Normalization in the Fitting of Data by Iterative Methods

APPLICATION TO TRACER KINETICS AND ENZYME KINETICS

By J. H. OTTAWAY

Department of Biochemistry, University of Edinburgh Medical School, Teviot Place, Edinburgh EH8 9AG, U.K.

(Received 23 March 1973)

1. The normalization of biochemical data to weight them appropriately for parameter estimation is considered, with reference particularly to data from tracer kinetics and enzyme kinetics. If the data are in replicate, it is recommended that the sum of squared deviations for each experimental variable at each time or concentration point is divided by the local variance at that point. 2. If there is only one observation for each variable at each sampling point, normalization may still be required if the observations cover more than one order of magnitude, but there is no absolute criterion for judging the effect of the weighting that is produced. The goodness of fit that is produced by minimizing the weighted sum of squares of deviations must be judged subjectively. It is suggested that the goodness of fit may be regarded as satisfactory if the data points are distributed uniformly on either side of the fitted curve. A chi-square test may be used to decide whether the distribution is abnormal. The proportion of the residual variance associated with points on one or other side of the fitted curve may also be taken into account, because this gives an indication of the sensitivity of the residual variance to movement of the curve away from particular data points. These criteria for judging the effect of weighting are only valid if the model equation may reasonably be expected to apply to all the data points. 3. On this basis, normalizing by dividing the deviation for each data point by the experimental observation or by the equivalent value calculated by the model equation may both be shown to produce a consistent bias for numerically small observations, the former biasing the curve towards the smallest observations, the latter tending to produce a curve that is above the numerically smaller data points. It was found that dividing each deviation by the mean of observed and calculated variable appropriate to it produces a weighting that is fairly free from bias as judged by the criteria mentioned above. This normalization factor was tested on published data from both tracer kinetics and enzyme kinetics.

This paper is a contribution to solving the problems of weighting experimental observations when using them to determine the values of parameters (constants) of biochemical interest. It is written from the point of view of one who uses iterative function-minimization methods for these determinations. Such methods are the best available for accurate estimation of the parameters of the exponential equations relating to tracer kinetics (Atkins, 1971*a,b*; Ottaway, 1971*a*; Davis & Ottaway, 1972). They also have advantages in dealing with the non-linear equations of enzyme kinetics (Kowalik & Morrison, 1968; Atkins, 1971*b*), because the linearizing techniques recommended by Cleland (1967) and others either make specific assumptions about the distribution of variance or are liable to introduce bias in the estimates unless carefully considered corrections are made (Wilkinson, 1961; Dowd & Riggs, 1965). Moreover, they depend on good initial estimates of the constants (Cleland, 1967), which are not very easily obtained for more complex rate-equations.

We know rather little about the distribution of error in experimental observations. It is often dependent on the size of the measured variable(s). Box (1970) has made more general observations, including the point that the variance of the so-called 'independent' variable (time, concentration etc.) is not really small enough to be neglected as it is in the standard treatments. This latter point will not be taken up here. The variance of specific activity measurements cannot be independent of their size, because specific activity is a ratio. According to Francis *et al.* (1959), the standard error of a specific activity A/B is:

$$\frac{A}{B} \left[\frac{a^2}{A^2} + \frac{b^2}{B^2} \right]^{\frac{1}{2}}$$

where a and b are the standard errors of A and B respectively. Thus even if the standard error of the estimation of B is very small, the standard error of the ratio is a/B . In any case it is evident that the variance of the data of a decay curve spanning 3 or 4 orders of magnitude must vary in some way with the size

of the data. The variance is often assumed to be linearly related to the logarithm of the estimate (Worsley & Lax, 1962), or to the expected value of the variable ('chi-square'; see Sauer *et al.*, 1970). With enzyme kinetic data, intuition may play a greater role, cf. Reich *et al.* (1972).

