UC Davis

IDAV Publications

Title

Analyte Identification in Multivariate Calibration

Permalink

<https://escholarship.org/uc/item/5ph8t7q1>

Journal

Biometrics, 57

Authors

Jones, Geoffrey
Rocke, David

Publication Date

2001

Peer reviewed

Analyte Identification in Multivariate Calibration

Geoffrey Jones

Institute of Information Sciences and Technology, Massey University,
Palmerston North, New Zealand

and

David M. Rocke

Center for Image Processing and Integrated Computing,
University of California, Davis, California 95616, U.S.A.
email: dmrocke@ucdavis.edu

SUMMARY. In calibration experiments, an estimated relationship between covariate information for a sample and an observed response is used to infer the covariate information for unknown samples from their responses. In some situations, this covariate information comprises a nominal variable (e.g., identity of a chemical, sex of an animal) and a real-valued variable (e.g., concentration of the chemical, age of animal). If the calibrating relationship can be estimated separately for each candidate identity, the first step in analyzing unknown samples is to correctly determine their identity. A discrimination statistic is suggested for use in this situation and its asymptotic distribution is derived. The investigation is motivated by the possibility of using multiple immunoassays in environmental monitoring to identify and quantitate contaminated samples in situations where there are several candidate pollutants that cross-react significantly to single assays. An example is given of the use of a four-antibody assay for the simultaneous monitoring of the levels in water samples of several of the commonly used triazine herbicides and their derivatives.

KEY WORDS: Cross-reactivity; Discrimination; Environmental monitoring; Four-parameter log-logistic model; Immunoassay; Multivariate calibration; Nonlinear regression.

1. Introduction

Suppose we have a multivariate bioassay system in which the individual component assays are not specific for a particular target analyte but respond with varying sensitivities to a number of different analytes. Such a situation can occur with immunoassays when antibodies developed for one target analyte cross-react significantly with other similar compounds. For example, in the environmental monitoring of the levels of s-triazine herbicides in water samples, it may be desirable to use an array of nonspecific immunoassays to detect, identify, and quantitate the members of this class of similar chemicals simultaneously. A simple example is illustrated in Figure 1, showing the bivariate response from a two-antibody system to varying concentrations of the herbicides atrazine and terbutryn together with the response from an unknown sample. We wish to identify the chemical in the sample and estimate its concentration. Intuitively one might use the distance from the sample response to the response path specific to each chemical, but the metric for this distance must be chosen to incorporate what is known, or can be assumed, about the uncertainty in the responses. In reality, the class of possible analytes might be much larger. We have used the methods described here to investigate, in collaboration with chemists

and toxicologists, the use of an array of four antibodies for identifying eight s-triazines in water samples in concentrations of 0.5–100 ppb (parts per billion). Because our analysis extends to the possibility of simple mixtures of analytes, we denote the concentration by the $r \times 1$ vector $\mathbf{x}$ so that, if we are looking for two-component mixtures, $r = 2$ and there are 28 possible identities.

We assume that the vector response of the assay system to concentration $\mathbf{x}$ of analyte j can be modeled as

$$\mathbf{y}_j = \mathbf{f}_j(\boldsymbol{\theta}_j, \mathbf{x}) + \boldsymbol{\epsilon}, \quad (1)$$

where $\mathbf{y}_j$ is a $q \times 1$ response vector, $\mathbf{f}_j$ is a function assumed known to within a fixed $p \times 1$ parameter vector $\boldsymbol{\theta}_j$, and $\boldsymbol{\epsilon}$ is a zero-mean error term. The dependence of $\mathbf{f}_j$ on the analyte identity j may be only through the parameter vector $\boldsymbol{\theta}_j$, or it may be more general. We shall assume that $\boldsymbol{\epsilon}$ is normally distributed with positive-definite covariance matrix $\boldsymbol{\Sigma}$.

Given training data $\{(\mathbf{y}_{ij}, \mathbf{x}_{ij}) : i = 1, \dots, n_j\}$ for each analyte identity $j = 1, \dots, J$, we can obtain consistent estimates $\hat{\boldsymbol{\theta}}_j$ of $\boldsymbol{\theta}_j$ and $\mathbf{S}$ of $\boldsymbol{\Sigma}$ provided that certain smoothness and identifiability conditions are met. If we have an observed response from a sample of known identity but unknown concentration, standard calibration theory may be used (Brown,

Figure 1. Response paths for atrazine and terbutryn assayed with two antibodies. Y_1 and Y_2 are the observed responses (an optical density) from each of the two assays on two sets of standard concentrations and one unknown sample.

