A Monte Carlo Study of Precision, Bias, Inconsistency, and Non-Gaussian Distributions in Nonlinear Least Squares

Joel Tellinghuisen

Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235

Received: September 14, 1999; In Final Form: January 11, 2000

The analysis of unbiased, normally distributed (Gaussian) data by linear least-squares yields unbiased, normally distributed parameter estimates, whose standard errors are known exactly if the error structure of the data is known. By contrast, nonlinear models generally yield nonnormal, biased parameter estimates, for which the standard errors depend on the parameter values, hence vary among statistically equivalent data sets. These properties of least squares are examined through Monte Carlo calculations on three important models: the straight-line linear relationship, the rectangular hyperbola of relevance to the analysis of equilibrium binding constant and Michaelis−Menten kinetics data, and the declining exponential. Parameter distributions are precisely characterized by processing 10^5 equivalent data sets in a typical Monte Carlo run. The main conclusions are as follows. (1) "Exact" parameter errors can be obtained easily from the variance−covariance matrix for a perfectly fitting data set having known data error. (2) The "normal" interpretation of these exact standard errors yields parameter confidence limits that are reliable to better than 10% if the relative standard error is $<1/10$. (3) Parameter bias scales with the data variance (σ_y^2) and is very nearly linear in n^{-1}, where n is the number of data points. (4) With proportional error structures ($\sigma_{y_i} \propto y_i$) and Poisson structures ($\sigma_{y_i}^2 \propto y_i$), there is an inherent and unavoidable ambiguity in assessing the weights; weights based on either observed or calculated y_i lead to bias that persists as $n \rightarrow \infty$, or *inconsistency*, in many of the nonlinear estimators. (5) Certain key parameters such as the binding constant and the exponential decay rate are consistent. (6) A common source of highly asymmetric parameter distributions is a parameter which is the reciprocal of a relatively uncertain normal or near-normal parameter; such reciprocal parameters have infinite variance, so when uncertain by $\gtrsim$20%, their signature in Monte Carlo calculations is unstable or divergent statistics.

Introduction

The method of least squares (LS) is the default data analysis tool in most of physical science. The statistical properties of *linear* least-squares estimators are well-known.[1−3] These include, importantly: If the data are distributed independently and normally (i.e., with the Gaussian distribution) about their true values, the estimators of the parameters will be *unbiased* and *minimum-variance* and will themselves be normally distributed about their true values, with variances that can be calculated directly from the *matrix of the normal equations*. The structure of the latter quantity depends only upon the distribution of values for the independent variable(s) and therefore can be calculated at the outset, permitting its use in *experimental design*.

More often than not, the theoretical expressions that predict the relationships among the experimentally measured variables are not linear in the desired parameters, necessitating analysis by *nonlinear* least squares. Included among these cases are most of the situations in which the directly measured quantities are transformed into the linear relation, $y' = A + Bx'$, where y' and x' are the transformed variables, and where the linear parameters A and B may depend in various ways on the original parameters, a and b. From a computational standpoint, the chief differences between linear and nonlinear LS are the following: Linear LS equations are, in principle, solvable in a single step and yield a single, minimum-variance solution; nonlinear problems must be solved iteratively and may converge on multiple solutions, or may not converge at all.

Another important difference between linear and nonlinear LS is that the guarantees of no bias, minimum variance, and normal parameter distributions that reassure practitioners of linear LS are absent from nonlinear LS.[3,4] This has led many to turn to Monte Carlo (MC) calculations for information about these properties, especially the confidence intervals to be expected for the parameter estimates. Among recent works of this sort are studies of exponential decay,[5−8] of the relationships $y = a/(b + x)$ and $y = ax/(b + x)$,[9−13] and of several more specialized functions.[14−16] Such studies have been facilitated by the increasing availability in recent years of microcomputer data analysis packages that include provision for fitting data to "user-defined" functions.

As already noted, a very important aspect of the method of least squares is that it provides for built-in estimation of the statistical errors in the parameters. This holds for both linear and nonlinear LS. Unfortunately the commercial data analysis programs do not always treat parameter errors with as much respect as they do the parameters themselves. A main purpose of the present work is to review this feature of least squares and, through examples, to suggest that the built-in estimates of parameter error will usually suffice for the estimation of confidence intervals. In fact, these "parametric" estimates may be even more reliable than results obtained from the Monte Carlo approach used in many recent studies, including most of those in refs 8−16.

The question of bias in nonlinear LS is another topic that has been addressed using Monte Carlo calculations. A statistical

10.1021/jp993279i CCC: $19.00
Published on Web 02/26/2000

estimator is said to be biased if its expectation (or average) value does not equal the true value of the quantity being estimated. If the bias persists in the limit $n \rightarrow \infty$, where n is the number of points in the data set, the estimator is *inconsistent*.[1] For the purpose of investigating bias and consistency, it does appear that some form of MC approach is needed in most cases. However, one must be careful in choosing which *form* of the parameter to study in this manner. In many of the cases of most significant apparent bias, it turns out that the parameter under study is actually the *reciprocal* of a normally or nearly normally distributed parameter.[5,17] The distribution function for such reciprocals has Lorentzian tails; this means that the *variance is not finite*, which in turn means that the central limit theorem does not apply and that, therefore, the parameter cannot be estimated reliably by MC averaging. This can be a serious matter any time the standard error in the parameter is greater than $\sim^1/_5$ its value but will be of little practical concern when the relative error is less than $^1/_{10}$.

In the following, I first review the formalism of linear and nonlinear least squares. Using MC calculations of 10^5 equivalent data sets, I then benchmark the computational procedures on a simple linear model before examining in detail two nonlinear models, the function $y = ax/(b + x)$, of relevance for the analysis of binding constant data, and the declining exponential, with and without a background. For these nonlinear models, the parameter distributions are fundamentally nonnormal; however, this nonnormality is generally negligible for the estimation of parameter confidence intervals, if the special problems of reciprocal parameters are eliminated. Methods for recognizing such reciprocal parameters are described. The nonlinear estimators generally exhibit bias, and for some of the estimators in cases of nonconstant data error the bias persists as $n \rightarrow \infty$. This inconsistency can be a more serious concern. Another potential source of difficulty, nonnormality and bias in the data themselves, is treated elsewhere[17] and will be considered only briefly here. If the data are not normally distributed about their true values, then many of the predictions of the least-squares method no longer hold, even for linear LS.

Theoretical Background

Linear Least Squares. *Matrix Formulation*. The least-squares equations are obtained by minimizing the sum of weighted squared residuals S,

$$S = \Sigma\, w_i \delta_i^{\,2} \tag{1}$$

with respect to a set of adjustable parameters $\boldsymbol{\beta}$, where δ_i is the residual (observed−calculated mismatch) for the ith point and w_i is its weight. For the purpose of the matrix notation that follows, $\boldsymbol{\beta}$ is a column vector containing p elements, one for each adjustable parameter. Thus, its transpose is a row vector: $\boldsymbol{\beta}^{\mathbf{T}} = (\beta_1, \beta_2, ..., \beta_p)$. The problem is a linear one if the measured values of the dependent variable (y) can be related to those of the independent variable(s) (x, u, ...) and the adjustable parameters through the matrix equation[2,18]

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \delta \tag{2}$$

where $\mathbf{y}$ and $\boldsymbol{\delta}$ are column vectors containing n elements (for the n measured values) and the *design matrix* $\mathbf{X}$ has n rows and p columns, and depends only on the values of the independent variable(s) (assumed to be error-free) and not on the parameters $\boldsymbol{\beta}$ or dependent variables $\mathbf{y}$. For example, a fit to $y = ax + b/x^2 + c\ \ln(3u)$ qualifies as a linear fit, with two independent variables (x, u), three adjustable parameters (a, b, c), and $\mathbf{X}$ elements $X_{i1} = x_i$, $X_{i2} = x_i^{-2}$, $X_{i3} = \ln(3u_i)$. On the other hand, the fit becomes nonlinear if, for example, the first term is changed to x/a, or the third to $3\ \ln(cu)$. It also becomes nonlinear if one or more of the "independent" variables is not error-free and hence is treated (along with y) as a dependent variable.

The solution to the minimization problem in the linear case is the set of equations

$$\mathbf{X^T W X}\boldsymbol{\beta} \equiv \mathbf{A}\boldsymbol{\beta} = \mathbf{X^T W y} \tag{3}$$

where the square weight matrix $\mathbf{W}$ is, in this case, diagonal, with n elements $W_{ii} = w_i$. (More generally, the measured y_i may not be mutually independent, rather may show correlation; in that case there are nonzero off-diagonal elements in $\mathbf{W}$.[2,18−20]) The matrix $\mathbf{A}$ is the previously mentioned *matrix of the normal equations*. Equations 3 are solved for the parameters $\boldsymbol{\beta}$, e.g.,

$$\boldsymbol{\beta} = \mathbf{A}^{-1}\mathbf{X^T W y} \tag{4}$$

where $\mathbf{A}^{-1}$ represents the inverse of $\mathbf{A}$. Knowledge of the parameters then permits calculation of the residuals $\boldsymbol{\delta}$ from eq 2 and thence of S, which in matrix form is

$$S = \boldsymbol{\delta}^{\mathbf{T}}\mathbf{W}\boldsymbol{\delta} \tag{5}$$

Importantly, the variances in the parameters are the diagonal elements of the *variance−covariance matrix* $\mathbf{V}$, which is proportional to $\mathbf{A}^{-1}$ (see below).