In handling tracer kinetic data the following difficulty sometimes arises. Many authorities recommend (cf. Berman *et al.*, 1962) that such data should be handled in the 'fractional turnover' form:

$$dR_i/dt = -\lambda_{ji} R_i + \dots$$

where R_i is the total radioactivity in the i th compartment, and λ_{ji} is the fraction of the radioactivity in that compartment which leaves it for compartment j in unit time. This formulation does not specify the existence of a steady state. The λ_{ji} values are formally equal to k_{ji}/S_i , where k_{ji} is the actual rate of transfer of chemical and S_i is the total amount of chemical in pool i . In the equation above the λ_{ji} values are required to be constants, but no restriction is placed on the k and the S values other than that of direct proportionality. Ottaway & Mowbray (1973) have shown that it is not safe to assume that in specific instances this lack of rigour may be ignored, and therefore in this formulation of the equations it is always desirable to use parallel sets of equations for R_i and S_i , even if the system is in a steady state; if the system is not in a steady state (Mowbray & Ottaway, 1973) two sets of equations are obligatory.

The parameters λ_{ij} thus appear in two sets of equations, and their optimal values are obtained by minimizing the sums of squares of residuals of all the variables, but it is clear that this will not be satisfactory unless the sums of squares are made in some way comparable. Not only are the sizes of the variables, radioactivity and chemical amount, normally very different, but in practice the size of the chemical pool does not vary very greatly, and it is essential for the success of the method that the radioactivity varies considerably in each pool during the course of the experiment. Dividing each observation by the value of the appropriate variable at zero time solves the first problem, but does not take care of the second. Moreover, some form of weighting must be included to allow for the fact that the variance of the radioactivity measurement (often derived by multiplying the specific activity ratio by the amount of chemical) is usually much greater than that of the amount of chemical, which is usually dependent on a single measurement. Consideration of these problems suggests that the normalized sum of squares for each variable should be a pure number (i.e., with a squared term in the denominator) rather than a 'chi-square' type of expression. The solution to the problem depends on whether or not there are replicate observations at each sampling point.

Multiple observations at each sampling point

Trial and error have shown that a satisfactory solution to the normalization and weighting problem can be obtained by dividing the sum of squares of deviations for each variable at each sampling point by the variance of the experimental observations at that point. Then the total sums of squares of residuals is:

$$\sum_{j=1}^t \frac{n \sum_{i=1}^n (x_{ij} - \hat{x}_j)^2}{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}$$

summed over all variables.

This method has the following advantages. The minimum value of the normalized sum at each point is obtained when the calculated value of the variable, $\hat{x}_j$, is equal to $\bar{x}_j$, the mean of the experimental observations. This minimum is numerically equal to n , the number of observations at that point; thus weighting according to the size of that set of samples in the total residual is automatically allowed for. If the reproducibility of the estimation of any variable at any sampling point is bad, for whatever reason, the large variance at that point automatically ensures that the local normalized sum of squares will be small, i.e. that sample will not play a large role in determining the final fit. Finally, a point important in iterative minimization is that the response surface comes smoothly and sharply down towards the minimum as the calculated values of the variables approach the local mean but does not rise to astronomical levels (where searching is very difficult) if the initial guesses of the parameters are bad.

Experience with this normalization factor (Ottaway, 1971b; Davis & Ottaway, 1972; Mowbray & Ottaway, 1973) has been so successful that no other is recommended here. It is one of the techniques recommended by Jacquez *et al.* (1968). Although it was developed for multi-variable systems, it can equally well be used with single-variable problems, e.g. enzyme velocity, so long as there are at least two replicates at each point. As the number of replicates increases, the estimate of the local variance obviously improves.

Single observations at each sampling point

The problem is more difficult in this case, because the experiment itself does not provide any information about the local variance. This must be provided by previous experiment, general principle, or intuition; Reich *et al.* (1972) make an explicit statement about the problem. Probably the best advice one can give is that parameters should never be estimated if replicate observations are not available (Ottaway, 1971b), but it is not always possible to adhere to this principle, for example if a tracer investigation is

Fig. 1. The fit of a curve representing the sum of two exponentials to experimental data
(a) Fit obtained by the procedure of Worsley & Lax (1962); (b) weighting by $1/(x+\hat{x})^2$.

carried out on a single patient, or if the choice lies between making one estimation of the velocity of an enzyme-catalysed reaction over a wide grid of concentrations of substrates (Ottaway & Apps, 1972), or duplicate estimations over a much smaller grid.

