Biochem J. (1981) **193**, 1005–1008
Printed in Great Britain

Fitting of enzyme kinetic data without prior knowledge of weights

Athel CORNISH-BOWDEN* and Laszlo ENDRENYI†
* *Department of Biochemistry, University of Birmingham, P.O. Box 363, Birmingham B15 2TT, U.K., and*
† *Department of Pharmacology, University of Toronto, Toronto, Ontario M5S 1A8, Canada*

(*Received 2 October 1980/Accepted 20 November 1980*)

A method is described for fitting equations to enzyme kinetic data that requires minimal assumptions about the error structure of the data. The dependence of the variances on the velocities is not assumed, but is deduced from internal evidence in the data. The effect of very bad observations ('outliers') is mitigated by decreasing the weight of observations that give large deviations from the fitted equation. The method works well in a wide range of circumstances when applied to the Michaelis–Menten equation, but it is not limited to this equation. It can be applied to most of the equations in common use for the analysis of steady-state enzyme kinetics. It has been implemented as a computer program that can fit a wide variety of equations with two, three or four parameters and two or three variables.

Under ideal conditions, the method of least squares is the preferred method for fitting theoretical equations to experimental data. It gives poor results, however, if the observations are incorrectly weighted or if the data contain 'outliers', i.e. very poor observations at higher frequency than allowed for by the normal distribution (Cornish-Bowden *et al.*, 1978; Atkins & Nimmo, 1980). In the case of the Michaelis–Menten equation and other two-parameter equations, one can obtain more reliable estimates in such circumstances by using a 'robust' method of fitting, such as that based on the direct linear plot (Cornish-Bowden & Eisenthal, 1978). This does not provide a general solution to the problem, however, as one is often concerned with non-linear equations with three or more parameters, such as the equations for inhibited reactions and two-substrate reactions. It is not, unfortunately, a practical proposition to generalize the direct linear plot to these equations, because the computational labour would increase extremely steeply with the number of parameters.

In an effort to develop robust methods that could be applied to many equations of biochemical importance, we have studied the method of 'biweight regression' (see Mosteller & Tukey, 1977) as applied to the Michaelis–Menten equation. Provided that independent information about the proper weights is available, this approach performs somewhat better than the direct linear plot: it gives excellent protection against outliers, and also compares well with ordinary least squares in the absence of outliers; unfortunately, however, it not only performs poorly when used with an incorrect weighting assumption, it simultaneously loses much of its robustness against outliers (Cornish-Bowden, 1980). In principle, this difficulty can be overcome by performing independent experiments to determine the correct weighting, as described, for example, by Storer *et al.* (1975) and by Askelöf *et al.* (1976). The necessary experiments are very tedious, however, and are very rarely performed. It is instead common practice to base weighting schemes on hypotheses about how the error variances vary with the true rates. Unfortunately, the hypotheses advocated, even in recent reviews (e.g. Cleland, 1979), take no account of published investigations of actual error behaviour in enzyme kinetics (Storer *et al.*, 1975); Siano *et al.*, 1975; Askelöf *et al.*, 1976; Nimmo & Mabood, 1979), and are consequently likely to be erroneous in a high proportion of circumstances. Any general hypothesis is anyway likely to be incorrect sometimes because of variations in experimental conditions, techniques and skills. We have therefore investigated whether it is possible to obtain satisfactory weights from internal evidence contained within small data sets.

Theory and methods

Error variance

To a first approximation, one can take the variance $\sigma^2(v_i)$ of a rate v_i, measured in a consistent way, either as a constant:

$$\sigma^2(v_i) = \sigma_0^2 \qquad (1)$$

0306-3275/81/0301005-04$01.50/1

or as proportional to v_i^2:

$$\sigma^2(v_i) = \sigma_2^2 v_i^2 \quad (2)$$

or as some combination of the two. For convenience, we shall refer to the first case as 'simple errors' and the second as 'relative errors'. We shall also consider 'complex errors', defined as resulting from the presence of additive simple and relative components:

$$\sigma^2(v_i) = \sigma_0^2 + \sigma_2^2 v_i^2 \quad (3)$$

If the value of σ_0/σ_2 is known, one can use it to define proper weights for v_i, because the weight w_i assigned to v_i should be inversely proportional to $\sigma^2(v_i)$, i.e.:

$$w_i = \frac{(\sigma_0/\sigma_2)^2 + \bar{v}^2}{(\sigma_0/\sigma_2)^2 + \hat{v}_i^2} \quad (4)$$

where $\bar{v}$ is a constant (independent of v_i), such as the mean of all the v_i values, and $\hat{v}_i$ is the best available estimate of the true rate; initially we put $\hat{v}_i$ equal to v_i, the observed rate, but after the first iteration $\hat{v}_i$ is calculated from the best available parameter values. The numerator of the expression for w_i is not strictly necessary, as it could be replaced by any constant other than zero without altering the final fitted parameters, but a numerator of the type shown is a convenient 'normalizing' factor that ensures that the calculated weights do not differ enormously from unity. Note that it is not necessary to know or be able to calculate the individual values of σ_0 and σ_2, but only their ratio, which we shall call the 'σ-ratio'. This has the dimensions of a rate, and can be regarded approximately as the rate where the S.D. of v_i changes from being independent of $\hat{v}_i$ to being proportional to $\hat{v}_i$ [see Fig. 2 in Cornish-Bowden (1980)].

Estimation of the σ-ratio

If observations are very abundant, and especially if there are many replicates, several ways of estimating σ_0/σ_2 are possible, but we shall be concerned with the common experimental circumstance in which there are no more than ten observations. Is it then possible to obtain a valid estimate of σ_0/σ_2 without prior knowledge? Provided that some initial parameter estimates exist, so that preliminary $\hat{v}_i$ values can be calculated, we can define squared residuals, e_i^2, as $(v_i - \hat{v}_i)^2$, and use them as estimates proportional to $\sigma^2(v_i)$. These are, however, grossly erratic estimates of the true variances, and our initial efforts to treat this as a linear-regression problem with e_i^2 as dependent variable and $\hat{v}_i^2$ as independent variable led to hopelessly inadequate results. We have had much more promising results with a method that seeks to find a value of σ_0/σ_2 that defines weights, w_i, such that $w_i e_i^2$ and $\hat{v}_i^2$ are uncorrelated, as measured by Kendall's (1970) τ, a rank correlation coefficient. In this method [see Cornish-Bowden *et al.* (1978) for a related interpretation of the median estimates of the Michaelis–Menten parameters in terms of Kendall's (1970) τ] one uses each pair of observations $\hat{v}_i$, $\hat{v}_j$ (apart from any pairs for which $\hat{v}_i = \hat{v}_j$) to estimate the σ-ratio as follows:

$$(\sigma_0/\sigma_2)_{ij}^2 = (\hat{v}_i^2 e_j^2 - \hat{v}_j^2 e_i^2)/(e_i^2 - e_j^2) \quad (5)$$

and defines σ_0/σ_2 as the square root of the median of all the $(\sigma_0/\sigma_2)_{ij}^2$ values.

An alternative form of eqn. (5):

$$(\sigma_0/\sigma_2)_{ij}^2 = [(\hat{v}_i^2/e_i^2) - (\hat{v}_j^2/e_j^2)]/[(1/e_j^2) - (1/e_i^2)] \quad (6)$$

illustrates that the σ-ratio can be thought of as the (median) slope of the stright line obtained by plotting $\hat{v}_i^2/e_i^2$ against $-1/e_i^2$. With simple errors, $1/e^2$ is approximately constant, the denominator of eqn. (6) fluctuates about zero, and the σ-ratio is $+\infty$. Conversely, in the presence of relative errors, the numerator fluctuates about zero and with it the σ-ratio.

