APPLICATION OF WEIGHT FACTORS IN THE DISCRIMINATION BETWEEN MATHEMATICAL MODELS OF ENZYME KINETICS

Tamás BÁRTFAI and Bengt MANNERVIK
Department of Biochemistry, University of Stockholm, Stockholm, Sweden

Received 2 March 1973

1. Introduction

A fundamental aim of experimentation in physical sciences is to explore the elementary components of an observed phenomenon, i.e. to define the physical model of the event. For a quantitative description, it is furthermore desirable to derive a mathematical model of the system investigated and to compute the set of parameter values, which gives the best fit of the model to the data. The use of weight factors is an integral part of the computation procedures for estimation of the parameter values [1] of the adopted model. However, the application of weight factors in the discrimination between alternative mathematical models has not, so far, been utilized. In the present paper it is demonstrated that weighting can serve this purpose, and it is furthermore illustrated by an example how the choice of weight factors improves the discrimination. It should be noted that this new principle for discrimination between alternative mathematical models is especially valuable when no or limited possibilities exist for obtaining new experimental data according to the optimal design for discrimination [2–4]. The procedure described here has been used in enzyme kinetics, but can equally well be applied in other fields of science, where alternative mathematical models are considered and the data can be subjected to regression analysis.

2. Basis for discrimination between alternative mathematical models

It is assumed that some of the powerful optimization techniques [5] has been used for fitting several plausible mathematical models to a given set of experimental data. The results of fitting different models to the same data can be used to select on a statistical ground the best model. A procedure for discrimination on this basis has recently been described [3, 4] and criteria are given to allow rejection of poor models. Models, which have not been rejected according to other criteria, are compared (criterion E [3, 4]) by pairwise examination of the ratio of their corresponding residuals sums of squares, Q^2,

$$\frac{Q_j^2}{Q_k^2} = \frac{\sum_i (v_i - \hat{v}_{ij})^2/(n-p_j)}{\sum_i (v_i - \hat{v}_{ik})^2/(n-p_k)} \qquad (1)$$

where Q_j^2 and Q_k^2 are the residual sums of squares, $\hat{v}_{ij}$ and $\hat{v}_{ik}$ the predicted values in the i-th experimental point, and p_j and p_k the number of parameters for models j and k, respectively; n is the number of experimental points and v_i the observed value of the independent variable (velocity in the example given below) in the i-th experimental point. The difference between the observed and predicted value of model $j(k)$ is the residual of model $j(k)$, in the i-th point $q_{ij}(q_{ik})$,

$$q_{ij} = v_i - \hat{v}_{ij} \qquad (2)$$

It is of utmost importance that outliers, which are single observations giving unexpectedly large residuals, are eliminated (cf. [6]), because they have a strong influence on the regression and may lead to erroneous conclusions in the discrimination procedure. A choice of the best model can be made if the Q^2-value for one

model is significantly smaller than the values of all other models, which have passed the earlier examination. It is considered that a significant difference exists if the ratio (1) exceeds the Fisher statistic (F-test) at the appropriate degrees of freedom, $(n-p_j)$ and $(n-p_k)$, and desired significance level. (It should be kept in mind, however, that nonlinear regression will not allow strict use of the F-test unless a linear approximation is applicable in the vicinity of the solution vector in the parameter space. Tests for linearity, cf. [7], can corroborate the use of the F-test.) The model with the smallest Q^2-value is considered to be the best of the alternatives.

If a choice between two alternative models cannot be made, because the Q^2-values are too similar, it is possible to increase Q^2 by collecting systematically large residuals by additional experimentation at points in the space of experimental variables, where a poor model fails to describe the system [2, 8]. The Q^2-value of the "true" model will not increase. We have earlier defined a discrimination function, g, which can help to make the proper experiment design for selection of the best model [3, 4]. This function is based on the residuals of the alternative models j and k, which are considered to consist of three components each:

$$q_{ij} = \epsilon_i + \epsilon_{ij} + \epsilon_{ijh} \quad (3)$$

$$q_{ik} = \epsilon_i + \epsilon_{ik} + \epsilon_{ikh} \quad (4)$$

where ϵ_i is the experimental or "pure" error (variance error), ϵ_{ij} and ϵ_{ik} are errors due to inadequacy of models j and k (bias error), and ϵ_{ijh} are errors due to the h-th regression procedure (computing error). Presently, we consider $\epsilon_{ijh} \approx \epsilon_{ikh}$, which implies that the only difference between the residuals of the two models is the bias error. The discrimination function, g, is defined as the absolute value of the difference between the residuals, and the value of g in the i-th experimental point is:

$$g_i = |q_{ij} - q_{ik}| = |\hat{v}_{ij} - \hat{v}_{ik}| \approx |\epsilon_{ij} - \epsilon_{ik}| \quad (5)$$