1993; Oman, 1999). Suppose, however, that we have an observed response y_0 from a sample for which the identity j and concentration x_0 are both unknown. We discuss in this article the problem of identifying the analyte by examining the compatibility of y_0 with estimated response curves $C_j = \{f_j(\hat{\theta}_j, x) : x \in \mathbb{R}^{+r}\}$, $j = 1, \dots, J$.

If we define the distance to the point on C_j at concentration x as $R_j(x) = [y_0 - f_j(\hat{\theta}_j, x)]^T S^{-1} [y_0 - f_j(\hat{\theta}_j, x)]$, then the conditional least squares estimator of x_0 , assuming analyte j , is the minimizer of this distance, $\hat{x}_{0j} = \text{argmin}_x R_j(x)$. The use of this minimum value $R_j(\hat{x}_{0j})$ as a diagnostic statistic has been discussed in the literature for the case of linear calibration (Brown and Sundberg, 1989; Martens and Naes, 1989), but the distributional results used do not allow for uncertainty in the estimate $\hat{\theta}$. We discuss in Section 2 the normalization of this distance to allow for parameter uncertainty in the linear case. Extension to nonlinear calibration curves is considered in Section 3. In Section 4, we illustrate the use of the minimum normalized distance R_j^* in the identification of s-triazine herbicides by multiple immunoarray.

The approach we take is to report all possible identities j for which R_j^* is probabilistically small. If a point estimate of j were required, it would be natural to choose j so that R_j^* is minimized. The concentration x_0 , conditional on the identity, could also be given as a point or interval estimate. These and other issues are discussed in Section 5.

Details of proofs and simulations are omitted. The interested reader is referred for these to Jones and Rocke (1997).

2. Linear Case

Suppose we have a linear multivariate calibration experiment with training data (X, Y) , where the responses Y form an $n \times q$ matrix, where each row is an independent observation of a q -variate response and the rows of the $n \times p$ design matrix X give the corresponding values of the p covariates. We assume the multivariate linear model $Y = 1_n \alpha^T + XB + E$, where $\alpha_{q \times 1}$ is a vector of unknown constants, $B_{p \times q}$ is the matrix of unknown slope parameters, and $E_{n \times q}$ is an error matrix whose rows are independently distributed as $N(0_q, \Sigma)$, with

Σ an unknown $q \times q$ positive definite covariance matrix. We assume further that the columns of X have been centered (i.e., that each covariate is measured from its mean) so that $X^T 1_n = 0_p$.

We denote the usual estimates of the model parameters by $\hat{\alpha}$, $\hat{B}$, and S (cf., Anderson, 1984, pp. 287–291). Now suppose we have ℓ replicates of an unknown with covariate values x_0 and mean response y_0 . Brown (1982) showed that

$$\frac{(y_0 - \hat{\alpha} - \hat{B}^T x_0)^T S^{-1} (y_0 - \hat{\alpha} - \hat{B}^T x_0)}{\frac{1}{\ell} + \frac{1}{n} + x_0^T (X^T X)^{-1} x_0} \sim (n - p - 1) \frac{q}{n - p - q} F_{n-p-q}^q \quad (2)$$

Williams (1959) suggested that, with the conditional least squares estimator $\hat{x}_0$ replacing x_0 , the above result holds with the numerator degrees of freedom in the F -distribution and scaling factor replaced by $q - p$. His suggestion seems to be based on the intuition that estimating the p parameters in x_0 will reduce the degrees of freedom by that amount. A more careful intuition notes the complementary effect on the denominator degrees of freedom, as in the derivation of the multivariate T^2 distribution (Anderson, 1984, pp. 161–162), and suggests the denominator degrees of freedom to be $n - q$. This is confirmed by the large- n asymptotic expansions for $R \equiv [y_0 - f(\hat{\theta}, \hat{x})]^T S^{-1} [y_0 - f(\hat{\theta}, \hat{x})]$ of Fujikoshi and Nishi (1984) and Davis and Hayakawa (1987). Their results, however, depend on x_0 , and they cannot justify replacing this by $\hat{x}_0$ since the variance of $\hat{x}_0$ does not decrease as n increases. We could circumvent this difficulty by insisting that ℓ be large, but this severely limits the practical applicability of the result. Instead, we follow Brown and Sundberg (1987) in using small-sigma asymptotics so that our result applies for moderate amounts of training data and any number of replicates of unknowns, provided that the errors are reasonably small.