For these equations to make sense, it is essential that the measurements y_i be drawn from parent distributions of finite variance.[2] (This, for example, excludes Lorentzian parent distributions.) If additionally they are *unbiased* estimates of the true means, then the LS equations will yield unbiased estimates of the parameters β. If the parent distributions are normal, then the parameter estimates will also be normally distributed. For these to be *minimum-variance* estimates as well, it is necessary that the weights be taken as proportional to the inverse variances,[1,2]

$$w_i \propto \sigma_i^{\,-2} \tag{6}$$

Under these conditions, least squares is also a *maximum likelihood* method, which is reassuring, since maximum likelihood is the more fundamental statistical principle behind data analysis in physical science. Note that it is possible to have LS estimators that are unbiased but not minimum-variance, or minimum-variance but not unbiased, or even unbiased and minimum variance, but nonnormal.

If the parent distributions for the data are normal and the proportionality constant in eq 6 is taken as 1.00, then the quantity S is distributed as a χ^2 variate for $\nu = n - p$ degrees of freedom.[1−3] Correspondingly, the quantity S/ν follows the *reduced chi-square* (χ_ν^2) distribution, given by

$$P(z)\, dz = Cz^{(\nu-2)/2} \exp(-\nu z/2)\, dz \tag{7}$$

where $z = \chi_\nu^2$ and C is a normalization constant. It is useful to note that a χ^2 variate has a mean of ν and a variance of 2ν,[21] which means that χ_ν^2 has a mean of unity and a variance of $^2/_\nu$. In the limit of large ν, $P(z)$ becomes Gaussian.

A Priori and a Posteriori Weighting. If the proportionality constant in eq 6 is taken as unity, then the proportionality constant connecting $\mathbf{V}$ and $\mathbf{A}^{-1}$ is likewise unity, giving

$$\mathbf{V} = \mathbf{A}^{-1} \tag{8}$$

If the parent data distributions are normal, the parameter

distributions are also normal, as already noted. Then, the confidence intervals for the parameters can be evaluated straightforwardly from the standard error function tables. For example, the 95% confidence interval on β_1 is $\pm 1.96 V_{11}^{1/2}$. Note that this is always the case in MC calculations on linear fit models with normally distributed errors, since the σ_{yi} are set by the computationalist. Accordingly, in such calculations 95% of the estimates of β_1 are expected within $\pm 1.96 V_{11}^{1/2}$ of the true value. Conversely, a significant deviation from this prediction indicates a flaw in the MC procedures.

There is some confusion in the literature regarding these matters. In general, the off-diagonal elements in **V** (the covariances) are nonzero, for both linear and nonlinear fits. This means that the parameters $\boldsymbol{\beta}$ are *correlated*. However, each of the parameters in a linear fit is distributed normally about its true value, with $\sigma_{\beta i} = V_{ii}^{1/2}$, irrespective of its correlation with the other parameters; and this goes for *all* the parameters, including, for example, all four parameters in a fit to a cubic polynomial. The correlation comes into play only when we ask for *joint* confidence intervals of two or more parameters, in which case the confidence bands become ellipsoids in two or more dimensions.[4]

The use of eq 8 implies prior knowledge of the statistics of the y_i. Accordingly, the weights obtained in this manner may be designated as a priori weights.[20] (The term "internal consistency" was attached to this case by Birge[22] and Deming.[23]) Note that this a priori **V** is also *exact*, not an estimate. Unfortunately, from the experimental side we never have perfect a priori information about the statistics of the data. However, there are cases, especially with extensive computer logging of data, where the a priori information may be good enough to make eq 8 the proper choice and the resulting **V** *virtually* exact. A good example is data obtained using counting instruments, which often follow Poisson statistics closely, so that the variance in y_i ($\sigma_{y_i}^2$) can be taken as y_i. (For large y_i Poisson data are also very nearly Gaussian.)

At the other extreme, we have the situation where nothing is known in advance about the statistics of the y_i, except that we believe the parent distributions all to have the same variance, independent of y_i. In this case, the weights w_i can all be taken to be the same constant, which without loss of generality we can take to be 1.00. This is the case of unweighted least squares. The variance in y is then estimated from the fit itself as

$$\sigma_y^2 \approx s_y^2 = \frac{\Sigma \delta_i^2}{n-p} = \frac{S}{\nu} \tag{9}$$

which is recognized as the usual expression for estimating a variance by sampling. The use of eq 9 represents an a posteriori assessment of the variance in y_i. (This was designated "external consistency" by Birge[22] and Deming.[23]) The variance−covariance matrix now becomes

$$\mathbf{V} = \frac{S}{\nu}\mathbf{A}^{-1} \tag{10}$$

Under the same conditions as stated before eq 7, s_y^2 is distributed as a scaled χ^2 variate. This means, for example, if the s_y^2 values from a Monte Carlo treatment of unweighted LS are divided by the true value σ_y^2 used in the MC calculations, the resulting ratios are distributed in accord with eq 7 for χ_ν^2.

Inclusion of the factor S/ν in the definition of **V** in the case of a priori weighting would be equivalent to scaling **V** by χ_ν^2. While this does not affect the statistical average of the estimated variances (since the average value of χ_ν^2 is 1), it does introduce a spread in the results that might be obtained in a given experiment. But this is an unnecessary and improper spread if the statistics of the y_i are truly known at the outset, and eq 8 is the correct expression for **V** in this case.[20,24] Then the value of χ_ν^2 ($=S/\nu$) returned by the fit can be used to assess the *goodness of fit*.

In the case of a posteriori assessment, the uncertainty in s_y does not greatly compromise the reliability of the parameter standard error estimates when the data set is large. For example, since the variance in χ_ν^2 is $2/\nu$, the relative standard deviation in s_y^2 is 0.1 when $\nu = 200$. This translates into a 5% relative standard deviation in s_y ($\sqrt{1/(2\nu)}$) and hence also in all the parameter standard error estimates ($V_{ii}^{1/2}$).

What about the confidence limits on the parameters in the case of a posteriori assessment? The need to rely upon the fit itself to estimate s_y means the parameter errors are no longer exact but are uncertain estimates. Accordingly, we must employ the *t*-distribution to assess the parameter confidence limits. Under the same conditions that yield a normal distribution for the parameters $\boldsymbol{\beta}$ and scaled χ^2 distributions for s_y^2 and for the V_{ii} from eq 10, the quantities $(\beta_i - \beta_{i,\mathrm{true}})/V_{ii}^{1/2}$ belong to the *t*-distribution for ν degrees of freedom,[1] which is given by

$$f(t)\,\mathrm{d}t = C'(1 + t^2/\nu)^{-(\nu+1)/2}\,\mathrm{d}t \tag{11}$$

with C' another normalizing constant. For small ν, the *t*-distribution is narrower in the peak than the Gaussian distribution, with more extensive tails. However, the *t*-distribution converges on the unit-variance normal distribution in the limit of large ν, making the distinction between the two distributions unimportant for large data sets (except in the prediction of outliers in the far wings).

The distinction between the use of eq 8 for **V** and the normal distribution for parameter confidence limits vs eq 10 and the *t*-distribution is often blurred in the literature. For example, MC calculations start with assumed values for the uncertainties in the data; and if these errors are normally distributed and the model is linear, the calculations will certainly yield normal distributions for the parameters, if correctly done. Introduction of a *t*-distribution-based enlarging factor for the confidence limits on the parameters is thus a sort of "apples vs oranges" procedure, in that it implicitly assumes that the data analyst will be working with similar data but without the benefit of any knowledge of the quality of those data.

The structure of **V** is such that it scales with σ_y^2 and with $1/n$. The latter, for example, means that the parameter standard errors ($V_{ii}^{1/2}$) do correctly go as $n^{-1/2}$, all other things being equal. This means that they are to be interpreted in the same manner as the standard deviation in the mean in the case of a simple average. (One can readily verify that, for a fit of data to $y = a$, the equations do yield for σ_a the usual expressions for the standard deviation in the mean.)

Intermediate Situations. Sometimes one has a priori information about the *relative* variation of σ_{y_i} with y_i but not a good handle on the *absolute* σ_{y_i}. For example, data might be read from a logarithmic scale, or transformed in some way to simplify the LS analysis. As a specific example of the latter, data might be fitted to $y = ax + bx^2$ by first dividing by x to yield $y' \equiv y/x$, then fitting to $y' = a + bx$. If the original y_i have constant standard deviation σ_y, then simple error propagation (with x still treated as error-free) shows that the standard deviations in the y_i' values are σ_y/x_i, meaning the weights w_i are $\propto x_i^2$.

Clearly, a weighted fit is called for in the latter "straight-line" analysis; neglect of the now-unequal weighting of the dependent variable will fail to yield the desired minimum-

variance estimates of the parameters (though will still produce unbiased estimates in this case). A check will show that the weighted fit to $y' = a + bx$ yields a set of equations (eqs 3) identical to those for the unweighted fit to $y = ax + bx^2$. This appears to be a general property of linear least-squares fits to alternative forms relatable by linear variable transformations (which preserve the normal structure of the original data). Also, the results for both $\boldsymbol{\beta}$ and $\mathbf{V}$ (through eq 10) are independent of arbitrary scale factors in the weights. If the latter are taken as simply $w_i = x_i^2$, then S/ν (with S given by eq 1 or 5) will be an estimate of σ_y^2. Since this is an a posteriori assessment, the t-distribution should be used to obtain the confidence limits on the parameters.

