20.00 © 2001 American Chemical Society
Published on Web 03/22/2001

$$\sigma_{f_1}{}^2 = \sigma_a{}^2 + \sigma_b{}^2 x^2 + 2\sigma_{ab}{}^2 x \tag{4a}$$

$$\sigma_{f_2}{}^2 = \sigma_a{}^2 + \sigma_b{}^2 \pm 2\sigma_{ab}{}^2 \tag{4b}$$

which differ from the predictions of eq 2 by the inclusion of the terms in the covariance, $\sigma_{ab}{}^2 = V_{12} = V_{21}$. Thus eq 2 yields correct results only when $V_{12} = 0$, which it does in this case when $\bar{x} = 0$ (or $\sum w_i x_i = 0$ for unequally weighted data). Since f_1 and f_2 are linear functions of $\boldsymbol{\beta}$, eq 1 is exact and f_1 and f_2 are both normally distributed. For the relative error in f_3, eq 1 yields

$$\left(\frac{\sigma_{f_3}}{f_3}\right)^2 = \left(\frac{\sigma_a}{a}\right)^2 + \left(\frac{\sigma_b}{b}\right)^2 - \frac{2\sigma_{ab}{}^2}{ab} \tag{4c}$$

which again differs from the predictions of eq 2 by the inclusion of the last term. Since f_3 is not a linear function of $\boldsymbol{\beta}$, f_3 is not normally distributed and eq 4c is not rigorous. A fit of the same data to $y = a + aBx$ will yield results for B ($=b/a$) and its variance (e.g., $\sigma_B{}^2 = V_{22}$) that are identical with those obtained via the error propagation approach of eq 1. This fit is a nonlinear fit to a straight line. Similarly, the linear fit can be redefined to yield f_1 and f_2 directly. For example a fit to $y = A + b(x - 1)$ will yield directly A ($=a + b$) and its error.

Theoretical Background

There is no special connection between the occurrence of correlation among the LS parameters and the error structure in the data, so for simplicity most of the present tests have involved unweighted least squares and hence the assumption of constant error in the data. As in the previous studies,[7,8] all error is assumed to reside in the response variable y.

In unweighted linear LS, the matrix $\mathbf{A}$ is given in terms of the design matrix $\mathbf{X}$ by $\mathbf{A} = \sigma_y{}^{-2}\mathbf{X}^T\mathbf{X}$, where $\sigma_y{}^2$ is the (constant) variance in y. This is a special case of weighted LS, where $\mathbf{A} = \mathbf{X}^T\mathbf{W}\mathbf{X}$, with $\mathbf{W}$ being diagonal and having elements $w_i = W_{ii} = \sigma_{yi}{}^{-2}$. Since the elements of $\mathbf{X}$ depend only on the independent variable x, $\mathbf{V}$ is known exactly once the model and the x-structure and error structure of the data are established, as already noted. A corresponding relation holds for $\mathbf{A}$ in nonlinear LS, except that the elements of the matrix $\mathbf{X}$ can depend on the parameters $\boldsymbol{\beta}$ and the response variable y (see below). However, an "exact" $\mathbf{V}$ can be defined here too, by simply employing exactly fitting data and the true parameter values.[7]

The rigorous validity of eq 1 for functions f that are linear functions of linear LS parameters follows from the linear transformation properties of such quantities.[9] In particular, if $\boldsymbol{\alpha}$ represents a set of quantities related to the linear LS parameters $\boldsymbol{\beta}$ by the linear transformation $\boldsymbol{\alpha} = \mathbf{L}\boldsymbol{\beta}$, then the values of the $\boldsymbol{\alpha}$ are the same as would be obtained by directly fitting the data to $\boldsymbol{\alpha}$; the corresponding variance-covariance matrix is given in terms of that for $\boldsymbol{\beta}$ by

$$\mathbf{V}_{\boldsymbol{\alpha}} = \mathbf{L}\,\mathbf{V}_{\boldsymbol{\beta}}\,\mathbf{L}^T \tag{5}$$

Equation 1 thus yields a selected diagonal element of $\mathbf{V}_{\boldsymbol{\alpha}}$ for $f = \alpha_i$, with $\mathbf{g}^T$ being the ith row of $\mathbf{L}$ and $\mathbf{g}$ the ith column of $\mathbf{L}^T$. Further, since the LS fits are linear, with normally distributed error in the data, both sets of parameters, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, are normally distributed. Equation 5 holds also for two sets of nonlinear LS parameters that are related through a linear transformation;[9] however, in this case the parameters are not normally distributed.