THEOREM 1: Let $\Sigma = m^{-1} \Sigma_0$, where Σ_0 is a fixed positive definite symmetric $q \times q$ matrix. Then, as $m \rightarrow \infty$, the statistic R^* defined by

$$R^* \equiv \frac{(y_0 - \hat{\alpha} - \hat{B}^T \hat{x}_0)^T S^{-1} (y_0 - \hat{\alpha} - \hat{B}^T \hat{x}_0)}{\frac{1}{\ell} + \frac{1}{n} + \hat{x}_0^T (X^T X)^{-1} \hat{x}_0} \quad (3)$$

where $\hat{x}_0 = (\hat{B}S^{-1}\hat{B}^T)^{-1}\hat{B}S^{-1}y_0$ has

$$R^* \xrightarrow{D} (n - p - 1) \frac{q - p}{n - q} F_{n-q}^{q-p} \quad (4)$$

and F_{n-q}^{q-p} denotes the F -distribution function with $(q - p)$ and $(n - q)$

The validity of this approximation depends on $B\Sigma^{1/2}$ being large. This is similar to Brown's (1982) condition for inevitability of a prediction region; intuitively, one requires the slope to be large relative to statistical uncertainty. Simulation suggests, however, that the approximation is remarkably good even for moderate-sized Σ . In practice, n will also be moderate sized so that the two asymptotic approaches work together.

If another estimator of Σ is available, such as is obtained by pooling S with the residuals from the unknowns, an analogous result is obtained by adjusting the denominator degrees of freedom in the F -distribution, provided that this estimator

is still independent of $\mathbf{y}_0$ and $\hat{\mathbf{B}}$. If there are many additional degrees of freedom, then the distribution of R^* can be reasonably approximated by χ^2_{q-p} .

To investigate the practical applicability of our asymptotic result, simulation was used to generate empirical distributions for R^* and R , and these were compared with postulated distributions using the Kolmogorov-Smirnov statistic $\max |\hat{F}(\cdot) - F(\cdot)|$ and the observed rates of exceedance of the nominal 95th and 99th percentiles for various values of n , q , p , and m . Williams's proposal for R^* was compared with ours and a large- n simplification $R^* \sim \chi^2_{q-p}$; we also investigated Brown and Sundberg's (1989) large- n asymptotic result $R \sim (1/l)\chi^2_{q-p}$.

Full details of the simulation results are available in a technical report (Jones and Rocke, 1997). To summarize, Williams's proposal behaves poorly but expectedly less so as n gets larger. Our approximation, on the other hand, performs remarkably well even for small m (corresponding to large Σ): if all the slope parameters are $O(1)$, the approximation is good for $m = 4$. The large- n approximation to our result turns out to be inadequate until n approaches 50 and then only for small q . Brown and Sundberg's (1989) asymptotic result using R is not applicable when there is only a moderate amount of training data.

3. Nonlinear Extensions

The general nonlinear multiresponse model (equation (1)) is very difficult to analyze since, unlike the linear case, estimation of θ cannot in general be separated from estimation of Σ and $\hat{\theta}$ and $\mathbf{S}$ are not independent. Analogous results to the linear case (equations (2) and (4)) appear to be intractable for the general nonlinear model. With appropriate regularity conditions (cf., Seber and Wild, 1989, pp. 581-585), the large-sample distribution of $n^{1/2}(\hat{\theta} - \theta)$ is asymptotically normal with zero mean and a covariance matrix that can be estimated consistently from the data. However, using small-sigma asymptotics, with $\Sigma = m^{-1}\Sigma_0$, the asymptotic covariance matrix, $\mathbf{V}_0$ say, of $m^{1/2}(\hat{\theta} - \theta)$ cannot be estimated consistently and neither can Σ_0 .

