A PROCEDURE FOR ESTIMATING PARAMETERS IN NONLINEAR RANDOM MODELS

Paulette Johnson, Texas A&M University
George Milliken, Kansas State University

ABSTRACT

Consider an experiment where a nonlinear continuous function has been fitted to the data from several treatments or populations where the function depends upon a random variable with unknown mean and unknown covariance matrix. A two-step procedure, based on asymptotic theory, is developed to obtain estimates of the elements of the covariance matrix. The procedure is then applied to a nonlinear one-way classification model and an example is presented.

1. INTRODUCTION

To describe various economic, demographic, or physiometric relationships where the coefficients of the model are affected by unobserved variables, the following linear model is frequently used,

\[ X = \beta_0 + \varepsilon \quad \text{where} \quad E(\varepsilon) = 0, \quad E(\varepsilon) = 0, \quad V(\varepsilon) = \sigma^2. \]  

(1.1)

If the columns of \( X \) are measured independent variables, model (1.1) is a random coefficient model. If, instead, \( X \) relates to an experimental design such as a one-way or nested classification, model (1.1) is a variance component model. For both of these cases, there are well-known methods of estimating \( \beta_0 \) and \( \sigma^2 \) in the linear case. However, there are many situations where the relationship between the observed random variable, \( y \), and \( \beta \) is not linear but is better described by some nonlinear function.

This paper deals with the problem of estimating \( \beta_0 \) and \( \sigma^2 \) for the nonlinear model

\[ y = f(X, \beta) + \varepsilon \quad \text{where} \quad E(\varepsilon) = 0, \quad E(\varepsilon) = 0, \quad V(\varepsilon) = \sigma^2. \]  

(1.2)

considered as either a random coefficient or variance component model.

As an economic example, consider a Cobb-Douglas production equation with additive error structure

\[ y = \alpha L^a K^b + \varepsilon \]  

(1.3)

where \( y \) is real output, \( L \) is labor input, \( K \) is capital input, \( \alpha \) is an unknown parameter, \( a \) and \( b \) are elasticity coefficients, and \( \varepsilon \) is a random disturbance with zero mean and variance \( \sigma^2 \). To study the productive relationship in a number of different market situations, suppose \( t \) types of firms are chosen at random from a population of firms. Each type of firm is then examined in \( n \) randomly selected market situations. The Cobb-Douglas model is

\[ y_{ij} = \alpha L^a K^b + \varepsilon_{ij} \]  

(1.4)

where \( E(\varepsilon_{ij}) = 0, \quad E(\varepsilon_{ij}) = 0, \quad V(\varepsilon_{ij}) = \sigma^2, \quad V(\varepsilon_{ij}) = \sigma^2, \quad \text{and} \quad V(\varepsilon_{ij}) = \sigma^2. \)

As a physiometric example, consider the enzyme kinetic model of the initial velocity, \( y \), of a reaction which obeys the Michaelis-Menten equation

\[ y = \frac{sX}{V_{max}} = \sigma + \varepsilon \]  

(1.5)

where \( X \) is the concentration of the substrate which reacts with the enzyme, \( \theta \) is the maximum velocity which is theoretically attained when the enzyme is saturated by an infinite concentration of substrate, \( \theta \) is the Michaelis-Menten constant which is the amount of substrate needed to attain half the maximum initial velocity, and \( \varepsilon \) is a random error with mean zero and variance \( \sigma^2 \). To determine the distribution in a given population of animals for a particular enzyme, a random sample of \( t \) subjects is chosen, and for each subject the initial velocity, \( y \), is found for \( n \) levels of substrate. The model is

\[ y_{ij} = \frac{sX_{ij}}{V_{max}} = \sigma + \varepsilon_{ij} \]  

(1.6)

where \( E(\varepsilon_{ij}) = 0, \quad E(\varepsilon_{ij}) = 0, \quad V(\varepsilon_{ij}) = \sigma^2, \quad V(\varepsilon_{ij}) = \sigma^2, \quad \text{and} \quad V(\varepsilon_{ij}) = \sigma^2. \)

Finding estimates of \( \beta_0 \) and \( \sigma^2 \) will characterize the distributions associated with the Michaelis-Menten constant for a given population.

The procedure developed herein estimates the variance components of \( \beta \) in the nonlinear model using two steps. First, a nonlinear least squares procedure is used to obtain an estimate of \( \sigma^2 \) and its asymptotic covariance matrix for each treatment or population. Second, a linear model with random coefficients is constructed from these estimates. Then, any of several variance component estimation procedures, such as maximum likelihood, analysis of variance, or MINQUE, is used to estimate the variance component.

