J. Enzyme Inhibition, 1992, Vol. 5, pp. 317–321
Reprints available directly from the publisher
Photocopying permitted by license only

ALGEBRAIC LEAST SQUARES ESTIMATES OF INHIBITOR CONSTANTS

M.E. JONES†

The National Centre for Epidemiology and Population Health, The Australian National University, GPO Box 4, Canberra ACT, Australia. 2601

(*Received October 12, 1990*)

When the data has a constant coefficient of variation, appropriately eighted least squares estimates of the parameters for competitive inhibition in both enzyme and binding experiments can be obtained algebraically. The algorithms are presented and justified.

KEY WORDS: Enzyme inhibition, nonlinear parameter estimation, weighted least squares.

INTRODUCTION

In the analysis of initial velocity data from experiments involving enzyme inhibition, the Dixon[1] plot is widely used. As in other aspects of the analysis of enzyme and binding data, reciprocal transforms can be used in two contexts; they can assist exploratory data analysis, and can be used to estimate parameters. Few would argue against the use of transform methods as an aid to exploratory data analysis, but there is increasing consensus that parameter estimation should depend on an appropriately weighted nonlinear least squares fit of the data to the model. If the analysis goes beyond point estimation to confidence intervals and hypothesis testing, then most methods depend on the assumption of an underlying normal distribution of errors, and the estimation of the variance then rests on the residual error following least-squares estimation of the parameters. In models of enzyme inhibition, the parameters enter the model nonlinearly. The estimation of the parameters in nonlinear models is not, in general, possible algebraically; the exact answer can not be obtained in a finite number of calculations, and estimation therefore depends on iteration and approximation. Commonly, the variance of the data is not constant. Small readings often have smaller variances than larger readings. Under such circumstances an appropriate assumption is often one of a constant coefficient of variation,[2] implying that the ratio of the standard deviation to the predicted velocity is constant.

Where the standard deviation depends on the predicted value, even linear regression models have the parameters entering the calculations nonlinearly, and again estimation must depend on iteration and approximation. A discussion of iterative techniques in parameter estimation often involves the Marquardt-Levenberg algorithm and the Cholesky decomposition of matrices. This is unfortunate because these topics lie outside the sphere of interest of many biologists. The iterative techniques require initial estimates of the parameters, and may fail to converge if these estimates are not close to the actual values. Because of their drawbacks, these techniques are better

†Correspondence – The School of Medicine, The Flinders University of South Australia, Bedford Park, SA 5042, Australia.

avoided if there is a direct algebraic approach to the same answer. Where the data has a constant coefficient of variation, the models involved in competitive enzyme inhibition (and equivalently, competitive binding) permit an algebraic solution to the weighted least sqaures estimation of the parameters. The algorithms which follow are not an approximation and do not involve transformation of the data. The estimates obtained will be exactly those towards which the iterative techniques are supposed to converge.

COMPETITIVE INHIBITION

The model being considered is that describing initial velocity data from a simple enzymatic reaction in the presence of a competitive inhibitor. Using standard nomenclature (see, for instance Lehninger[3]),

$$V = \frac{V_{max}[S]}{K_s + [S] + [I]K_s/K_i},$$

where $[S]$ and $[I]$ are substrate and inhibitor concentrations, V is the predicted velocity of the reaction and the parameters to be estimated are maximum velocity (V_{max}), substrate constant (K_s), and inhibitor constant (K_i). The experimental data consists of n triplets (v_j, s_j, h_j) $j = 1,2, \cdots, n$ from n incubations in which the j^{th} incubation has substrate concentration s_j and inhibitor concentration h_j and has an observed initial velocity v_j. With standard deviation proportional to the mean, the appropriate least squares estimates $\hat{V}_{max}$, $\hat{K}_s$, and $\hat{K}_i$ will be those which minimize the weighted sum of squares, Ω, defined

$$\Omega = \sum_{j=1}^{n} \left[\left(v_j - \frac{V_{max} s_j}{K_s + s_j + K_s h_j / K_i} \right) \Big/ \frac{V_{max} s_j}{K_s + s_j + K_s h_j / K_i} \right]^2 \qquad (1)$$

Then under the assumption of constant coefficient of variation the values $\hat{V}_{max}$, $\hat{K}_s$ and $\hat{K}_i$ minimizing Ω are derived by the following algebraic algorithm.

Algorithm 1

Nine quantities A to I are calculated from the experimental data;

$$\begin{aligned} A &= \sum v_j^2, \\ B &= \sum v_j^2/s_j, \\ C &= \sum v_j^2 h_j/s_j, \\ D &= \sum v_j^2/s_j^2, \\ E &= \sum v_j^2 h_j/s_j^2, \\ F &= \sum v_j^2 h_j^2/s_j^2, \\ G &= \sum v_j, \\ H &= \sum v_j/s_j, \\ I &= \sum v_j h_j/s_j. \end{aligned}$$

Then let

$$\begin{aligned} J &= A(DF - E^2) - B(BF - CE) + C(BE - CD), \\ L &= [G(DF - E^2) + H(CE - BF) + I(BE - CD)]/J, \\ M &= [G(CE - BF) + H(AF - C^2) + I(BC - AE)]/J, \\ N &= [G(BE - CD) + H(BC - AE) + I(AD - B^2)]/J, \end{aligned}$$

in terms of which the required parameter estimates are,

$$\begin{aligned} \hat{V}_{max} &= 1/L, \\ \hat{K}_s &= M/L, \\ \hat{K}_i &= M/N. \end{aligned}$$

RADIOTRACER DILUTION

Radiotracer dilution represents a particular case of competitive inhibition in which $K_s = K_i$. It is perhaps more commonly employed in binding, rather than enzyme, experiments, but for simplicity we present the enzymatic equivalent. We equate hot (radiolabelled) substrate with s_j and cold (unlabelled) with h_j. Then v_j is the velocity of incorporation of the radiolabelled substrate.