However, because of our assumption (equation (1)) that Σ does not depend on j , the calibration data for each analyte may be combined to give a pooled estimate of the variance. If there are many replicated unknowns, this further increases the degrees of freedom for $\mathbf{S}$. Thus, the effective sample size for the estimation of Σ will often be much larger than that used in estimating θ . In such cases, we may treat Σ_0 as known, in which case $\mathbf{V}_0$ is consistently estimated by $m\hat{\mathbf{V}}$, say, from the calibration data. The asymptotic distribution of $m^{1/2}[\mathbf{f}(\hat{\theta}, \mathbf{x}_0) - \mathbf{f}(\theta, \mathbf{x}_0)]$ is given via the delta method as $\mathbf{N}_q(\mathbf{0}, \mathbf{G}_0)$, where

$$\mathbf{G}_0(\mathbf{x}_0) = \left(\frac{\partial \mathbf{f}^T(\theta, \mathbf{x}_0)}{\partial \theta} \right) \mathbf{V}_0 \left(\frac{\partial \mathbf{f}(\theta, \mathbf{x}_0)}{\partial \theta^T} \right)$$

is consistently estimated by replacing θ by $\hat{\theta}$ and $\mathbf{V}_0$ by $m\hat{\mathbf{V}}$. Writing the estimator of the variance of $\mathbf{f}(\hat{\theta}, \mathbf{x}_0)$ as $\hat{\mathbf{G}}$, it is easily seen that

$$R^*(\mathbf{x}_0) \equiv [\mathbf{y}_0 - \mathbf{f}(\hat{\theta}, \mathbf{x}_0)]^T (\ell^{-1}\Sigma + \hat{\mathbf{G}})^{-1} [\mathbf{y}_0 - \mathbf{f}(\hat{\theta}, \mathbf{x}_0)] \xrightarrow{\mathcal{D}} \chi^2_q, \quad (5)$$

which gives a nonlinear equivalent of equation (2). However,

the distributional result is now only approximate and depends on Σ being small.

Next we consider replacing $\mathbf{x}_0$ by the conditional least squares estimator $\hat{\mathbf{x}}_0$ to derive the distribution of a statistic for discriminating between candidate calibration models. In the simpler case where θ is known, or estimated precisely, we have the following.

THEOREM 2: Let $\Sigma = m^{-1}\Sigma_0$, where Σ_0 is a positive definite covariance matrix. Then, as $m \rightarrow \infty$, $R \equiv [\mathbf{y}_0 - \mathbf{f}(\theta, \hat{\mathbf{x}}_0)]^T \Sigma^{-1} [\mathbf{y}_0 - \mathbf{f}(\theta, \hat{\mathbf{x}}_0)] \xrightarrow{\mathcal{D}} (1/\ell)\chi^2_{q-r}$.

Finally, we consider how the above result is adapted to allow for uncertainty in the parameter estimate $\hat{\theta}$. Letting

$$\hat{\mathbf{G}} = \left(\frac{\partial \mathbf{f}^T(\hat{\theta}, \hat{\mathbf{x}}_0)}{\partial \theta} \right) \hat{\mathbf{V}} \left(\frac{\partial \mathbf{f}(\hat{\theta}, \hat{\mathbf{x}}_0)}{\partial \theta^T} \right)$$

where $\hat{\mathbf{V}}$ is the estimated covariance of θ , we have the following for the situation of Theorem 2.

THEOREM 3: $R^* \equiv [\mathbf{y}_0 - \mathbf{f}(\hat{\theta}, \hat{\mathbf{x}}_0)]^T (\ell^{-1}\Sigma + \hat{\mathbf{G}})^{-1} [\mathbf{y}_0 - \mathbf{f}(\hat{\theta}, \hat{\mathbf{x}}_0)] \xrightarrow{\mathcal{D}} \chi^2_{q-r}$.

The statistic R^* again represents a normalized distance from the response of an unknown to an estimated calibration model. It can be used as a diagnostic statistic to decide whether an unknown sample belongs to the same population as a set of calibration data, the population here being $\{\mathbf{f}(\mathbf{x}) \in \mathbb{R}^q : \mathbf{x} \in \mathbb{R}^{+r}\}$. In the next section, we illustrate an application of the above results to the problem of determining the identity and concentration of contaminated environmental samples by multiple immunoassay.

4. Multiple Immunoassay

4.1 Single Unknown Analyte

Immunoassay is a form of chemical analysis that uses antibodies to detect and quantitate a target compound or analyte. A common model (Rodbard, 1981) for the response Y at concentration x is

$$\log Y = \log \left(\frac{A - D}{1 + (x/C)^B} + D \right) + \epsilon, \quad (6)$$

where $\epsilon \sim N(0, \sigma^2)$, the log transform being applied to correct for heteroscedasticity and skewness.