The accepted criterion for fit of the observations to a model is that the weighted sums of squares of deviations should be a minimum. Since there is no objective criterion governing the choice of weighting factors, in the absence of external information the choice must be an aesthetic one, namely a factor that appears to the experimenter to give the best fit of calculated values to experimental observations over the whole range, or a factor whose logic is pleasing. One or two provisos may be added to this: for instance if the parameters are being optimized by a

digital-computer search algorithm, it is undesirable, other things being equal, that the calculated value of the dependent variable should stand alone in the denominator of the weighted sums of squares expression. If any calculated value at a single sample point becomes very small, as it may well do during a perfectly efficient search, the computer will terminate the operation because dividing by very small or zero dividends leads to quotients too large to be accommodated in the registers of the computer. Further, one or two possible treatments of the sums of squares of deviations lead to consistent bias in the fitting (see below).

Specific examples of non-replicate data fitting are discussed below to illustrate the points involved. For the tracer data a set of results published by

Fig. 2. The fit of a curve representing the sum of two exponentials to experimental data
(a) Weighting each observation by $1/x$; (b) weighting by $1/\hat{x}$.

Lippincott *et al.* (1965), which covers almost 4 orders of magnitude, has been selected. Lippincott *et al.* (1965) fitted the equation:

$$x = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} \quad (1)$$

to these points (see Fig. 1, curve A). If the unweighted sums of squares of residuals are minimized, the second exponential term becomes negligibly small. Lippincott *et al.* (1965) used an analytical method of Worsley & Lax (1962), which assumes that the error $x - \hat{x}$ is proportional to $\ln x$. This is, in effect, the weighting that would be used in traditional 'curve-peeling', in which straight lines are fitted to the logarithms of the data by unweighted regression analysis. Fig. 1 shows that this method fits the experimental points rather well, using for example the criterion of the distribution of the points on

either side of the fitted line: nine are above and nine below. When the data are fitted by function minimization, on the other hand, two fairly obvious devices for normalizing the squares of the deviations give curves that fit the data distinctly less well. If each deviation is divided by the corresponding experimental values before squaring, fourteen of the points are above the fitted line and only six below. If each deviation is divided by the calculated value before squaring, six of the data points are above the fitted curve and twelve below (Fig. 2). Table 1 shows the actual values of the parameters produced by these and other procedures. For example, the 'chi-square' normalization, $(x - \hat{x})^2 / \hat{x}$, produced an even worse fit than those discussed.

In the course of fitting many different sets of data, it was found that use of the normalization $(x - \hat{x})/x$

Table 1. *Parameter values calculated by various procedures*

These parameters are those of eqn. (1), calculated from the data displayed in Figs. 1 and 2 by the procedures described in the text.

Method	A_1	A_2	λ_1	λ_2
Worsley & Lax (1962)	4.456	0.0110	1.398	0.1252
Optimization: normalization by $1/(x+\hat{x})^2$	6.988	0.0123	1.543	0.1332
Normalization by $1/\hat{x}^2$	7.923	0.00524	1.287	0.0597
Normalization by $1/x^2$	742.2	0.0447	6.731	0.2905
Normalization by $1/\hat{x}$ ('chi-square')	105.25	0.3043	4.158	0.4195

consistently gave curves that were below most of the data points, particularly at the numerically smallest end of the range of experimental variables, whereas normalization by $(x-\hat{x})/\hat{x}$ led to curves that were above the data at the low end of the range. These observations can be expressed more formally. For example, the numerical size of the expression $(x-\hat{x})/x$ depends inversely on the size of x . If x is small, $x-\hat{x}$ must be very small if the quotient is to have the same average size as the others in the data set. Thus the curve fits rather closely to the smallest data point, even if this lies well away from the line through the rest of the points. Hebbard (1969) has drawn attention to the bias introduced by putting x in the denominator. It is not so easy to see why bias may be introduced if $\hat{x}$ is put in the denominator. The reason is that unlike the local variance, discussed in the preceding section, or the simple observed value, $\hat{x}$ is a variable. The presumable intention in using $1/\hat{x}$, or some function of it, as a weighting factor, is to use an estimate of the expected value of the variable at that point, but in fact $(x-\hat{x})/\hat{x}$ and similar expressions, treated as functions of $\hat{x}$, have a rather complicated shape. Taking the usual experimental situation when $\hat{x} > 0$, the expressions all tend to infinity as $\hat{x} \rightarrow 0$, fall to zero when $\hat{x} = x$, and rise slowly to an asymptotic value of unity as $\hat{x} \rightarrow \infty$. They are thus very asymmetric functions, and when $\hat{x}$ itself is small, the slope of the left-hand limb may be very steep. If $\hat{x}$ is forced away from x at any point by the pressure of deviations at other points on the curve, it will frequently happen that the absolute value of $(x-\hat{x})/\hat{x}$ is smaller when $\hat{x} > x$ than when $\hat{x} < x$, for a given ratio of $x/\hat{x}$ or $\hat{x}/x$ (say $1/2$ or $2/1$). The fitted curve will then tend to be above the data points; the effect is probably more likely to be seen at the tail of a decay curve.