A result analogous to eq 5 can also be obtained for two different sets of nonlinear LS parameters that are not linearly related. For this purpose note that, while the matrix $\mathbf{A}$ is still given by $\mathbf{A} = \sigma_y{}^{-2}\mathbf{X}^T\mathbf{X}$, the elements of $\mathbf{X}$ are now partial derivatives of the fit function F with respect to the parameters, namely $X_{ij} = (\partial F_i/\partial\beta_j)$; these are evaluated at x_i using the converged values of the parameters $\boldsymbol{\beta}$.[7] Thus, in this fit $\mathbf{V}_{\boldsymbol{\beta}} = \sigma_y{}^2(\mathbf{X}^T\mathbf{X})^{-1}$. Now suppose that the fit is carried out for an alternatively defined set of parameters $\boldsymbol{\gamma}$. The new $\mathbf{V}$ is $\mathbf{V}_{\boldsymbol{\gamma}} = \sigma_y{}^2(\mathbf{Y}^T\mathbf{Y})^{-1}$, with the elements of $\mathbf{Y}$ defined analogously, $Y_{ij} = (\partial F_i/\partial\gamma_j)$. However, the partial derivatives with respect to one set of variables can be related to those with respect to the other by, for example

$$\left(\frac{\partial F_i}{\partial\beta_j}\right) = \sum_{k=1}^{p}\left(\frac{\partial F_i}{\partial\gamma_k}\right)\left(\frac{\partial\gamma_k}{\partial\beta_j}\right) \tag{6}$$

where the sum runs over the p adjustable parameters. Thus the matrix $\mathbf{X}$ can be related to $\mathbf{Y}$ by

$$\mathbf{X} = \mathbf{Y}\,\mathbf{U} \tag{7}$$

where the Jacobi matrix $\mathbf{U}$ is $p \times p$, with elements $U_{ij} = (\partial\gamma_i/\partial\beta_j)$. Accordingly

$$\mathbf{V}_{\boldsymbol{\gamma}} = \mathbf{U}\,\mathbf{V}_{\boldsymbol{\beta}}\,\mathbf{U}^T \tag{8}$$

and eq 1 can again be seen to give the ith diagonal element of $\mathbf{V}_{\boldsymbol{\gamma}}$, with $\mathbf{g}^T$ being the ith row of $\mathbf{U}$ and $\mathbf{g}$ the ith column of $\mathbf{U}^T$.

Equations 5 and 8 apply not just to different sets of parameters defined in terms of each other alone, but also to functions which include a dependence on the independent variable, e.g., the fit function itself at x_0 or its derivative. However, in such cases it is usually more efficient to use eq 1, because the direct fitting approach requires repeating the fit at different selected values of the independent variable, as is illustrated below. Equation 1 is also preferred in cases where the relation between the derived property and the originally fitted parameters is complex, as in the computation of RKR potential curves for diatomic molecules.[10]

Computational Methods

The partial derivatives required in eq 1 are evaluated numerically. For accuracy, these are estimated centrally, e.g., for three parameters ($\beta_1 = a$, etc.)

$$\left(\frac{\partial f(x;a,b,c)}{\partial a}\right) \approx \left(\frac{f(x;a + \Delta a/2,b,c) - f(x;a - \Delta a/2,b,c)}{\Delta a}\right) \tag{9}$$

In double precision arithmetic, Δa is usually set to 10^{-5}–10^{-7} a. This numerical approach makes the use of eq 1 straightforward even in cases where the derivatives cannot be expressed easily in closed form, e.g., in the aforementioned case of RKR potential curve calculations.

The Monte Carlo (MC) calculations employed routines like those described in the earlier works.[7,8] For the error propagation tests, the targeted quantities f were calculated using the results of each MC fit and then were binned and statistically evaluated along with the fit parameters.