It often happens that an antibody developed to bind specifically to a particular target molecule will also bind, perhaps less strongly, to similar molecules in the same class of compounds. The s-triazine herbicides, e.g., are a commonly used class of similar compounds, a number of which have been finding their way into ground water and may pose a human health hazard. Large-scale monitoring requires an assay methodology that, like immunoassay, is quick, precise, and cheap, but the development of completely specific antibodies for each member of the class is extremely costly and the subsequent monitoring of each by a different assay inefficient and slow.

One alternative is to use a panel of less specific antibodies to achieve simultaneous identification and quantitation. Our example uses four antibodies to discriminate between eight analytes. The format is such that standard concentrations of each candidate analyte are assayed together with unknown samples on a microtiter plate, as in Figure 2. One such

Table 2
Binary mixture analysis for 1
ppb atrazine + 1 ppb cyanazine

Analyte 1	Analyte 2	$\hat{x}_{01}$	$\hat{x}_{02}$	R^*	p -Value
Atrazine	Cyanazine	0.35	1.61	1.41	0.494
Simazine	Cyanazine	0.38	1.84	1.28	0.527
Cyanazine	DIatrazine	2.03	2.46	1.16	0.560

$$\log(Y_i) = \log \left(\frac{A_i - D_i}{1 + \tilde{x}^{B_i^*}} + D_i \right) + \epsilon_i$$

with

$$\tilde{x} = \sum_{j=1}^r \left(\frac{x_j}{c_{ij}} \right)^{B_{ij}/B_i^*}$$

where B_i^* is the geometric mean of the slope parameters.

If no single-analyte identity is found to be plausible for a sample using the method of Section 4.1, we can next examine the possibility of two-component mixtures. We again search through the possible identities and test their plausibility using R^* as defined in Theorem 3, now with $r = 2$.

This procedure has been investigated using binary mixtures of 1 ppb of each of two analytes chosen from eight candidate triazine compounds. Samples containing binary mixtures of analytes are, as expected, more difficult to identify. Often there were a number of possible identities for the samples, with the problem of confusion within subgroups being compounded. Thus, e.g., a mixture of atrazine and prometon might look like simazine and terbutryn or atrazine and prometryne. In most cases, mixtures were clearly identified as such, i.e., not as single analytes, although mixtures of prometon and terbutryn or prometryne and terbutryn were incorrectly classified as containing terbutryn only. Table 2 shows the acceptable results for a mixture of 1 ppb atrazine with 1 ppb cyanazine. There were 8 candidate analytes and thus 28 possible binary mixtures. The computer searches through all 28, looking for acceptable solutions. There were three acceptable solutions found, one being the correct identity.

5. Discussion

We have presented a statistic for discriminating between analytes in multivariate calibration and have derived an approximating distribution using small-sigma asymptotics. This statistic incorporates uncertainty in the estimated model parameters. In the nonlinear case, we have assumed a known covariance matrix, which in some applications is not altogether unreasonable since it can be estimated much more precisely than can the individual model parameters. Extension to incorporate the uncertainty in the covariance does not here seem to be as simple as merely substituting an F -distribution for a chi-squared, as is done in the linear case.

Our approach is based on the use of the classical or conditional least squares estimator for the unknown covariate x_0 . Brown and Sundberg (1987) showed that, in the linear case, the minimizer of equation (2) is actually the maximum likelihood estimator, but simulations suggest that this is not a reliable way of getting the maximum likelihood estimation

(MLE) for nonlinear $f(\cdot)$. Clarke (1992) uses a profile likelihood method for the nonlinear case. A full maximum likelihood analysis would be extremely complex with multiple unknowns, especially since the correct model for each is unknown. One could use the solution from our procedure as the starting point of a maximum likelihood analysis, perhaps proceeding sequentially by adding each unknown to the appropriate calibration set to reestimate the parameters, then reapplying our procedure. Given, however, the discrete nature of the decisions being made (choice of analyte), it is easy to imagine such a system failing to converge.