2. PROCEDURE

First consider the simple case of a nonlinear model as a function of one random variable defined by

\[ y_{ij} = f(X_{ij}, \theta_j) + \varepsilon_{ij} \]  

(2.1)
where \( y_{ij} \) is the \( j \)th observation of the \( i \)th treatment of a known, nonlinear, continuous function satisfying certain regularity conditions (Jennrich (1969)), and \( \eta_{ij} \) is an \( m \times 1 \) vector of known constants \( (\eta_1) \). Assume that the treatments are a random sample from a larger population of treatments, then \( \theta_i \) is an unknown random variable with mean \( \theta_0 \) and variance \( \sigma^2_0 \). Also assume that the \( \epsilon_{ij} \) are independent, identically distributed random variables with mean zero and variance \( \sigma^2 \), and are independently distributed of \( \theta_i \).

In the first step, a least squares estimate of \( \theta_i \) is obtained from the 1st population by application of a nonlinear procedure such as Marquardt's (1963) or Hartley's (1961) modified Gauss-Newton method to the ith model defined in (2.1). Assuming Jennrich's regularity conditions (which include the existence and continuity of the first and second order derivatives, \( f' \) and \( f'' \)), the conditional distribution of the estimator \( \hat{\theta}_i \) given \( \theta_i \) (that is, given that the ith treatment was observed) satisfies

\[
\sqrt{n_i}(\hat{\theta}_i - \theta_0) \sim N(0, \sigma^2 C(\theta_0)) \tag{2.2}
\]

where \( C(\theta_0) = \lim_{N \to \infty} C_n(\theta_0) \) and \( C_n(\theta_0) = \frac{1}{n_i} \left[ \sum_{i=1}^{n_i} \frac{2f(\eta_{ij})}{\eta_{ij}} \right]^{-1} \). Next, obtain the variance of the unconditional distribution of \( \hat{\theta}_i \) as

\[
\text{Var}(\hat{\theta}_i) = E[(\hat{\theta}_i - \theta_0)^2] = E(\hat{\theta}_i - \theta_0)^2 = E_0^{1/n_i} E_n(\hat{\theta}_i - \theta_i)^2 + \sigma^2 \delta_i. \tag{2.3}
\]

Assuming \( n \) is large enough for the asymptotic results to hold, the conditional variance of \( \hat{\theta}_i \) given \( \theta_i \) is \( \delta_i = E_0(\hat{\theta}_i - \theta_i)^2 = n_i \sigma^2 C(\theta_0) \). So the unconditional variance of \( \hat{\theta}_i \) becomes

\[
\text{Var}(\hat{\theta}_i) = n_i \sigma^2 C_n(\theta_i) + \sigma^2 \delta_i \tag{2.4}
\]

Jennrich (1969) states that the least squares estimator \( \hat{\theta}_i - \theta_0 \) almost surely as \( n \to \infty \) and that \( C_n(\theta_i) \) converges uniformly to \( C(\theta) \). The \( C_n(\theta_i) \) are integrable on the compact set \( \theta \) where \( \hat{\theta}_i \) is defined, so that (Rudin (1976)) \( E_0[C_n(\hat{\theta}_1)] \to E_0[C(\theta_0)] \) almost surely as \( n \to \infty \).

The second step involves constructing a linear model for \( \hat{\theta}_i \) from the unconditional distribution of \( \theta_i \). The linear model is

\[
\delta_i = \theta_0 + \epsilon^*_i = \theta_1, \ldots, \theta_t \tag{2.5}
\]

where the \( \epsilon^*_i \) are independent unobservable random variables with mean zero and variance \( n_i \sigma^2 C(\theta_i) \). The residual sum of squares from fitting model (2.1) to the data from each of the \( t \) treatments are pooled to provide the estimate of \( \sigma^2 \). Any of several variance component estimation techniques can be used to estimate \( \sigma^2_1 \) and \( \sigma^2_2 \) by using model (2.5) and the pooled residual mean square.