One function containing $\hat{x}$ in the denominator that is free from the asymmetry described above is that used by Worsley & Lax (1962), namely $(\ln x/\hat{x})^2$. Because of the properties of logarithms, this function rises symmetrically from a minimum at $\hat{x} = x$ (when $x \neq 0$). It may be shown that for small errors, $(\ln x/\hat{x})^2$ is almost identical with $[(x-\hat{x})/\hat{x}]^2$, and it might therefore be expected that the two methods of weighting

the data would produce very similar estimates of the parameter values. Table 1 shows that this is not the case with the data set selected for testing, and Fig. 2 (curve *B*) shows that the curve fitted by using $1/\hat{x}^2$ as a weighting factor is above the data points, particularly at the later sampling times. Moreover, Table 2 indicates a very uneven distribution of the residual weighted variance for this weighting factor, 90% of the residual being associated with data points below the fitted curve, which is to be expected if the arguments developed above are correct. Study of results derived from other sets of data shows that this behaviour is quite frequent, and normalization by $1/\hat{x}$, or $1/\hat{x}^2$, is not a safe procedure.

The expression $(\ln x/\hat{x})^2$ is not directly related to the minimization of the squares of the deviations of the untransformed data. It has no rationale if the function underlying the data has no exponential content, and it fails completely if either x or $\hat{x}$ is zero. A more general expression, which is symmetrically distributed around a minimum value found when $\hat{x} = x$, is obtained if the deviation $(x-\hat{x})$ is divided by the sum of observed and calculated value, i.e. $(x+\hat{x})$, given that $\hat{x} > 0$ and $x \geq 0$. The more convenient form $[(x-\hat{x})/(x+\hat{x})]^2$ has been used in this laboratory with considerable success for treating various kinds of data whose variance is suspected not to be constant. Its effectiveness in treating the data of Lippincott *et al.* (1965) is shown in Table 1 and Fig. 1 (curve *A*). Table 2 shows that the distribution both of the residual variance, and of the data points around the fitted line, is uniform. To some extent the correspondence between the parameter values produced by this weighting device, used in a direct optimization algorithm, and the method of Worsley & Lax (1962) is to be expected, since for relatively small errors the values of the weighted residuals are very similar, but for the deviations of 200% or more associated with some of the data points, the similarities are quite small.

In judging the fit produced by these methods of treating the raw single data, the implicit criterion has been that the fitted curve should go through the middle of the points. That is, if points on the curve itself are neglected, a roughly equal number of

Table 2. *Distribution of squares of deviations about a fitted curve, obtained with various weighting factors*

Data are those of Figs. 1 and 2; the weighting factors are those in Table 1. See also the text.

Weighting factor ...	$1/x^2$	$1/\bar{x}^2$	$1/(x+\bar{x})^2$
Number of points above curve	14	6	9
Percentage of total weighted sums of squares of deviations associated with these points	88%	10%	35%
Mean percentage deviation per point	6.3%	1.6%	4%
Number of points below curve	4	12	9
Percentage of total weighted sums of squares of deviations associated with these points	12%	90%	65%
Mean percentage deviation per point	3%	7.5%	7%

points should fall on either side of the curve. By this test, which is an intuitive one, weighting by Worsley & Lax's (1962) procedure and by $1/(x+\bar{x})^2$ perform well, whereas the other procedures do not. If the hypothesis of equal distribution around the fitted curve is accepted, the chi-square test can be used to make the decision a quantitative one. For 18 points, chi-square reaches a 5% level of significance if the distribution of points is 13:5, or worse, on one or other side of the fitted curve. By this test, weighting by $1/x^2$ performs significantly worse than any of the other procedures.