The investigated models include (1) polynomial representations, linear through cubic, (2) an exponentially limiting function of form $a + b(1 - e^{-cx})$ (which is a special case of an exponential plus a background), and (3) a spectral band resolution model involving two Gaussian bands nearly coincident in wavelength. Model 1 is linear, so the parameters and linear functions thereof are rigorously normal, as already noted. These cases were used to validate the computational methods. The second model is linear if the constant c is fixed, nonlinear otherwise. The third is nonlinear unless the peak positions and widths are fixed. Functions of the parameters that involve products and ratios are nonlinear and yield nonnormal distributions for both linear and nonlinear LS parameters. In this category are useful quantities such as the areas under the two bands or the band area ratios and fractions in model 3.

To further illustrate how error propagation can be bypassed by redefining the fit parameters, suppose that data are fitted to a cubic polynomial, $y = b_0 + b_1x + b_2x^2 + b_3x^3$, and that the function f is the fit function itself. Here $\mathbf{g}^T = (1,x,x^2,x^3)$, and σ_f at any $x = x_0$ is readily obtained by numerical evaluation of the matrix product in eq 1. Alternatively, an equivalent fit can be obtained using the argument $z = (x - x_0)$. If the fit relation is defined as $y = c_0 + c_1z + c_2z^2/2 + c_3z^3/6$, the fit yields directly the values and errors for the fit function and all its derivatives at x_0. This recentering method can also be used on many nonlinear models, including model 2 above.

All of the LS fits studied in this work can be done with a number of microcomputer data analysis programs; I have used the KaleidaGraph program (Synergy Software).[11] Such programs do not normally include provision for evaluating eq 1, so the user must either write a short macro for this calculation or be able to define the desired quantity as a parameter in the fit. Still, the latter approach works in many cases that might not seem amenable to it at first thought. For example, the individual band intensities, band areas, and band-area ratios in model 3 can be handled this way, as is discussed below.

Results and Discussion

Linear Models.

As already noted, linear LS models yield particularly simple results and thus provide useful tests on the computational procedures. The cubic model used in a recent study[12] was employed in these preliminary checks. It had eight data values at $x = 1, 2, ..., 8$ generated as $y = 1 + 5x + 0.01x^2 - 0.025x^3$. In all checks the four fit parameters and the functions $f(x)$ and $f'(x)$ were normally distributed, with variances as predicted.

Interestingly, these predictions of linear LS apply even for fits to the *wrong* model. This statement holds not just for "slightly" wrong models, such as omitting the statistically ill-defined quadratic coefficient in the present cubic example, but also for drastically wrong models, including fitting the cubic data to a straight line. In all cases the parameters and derived functions are distributed as predicted for the respective fits. This result is at odds with naive anticipation. Of course χ^2 for wrong-model fits is systematically too large, being augmented by a variance term for the model error.[6] In the case of the straight-line fit of the present cubic data, the increase amounts to 49%, which is determined from a fit of the error-free data to a straight line.

Figure 1 illustrates distributions for products and ratios of the first two parameters in the cubic model. All of these are far from normal but become closer to normal as σ_y is reduced from 0.5 to 0.1, as shown in Figure 2. The most anomalous distribution is that associated with $f = b_1/b_0$, which exhibits "reciprocal statistics,"[7,8] indications of which persist even when σ_y is reduced by the factor of 5. The reason this behavior is so much more pronounced for b_1/b_0 than for its reciprocal is that the relative error in b_0 (σ_1/β_1) is much larger—1.23 vs 0.223 for b_1. Since the standard errors in the parameters scale with σ_y, both ratios drop by a factor of 5 as σ_y is decreased from 0.5 to 0.1.

A Nonlinear Example: Exponentials.

A useful function for data that have a nonzero large-x asymptotic limit is the form $y = a + b(1 - e^{-cx})$. As already noted, this model becomes linear when c is fixed, whereupon linear functions of a and b follow eq 1 rigorously. That includes the fit function itself, as was readily verified through the MC calculations.