In general, the nonlinear model is a function of a \( p \times 1 \) vector random variable \( \eta_{ij} \), \( i=1, \ldots, t \) which is assumed to have mean \( \theta_0 \) and covariance matrix \( V \). The unconditional covariance matrix, analogous to (2.1), is

\[
\text{Var}(\hat{\theta}_i) = \sigma^2 V(\hat{\theta}_i) + D \tag{2.6}
\]

where \( \hat{\theta}_i = (\hat{\theta}_1, \ldots, \hat{\theta}_t), \eta^* = (\eta_1, \ldots, \eta_t) \)

\[
U = \begin{bmatrix}
E_n[C_n(\theta_2)] & \cdots & 0 \\
0 & \ddots & \vdots \\
\vdots & \ddots & \ddots \\
0 & \cdots & n_t E_n[C_n(\theta_t)]
\end{bmatrix},
\]

\[
D = \begin{bmatrix}
0 & \cdots & 0 \\
0 & V & \cdots \\
\vdots & \ddots & \ddots \\
0 & \cdots & V
\end{bmatrix},
\]

and \( V^{-1}(\hat{\theta}_i) \) is a \( p \times p \) matrix with elements

\[
n_i j \delta_i - \frac{\partial}{\partial \eta_{ij}} C_n(\hat{\theta}_i) + \frac{\partial}{\partial \eta_{ij}} C_n(\hat{\theta}_i)
\]

In order to estimate a particular variance component, say \( \sigma^2_i \), where \( 1 \leq \tau \leq p \), use the information from the unconditional distribution for each \( \eta_{ij} \) to construct a linear model similar to (2.5). The linear model is

\[
\hat{\delta}_i = \theta_0 + \epsilon^*_i = \theta_1, \ldots, \theta_t \tag{2.7}
\]

where the \( \epsilon^*_i \) are independent random variables with mean zero and variance \( n_j \sigma^2 E_n[C_n(\theta_0)] + \sigma^2_\tau \) where \( C_n(\theta_0) \) is the \( \tau \)th diagonal element of \( C_n(\theta_0) \) is the \( \tau \)th diagonal element of \( C_n(\theta_0) \).

If \( \eta_{ij} \) is assumed to be normally distributed, then the method of maximum likelihood could be used to estimate the parameters. Under the regularity conditions (Jennrich (1969)), the unconditional asymptotic distribution of \( \hat{\theta}_i \) is \( N(\theta_0, \sigma^2 C_n(\eta_{ij})) \), \( i=1, \ldots, t \).

3. ONE-WAY CLASSIFICATION MODEL

To illustrate this procedure for a particular case, consider the model

\[
y_{ij} = f(\alpha_{ij}, \beta_{ij}, x_{ij}) + \epsilon_{ij}
\]

\( i=1, \ldots, t; j=1, \ldots, n_i \) \tag{3.1}

where \( f(\cdot) \) satisfies Jennrich's regularity conditions, \( n_i \) is large, and \( E(\alpha_{ij}) = \alpha_0, V(\alpha_{ij}) = \sigma^2_0 \),
\( E(\delta_t) = \delta_t \) and \( V(\delta_t) = \delta_t^2 \) for all \( i, j \). First, a nonlinear least square procedure applied to each of the \( t \) samples yields \( \hat{\delta}_t = (\hat{\delta}_1, \hat{\delta}_1, \ldots, \hat{\delta}_t) \), and \( C_n(\hat{\delta}_t) \). The estimate of the variance component \( \sigma^2 \), is obtained by pooling \( \hat{\sigma}_1^2, \ldots, \hat{\sigma}_t^2 \).

If interest centers on estimating \( \sigma^2 \), a linear model based on the \( a_i \), \( i=1, \ldots, t \), is constructed

\[
\hat{a}_i = a_0 + \hat{e}_i^2 \quad i=1,\ldots,t \tag{3.2}
\]

where \( V(\hat{e}_i^2) = n_i \sigma^2 E[C_n(a_i)] + \sigma^2 \).

To use the analysis of variance technique to estimate \( \sigma^2 \), compute the sample variance of the \( a_i \)'s and evaluate its expectation. The sample variance is

\[
\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (\hat{a}_i^2 - \bar{a}^2)/(t - 1) \tag{3.3}
\]

where \( \bar{a} = \frac{1}{n} \sum_{i=1}^n \hat{a}_i \). The expectation of \( \hat{\sigma}_n^2 \) is

\[
E[\hat{\sigma}_n^2] = \sigma^2 \left[ \frac{1}{t} \sum_{i=1}^n n_i E[C_n(a_i)] \right] + \sigma^2. \tag{3.4}
\]

Substitute \( \sigma^2 \) for \( \sigma^2 \), \( \sigma^2 \) for \( \sigma^2 \), \( E[C_n(a_i)] \) for \( E[C_n(a_i)] \) and \( s_i^2 \) for \( E[s_i^2] \) in (3.4). Solving (3.4) for \( \sigma^2 \) provides

\[
\sigma^2 = \hat{\sigma}_n^2 - \sigma^2 \left[ \frac{1}{t} \sum_{i=1}^n n_i E[C_n(a_i)] \right]. \tag{3.5}
\]