Numerical evaluation of g is based on the current estimates of the parameter values of the corresponding models.

The value of g increases with the information for discrimination between two models gained by experimentation in a particular point in the space of independent variables (reactant concentrations in the example given below). Thus the data set should be enlarged by experiments designed to give high g-values, which correspond to large residuals of the poor model. This is a natural weighting procedure and results in a higher Q^2-value of the poor model, which will allow discrimination if the information content of the data is sufficient.

3. Weighting

3.1. *Weight factors for parameter estimation*

Weight factors should be used in the parameter estimation based on the least squares method, because the terms in the regression function

$$Y = \sum_i (v_i - \hat{v}_{ij})^2 \quad (6)$$

should have equal relative accuracy, [1]. Normally, the weighting is based on the experimental error in the i-th point taking the weights (w_i) inversely proportional to the square of the experimental error (the variance)

$$Y = \sum_i w_i (v_i - \hat{v}_{ij})^2 = \sum_i \frac{1}{\epsilon_i^2} (v_i - \hat{v}_{ij})^2 \quad (7)$$

This weighting procedure reduces bias in the parameter estimation.

3.2. *Weight factors for discrimination*

The use of weight factors in parameter estimation is common practice, but weighting for discrimination has to our knowledge not heretofore been used. The basic idea is the same, viz. to give the greatest importance to values offering most of the information for the particular purpose, i.e. parameter estimation or model discrimination. However, the information for the two separate purposes is distributed in different ways in the space of independent experimental variables; the experimental points minimizing the volume of an ellipsoidal confidence region about the estimated parameter vector serve the first purpose [9], whereas the experimental points maximizing g serve the second goal.

Consider models j and k, which predict $\hat{v}$-values,

which are the same (within the significance level adopted) everywhere in the space of experimental variables except in a region containing experimental point m. A discrimination between models j and k is feasible only on the basis of experiments in the neighborhood of point m. Only for points in this region will the discrimination function g assume a value significantly greater than zero. If experiments are repeated exclusively near point m, a significant difference between the Q^2-values for the two models will develop after a number of experiments. The same effect is achieved if point m is given a sufficiently heavy weight in comparison with other points.

3.2.1. Weighting based on the discrimination function

The weight factors should preferably have a direct functional relation to the information content for discrimination in a particular experimental point. Therefore, the value of g offers a natural weight factor. The weighted Q^2-values of models j and k are:

$$Q_j^2 = \sum_i g_i(v_i - \hat{v}_{ij})^2/(n-p_j) \quad (8)$$

$$Q_k^2 = \sum_i g_i(v_i - \hat{v}_{ik})^2/(n-p_k) \quad (9)$$

$$g_i \epsilon [0, \max(\hat{v}_{ij}, \hat{v}_{ik})]$$

The weighted Q^2-values permit discrimination with a smaller number of experimental data than non-weighted values. The only difference between making new experiments with replicates in the experimental points in proportion to the corresponding g values and calculating the weighted Q^2-values is the assumption that the mean of the observed values in the i-th experimental point will not change after g replicates. As the residuals of both models are multiplied by the same factor, the original information of the measurements will not be distorted by the weighting. The degrees of freedom (d.f.) of the Q^2-values for the models will not change by the weighting, because the number of independent pieces of information has not been altered. The numerical values of g are readily available if computer programs calculating residuals or predicted $\hat{v}$-values are used. The time and cost of calculating g_i-values are negligible in comparison with the requirements for new experiments to reach the same goal.

A heavier weighting for discrimination is achieved by using g_i^2 as a weight factor. Under certain conditions g_i^2 is directly proportional to the information for discrimination [10].