One may also consider whether the weighted variance itself is distributed equally on either side of the curve. Clearly it ought to be so distributed, if the variance is normally distributed, over the whole length of the curve, and if the curve itself passes through the expected means of a large series of observations from the experimental population. The justification for considering the distribution of the variance around the curve when one knows nothing about its normality is the sensitivity of the partial sum of squares of deviations on either side of the curve to slight changes in the position of the curve.

Table 2 shows that the distribution of the residual variance was very far from uniform in the test example, when either $1/x^2$ or $1/\bar{x}^2$ was used as a weighting factor. It should be stressed that these tests of the efficiency of weighting are only valid if the model equation genuinely fits the data, as eqn. (1) appears to do. If the model only fits a fraction of the data points, inequalities in the distribution of the points around the fitted curve are to be expected.

The recommendation made here for the use of $1/(x+\bar{x})^2$ does not depend solely on its behaviour with the data shown in Figs. 1 and 2. It has been tried out on many data sets, particularly ones with replicate values (see, for example, Davis & Ottaway, 1972;

Mowbray & Ottaway, 1973) in which comparisons with weighting by the local variance can be made directly.

Weighting by $1/(x+\bar{x})^2$ tends towards weighting by $1/x^2$ as the variability of the data decreases, so that in the limit its use implies that the variance is proportional to the square of the observed variable, i.e. that the coefficient of variation is constant. This is a proposition put forward as a reasonable alternative to a constant variance by many authors who have considered data of a biochemical nature (e.g. Wilkinson, 1961; Dowd & Riggs, 1965; Jacquez *et al.*, 1968). There is thus reasonable ground for trying $1/(x+\bar{x})^2$ for any data without replicates for which weighting may be thought to be necessary. For example, its use may be considered for the direct fitting of rate equations to enzyme kinetic data. (This paper is not concerned with obtaining estimates of enzyme parameters from inverted or other linear forms of the hyperbolic rate equation.)

No thorough investigation of the actual distribution of error in enzyme work seems to have been made, although various authors (e.g. Wilkinson, 1961; Cleland, 1967) give detailed procedures for estimating the enzyme parameters under the assumption that the weighting is known. Dowd & Riggs (1965) surveyed 28 papers in which Lineweaver-Burk plots were used in six successive issues of the *Journal of Biological Chemistry*, and concluded that 'it was impossible to make more than an educated guess about the magnitude of the error to which v was subject ... nor to tell whether the absolute magnitude of the error in a given experiment was reasonably constant for various values of v , or whether the error tended to vary with v so that the percentage error remained constant'. It seems unlikely that the situation has changed markedly in the ensuing 8 years, but some protection against the possibility that the

variance is not constant over the range of v observed seems necessary. Jacquez *et al.* (1968) stress that a constant variance is an unlikely experimental situation.

Ottaway & Apps (1972) used the method of normalization proposed here in fitting a grid of velocities of NAD kinase (EC 2.7.1.23) obtained when $[Mg^{2+}]$ and $[ATP^{4-}]$ were simultaneously varied, to a rate equation by direct optimization. The measured velocities varied over a 50-fold range and the substrate and activator over a 500-fold range. It was shown that the percentage error (which was not the same as the weighted sum of squares of deviations being minimized) was constant when the observed velocities were fitted to a rate equation which was expected from previous work (Apps & Marsh, 1972) to describe the behaviour of the enzyme, over the whole range of substrate/activator concentrations used. This seems a fairly stringent test. It was also shown (J. H. Ottaway, unpublished work) that if the unweighted squares of the deviations were used for fitting the rate-equation to the data, not only were there more large relative errors, but they were significantly associated with small values of v . Moreover, in the weighted fitting, 50% of the errors ($v - \hat{v}$) were positive and 50% were negative, whereas, in the unweighted fitting, two-thirds of the errors were negative and only one-third positive. The probability of this occurring by chance is less than 0.05%, by the chi-square test (117 observations). It was concluded that weighting the velocities, in direct fitting of the rate equation without transformation, significantly improved the fit in this instance.