Another situation is the case where data come from two or more parent distributions of differing σ_y, but again known in only a relative sense. As before, the results of the calculations are independent of an arbitrary scale factor in the weights. However, to obtain meaningful estimates of the parent variances, it is customary to designate one subset as reference and assign $w_i = 1$ for these data, with all other weights taken as s_{ref}^2/s_i^2 (hence the need for knowledge of the relative precisions). Then, the quantity S/ν ($=s_{\mathrm{ref}}^2$) obtained from the fit is more properly referred to as the "estimated variance for data of unit weight," and the estimated variance for a general point in the data set is then s_{ref}^2/w_i.

A cautionary note is in order for users of commercial data analysis programs. Those programs which do provide estimates of the parameter errors do not always make clear which equation, (8) or (10), is being used. For example, the program KaleidaGraph (Synergy Software) uses (10) in unweighted fits to user-defined functions, but (8) in all weighted fits. This means that in cases such as those just discussed, where the weights are known in only a relative sense, the user must scale the parameter error estimates by the factor $(S/\nu)^{1/2}$ to obtain the correct a posteriori values. (In the KaleidaGraph program the quantity S is called "Chisq" in the output box but is a scaled χ^2 variate except in cases where the input σ_i values are valid in an absolute sense.)

Nonlinear Least Squares. In nonlinear fitting the quantity minimized is again S, and the least-squares equations take a form very similar to eq 3 but must be solved iteratively. The search for the minimum in S can be carried out in a number of different ways;[3,4] but sufficiently near this minimum, the corrections $\Delta\boldsymbol{\beta}$ to the current values $\boldsymbol{\beta}_0$ of the parameters can be evaluated from[3,4,6,23]

$$\mathbf{X^T W X}\Delta\boldsymbol{\beta} \equiv \mathbf{A}\Delta\boldsymbol{\beta} = \mathbf{X^T W}\boldsymbol{\delta} \tag{12}$$

leading to improved values,

$$\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0 + \Delta\boldsymbol{\beta} \tag{13}$$

The quantities $\mathbf{W}$ and $\boldsymbol{\delta}$ have the same meaning as before, but the elements of $\mathbf{X}$ are $X_{ij} = (\partial F_i/\partial\beta_j)$, evaluated at x_i using the current values $\boldsymbol{\beta}_0$ of the parameters. The function F expresses the relations among the variables and parameters in such a way that a perfect fit yields $F_i = 0$ for all i. For the commonly occurring case where y can be expressed as an explicit function of x, F takes the form

$$F_i = y_{\mathrm{calc}}(x_i) - y_i = -\delta_i \tag{14}$$

In the case of a linear fit, starting with $\boldsymbol{\beta}_0 = 0$, these relations yield for $\boldsymbol{\beta}_1$ equations identical to eqs 3 and 4 for $\boldsymbol{\beta}$. In the more general case where y cannot be written explicitly in terms of the other variables, these equations still hold, but with $\boldsymbol{\delta}$ in (12) replaced by $-\mathbf{F}_0$, where the subscript indicates that the F_i values are calculated using the current values $\boldsymbol{\beta}_0$ of the parameters.

Regardless of how convergence is achieved, the variance−covariance matrix is again given by eq 8 in the case of a priori weighting and eq 10 for a posteriori weighting, with $\mathbf{X}$ as redefined just below eq 13. However, there is an important distinction between $\mathbf{V}$ in the general nonlinear case vs the linear case: The matrix $\mathbf{A}$ now contains a dependence on the parameters. Also, in general there is no need to distinguish between dependent and independent variables in nonlinear fitting, as all variables may be taken to be uncertain.[23] In that case $\mathbf{A}$ may also depend on the values of all the variables, not just the (previously) independent variables. Thus, even in the case of a priori weighting, $\mathbf{V}$ from eq 8 will vary from data set to data set. However, one can extract estimates of $\mathbf{V}$ from a perfectly fitting theoretical curve and use this $\mathbf{V}$ in the same fashion as in the case of linear fitting.[6] In discussions below, I will refer to this $\mathbf{V}$ as the "exact" nonlinear $\mathbf{V}$.

This ability to predict nonlinear parameter variances at the outset appears not to be fully appreciated. While it is true that the parameter distributions in nonlinear fitting are generally not normal, in many situations they are close enough thereto to permit estimation of confidence intervals in this a priori fashion with a reliability that exceeds that achievable in typical Monte Carlo calculations. This is because the MC variance estimates are subject to the previously noted statistics of a χ^2 variate, which means for a 1000-set MC calculation a relative standard deviation of about 4.5% in the variances, or half that in the standard errors. And many published studies have employed far fewer than 1000 data sets, with concomitant loss in error precision as $N^{-1/2}$.

Error Propagation. One property of linear fitting is particularly appealing: Provided the adjustable parameters are truly overdetermined by the data at hand, one is assured of a numerical solution to the problem. This does not hold for nonlinear fitting, in which a poor choice of initial parameter estimates $\boldsymbol{\beta}_0$ can lead to divergence or slow convergence. Thus, there are times when a transformation to a linear form is a practical convenience. However, in such cases one must take care in estimating the errors in the desired (nonlinear) parameters, because the simple rules for error propagation do not apply.

For example, suppose the data are to be fitted to $y = 1/a + bx/a$. This is a nonlinear fit, and if pursued as such will yield proper estimates of σ_a^2 and σ_b^2 as the appropriate diagonal elements of $\mathbf{V}$. Alternatively, one might choose to fit to $y = A + Bx$, which is linear and with all the usual assumptions will yield normally distributed estimates of A and B. But A and B are correlated parameters, so the calculation of the estimated error in any function f of A and B must employ the full expression,[2,18,23,25]

$$s_f^{\,2} = \mathbf{g^T V g} \tag{15}$$

in which the elements of $\mathbf{g}$ are $g_i = \partial f/\partial\beta_i$, and $\mathbf{V}$ is the variance−covariance matrix obtained from the linear fit to A and B. In this case, a is a function of A alone, and the usual rules of error propagation apply. However, b is a function of both A and B ($b = B/A$), so the full expression must be used. One can verify that the estimates of σ_a^2 and σ_b^2 obtained from σ_A^2 and σ_B^2 using eq 15 are identical to those obtained directly from the nonlinear $\mathbf{V}$. Also, for any given data set, the nonlinear fit will yield a and b values identical to those obtained from the linear estimates of A and B. On the other hand, whereas the

a priori values of σ_A^2 and σ_B^2 are constant (i.e., independent of A and B), the a priori values of σ_a^2 and σ_b^2 depend explicitly on a and b. For example, in the present case $\sigma_a^2 = a^4\sigma_A^2$, so in MC calculations it will have the distribution of $a^4 = A^{-4}$ (see below).

Derived Parameter Distributions. If we know the probability distribution function $P(x)$, we can obtain the distribution function $Q(u)$ for a second variable u related to x by the function $u(x)$ using[1]

$$Q(u) = \Sigma\, P[x_i(u)]\left|\frac{\mathrm{d}x_i(u)}{\mathrm{d}u}\right| \tag{16}$$

where the sum is over all points x_i that solve $u = u(x_i)$. In many of the cases of interest in physical science, the mapping between u and x is one-to-one, and the sum in (16) becomes a single term. If $P(x)$ is the normal distribution, $Q(u)$ will not be normal unless $u(x)$ is linear in x.

Returning to the example in the previous section, suppose that data are fitted to $y = 1/a + bx/a$. Under the usually stipulated premises, we know that A and B from a fit of the same data to $y = A + Bx$ are normally distributed. Applying the two fits to a given data set yields $a = 1/A$, so the distribution of a is that of $1/A$. The mapping between A and a is one-to-one, and eq 16 yields

$$Q(a) = \frac{C}{a^2}\exp\left[-\frac{1}{2\sigma_A^2}\left(\frac{1}{a} - A_0\right)^2\right] \tag{17}$$

where A_0 is the true value of A and C is a normalizing constant. This function has Lorentzian wings; therefore, *the variance of a is infinite*. (This may be seen also from $\sigma_a^2 = \langle A^{-2}\rangle - \langle A^{-1}\rangle^2$, evaluated using $P(A)$.)

From a practical standpoint, the infinite variance of a may not be a problem if σ_A/A is sufficiently small. Then a is approximately normal about $a_0 = 1/A_0$, with $\sigma_a = a_0^2\sigma_A$, as predicted from simple error propagation. For example, if $\sigma_A/A = {}^1/_5$, the probability of a negative value of A is $<2 \times 10^{-7}$. Thus, in Monte Carlo calculations, the chances of hitting a value of A near zero are sufficiently small for nominal sample sizes that the sampling statistics for a might appear normal or near-normal. On the other hand, when $\sigma_A/A \gtrsim {}^1/_3$, the chances of such hits are large enough to destroy the sampling statistics for a, yielding large apparent biases in $\langle a\rangle$ and poorly defined (and nonconvergent) estimates of σ_a^2. Correspondingly, the sampling estimates of $\langle a\rangle$ do not converge, since the central limit theorem does not apply, even though $\langle a\rangle$ *is* mathematically defined in this case.