3.2.2. Weighting based on normalized residuals

The weighting based on the discrimination function (cf. sect. 3.2.1) has the uncertainty that large residuals of two rival models will give a heavy weight if the difference between the predicted values is large in the same experimental point. This causes undesirable emphasis on points, in which both models are poor. To avoid this feature, a normalized weight factor was defined

$$w_i = \frac{|q_{ij}^2 - q_{ik}^2|}{q_{ij}^2 + q_{ik}^2} \quad (10)$$

$$w_i \epsilon [0,1]$$

This weight factor is maximal when one model coincides with the experimental point, and equal to zero when both models have equal deviations, i.e., information for discrimination is lacking. This weighting procedure is reliable, but improves the discrimination only moderately. A similar weight factor has been proposed [11] for parameter estimation.

3.2.3. Model selection based on the information theory

The approach to the discrimination between models afforded by the discrimination function g is closely related to the solution given by the information theory of the problem how probabilities are distributed among alternative models [12]. Under the assumptions of normal distribution, constant experimental error, and linearity of the models, the posterior probability density function is for model j in experimental point $(i+1)$ (cf. [8]).

$$p_{i+1,j} = (2\pi(\sigma^2+\sigma_j^2))^{-1/2} \exp\left(\frac{-(v_{i+1}-\hat{v}_{i+1,j})^2}{2(\sigma^2+\sigma_j^2)}\right) \quad (11)$$

where σ^2 and σ_j^2 are exp imental variance and variance according to model j, res, ectively; σ_j^2 and predicted values are based on the first i experimental points.

The prior probability in point $(i+1)$ for model j can be expressed as

$$\Pi_{j+1,j} = \frac{p_{i,j}\Pi_{i,j}}{p_{i,j}\Pi_{i,j} + p_{i,k}\Pi_{i,k}} \quad (12)$$

The best model is chosen when the information entropy (S) of the system of alternative models

$$S = -(\Pi_j \ln \Pi_j + \Pi_k \ln \Pi_k) \quad (13)$$

is minimal, i.e. either Π_j or Π_k is zero [8]. Π_j and Π_k denote the probabilities associated with models j and k.

Evaluation of Π_j and Π_k leads to a selection of one model in few steps of calculation, but this method has the following drawbacks. First, the discrimination is very much influenced by the estimate of the experimental variance (σ^2), and the experimental error should be examined in every point to get a reliable estimate. Second, the method is very sensitive and only a few experimental data are effective in the discrimination. This feature favors a model giving an excellent fit to the experimental points first used in the computation, and makes the model selection insensitive to inadequacy of the model in subsequent points. (When $p_{i,j} = 1.0$, a very large residual of model j in point $(i+1)$ will not alter the choice already made.) Application of this method requires particular precautions, e.g. randomization of the data set and giving different probabilities to the models as starting values.

4. An example of the use of weighting for discrimination between alternative steady state kinetic models

The weighting methods for discrimination outlined above are applicable in all cases where regression analysis is used in the selection of the best mathematical model. As an example is given the fitting of two steady state kinetic models of yeast glyoxalase I to initial velocity data (experiments in collaboration with Dr. K. Ekwall and Mrs. B. Górna-Hall; cf. [13]). The enzyme acts on an equilibrium system, and according to model 1, it is assumed that only one of the components influences the velocity. This corresponds to the ordinary Michaelis–Menten rectangular hyperbola. According to model 2, an additional component is required to describe the kinetics. This model corresponds to linear competitive inhibition by the second component (cf. [13]). The first experimental data (data set A) were obtained under conditions suitable for parameter estimation under the assumption that model 1 was valid. Fitting model 2 to these data gave a slight decrease in the Q^2-value, but the difference between the Q^2-values for the alternative models was not significant (cf. table 1). Use of the normalized weight factors hardly affected the ratio of the Q^2-values, whereas the weight factors based on the discrimination function (i.e. g_j and g_j^2) increased the ratio somewhat. However, the difference was not significant (at 95%) according to the F-test. Discrimination according to the information theory (eqn. (11)) selected model 2 (i.e. $\Pi_2 = 1.0$) in less than 10 steps of computation, but in some cases, where the order of calculation was changed by permuting the serial numbers within the data set, model 1 was selected. We conclude that the information for discrimination in data set A was not rich enough for discrimination according to the weighted regression procedures. On the other hand, model selection based on the information theory is powerful, but is founded on more restricted assumptions, and has to include calculations on a number of permutations of the order within the data set.