To work satisfactorily this approach to estimating σ_0/σ_2 requires modification to cope with the fact that many $(\sigma_0/\sigma_2)_{ij}^2$ values turn out to be negative when calculated from eqn. (5). This can happen either because $(e_i^2 - e_j^2)$ has the opposite sign from $(\hat{v}_i^2 - \hat{v}_j^2)$, in which case $(\sigma_0/\sigma_2)_{ij}^2$ is interpreted as $+\infty$, or because $[(e_i/\hat{v}_i)^2 - (e_j/\hat{v}_j)^2]$ has the same sign as $(\hat{v}_i^2 - \hat{v}_j^2)$, in which case it is interpreted as 0. The rationale for these interpretations is the supposition that $\sigma^2(v_i)$ is likely neither to decrease as $\hat{v}_i$ increases nor to increase more steeply than in proportion to $\hat{v}_i^2$.

Biweight regression

The essence of biweight regression (Mosteller & Tukey, 1977) is that after the first iteration observations with large values of e_i^2 are assigned decreased weight. In conjunction with the weighting scheme discussed above, we define scaled deviations u_i as follows:

$$u_i = w_i^{\frac{1}{2}} e_i / cS \quad (7)$$

where w_i and $e_i = v_i - \hat{v}_i$ are defined as above, c is a constant and S is a robust measure of the scale of the deviations. Various ways of defining c and S have been suggested; here we shall follow Mosteller & Tukey (1977) in defining $c = 6$ and S as the median value of $|w_i^{\frac{1}{2}} e_i|$. We now define new weights, W_i, as follows:

$$W_i = \begin{cases} w_i(1-u_i^2)^2 & \text{if } |u_i| \leqslant 1 \\ 0 & \text{if } |u_i| > 1 \end{cases} \quad (8)$$

These modified weights are then used instead of w_i in the next iteration. Note that $W_i \simeq w_i$ if $|u_i|$ is moderate or small, but $W_i \ll w_i$ (and ultimately $W_i = 0$) if $|u_i|$ is large.

Implementation

In the first step of fitting a set of data by the methods outlined above, we obtain a preliminary least-squares fit by a suitable regression method using weights (w_i) for the v_i calculated from eqn. (4). Unless there is prior information about the likely value of σ_0/σ_2, we begin by putting $\sigma_0/\sigma_2 = \bar{v}$, the mean value of v_i. The preliminary kinetic parameters provided by this regression generate calculated rates, $\hat{v}_i$, which provide a median estimate of σ_0/σ_2 and thence new weights, w_i. These are then inserted into eqns. (7) and (8) to give modified weights W_i, which are used as weights for v_i in a new regression. The improved parameter values generate improved estimates of σ_0/σ, w_i and W_i, and the iterative process is continued until the estimates are consistent from one iteration to the next. This usually requires no more than ten iterations.

This general approach can be applied not only to the Michaelis–Menten equation, but also to any equation that can be expressed in linear form. Most of the equations commonly used in steady-state enzyme kinetics are of this kind, e.g. the simpler equations for inhibited reactions and two-substrate reactions. We have therefore implemented the procedure as a computer program in FORTRAN that can be applied to any linearizable equation that contains no more than three observed variables (e.g. a rate and one or two concentrations) and no more than four parameters. This program is available from one of us (A. C.-B) on request. It can also be used as a conventional least-squares program, because the biweight feature can be suppressed if desired. The value of σ_0/σ_2 can be adjusted by the program as suggested above, or it can be specified by the user if prior information about its value exists. Furthermore, it allows complex errors to be defined as advocated by Askelöf *et al.* (1976) a different approach from that embodied in eqn. (3).

Results and discussion

Table 1 shows a small, but typical, proportion of the results obtained in Monte Carlo trials of the method described above as applied to the Michaelis–Menten equation with data simulated under various conditions. Although Table 1 shows results for K_m only, other parameters, such as V, V/K_m, $1/V$ and K_m/V, were also studied, and they behaved similarly. In general, K_m is a convenient parameter to consider for this purpose because it is the least precisely defined and is sensitive to the quality of the data over the whole range, in contrast, for example, with V, which is defined mainly by the larger v_i values.