These considerations apply in similar fashion to data y_i that are normally distributed about their true values but are transformed nonlinearly, e.g., by inversion or logarithmic conversion.[17] The transformed data are biased estimators of the original quantities, so that even a properly weighted linear LS fit may yield biased estimates of the parameters β.

From an empirical standpoint, slightly nonnormal distributions can often be represented adequately as skewed Gaussians,

$$P(x) = Cf(x - x_0)\exp\left[-\frac{(x - x_0)^2}{2\sigma_x^2}\right] \tag{18}$$

where in simplest form the asymmetry function $f(x - x_0)$ can be taken as $1 + q(x - x_0)$, with q an adjustable parameter. In cases of more extreme distortion, an extra term in $(x - x_0)^3$ helps. With just the linear correction term the bias in x is

$$\langle x - x_0\rangle = q\sigma_x^2 \tag{19}$$

according to which the bias scales as the variance.

Computational Methods

The Monte Carlo calculations were carried out on a mainframe computer (DEC AlphaServer 2100A 4/275) using programs coded in FORTRAN. The built-in random number generator was used in most of the calculations, but with the additional "shuffle" of the RAN0 routine of Press et al.[4] incorporated at times. The uniform random deviates were converted to Gaussian using the Box−Muller method.[4] To minimize postprocessing of the very large files that would normally be produced in a run of 10^5 data sets, the distributional information was obtained by binning "on the fly." The statistical averages and higher moments were similarly computed by running accumulation. For the linear and near-linear models on five-point data sets, a typical run of 10^5 sets required only ~1 s of CPU time; this increased to several minutes for 10^5 60-point sets on some of the nonlinear models.

The statistics for the various quantities from the MC calculations (including the Gaussian random deviates themselves) were calculated by accumulating the appropriate sums and then dividing by the number of sets N at the conclusion. The variances were then calculated as, e.g., $s_a^2 = \langle a^2\rangle - \langle a\rangle^2$. For assessing the significance of bias, it is necessary to know the precision of the MC parameter estimates, which (at the 68.3% or 1σ level, provided they are normal) is their estimated standard error, $s_a/\sqrt{N}$. On the other hand, the sampling estimates of the parameter variances are subject to the previously mentioned properties of the χ^2 distribution, for N degrees of freedom in this case. Thus, their relative standard errors are $(2N)^{-1/2}$.

The histogrammed data were analyzed by fitting to the appropriate models using the user-defined curve-fitting function in the KaleidaGraph microcomputer program. The uncertainties in the binned values were taken as their square roots, in keeping with the Poisson nature of the binning process. Bins containing fewer than eight counts were normally omitted. For the most part, the values were fitted simply as sampled points. However, technically the bin counts represent integrals over the specified intervals. For the χ^2 distributions in particular, it was necessary to approximate this integral more accurately by breaking each binning interval into subintervals (usually 10) in the fitting function.

Results and Discussion

A Linear Model. To check out the computational procedures, I first conducted calculations for a straight-line linear model, $y = A + Bx$, with $A = 1$, $B = 5$, and five x values extending from 1.1 to 12 (other values: 3.3, 5.5, 8.3). Some tests employed constant uncertainty in y, but most used proportional uncertainty, with $\sigma_y = 0.04y$. These conditions were designed to make the relative precision in the intercept much lower than that in the slope, to facilitate comparisons that might relate to the relative precision. The results of these tests verified expectations for the case where all the usual assumptions are valid. (1) Unbiased, finite-variance data, properly weighted, yielded unbiased parameter estimates, with variances given by the exact **V** (eq 8). (2) Normally distributed data yielded normally distributed LS parameter estimates. (3) The parameter residuals divided by the corresponding a posteriori error estimates (from eq 10) were

t-distributed, and the quantities S/ν followed the reduced χ^2 distribution for the appropriate number of degrees of freedom ν.

If the usual assumptions are not valid, some of these predictions fail. Perhaps the most important such failure concerns the neglect of weights for data which should properly be weighted (heteroscedastic data). If unweighted regression is applied to such data, the parametric error estimates from eqs 8 and 10 are *completely unreliable* and may be either optimistic or pessimistic as compared with the "true" values, which can only be obtained from the MC calculations in this case. Although the parameter estimates remain unbiased, their errors are always larger than the values obtained with proper weighting (the minimum-variance values). Also, the S/ν values are no longer χ^2 distributed, nor are the a posteriori-assessed parameters *t*-distributed.

As an application of these results, consider a study published in this journal by Gonçalves et al.[16] These authors fitted reported experimental rate constants and their errors to $\ln k = a_1 + a_2/T + a_3 \ln T$, with T treated as error-free. While it is true that the logarithmic conversion of the raw data introduces bias, the effects of this on the parameters and their errors are nominal, provided the data are properly weighted for the log transformation. The fit is a linear one, so with neglect of this data transformation bias, the parameter estimates are normally distributed, with errors given exactly by the a priori **V**, eq 8. MC calculations should validate this result and in fact do so, whether one starts with a perfectly fitting theoretical set of ln *k* values or with the reported experimental values (Table 1 in ref 16). On the other hand, the latter yield a χ_ν^2 value of 9.29, which means that the a posteriori error estimates (eq 10) are a factor of 3.05 larger than the a priori. This appears to be the source of the factor of 3 discrepancy reported by these authors for parametric and MC statistics.[26] Incidentally, the very large χ_ν^2 value is an indicator of problems with the data or model or both; inspection of a graphical display of the data suggests that some of the low-T values may be erroneous.

Bias in Data and Weights. In the case of the linear model with proportional uncertainty, being able to "play God" in the MC calculations permitted proper weighting using $w_i = 1/\sigma_{yi}^2 = (0.04y_i)^{-2}$, with y_i being the *true* value of y at x_i. However, in the real world we would not know the true y_i values so would have to compute the weights either (1) using the *observed* y_i values or (2) using the *calculated* y_i at x_i. In the latter case, the weights become a part of the adjustment process and the fit therefore becomes nonlinear. Both choices result in biased estimates of the parameters. Moreover, as shown in Figure 1, the biases are very nearly linear in $1/n$ and extrapolate to nonzero values in the limit $n \rightarrow \infty$, for both choices. Thus, the estimators are inconsistent. Even with this bias and inconsistency, however, the parameter distributions remain normal, at least for the precision obtained from 10^5 equivalent data sets. Also, the S/ν distributions are adequately fitted by the reduced χ^2 function for $\nu = 3$; but interestingly, the $\delta\beta_i/s_{\beta i}$ values do not follow the *t*-distribution for *A* and are borderline *t*-distributed for the slope *B*. (Since the relative error in *A* is much larger than that in *B*, it is reasonable to conclude that the *t*-distribution does not hold for either.)

For a given n, the bias scales as σ_y^2, as was alluded to connection with eqs 18 and 19. This behavior is widely observed for both linear and nonlinear fits and is illustrated further below.

As was noted earlier, if the data that are fitted are themselves the nonlinear transforms of normally distributed quantities (e.g., reciprocals or logs), they become biased estimators of the true data and yield nonnormal parameter estimates that are biased for all n. Nevertheless, in the case of linear fits, the effect on the error estimates is generally nominal (<3% for ~10% data error).[17]

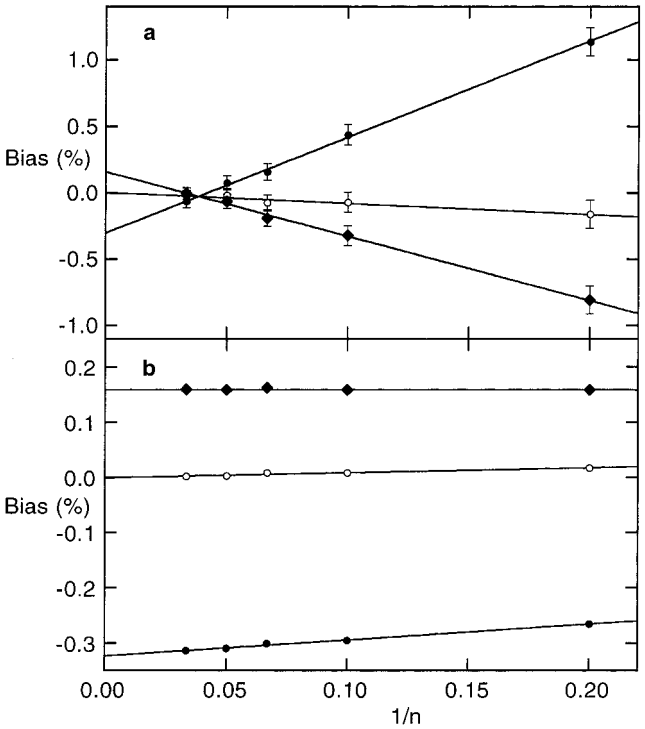

Figure 1. Bias in linear LS estimates of intercept A (a) and slope B (b), as a function of $1/n$. The linear model $y = A + Bx$ had $A = 1$ and $B = 5$, with the x-structure described in the text, and the y values were given proportional, normally distributed error ($\sigma_y = 0.04y$). The open points were obtained by evaluating the weights using the theoretical y_i values, the solid circles using the "observed" y_i, and the diamonds using the calculated y_i (optimized individually in each fit). Each point is the result of 10^5 fits; the error bars represent 1σ and are too small to appear in the lower plot. For the theoretical weighting, neither the intercept nor the slope is statistically significant at the 2σ level for either A or B; the bias is significant for all n in the other two cases.