Discussion

In assessing the usefulness of the procedures suggested in this paper, it is important to get their probable applications in the correct order of importance. Most attention may well be paid to the observations on enzyme kinetics, but for simple purposes the effects are probably marginal. Dowd & Riggs (1965) suggested that a disregard of proper weighting seemed to be in accord with the current practice of enzymologists, since in only one of the papers that they surveyed was any attention given to the problem. There is certainly no lack of advice about the treatment of standard kinetic data, but if drawing the best straight line by eye on a Lineweaver-Burk plot produces satisfying results, it probably indicates that the information normally sought does not have to be very accurate or very reliable. Such a conclusion does not, however, cover the field completely. There is now relatively less interest in the determination of K_m and V_{max} , in what are effectively single-substrate situations, and more emphasis on assessing the validity of quite complex rate-equations. This tendency has been accentuated by the ability of modern

optimization algorithms, programmed for the digital computer, to fit rate equations of any order of complexity directly to the data, with no need for any form of linearization (Kowalik & Morrison, 1968; Morrison, 1969; Atkins, 1971*b*; Ottaway, 1971*a*; Davis & Ottaway, 1972; Reich *et al.*, 1972). If, in particular, attempts are made to discriminate between models on the basis of minimizing the residual error, the nature of that error, in the sense of the weighting to which it has been subjected, is very important, particularly if a wide range of substrate or other ligand concentrations was used to ensure that the model fits the data over a wide range of conditions (see, for example, Ottaway & Apps, 1972). Replication of the data would seem to provide the only satisfactory solution to this problem. The methods discussed in this paper, which one might adopt or should avoid in treating replicate data, may be helpful.

The chief value of this study is thought to lie in the recommendations for treating tracer kinetic data. No general consideration of the weights that should be given to such data appears to be available, although statistical methods of treating data that have already been weighted certainly exist (e.g. Berman & Weiss, 1967). What should be done when tracer and chemical data must be combined in the estimation of common parameters has not been discussed. The deviations between the observed and calculated value of the variable at each sampling point (time) ought to be given a weight inversely proportional to the variance at that point. If replicate observations are available, the variance of each experimental variable at each sampling point can be computed directly and it seems logical to use this for normalizing the squared deviations. It has the advantage of allowing for the fact that some observations may be lost during the experimental procedure, and it also reduces the weighting at sampling points where reproducibility is low (as for instance soon after the addition of tracer to the system). It also reduces the weight attached to variables in which the error of measurement is unavoidably high. It is possible that more unbiased estimates would be obtained if n (in the numerator of the expression given in the section on multiple observations at each sampling point) were replaced by $(n-1)$.

One certainly should not, when replicates are available, divide each individual deviation by a normalizing factor. Not only will this lead to bias, if the factor is x_i or $\hat{x}_i$, as shown here, but the information contained in the scatter of the replicates will be lost.

There is an important group of tracer measurements in which it is almost impossible for replicate estimates to be obtained, namely those in which labelled metabolites, tracers or drugs are administered

to human subjects, whether these are thought to be normal or diseased. Only in exceptional cases is it possible to assemble a remotely homogeneous group of normal healthy volunteers (e.g. Neer *et al.*, 1967). On the other hand, reliable analysis of data obtained from single subjects may be very important, as for example when using ^{133}Xe washout in the detection and localization of brain tumours (Handa *et al.*, 1970). It has therefore seemed desirable to me to put forward a helpful suggestion, even though I feel that accurate parameter estimation really requires replicate observations. The normalization factor $1/(x+\hat{x})^2$, which it has been suggested is simple to apply, is equivalent to the fairly logical assumption that the standard deviation of the observation is proportional to its size as the estimate of $\hat{x}$ comes closer to the observed value of x , and avoids the deviations from closest fitting to the observations, principally in the tail of the curve, which are liable to occur when normalization either by x or by $\hat{x}$ alone is used (cf. Fig. 2). The factor is entirely empirical in origin, and its validity ought to be tested, for each kind of data used, by one of the tests suggested in this paper, or by any other that is suitable. However, it has been shown in this laboratory, by partial differentiation with respect to the parameters near the point of best fit, that weighting by $1/(x+\hat{x})^2$ does give roughly equal value to all parameters in exponential and Michaelis-Menten expressions, both for high and low values of the observed variables (i.e. both at small t and large t in exponential expressions, and at low $[S]$ and high $[S]$ in enzyme rate equations). This gives one some confidence that, so long as the coefficient of variation of the error is approximately constant, this weighting will provide a good response surface, with the same degree of slope with respect to all the parameters, in multi-parameter search procedures. This is an important aspect, since much tracer work is directed towards estimating the values of rate constants and pool sizes in models, and it is desirable to use methods that allow these estimates to be obtained directly, rather than through the often doubtful resolution of decay curves into component terms.