Algorithm 2

Define the quantities A to I as in the first algorithm. Then proceed,

$$\begin{aligned} J &= D(A + 2C + F) - (B + E)^2, \\ L &= [D(G + I) - H(B + E)]/J, \\ M &= [H(A + 2C + F) - (G + I)(B + E)]/J, \end{aligned}$$

in terms of which the required estimates are $\hat{V}_{max} = 1/L$ and $\hat{K}_s = M/L$ as previously.

THEORY

The weighted sum of squares 1 may be recast,

$$\Omega = \sum_{j=1}^{n} \left(\frac{v_j}{V_{max}} + \frac{K_s}{V_{max}} \frac{v_j}{s_j} + \frac{K_s}{V_{max} K_i} \frac{v_j h_j}{s_j} - 1 \right)^2,$$

and accordingly the problem may be rephased in terms of a multiple linear regression. In its more familiar context, multiple linear regression seeks to describe some vector of dependent variable readings, or 'observations' in terms of a linear combination of several explanatory variables. In the present context, the vector of 'observations' is replaced by a vector in which all the entries are unity. The explanatory variables are v_j, v_j/s_j, and $v_j h_j/s_j$, the vectors of which define a three-dimensional model space. As in any multiple linear regression, we seek the vector in model space which is, in a least-squares sense, closest to the vector of observations. In brief, the explanatory

variables v_j, v_j/s_j and v_jh_j/s_j are regressed on the unit vector. The coefficients associated with the three variables will then be the least squares estimates of $1/V_{max}$, K_s/V_{max} and $K_s/(K_iV_{max})$ respectively.

Readers with access to general purpose regression programs such as MINITAB[4] or GLIM[5] may wish to use these facilities rather than to program the algorithm as presented. In the following development we let $L = 1/V_{max}$, $M = K_s/V_{max}$ and $N = K_s/(K_iV_{max})$. Then using A to I as defined earlier, the normal equations $\partial\Omega/\partial L = 0$, $\partial\Omega/\partial M = 0$ and $\partial\Omega/\partial N = 0$ may be written in matrix form,

$$\begin{pmatrix} A & B & C \\ B & D & E \\ C & E & F \end{pmatrix} \begin{pmatrix} L \\ M \\ N \end{pmatrix} = \begin{pmatrix} G \\ H \\ I \end{pmatrix},$$

and the equations as presented in algorithm 1 amount to the use of Cramer's rule (see, for instance, Noble[6]) to invert the matrix, J being its determinant. With the assumption that $K_s = K_i$ in algorithm 2, the objective function simplifies to

$$\Omega = \sum_{j=1}^{n} \left(\frac{1}{V_{max}} \left[v_j + \frac{v_jh_j}{s_j} \right] + \frac{K_sv_j}{V_{max}s_j} - 1 \right)^2,$$

and the problem remains one of multiple linear regression on the unit vector. With L and M as defined previously, the normal equations become,

$$\begin{pmatrix} A + 2C + F & B + E \\ B + E & D \end{pmatrix} \begin{pmatrix} L \\ M \end{pmatrix} = \begin{pmatrix} G + I \\ H \end{pmatrix},$$

and as before, the solution uses Cramer's rule with J being the determinant of the matrix.

DISCUSSION

Although these algorithms use the standard methods of linear algebra, they do not amount to the imposition of a linearizing transform. The objective function Ω minimized is precisely that which iterative methods minimize. The algorithms depend on the fortuitous circumstance that, in this particular nonlinear model, the nonlinearities of the model and of the weighting essentially cancel out, leaving a problem which can be solved algebraically. Accordingly this communication does not propose any new or improved estimator. Both the estimates and their statistical properties will be identical to those arrived at by more cumbersome iterative algorithms. Only the method of arriving at the estimates is different. The method is free of convergence problems, and does not require initial estimates. It offers, therefore, the statistical efficiency of the direct least squares methods together with the convenience of the transform methods. A constant coefficient of variation is often an appropriate assumption,[2] but if some alternative weighting is imperative, then the algorithms above offer useful initial parameter estimates with which to begin any iterative search.

Acknowledgements

This work was supported by the Clive and Vera Ramaciotti Foundations of Australia and by the National Health and Medical Research Council of Australia.

References

1. Dixon, M. (1953) *Biochem. J.*, **55,** 170–171.
2. Zivin, J.A. and Waud, D.R. (1982) *Life Sciences*, **30,** 1407–1422.
3. Lehninger, A.L. (1982) *Principles of Biochemistry*, pp. 213–215. New York: Worth Publishers.
4. Ryan, B.F., Joiner, B.L. and Ryan, T.A. jnr. (1976) *Minitab Handbook*, pp. 236–238. Boston: PWS-Kent.
5. Payne, C.D. (1987) Editor. *The Generalized Linear Interactive Modelling System, Release 3.77*. Royal Statistical Society. Oxford: Numerical Algorithms Group.
6. Noble, B. (1969) *Applied Linear Algebra*, pp. 209–211. New Jersey: Prentice-Hall.