A Nonlinear Model: Binding Constant Data. Monte Carlo methods have been employed to estimate parameter errors in a number of recent studies of rectangularly hyperbolic data, of relevance to the analysis of Michaelis−Menten kinetics and binding constant data.[10−13] Such data can be expressed in various ways, including (for binding constants)

$$y = \frac{aKx}{1 + Kx} \tag{20}$$

where K is the binding constant, x the prepared concentration of ligand, and a a scaling parameter. This expression assumes the ligand concentration is in great excess (although this is not necessary, since an exact treatment can be handled easily in nonlinear analysis). Equation 20 can be recast as a straight-line relationship a number of ways,[12] of which I will consider only

$$\frac{x}{y} = \frac{1}{aK} + \frac{x}{a} \equiv A + Bx \tag{21}$$

In both (20) and (21), x is customarily treated as error-free, and I will assume so also. Note that while the fit to A and B via (21) is linear (though inherently "data biased" if the y data are normal), the fit to a and K is not. However, for a given data set, the values of A and B returned by a linear fit will yield exactly the a and K produced by a nonlinear fit, and the errors

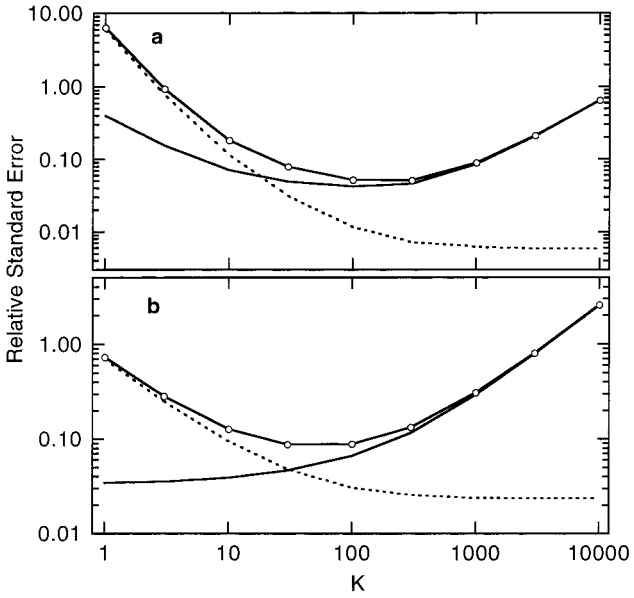

Figure 2. Relative standard errors in K (open points), A (solid line), and a (dashed), from fits of binding constant data (five points) to eqs 20 and 21 for $\sigma_y = 0.01$ (a) and $\sigma_y = 0.04y$ (b). The calculations employed eq 8 for **V** and perfectly fitting data having $a = 1$. Note that for this choice of a, $\sigma_B = \sigma_a$.

in a and K can be correctly evaluated from those in A and B using eq 15. Thus, we anticipate that the not-quite-normal (thanks to the inversion bias) distributions of A and B should translate into near-normal distributions for a and K, as long as both of these are determined with ~20% or better precision. Accordingly, the predictions from eqs 8 and 10 should suffice for estimating the parameter errors, obviating MC calculations for this purpose.

The only likely breakdown of the above predictions is the situation where one or both parameters display grossly nonnormal behavior, such that knowledge of the variance alone does not suffice to determine the confidence limits. In the mathematically similar case of vapor/solution equilibrium partitioning data,[9] the source of such failure is the occurrence of "reciprocal" behavior.[17] For example, since $a = 1/B$, a normally distributed B leads to a significantly nonnormal a if B (hence a) is relatively imprecise (say $\sigma_B/B > {}^1/_5$). Similarly, $K = B/A$, so if the intercept A is relatively imprecise, K will display reciprocal behavior. In this case, it may happen that its reciprocal K_d (the dissociation constant) is a statistically better defined quantity. Then it is necessary that B be relatively precise; however, imprecision in B is relatively unimportant in its effect on the distribution of K, as is imprecision in A for K_d.

To further illustrate these points, Figure 2 shows the exact relative standard errors computed for a, K, and A from fits of data to eqs 20 and 21, for two different error structures, constant and proportional. For the purpose of these calculations, I adopted the x-structure used in a recent MC study of this problem:[12] $x = 0.005$, 0.02375, 0.0425, 0.06125, and 0.08. The calculations can be conducted easily using some microcomputer data analysis programs; I used the KaleidaGraph program. The scaling parameter a was held at 1.00. It is easy to show that σ_K remains constant when a is scaled, as long as σ_y is similarly scaled, i.e., σ_K is independent of a for fixed σ_y/a.

From Figure 2, we see that for a fixed x-structure the relative error in K rises in the extremes of small and large K, a well-known result from many MC studies of this problem. At small K, nonnormality of K should never be a problem for proportional

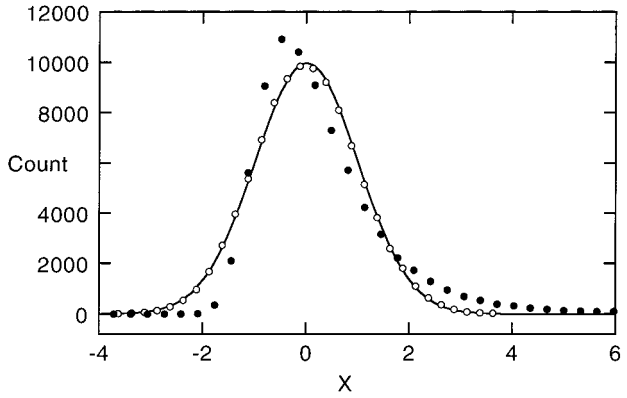

Figure 3. Histogrammed results for K (filled points) and $K^{-1} = K_d$ from 10^5 fits to eq 20 of five-point data sets having true values $a = 1$, $K = 3500$, $\sigma_y = 0.01$, and the x-structure given in text. The histogrammed quantity is $X = (\beta - \beta_{true})/\sigma_\beta$, with the true values being 3500 and 3500^{-1}, $\sigma_K = 855$, and $\sigma_{K_d} = 6.99 \times 10^{-5}$. The smooth curve is a Gaussian fitted to the K_d distribution; it yielded $\chi^2 = 86.1$ (30 points, $\nu = 29$).

error, because A is precisely determined in this regime. For constant error, σ_A does increase with decreasing K; however, A remains precise to a point where K itself is so uncertain as to render its nonnormality a secondary concern. At large K, on the other hand, the error in K is almost entirely due to the error in A for both error structures. Thus, in this regime, K_d is the statistically preferred quantity.

To clarify this last point, MC calculations were carried out for 10^5 data sets having $\sigma_y = 0.01$, $a = 1$, and $K = 3500$ (for which the exact calculations yield $\sigma_K = 854.8$). The statistics on K from the MC calculations were so unstable as to be worthless. (For $K = 4000$ they always crashed the program through overflows, due to a few very large K estimates.) However, K_d was well-behaved, with statistically insignificant bias and a relative error in full agreement with the exact predictions. The distributions of K and K_d are illustrated in Figure 3. Although K_d is not normal, it is close enough thereto to meet demands well beyond those of the proverbial "government work". The same cannot be said of K, which clearly demonstrates reciprocal behavior. The distribution of a (not shown), for which $\sigma_a/a = 0.0059$, is much closer to normal ($\chi^2 = 41.6$ for $\nu = 29$).

Next consider the behavior of the binding constant parameters in the more suitable operating regime near the middle of Figure 2. When $K = 30$, its exact relative standard error is 0.078 when $\sigma_y = 0.01$. Yet Figures 4 and 5 show that (1) K is clearly nonnormal; (2) nor are its a posteriori-normalized residuals t-distributed, though the agreement is closer in this case; and (3) the deviations from both distributions increase with increasing σ_y (and hence increasing σ_K and σ_a). Similar behavior is observed for a (not shown). The reciprocal K_d (also not shown) exhibits behavior comparable to that for K. On the other hand, S/ν does satisfactorily follow the reduced χ^2 distribution. Despite the substantial deviations from normality, the practical implications for confidence limits remain minor. For example, consider the results for $\sigma_y = 0.025$, where the relative error in K is 0.20. The statistics from the MC calculations give $\sigma_K = 6.03$, which is ~3% greater than the exact value from the a priori **V** (eq 8). (For comparison, the bias in K is +1.8%.) From the histogrammed data, the 90% range is 21.6−41.2, while the "exact" treatment yields $30 \pm 1.645 \times 5.867$, or 20.3−39.7, which differs mainly by an asymmetry shift of ~1.4.