This work was assisted by a grant from the Science Research Council. I am grateful to Dr. J. M. Aitchison, Atlas Computer Laboratory Extra-Mural Research Fellow, for many helpful discussions.

References

- Apps, D. K. & Marsh, A. (1972) *European J. Biochem.* **28**, 12–19
- Atkins, G. L. (1971a) *Biochim. Biophys. Acta* **252**, 405–420
- Atkins, G. L. (1971b) *Biochim. Biophys. Acta* **252**, 421–426
- Berman, M., Weiss, M. F. & Shahn, E. (1962) *Biophys. J.* **2**, 289–316
- Berman, M. & Weiss, M. F. (1967) *SAAM Manual*, U.S. Public Health Service Manual no. 1703, National Institutes of Health, Bethesda, Md.
- Box, M. J. (1970) *Technometrics* **12**, 219–229
- Cleland, W. W. (1967) *Advan. Enzymol.* **29**, 1–32
- Davis, R. H. & Ottaway, J. H. (1972) *Math. Biosci.* **13**, 265–282
- Dowd, J. E. & Riggs, D. S. (1965) *J. Biol. Chem.* **240**, 863–869
- Francis, G. E., Mulligan, W. & Wormall, A. (1959) *Isotopic Tracers*, 2nd edn., p. 218, Athlone Press, London
- Handa, J., Handa, H., Kanji, T., Hamamoto, K. & Kousaka, T. (1970) *Amer. J. Roentgenol.* **109**, 701–706
- Hebbard, D. F. (1969) in *Least Squares Methods in Data Analysis* (Anderssen, R. S. & Osborne, M. R., eds.), pp. 1–9, Australian National University Computer Centre Publication CC 2/69, Canberra
- Jacquez, J. A., Mather, F. J. & Crawford, C. A. (1968) *Biometrics* **24**, 607–626
- Kowalik, J. & Morrison, J. F. (1968) *Math. Biosci.* **2**, 57–66
- Lippincott, S. W., Korman, S., Lax, L. C. & Corcoran, C. (1965) *J. Nuclear Med.* **6**, 632–644
- Morrison, J. F. (1969) in *Least Squares Methods in Data Analysis* (Anderssen, R. S. & Osborne, M. R., eds.), pp. 63–69, Australian National University Computer Centre Publication CC 2/69, Canberra
- Mowbray, J. & Ottaway, J. H. (1973) *Eur. J. Biochem.* in the press
- Neer, R., Berman, M., Fisher, L. & Rosenberg, L. E. (1967) *J. Clin. Invest.* **46**, 1364–1379
- Ottaway, J. H. (1971a) in *Principles & Practice of Medical Computing* (Whitby, L. G. & Lutz, W., eds.), pp. 331–349, Churchill-Livingstone, Edinburgh
- Ottaway, J. H. (1971b) *Biochem. J.* **125**, 45P
- Ottaway, J. H. & Apps, D. K. (1972) *Biochem. J.* **130**, 861–870
- Ottaway, J. H. & Mowbray, J. (1973) *Eur. J. Biochem.* in the press
- Reich, J. G., Wangermann, G., Falck, M. & Rohde, K. (1972) *Eur. J. Biochem.* **26**, 368–379
- Sauer, F., Ertle, J. D. & Binns, M. R. (1970) *Eur. J. Biochem.* **17**, 350–363
- Wilkinson, G. N. (1961) *Biochem. J.* **80**, 324–332
- Worsley, B. H. & Lax, L. C. (1962) *Biochem. Biophys. Acta* **59**, 1–24