As expected, conventional least squares gives satisfactory results only when correctly weighted and in the absence of outliers. The biweight method with fixed σ-ratio gives good results with and without outliers if the σ-ratio is fixed at the correct value, but poor results if it is fixed at an incorrect value. Conversely, least squares with adjustable σ-ratio gives good protection against the effects of ignorance of the correct value, but does not protect against outliers. Only by combining the biweight method with the capacity to adjust the σ-ratio can one obtain satisfactory results in all circumstances. The combined calculation gives a similar performance to median estimation on the basis of the direct linear plot. As it is much more tedious to

Table 1. *Comparison of methods for fitting the Michaelis–Menten equation*

The Table shows the mean square deviations of K_m given by six methods in simulated experiments in which the true value of K_m was 10.0 and there were ten observations at substrate concentrations evenly spaced from $0.2K_m$ to $2K_m$. In experiments with 'relative errors', each rate had the same coefficient of variation of 5%; in experiments with 'simple errors', each rate had the same S.D. of $0.025V$. In experiments with 'no outliers', the errors were normally distributed with mean zero; in experiments 'with outliers', they were drawn with probability 80% from one normal distribution or with 20% probability from a second with a 4-fold larger S.D., in both cases with zero mean. There were 1000 experiments for each type of error. The numbers shown in parentheses after the mean square deviations in the Table are efficiencies of the methods in comparison with the best method for each type of error. In all cases the mean deviation (bias) was of the order of 0.1; it was thus trivial in comparison with the random error, and the mean square deviations are effectively variances.

Type of error	Simple least squares ($\sigma_2 = 0$)	Relative least squares ($\sigma_0 = 0$)	Least squares σ_0/σ_2 adjustable	Relative biweight ($\sigma_0 = 0$)	Biweight σ_0/σ_2 adjustable	Median
Relative errors, no outliers	1.84 (55%)	1.02 (100%)	1.31 (78%)	1.04 (98%)	1.32 (77%)	1.21 (84%)
Relative errors, with outliers	1.77 (38%)	1.01 (66%)	1.16 (58%)	0.76 (88%)	0.73 (92%)	0.67 (100%)
Simple errors, no outliers	2.36 (100%)	4.69 (50%)	2.45 (96%)	3.93 (60%)	2.95 (80%)	3.01 (78%)
Simple errors, with outliers	2.42 (68%)	4.32 (38%)	2.20 (75%)	2.64 (62%)	1.64 (100%)	1.71 (96%)

compute, it would offer no advantage over the direct linear plot if the Michaelis–Menten equation were the only model that needed to be considered. Unlike the direct linear plot, however, it can be applied with little extra effort to three- and four-parameter models and offers, for the first time, robust approach to the fitting of such models.

This work was supported by grant no. 1851 from the North Atlantic Treaty Organization and grant no. MA-3163 from the Medical Research Council of Canada.

References

Askelöf, P., Korsfeldt, M. & Mannervik, B. (1976) *Eur. J. Biochem.* **69**, 61–67

Atkins, G. L. & Nimmo, I. A. (1980) *Anal. Biochem.* **104**, 1–9

Cleland, W. W. (1979) *Methods Enzymol.* **63**, 103–138

Cornish-Bowden, A. (1980) in *Kinetic Data Analysis: Design and Analysis of Enzyme and Pharmacokinetic Experiments* (Endrenyi, L., ed.), pp. 105–120, Plenum, New York

Cornish-Bowden, A. & Eisenthal, R. (1978) *Biochim. Biophys. Acta* **523**, 268–272

Cornish-Bowden, A., Porter, W. R. & Trager, W. F. (1978) *J. Theor. Biol.* **74**, 163–175

Kendall, M. G. (1970) *Rank Correlation Methods*, 4th edn., Griffin, London

Mosteller, F. & Tukey, J. W. (1977) *Data Analysis and Regression*, pp. 333–379, Addison–Wesley, Reading, MA

Nimmo, I. A. & Mabood, S. F. (1979) *Anal. Biochem.* **94**, 265–269

Siano, D. B., Zyskind, J. W. & Fromm, H. J. (1975) *Arch. Biochem. Biophys.* **170**, 587–600

Storer, A. C., Darlison, M. G. & Cornish-Bowden, A. (1975) *Biochem. J.* **151**, 361–367