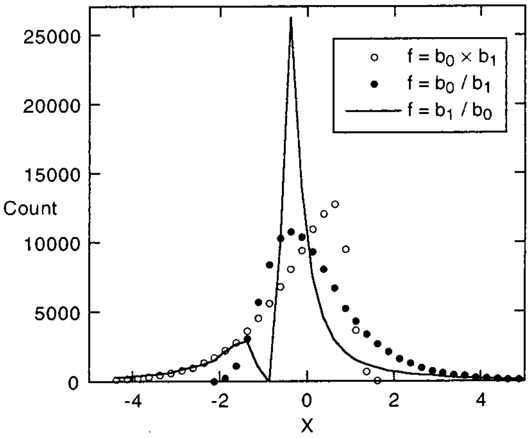

Figure 1. Histogrammed results of 10^5 Monte Carlo estimates of the product and ratios of the first two parameters (the constant and linear coefficients), as obtained from linear LS fits to the cubic model described in text. The binning argument X in this and subsequent histogram plots is $(f - f_{\rm true})/\sigma_f$. The error in the fitted data is $\sigma_y = 0.5$; the predicted relative errors (σ_f/f) are 1.02 (product) and 1.45 (ratios). (The statistical errors in the counts are smaller than the plotted points in this and subsequent figures.)

Interestingly, when c is included as an adjustable parameter, the most nonnormal parameter is b, and it is much more nonnormal that the fit function itself, as is illustrated in Figure 3. The asymmetry in the distribution of b is even more surprising, given that its relative standard error is only 4.3% here. Its MC sampled error is 10% larger than predicted, in mild violation of the "10% rule of thumb" stated earlier. The discrepancy drops to 2.4% when σ_y is reduced by a factor of 2, indicating that the parameter b is exhibiting divergent sampling statistics.[7]

Band Analysis Model.

The model of two nearly coincident Gaussian bands is illustrated in Figure 4, with results summarized in Table 1. The model was intentionally construed to yield large uncertainty in the component bands, even though the total is precise to within the width of the plotted curves. Calculations were done for two error structures: constant and proportional error. (These two mark the usual extremes in physical measurements.)

For both error structures the parameter distributions deviated only modestly from normal, with one exception: The centroid of the stronger component showed pronounced asymmetry, even though its inherent imprecision is small (see Figure 5). All parameters displayed biases that were statistically significant from the standpoint of the MC determinations, in both weighting schemes. These still amounted to at most ~10% of the corresponding exact standard errors, so they would not be of great practical import in an actual analysis. On the other hand, the biases scale with σ_y^2 while the parameter errors scale with σ_y,[7] so a tripling of the data error would make the biases a more significant 30% of the parameter errors. In this regard the smaller biases for the weighted model in Table 1 are misleading: if the data error is scaled to make each parameter error equal to that in the unweighted analysis, the biases in the proportional-error model exceed those in the constant-error model for four of the six parameters.

Among the examined properties from the analysis were the component band strengths at selected x values, the band areas,

TABLE 1: Comparison of "Exact" and Monte Carlo Results for Band Analysis Model

		biases and standard errors[b]					
		unweighted[c]			weighted[d]		
parameter[a]	value	bias[e]	σ_{exact}	σ_{MC}[f]	bias[e]	σ_{exact}	σ_{MC}[f]
1 (a_1)	300	5.1	66.35	65.5	3.8	39.32	40.0
2 (Δx_1)	75	−0.089	1.7008	1.698	−0.013	1.3127	1.306
3 ($x_{0,1}$)	520	0.018	0.4886	0.486	0.007	0.4689	0.469
4 (a_2)	500	−5.1	66.41	65.6	−3.8	39.87	40.6
5 (Δx_2)	90	0.126	1.0112	1.057	0.060	0.5104	0.534
6 ($x_{0,2}$)	515	−0.058	0.4067	0.432	−0.021	0.1754	0.183
$y_1(440)$	12.80	0.71	5.120	5.22	0.46	3.517	3.67
ratio[g]	2.000	0.113	0.7308	0.774	0.021	0.4447	0.443

[a] Two Gaussian bands: $y(x) = a \exp[-4 \ln 2((x - x_0)/\Delta x)^2]$. Last two rows give derived quantities and their propagated errors. [b] 4×10^4 spectra employed in Monte Carlo calculations. [c] $\sigma_y = 1.0$. [d] $\sigma_y = y/100$ and $w_i = \sigma_{yi}^{-2}$, evaluated using the true rather than the randomized y_i. [e] $\langle\beta\rangle_{\mathrm{MC}} - \beta_{\mathrm{true}}$; errors = $\sigma_{\mathrm{MC}}/200$. [f] Relative precision of MC σ values = $(2N)^{-1/2} = 0.0035$. [g] (Band 2 area)/(band 1 area).