To supply a more informative data set, experiments were designed to better discriminate between the two models. The results of these experiments (data set B) gave a greater difference between the Q^2-values of the two models (cf. table 1). Under the assumption of linearity of the models (cf. sect. 2), model 2 is significantly better than model 1 even without weighting ($w_j = 1.0$), and introduction of weight factors significantly increases the ratio of the Q^2-values. Again the model selection based on the information theory gives the same result as the weighting, but several computations are required to ascertain that the choice is correct.

5. Conclusion

It has been demonstrated that weighting can serve the purpose of discrimination between alternative mathematical models, which can be subjected to regression analysis. Three weight factors have been presented; one is derived from normalized residuals and two are based on the discrimination function. The first of these is weak but reliable, whereas the latter

Table 1
Effect of weighting on the quotient of the residual sums of squares, Q_1^2/Q_2^2, of two rival models.

Data set	Weights			
	1.0 (= no weights)	g_i	g_i^2	$\frac{\lvert q_{ij}^2 - q_{ik}^2 \rvert}{q_{ij}^2 + q_{ik}^2}$
(d.f. $\overset{A}{=}$ 25/24)	$\frac{0.000080}{0.000063} = 1.270$	$\frac{0.00001341}{0.00000892} = 1.503$	$\frac{0.00001214}{0.00000689} = 1.762$	$\frac{0.000648}{0.000503} = 1.288$
(d.f. $\overset{B}{=}$ 69/68)	$\frac{0.001614}{0.000253} = 6.379$	$\frac{0.006392}{0.000432} = 14.80$	$\frac{0.044683}{0.001848} = 24.18$	$\frac{0.095387}{0.007697} = 12.39$

Discrimination is obtained at the 95% confidence level for data set A if $Q_1^2/Q_2^2 > 1.98$ and for data set B if $Q_1^2/Q_2^2 > 1.53$ under the assumption of linearity of the models.

are stronger. The most powerful discrimination technique, which is based on the information theory, seems to over-estimate the information content and sometimes selects the wrong model. We suggest that increasingly stronger methods are used in the weighting for discrimination, and if only the method based on the information theory allows selection of the best model, the result should be confirmed by a series of computations on different permutations of the data set, and by use of different initial probabilities for the models considered.

It should also be noted that evaluation of the weight factors based on the discrimination function g also serves the purpose of experiment design. Thus, if the discrimination problem cannot be solved, new experiments should be made in the experimental region giving high g-values.

Finally, when the best model has been selected, the parameter values should be re-evaluated by a new computation according to eqn. (7).

References

[1] G.N. Wilkinson, Biochem. J. 80 (1961) 324–332.
[2] W.G. Hunter and A.M. Reiner, Technometrics 7 (1965) 307–323.
[3] T. Bártfai and B. Mannervik, FEBS Letters 26 (1972) 252–256.
[4] T. Bártfai and B. Mannervik, in: Analysis and Simulation of Biochemical Systems, eds. H.C. Hemker and B. Hess (North-Holland, Amsterdam, 1972) p. 197–209.
[5] S.L.S. Jacoby, J.S. Kowalik and J.T. Pizzo, Iterative Methods for Nonlinear Optimization Problems (Prentice-Hall, Englewood Cliffs, New Jersey, 1972).
[6] F.J. Anscombe, Technometrics 2 (1950) 123–147.
[7] I. Guttman and D.A. Meeter, Technometrics 7 (1965) 623–637.
[8] W.J. Hill, W.G. Hunter and D.W. Wichern, Technometrics 10 (1968) 145–160.
[9] G.E.P. Box and H.L. Lucas, Biometrika 46 (1959) 77–90.
[10] Y. Bard and L. Lapidus, Catalysis Reviews 2 (1968) 67–112 (see p. 104).
[11] J.H. Ottaway, Biochem. J. 125 (1971) 44p–45 p.
[12] C.E. Shannon, Bell System Tech. Journal 27 (1948) 379–423 and 623–656.
[13] T. Bártfai, K. Ekwall and B. Mannervik, Biochemistry 12 (1973) 387–391.