Given the asymmetry in the parameter distributions, it is not surprising that the parameter estimates are also biased. Figure

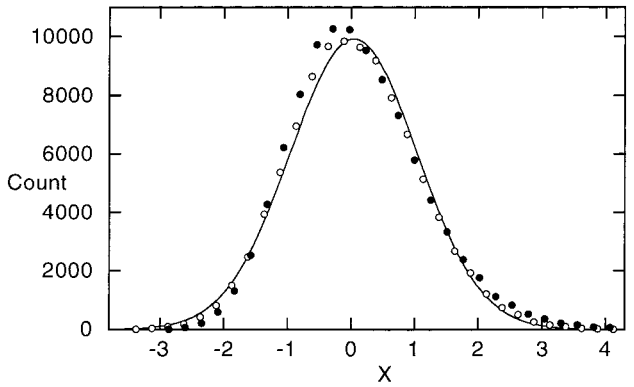

Figure 4. Histogrammed results for K from 10^5 fits to eq 20 of five-point data sets having true values $a = 1$, $K = 30$, and $\sigma_y = 0.01$ (open points) or $\sigma_y = 0.025$ (filled points). The x-structure of the data is the same as for Figure 3 (see text). The histogrammed quantity is $X = (K - 30)/\sigma_K$, with $\sigma_K = 2.347$ and 6.028, respectively. The smooth curve is a Gaussian fitted to the open points. It yielded $\chi^2 = 6.6 \times 10^2$, while a fit to the filled points gave $\chi^2 = 4.1 \times 10^3$.

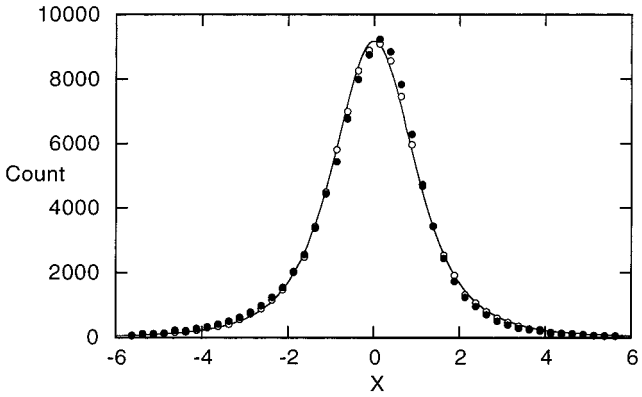

Figure 5. The same results as displayed in Figure 4, but histogrammed as the t-variable, $X = (K - 30)/s_K$, with s_K being the a posteriori estimate obtained along with K from each fit. The fits to the t-distribution (eq 11), shown for the open points, yielded $\chi^2 = 168$ and 906.

6 illustrates the dependence of the bias in both parameters on σ_y and on n. (For the latter it is important to preserve the x-structure of the data, which is accomplished by simply doubling, tripling, etc. the number of points at each x_i.). At the precision of this study, the bias in both parameters vanishes as $n \rightarrow \infty$, meaning the estimators are consistent. If the error structure is proportional rather than constant, we encounter the same problems implementing the weighting as discussed earlier in conjunction with Figure 1. Figure 7 illustrates the biases in both parameters as functions of $1/n$, for weighting using the theoretical y values, the "observed" and the "calculated", with the last of these being a part of the optimization process for each fit. All three K estimators appear to be consistent; however, the two a estimators that are actually available to the experimentalist are both inconsistent.

The Declining Exponential. *Background-Free*. An exponential decay without background can be linearized by logarithmic transformation. Let the fit relationship be

$$y = a \exp(-\Gamma t) = a \exp(-t/\tau) \qquad (22)$$

Neglecting the data bias introduced through the (properly weighted) log conversion, the fit to $\ln y = \ln a - \Gamma t$ should yield normally distributed estimates of $\ln a$ and Γ. Accordingly, the distribution of the lifetime τ should follow eq 17, and for

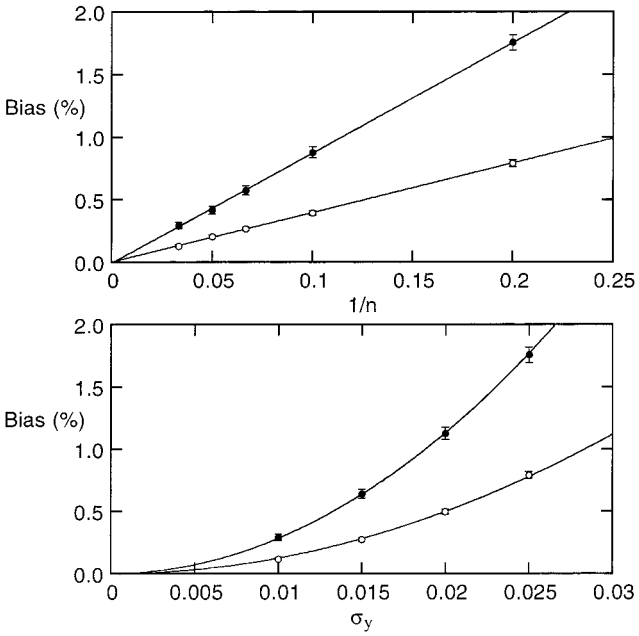

Figure 6. Bias in binding constant parameter estimates, for a constant y-error structure in the data: open points, a; filled points, K. Five-point data sets were used to study the σ_y dependence; for the n dependence, σ_y was fixed at 0.025. Error bars represent 1σ, from 10^5 data sets for each displayed point. The curves illustrate LS fits to $Y = c\sigma_y^2$ (lower) and $Y = c + d/n$ (upper). In the latter the intercept was not statistically significant for either parameter.

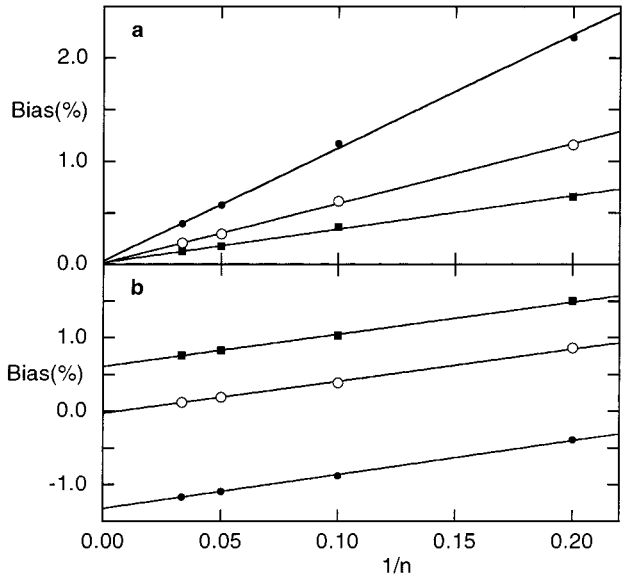

Figure 7. Bias in binding constant parameter estimates K (a) and a (b), for an 8% proportional y-error structure in the data: open points, weights evaluated using theoretical y values; filled circles, "observed" y; filled squares, "calculated" y. Each displayed point represents the results from 10^5 data sets. None of the intercepts in the upper plot is statistically significant at the 2σ level.

relatively uncertain Γ, Γ is the statistically preferred parameter. This conclusion was reached also in a study of maximum likelihood estimators for exponential decay.[5]

Monte Carlo calculations suggest that neither a nor Γ from nonlinear fits to (22) is normal, though a is quite close under some circumstances. (Note that if Γ is fixed rather than adjustable, the fit is linear in a and it becomes rigorously normal.) As an extreme example, MC calculations were done for a model having $a = \Gamma = 1$, with four evenly spaced t values

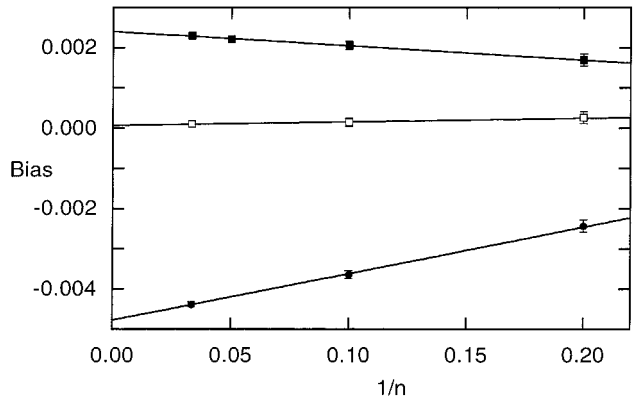

Figure 8. Bias in the pre-exponential factor a (true value 1.5) as a function of $1/n$, from nonlinear LS fits to exponential decay data having 4% proportional y-error and no background: open points, weights evaluated using theoretical y values; filled circles, "observed" y; filled squares, "calculated" y. Each displayed point represents the results from 10^5 data sets. Error bars, where discernible, represent 1σ.