Our approach has been to report all possible identities compatible with the data, i.e., for which R_j^* is probabilistically small. Using this approach, we can find the range of concentrations of each analyte for which reliable identification is possible. For example, from the results of Table 1, we see that our immunoarray can distinguish prometryne at some concentrations but not others. In some applications, it may be necessary to make an unequivocal choice of identity, in which case one would choose the smallest R_j^* , but the error rate will be high in cases where there is more than one plausible candidate. If prior information is available on the likelihood of the different analytes (e.g., application records in the herbicide example) this could easily be incorporated into the analysis. If interval estimates are required for the concentration in addition to the list of plausible analytes, $R^*(x_0)$ of equation (5) could be used to determine a plausible range of concentrations for each analyte. Perhaps here a full Bayesian analysis would be preferable, with priors specified on the model parameters and on x_0 as well as on the candidate analytes themselves. This is one direction of our future research in this area.

ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation (DMS 95-10511) and the National Institute for Environmental Health Science (P42 ES04699). We would like to thank Bruce Hammock and Monika Wortberg for providing ELISA data. We are grateful to two referees and an associate editor for their careful reading of the original manuscript and for their helpful comments.

RÉSUMÉ

Dans des expériences de calibration, une relation estimée entre une information de type covariable pour un échantillon et une réponse observée est utilisée pour inférer l'information covariable pour des échantillons pour lesquelles elle n'est pas connue à partir de leurs réponses. Dans certaines situations, l'information covariable comprend une variable nominale (par exemple, identité d'un corps chimique, sexe d'un animal) et une variable à valeurs réelles (par exemple, concentration du corps chimique, âge de l'animal). Si la relation de calibration peut être estimée séparément pour chaque identité potentielle, la première étape dans l'analyse d'échantillons inconnus est de déterminer correctement leur identité. Une statistique de discrimination est proposée pour traiter ce cas, et sa distribution asymptotique est calculée. Le but de ce travail a été motivé par la possibilité, en suivi de l'environnement, d'utiliser des tests immunologiques multiples pour identifier et quantifier des échantillons contaminés, dans des situations où il existe plusieurs polluants potentiels qui interagissent significativement dans des tests monofactoriels. Un exemple d'utilisation est donné, avec un test utilisant quatre anticorps pour le suivi

simultané du niveau de plusieurs triazines (herbicides) communément utilisées et de leurs dérivés dans des échantillons d'eau.

REFERENCES

- Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd edition. New York: Wiley.
- Brown, P. J. (1982). Multivariate calibration. *Journal of the Royal Statistical Society, Series B* **44**, 287–321.
- Brown, P. J. (1993). *Measurement, Regression and Calibration*. Oxford: Clarendon.
- Brown, P. J. and Sundberg, R. (1987). Confidence and conflict in multivariate calibration. *Journal of the Royal Statistical Society, Series B* **49**, 46–57.
- Brown, P. J. and Sundberg, R. (1989). Prediction diagnostics and updating in multivariate calibration. *Biometrika* **76**, 349–361.
- Clarke, G. P. Y. (1992). Inverse estimates from a multiresponse model. *Biometrics* **48**, 1081–1094.
- Davis, A. W. and Hayakawa, T. (1987). Some distribution theory relating to confidence regions in multivariate calibration. *Annals of the Institute of Statistical Mathematics* **39**, 141–152.
- Fujikoshi, Y. and Nishi, R. (1984). On the distribution of a statistic in multivariate inverse regression analysis. *Hiroshima Mathematical Journal* **14**, 215–225.
- Jones, G. and Rocke, D. M. (1997). *A discrimination statistic in multivariate calibration*. Massey University Information and Mathematical Sciences Reports. Massey University, New Zealand.
- Jones, G., Wortberg, M., Kreissig, S. B., Gee, S. J., Hammock, B. D., and Rocke, D. M. (1994). Extension of the four-parameter logistic model for ELISA to multianalyte analysis. *Journal of Immunological Methods* **177**, 1–7.
- Martens, H. and Naes, T. (1989). *Multivariate Calibration*. New York: Wiley.
- Oman, S. D. (1999). Multivariate calibration. In *Multivariate Analysis, Design of Experiments, and Survey Sampling*, S. Ghosh (ed). New York: Marcel Dekker.
- Rodbard, D. (1981). Mathematics and statistics of ligand assay: An illustrated guide. In *Ligand Assay: Analysis of International Developments on Isotopic and Nonisotopic Immunoassay*, J. Langan and J. J. Clapp (eds). New York: Masson.
- Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*. New York: Wiley.
- Williams, E. J. (1959). *Regression Analysis*. New York: Wiley.

Received October 1999. Revised August 2000.
Accepted September 2000.