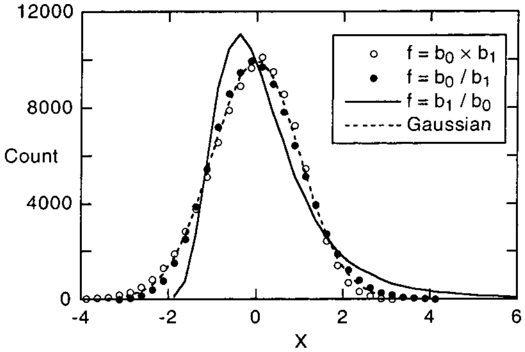

Figure 2. Results obtained as in Figure 1, but with the data error reduced by a factor of 5, to $\sigma_y = 0.1$. A unit-variance Gaussian is included for comparison.

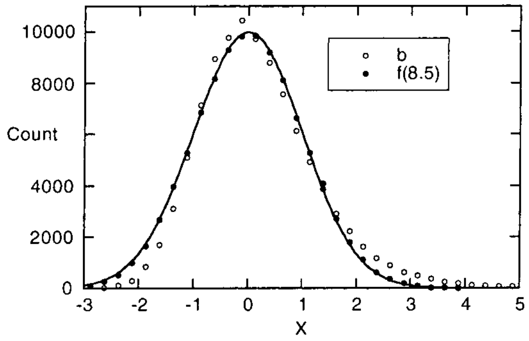

Figure 3. Results of 10^5 Monte Carlo estimates of the parameter b and the fit function $f = a + b(1 - e^{-cx})$ evaluated at $x_0 = 8.5$. The model had true values of 1, 35, and 0.2 for a, b, and c, respectively, with eight points at $x = 1, 2, ..., 8$ and $\sigma_y = 0.5$. The "exact" standard errors for b and $f(x_0)$ are 1.488 and 0.451, respectively. Despite the good visual agreement between the solid points and the Gaussian curve, the weighted fit of these data fails a chi-square test ($\chi^2 = 90.2$ for 28 degrees of freedom).

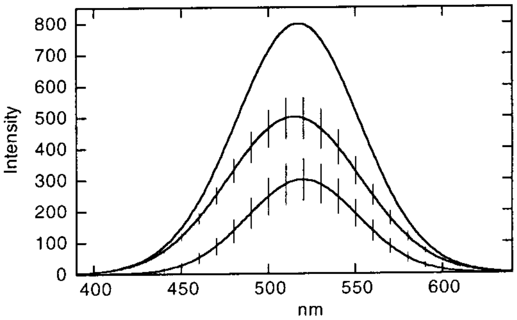

Figure 4. Band analysis model, showing component bands and their errors (1σ), as calculated using eq 1 for the case of constant error ($\sigma_y = 1.0$). Points were generated from $x = 400$ to $x = 630$ at intervals of 2.0.

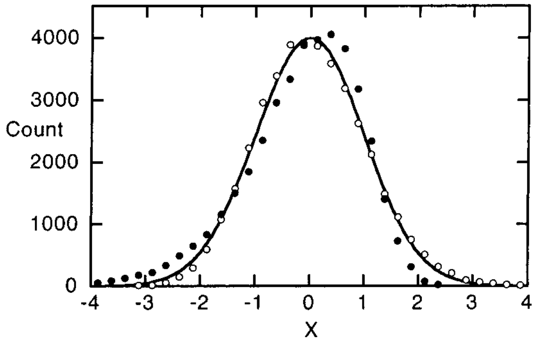

Figure 5. Results (histogram counts) of 4×10^4 Monte Carlo estimates of the centroids for the weak (open points) and strong component bands in the constant-error model. Similar results were obtained for the proportional-error model.

band-area ratios, and fractional band area. Not surprisingly, all of these displayed bias and nonnormality on a scale comparable to that exhibited by the fit parameters themselves. Figure 6 shows that in the wings of the spectrum the two components are quite nonnormal, with the skewness of the distributions reflecting the anticorrelation of the components, needed to preserve the precise total. Nonetheless, the sampled standard errors are within 2% of predictions and the biases are moderate (see Table 1).