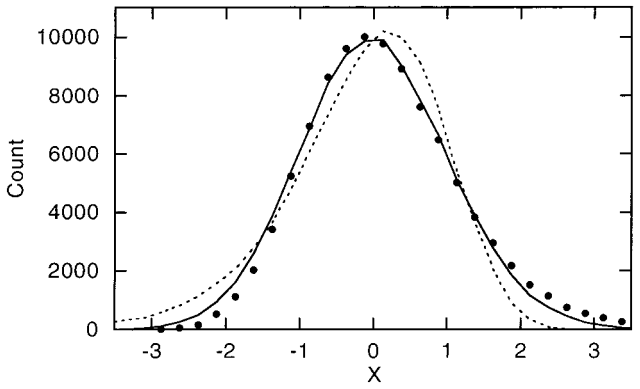

Figure 9. Histogrammed results from 10^5 fits of five-point data sets for exponential decay with a background: $y = 1.5e^{-t} + 1$. The data points were spread evenly over the range $t = 0-2$, with constant error, $\sigma_y = 0.04$. The quantity $X = (\beta_i - \beta_{i,\text{true}})/\sigma_i$, where the exact nonlinear standard errors are $\sigma_a = 0.143$ (points), $\sigma_\Gamma = 0.335$ (solid curve), and $\sigma_b = 0.149$ (dashed curve).

extending from $t = 0$ to 1.2. With constant $\sigma_y = 0.16$ [cf. $y(1.2) = 0.301$], the exact standard errors in a and Γ are 0.150 and 0.345, respectively. The histogrammed distributions resemble those in Figures 3−5, with Γ being nonnormal beyond question ($\chi^2 = 2.3 \times 10^3$) and τ exhibiting reciprocal behavior, but with a marginally normal ($\chi^2 = 42.5$, $\nu = 28$). The a posteriori-assessed residuals do satisfactorily obey the t-distribution ($\nu = 2$) for a ($\chi^2 = 42.3$, $\nu = 44$), but not for Γ ($\chi^2 = 520$). As before, the S/ν values are χ_ν^2-distributed ($\chi^2 = 39.6$, $\nu = 46$). Both parameter estimates are biased, by +0.34(5)% for a, +2.8(1)% for Γ.

Despite the sizable deviations from normality for Γ, the implications for estimating confidence limits are nominal. The standard error in Γ from the statistics of the MC calculations is 6.6% greater than the "exact" value. The 90% range from the MC calculations is 0.48−1.67, while the "exact" treatment yields $1.0 \pm 1.645 \times 0.3448$, or 0.43−1.57. As was found for binding constant fitting, the main difference is an asymmetry shift. Given that both ranges exceed the parameter value, it seems unlikely that any important decision will stand or fall based on these differences.

To investigate the bias more thoroughly, a large number of MC calculations were run for a basic five-point model, with $a = 1.5$, $\Gamma = 1$, and t values spaced evenly between $t = 0$ and 2. Both constant ($\sigma_y = 0.10$) and proportional ($\sigma_y = 4\%$ of y) error structures were examined. For constant error, the results resembled those already discussed, showing small but statistically significant bias in both estimators at finite n (e.g., for $n = 5$, $+1.1(3) \times 10^{-3}$ for a, $+5.2(4) \times 10^{-3}$ for Γ). However, both biases were linear in $1/n$ and extrapolated to statistically insignificant values at $1/n = 0$, meaning the estimators are consistent. For proportional error, on the other hand, the problem already encountered in the linear and binding constant models arises again: For the purpose of calculating the weights, $w_i = \sigma_{yi}^{-2}$, the true y_i can never be known to the analyst, who must choose between the observed and adjusted (calculated) y_i for this calculation. As Figure 8 illustrates, not only is there significant bias in the pre-exponential estimator for both of these choices, but it actually increases in magnitude with increasing n. For theoretical weighting this bias is not significant at the 2σ level for any n. Interestingly, for Γ (results not shown) none of the three weighting choices for proportional error yields a statistically significant bias for any n.

Exponential Decay with Constant Background. When a background is added to a declining exponential, the first consequence is a reduction in the precision of the other two parameters.[6] For example, for the five-point model just discussed, with constant σ_y, addition of a background parameter increases the errors in a and Γ by factors of 2.6 and 3.1, respectively, independent of the actual magnitude of the background. With proportional error, the magnitude of the background matters: For no actual background in the data, σ_a rises only slightly, but σ_Γ increases by a factor of 3.6, with a constant background equaling a in magnitude, both errors increase by another factor of ~4.

Other differences are revealed in the parameter distributions. Figure 9 shows these for the same five-point model just discussed, with addition of a background of magnitude 1.0, for constant error, $\sigma_y = 0.04$. None of the distributions comes close to normal, but Γ is actually more nearly normal than for the background-free case of comparable σ_Γ. The background exhibits a negative skewness, a property not observed previously in this study. Since the pre-exponential parameter shows the opposite skewness, the distributions seem to be reflecting the compensating nature of these parameters.

Parameter bias was investigated as a function of n in the same manner as before. Results are shown in Figure 10, the main features of which can be summarized as follows. (1) Statistically significant bias is present at all n in every case but one—Γ as estimated using weights based on calculated y_i for proportional error. (2) All Γ estimators are consistent. (3) All estimators are consistent for constant σ_y and for proportional error with weights based on the theoretical y_i. (4) The pre-exponential and background estimators are inconsistent for proportional weighting when the weights are assessed using either the observed or the calculated y_i values.

The last of these results agrees with observations for all the other models discussed earlier. For proportional error with weights calculated using the theoretical y_i, the results initially indicated inconsistency here too. With the inclusion of additional MC results, the fits of bias as a function of $1/n$ supported the addition of a quadratic term, and the intercepts then were within 1σ of zero. This treatment did not remove the apparent inconsistency in the other two weightings for proportional error. The nonzero intercepts for weights based on observed y_i are obvious in the plotted displays. Although not so obvious, the same is true for weighting on the calculated y_i, where the

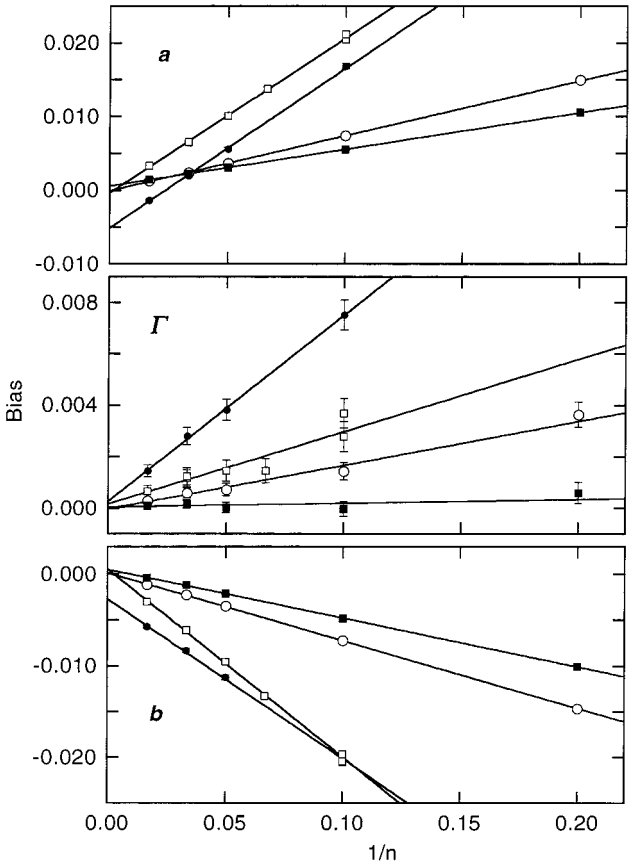

Figure 10. Bias as a function of $1/n$ for nonlinear estimators of exponential decay with a background, $y = ae^{-\Gamma t} + b$. The data structure was as described in Figure 9 and the text. Open circles, $\sigma_y = 0.04$ (constant); others, proportional error, with weights calculated using theoretical (open squares), observed (filled circles), and calculated (filled squares) y_i values. The proportional error was 4% in the first two cases but was reduced to 2% for weighting on the calculated y_i values in order to achieve 100% convergence in the MC calculations.

intercepts differ from zero by 6σ for the pre-exponential parameter and 4.5σ for the background.

An error structure intermediate between the constant and proportional errors is the Poisson structure, $\sigma_{yi}^2 \propto y_i$. This case is of considerable practical importance, as many experimental lifetime methods employ counters. Since the counts are usually binned into time windows, the fit should employ an integral model rather than a sampling one,[6−8] i.e., the calculated value for y_i should be the integral of the exponential plus background over the time window. However, in the large-n limit, this distinction vanishes, and for consistency with the treatment of constant and proportional errors, I have retained the same sampling model. To achieve comparable parameter errors, I have used $\sigma_{yi} = 0.05\sqrt{y_i}$, which, for the same five-point t-structure, yields the exact errors, $\sigma_a = 0.1467$, $\sigma_\Gamma = 0.2473$, $\sigma_b = 0.1525$.

The unavoidable ambiguity in assessing the weights that arose for proportional error is present again for Poisson error. Not surprisingly, the MC results for the bias and consistency of the various estimators resemble the behaviors found for proportional error. (1) For finite n, bias is present for all three ways of assessing the weights, for all three parameters. (2) For small n, the bias is positive in a and Γ, negative in the background b. (3) All three estimators for Γ are consistent. (4) The infinite-n bias (the inconsistency) in b is positive for weights based on the calculated y_i (optimized within each fit), while that for observed y_i is negative and somewhat larger in magnitude. (5) All three estimators give a bias of -4% for b at $n = 5$; the inconsistency for weights evaluated using observed y_i is $-0.25(4)\%$. In one clear departure from the properties for proportional error, all three estimators for the pre-exponential parameter a are statistically equivalent, exhibiting $\sim$2.7% bias for $n = 5$ but no bias (hence consistency) in the infinite-n limit.