The best linear unbiased estimate (BLUE) of \( \sigma^2 \) from (3.2) is

\[
\hat{\sigma}_0 = \left( \frac{1}{t} \sum_{i=1}^n \frac{1}{V(\hat{e}_i^2)} \right)^{-1} \left( \frac{1}{t} \sum_{i=1}^n \frac{\hat{a}_i^2}{V(\hat{e}_i^2)} \right) \tag{3.6}
\]

where \( V(\hat{e}_i^2) \) is defined in (3.2). The value of \( \hat{\sigma}_0 \) is not computable since \( V(\hat{e}_i^2) \) depends on the unknown quantities \( \sigma^2 \), \( \sigma^2 \) and \( E[C_n(a_i)] \). To get the estimate of the BLUE of \( \sigma^2 \), estimate the \( V(\hat{e}_i^2) \) by

\[
\hat{V}(\hat{e}_i^2) = n_i \hat{\sigma}_n^2 C_n(a_i) + \hat{\sigma}_n^2 \tag{3.7}
\]

and compute the estimate of the BLUE as

\[
\hat{\sigma}_0 = \left( \frac{1}{t} \sum_{i=1}^n \frac{1}{\hat{V}(\hat{e}_i^2)} \right)^{-1} \left( \frac{1}{t} \sum_{i=1}^n \frac{\hat{a}_i^2}{\hat{V}(\hat{e}_i^2)} \right). \tag{3.8}
\]

Similarly, if interest centers on \( \sigma^2 \), equations like (3.3), (3.4), and (3.8) can be obtained by substituting \( \hat{\delta}_1, \hat{\delta}_2, \hat{\delta}_3 \) and \( C_n(\hat{\delta}_i) \) for \( a_1, \sigma^2, s_i^2 \) and \( C_n(a_i) \) respectively.

4. EXAMPLE OF THE ONE-WAY CLASSIFICATION MODEL

The Michaelis-Menten model for the enzyme glucose-6-phosphate was fit to 8 randomly selected sheep, using data similar to that obtained in Engeling, et al., (1974). Initial velocities were determined in duplicate (which were averaged) for six substrate concentrations for each sheep. The Michaelis-Menten constant, \( \nu_1 \), in the model

\[
Y_{ij} = \frac{\delta_{ij} X}{\nu_{ij} + \delta_{ij}} + \epsilon_{ij} \quad i=1,\ldots,8; \, j=1,\ldots,6 \tag{4.1}
\]

was estimated using the modified Gauss-Newton method of SAS NLIN. Table 1 contains the animal number, \( i \), the residual sum of squares for each animal based on four degrees of freedom and the \( n_i C_n(\hat{\nu}_1) \).

From equation (3.3), \( \hat{\sigma}_n^2 = 4.587 \) and from the pooled residual sums of squares \( \hat{\sigma}_n^2 = 11.215 \).

Then

\[
\frac{\hat{\sigma}_n^2}{t} = \frac{0.5056}{8} = 0.0632.
\]

To compute the estimate of the BLUE of \( \nu_0 \), first compute \( V(\hat{e}_i^2) \) for each animal (also in Table 1) and then compute \( \nu_0 \) from (3.8) as \( \nu_0 = [1.748882]^{-1}[43.2868] = 24.75 \).

ACKNOWLEDGEMENT

This is part of Paulette Johnson's dissertation presented in partial fulfillment of the requirements for the Ph.D. degree in Statistics at Kansas State University.

TABLE 1

<table>
<thead>
<tr>
<th>Animal</th>
<th>( \hat{\nu}_1 )</th>
<th>( n_i C_n(\hat{\nu}_1) )</th>
<th>RSS_i ( \hat{V}(\hat{e}_i^2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22.69</td>
<td>0.07847</td>
<td>33.75</td>
</tr>
<tr>
<td>2</td>
<td>27.78</td>
<td>0.04478</td>
<td>45.03</td>
</tr>
<tr>
<td>3</td>
<td>23.05</td>
<td>0.07189</td>
<td>72.36</td>
</tr>
<tr>
<td>4</td>
<td>27.78</td>
<td>0.05087</td>
<td>81.21</td>
</tr>
<tr>
<td>5</td>
<td>22.70</td>
<td>0.05758</td>
<td>31.68</td>
</tr>
<tr>
<td>6</td>
<td>24.34</td>
<td>0.09846</td>
<td>13.23</td>
</tr>
<tr>
<td>7</td>
<td>25.79</td>
<td>0.07719</td>
<td>20.37</td>
</tr>
<tr>
<td>8</td>
<td>23.67</td>
<td>0.02641</td>
<td>61.26</td>
</tr>
</tbody>
</table>

BIBLIOGRAPHY