Figure 7 displays results for the ratio of band areas. In this case the distributions for the two error structures are quite different, with both the asymmetry and bias being much smaller for proportional error. Still, the MC standard errors are close to the predicted values (Table 1), despite the large relative error in this quantity. For comparison, the errors predicted by eq 2 are 30% too small in both cases.

All of the errors in derived properties obtained here via eq 1 can just as well be obtained directly from the fit through a redefinition of the fit parameters. For example, if a component band is defined as

$$y(x) = A \exp\{4 \ln 2(\Delta x)^{-2}[(x_1 - x_0)^2 - (x - x_0)^2]\} \quad (10)$$

the fit yields directly the amplitude A of the band at $x = x_1$ and its error. Thus one can generate complete error bands on the components, as shown in Figure 4, by varying x_1 systematically and rerunning the fit. Similarly, since the band-area ratio $R = (a_2\Delta x_2)/(a_1\Delta x_1)$, reexpressing the amplitude parameter for the stronger band as $R(a_1\Delta x_1)/\Delta x_2$ will yield directly R and its error.

Applications.

Linear fitting is widely used in the construction of calibration curves[12] and in the empirical representation of data as functions

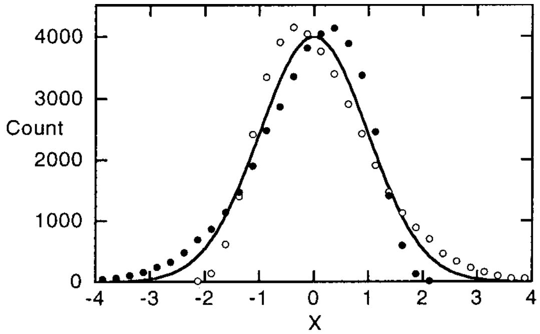

Figure 6. Histogram counts for the two component bands at $x = 440$, identified as in Figure 5. The predicted values are 12.8 ± 5.1 and 72.9 ± 5.1. Again, similar results were obtained for the proportional-error model.

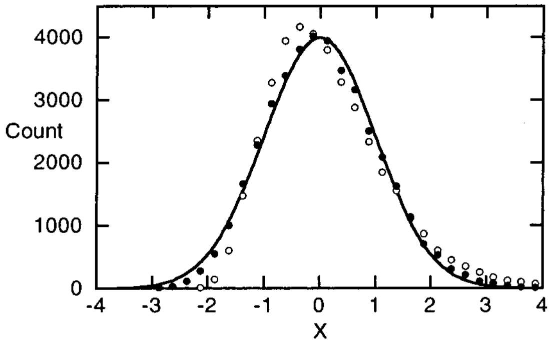

Figure 7. Count distribution for the strong:weak band area ratio in the constant-error (open points) and proportional-error models.

of their independent variable, e.g., thermochemical and kinetics data as functions of the temperature T and spectroscopic properties as functions of vibrational and rotational quantum numbers. Sometimes it is the first derivative that is sought, for example in the extraction of partial molar quantities for solutions, or ΔH from the T-dependence of equilibrium constants or vapor pressures. Equation 1 is required for the proper computation of errors in these cases, but often it can be avoided by the recentering approach, or by some other redefinition of the fit parameters. When abundant, very precise data are involved, such fits may require many adjustable parameters. These are normally highly correlated, so that the error bands on the functions as correctly evaluated by eq 1 are invariably much smaller than those obtained by the (incorrect) use of eq 2. For example, eq 2 gives a computed standard error a factor of 48 too large at $x_0 = 8$ in the cubic model discussed above.