Conclusion

Using Monte Carlo calculations of typically 10^5 equivalent data sets at a time, I have investigated the statistical properties of the least-squares estimators for the parameters in three important fit models: the straight-line linear fit, the rectangular hyperbola used in binding constant determinations, and the declining exponential. The results for the linear model confirm expected behavior—normal, unbiased, minimum-variance estimates when the data are normal and properly weighted. Under these circumstances, the t-distribution also holds for the parameter residuals divided by their a posteriori standard error estimates, and the sum of weighted squared residuals (S) follows the χ^2 distribution. If the data are not unbiased and normal, these predictions fail, even for linear LS. For nonlinear LS, nonnormal parameter distributions and bias are the rule rather than the exception, even for unbiased, normal data. However, properly weighted nonlinear fit models appear to yield S values distributed as χ^2, and generally the a posteriori-assessed parameters follow the t-distribution more closely than the a priori-assessed parameters follow the Gaussian distribution.

For small data error σ_y, the parameter bias scales with the variance σ_y^2. It is also very nearly linear in $1/n$, where n is the number of points in the data set. This behavior permits extrapolation to the infinite-n limit, where persistent bias constitutes inconsistency. Using this empirical approach, I have found that many of the estimators are inconsistent when the data error structure is y-proportional or Poisson in nature. The source of inconsistency is an inherent ambiguity in computing the weights in these cases. No inconsistency is found when the weights are assessed using the true y_i. But unfortunately the experimentalist cannot know the true values so must use either the observed y_i or the adjusted values from the fit itself. Both of these options yield inconsistency for many of the estimators studied here. Fortunately, some of the most sought-after parameters, the binding constant K and the decay rate Γ, are consistent for both choices.

Monte Carlo calculations have been widely used in the past to study bias in least squares. However, I am unaware of any previous demonstration of inconsistency based on an MC approach. It should be emphasized that my method is phenomenological, based as it is on the observed near-linear behavior of the bias as a function of $1/n$. However, the demonstration of consistency via this approach in the cases where it is most expected, for constant error structure, or when data having proportional error are weighted using the theoretical y_i values, lends credence to the findings of inconsistency in the other cases. In the literature, there is at least one formal determination of inconsistency in nonlinear LS; Bevington[3] showed that in the case of Poisson error structure there is a systematic difference between the observed and calculated total signal count. From his derivation, it can be seen that this bias will not vanish as $n \rightarrow \infty$. Technically, however, his demonstration was based on a sampling fit model when it should have employed an integral model. Still, based on the current results, his inconsistency is expected to hold for the proper model as well.

It is worth emphasizing that the inconsistency here is truly unavoidable from an experimental standpoint. In this respect it

differs from that which results from transforming the data via inversion or logarithmic conversion,[17] in which the transformation (1) produces "biased" data and (2) forces a change in the weighting of the data, both of which can contribute to inconsistency.[27] The implications at large n are the same: At some point the bias must exceed the parameter standard error. For example, in the case of the simple linear fit discussed in connection with Figure 1, for the slope B with weights assessed using the observed y_i, that point comes at $n \approx 360$ (see Figure 1b). Interestingly (and unique among the examples studied here), for the intercept in this case there appears to be a sort of "isobias" point for $n \approx 50$, where further the bias appears to vanish.

Many Monte Carlo studies have devoted attention to the effects of improper weighting of the data, often in conjunction with transformation to a "linear" (actually, straight-line) relationship. It has long been known that improper weighting, which means any weighting other than $w_i \propto \sigma_i^{-2}$ (eq 6), will fail to yield minimum-variance estimates of the parameters.[2,28,29] It is less recognized that in linear models with normal data, improper weighting biases the error estimates but *not* the parameter estimates. In cases where the parameter estimates are inherently biased, like nonlinear models or linear models with nonnormal data, we can anticipate that improper weighting will exacerbate the parameter bias in accord with the general σ_y^2 dependence of bias. In any event, it is to be emphasized that eqs 8 and 10 are completely unreliable for predicting parameter errors if the data are not weighted in accord with eq 6.

The purpose for which MC calculations have been most widely used in data analysis appears to be the one where they are least needed. That is in experimental design, specifically the study of the dependence of parameter error on the structure of the data, and the estimation of parameter confidence limits. For linear fit models with normal data, the parameter standard errors are known *exactly* from theory. For nonlinear models, the errors do depend on the actual parameter values. However, one can define an "exact" variance−covariance matrix **V** as that obtained for a perfectly fitting function with known error structure. In all the cases studied here, these exact standard errors have agreed with the MC statistical estimates to within 10%, even where the resulting errors were comparable in magnitude to the parameters. Moreover, results such as those displayed in Figure 2 for binding constant analysis can be routinely obtained using packaged microcomputer programs (here, KaleidaGraph). For example, most of the results in refs 8−16 can be obtained this way. Of course the exact errors cannot predict asymmetry in the parameter distributions. However, if the relative parameter error is <10%, the analyst is probably safe in trusting the normal approximation in assigning confidence limits. If >20% (and if that is deemed acceptable), it is necessary to look more closely, with an eye toward identifying especially nonnormal distributions, like those for parameters that are the reciprocals of normal or near-normal quantities. In some such cases, MC calculations may remain necessary for an adequate assessment, as they surely will for estimating bias.

It follows from the foregoing that studies that claim gross disparities between parametric (**V**-based) and MC estimates of parameter errors are suspect. For the most part, MC practitioners have not made such comparisons. However, in a few cases they have. Notable in this regard are the studies of spectrophotometric models of binding constants by Alper and Gelb,[14,15] which found the MC confidence ranges to be 2−3 times narrower than predicted by the parametric **V**s. Unlike all the cases examined in the present work, their models involved error in both x and y. However, a Monte Carlo re-examination of these cases has yielded much broader distributions for the parameters, with the result that the parametric and MC error estimates agree within the same ∼10% guideline as found for all the cases treated in the present paper.[30]

References and Notes

(1) Mood, A. M.; Graybill, F. A. *Introduction to the Theory of Statistics*, 2nd ed.; McGraw-Hill: New York, 1963.

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

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

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

(5) Tellinghuisen, J.; Wilkerson, C. W., Jr. *Anal. Chem.* **1993**, *65*, 1240−1246.

(6) Tellinghuisen, J. *Anal. Chem.* **1993**, *65*, 1277−1280.

(7) Tellinghuisen, J.; Goodwin, P. M.; Ambrose, W. P.; Martin, J. C.; Keller, R. A. *Anal. Chem.* **1994**, *66*, 64−72.

(8) Sharman, K. K.; Periasamy, A.; Ashworth, H.; Demas, J. N.; Snow, N. H. *Anal. Chem.* **1999**, *71*, 947−952.

(9) Ramachandran, B. R.; Allen, J. M.; Halpern, A. M. *Anal. Chem.* **1996**, *68*, 281−286.

(10) Sagnella, G. A. *Clin. Sci.* **1994**, *87*, 371−381.

(11) Ritchie, R. J.; Prvan, T. *J. Theor. Biol.* **1996**, *178*, 239−254.

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

(13) Bowser, M. T.; Chen, D. D. Y. *J. Phys. Chem. A* **1999**, *103*, 197−202.

(14) Alper, J. S.; Gelb, R. I. *J. Phys. Chem.* **1990**, *94*, 4741−4751.

(15) Alper, J. S.; Gelb, R. I. *J. Phys. Chem.* **1991**, *95*, 104−108.

(16) Gonçalves, R. M. C.; Martins, F. E. L.; Leitão, R. A. E. *J. Phys. Chem.* **1994**, *98*, 9537−9540.

(17) Tellinghuisen, J. To be published.

(18) 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*; K. Narahari Rao, Ed.; Academic Press: New York, 1976; pp 1−67.

(19) Tellinghuisen, J. *J. Mol. Spectrosc.* **1994**, *165*, 255−264.

(20) Tellinghuisen, J. *J. Mol. Spectrosc.* **1996**, *179*, 299−309.

(21) Abramowitz, M.; Stegun, I. A. *Handbook of Mathematical Functions*; Dover: New York, 1965.

(22) Birge, R. T. *Phys. Rev.* **1932**, *40*, 207−227.

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

(24) Hayashi, Y.; Matsuda, R.; Poe, R. B. *Analyst* **1996**, *121*, 591−599.

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

(26) The values reported in Table 4 of this work actually appear to be the 95% (not 68%) confidence limits, with inclusion of the *t*-factor for 15 degrees of freedom.

(27) Data transformation *can* convert a weighted fit to an unweighted one. For example, logarithmic transformation of data with proportional error yields constant error in the transformed variable. However, if the original data are normal, the transformed data are nonnormal and can still produce bias and inconsistency in the resulting unweighted LS analysis.

(28) Cvetanovic, R. J.; Singleton, D. L.; Paraskevopoulos, G. *J. Phys. Chem.* **1979**, *83*, 50−60.

(29) Kalantar, A. H. *J. Phys. Chem.* **1986**, *90*, 6301−6303.

(30) Tellinghuisen, J. Unpublished.