Traditionally physical scientists have preferred straight-line relations for interpreting data, and linear fitting is often still used in cases where the relations among the desired quantities are nonlinear. Included in this category are equilibrium binding constant data, some kinetics data (enzyme kinetics via the Lineweaver−Burk equation, unimolecular conversion), and adsorption data. For example, binding constant data can take the form[13]

$$y = \frac{aKx}{1 + Kx} \tag{11}$$

where K is the binding constant, x the prepared concentration of ligand, and a a scaling parameter. A nonlinear fit to eq 11 will yield directly K and its error,[7] whereas a linear fit to the relation

$$\frac{x}{y} = \frac{1}{aK} + \frac{x}{a} \equiv A + Bx \tag{12}$$

yields K as B/A and therefore requires eq 1 for proper assessment of the error in K. Fits to eqs 11 and 12 will not yield identical values of K, because the data inversion process in eq 12 leads to biased estimates.[8] However, if this data bias is neglected (e.g., for error-free data), the (nonlinear) first form of eq 12 yields a K identical with that obtained from B/A, and also yields directly a correct value of σ_K.

Conclusion

Least-squares parameters are normally correlated, and in the calculation of statistical errors in functions of the parameters, this correlation must be taken into account. This is easily done using the underutilized matrix expression of eq 1. In many cases the same can be accomplished through a judicious definition of the adjustable parameters in the least-squares model itself. Monte Carlo calculations verify the expected normal distributions in linear functions of normal parameters but demonstrate pronounced nonnormality in some nonlinear functions of the parameters.

Even though nonlinear parameters and nonlinear functions of linear parameters are not normally distributed, many cases in practice are likely to fall under the validity of the 10% "rule of thumb":[7] if the parameter or derived property has a standard error less than 10% of its magnitude, its directly estimated error (from **V**) or its propagated error (from eq 1) should prove reliable for estimating confidence limits within 10%. In this regard it should be noted that, in many cases, the data error is not known at the outset and must be assessed from the fit itself. This leads to a relative uncertainty of $(2\nu)^{-1/2}$ in the estimates of the parameter standard errors, where ν is the number of degrees of freedom in the fit. This uncertainty will often match or exceed the errors in confidence limits stemming from the Gaussian interpretation of non-Gaussian distributions.

Of course there is no way that the variance-covariance matrix or eq 1 can convey any information about the extent of deviations from normality in the distributions, so for cases where these are of interest, the Monte Carlo method will remain indispensable, as it will also for the assessment of bias.

References and Notes

(1) Hamilton, W. C. *Statistics in Physical Science: Estimation, Hypothesis Testing, and Least Squares*; The Ronald Press Co.: New York, 1964.

(2) Deming, W. E. *Statistical Adjustment of Data*; Dover: New York, 1964.

(3) Albritton, D. L.; Schmeltekopf, A. L.; Zare, R. N. An introduction to the least-squares fitting of spectroscopic data. In *Molecular Spectroscopy: Modern Research II*; Narahari Rao, K,. Ed.; Academic Press: New York, 1976; pp 1−67.

(4) Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences*; McGraw-Hill: New York, 1969.

(5) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes*; Cambridge University Press: Cambridge, UK, 1986.

(6) Draper, R. N.; Smith, H. *Applied Regression Analysis*, 3rd ed.; Wiley: New York, 1998.

(7) Tellinghuisen, J. *J. Phys. Chem. A* **2000**, *104*, 2834−2844.

(8) Tellinghuisen, J. *J. Phys. Chem. A* **2000**, *104*, 11829−11835.

(9) Albritton, D. L.; Schmeltekopf, A. L.; Tellinghuisen, J.; Zare, R. N. *J. Mol. Spectrosc.* **1974**, *53*, 311−314.

(10) Tellinghuisen, J. *J. Mol. Spectrosc.* **1990**, *141*, 258−264.

(11) Tellinghuisen, J. *J. Chem. Educ.* **2000**, *77*, 1233−1239.

(12) Tellinghuisen, J. *Analyst* **2000**, *125*, 1045−1048. The first end note in this paper incorrectly states that in nonlinear LS, error propagation via eq 1 is not formally equivalent to directly fitting the target quantity. The small numerical discrepancies on which this conclusion was founded were artifactual and disappeared when the convergence criterion in the nonlinear fits was changed from 10^{-6} relative change in the summed squared residuals to 10^{-12}. Equations 6−8 in the present work show the formal equivalence of the two approaches.

(13) Bowser, M. T.; Chen, D. D. Y. *J. Phys. Chem. A* **1998**, *102*, 8063−8071.