한국통계학회논문집 제2권 1호
1995년 4월 pp. 48-54

On the Fitting ANOVA Models to Unbalanced Data

Jong-Tae Park, Jae-Heon Lee, Byung-Chun Kim[1)]

Abstract

A direct method for fitting analysis-of-variance models to unbalanced data is presented. This method exploits sparsity and rank deficiency of the matrix and is based on Gram-Schmidt orthogonalization of a set of sparse columns of the model matrix. The computational algorithm of the sum of squares for testing estimable hyphotheses is given.

1. Introduction

The computation involved in fitting an analysis-of-variance(ANOVA) model to unbalanced data is an explicit solution of a least-squares problem

$$\min_b \|y - Xb\|_2, \tag{1.1}$$

where y is the response vector, X is the model matrix (often called the design matrix), b is a solution vector, and $\|\cdot\|_2$ is the Euclidean norm. Kennedy and Gentle(1980) discussed some computational methods for this problem as well as the computation of associated sums of squares. The available direct methods either employ an orthogonal factorization of the model matrix X or solve the normal equations by factoring or sweeping X^TX, where T denotes the transpose. Generally, methods utilizing the normal equations require less computation; however, they are less numerically stable and some precision is lost by the formation of X^TX. Although numerical stability considerations are usually not of concern in small ANOVA problems, they may become important in large problems.

For models involving several factors at several levels each, the model matrix X is large, so current direct methods require large amounts of time and storage. The use of iterative methods has been shown to reduce the time and storage requirements by Golub and Nash(1982), Hemmerle(1974), Kim, Marasinghe and Kennedy(1984), Jamshidian and Jennrich(1988). The aim in this paper is to develop a fast and storage-efficient direct algorithm by exploiting the sparsity and rank deficiency of the model matrix. Sums of squares for model terms are obtained via the $R(\ )$-notation in Speed and Hocking(1976) by fast

1) Department of Mathematics, KAIST, Gusung-dong, Yusung-gu, Taejon, 305-701, KOREA

fitting of appropriate models. Fast computation of sums of squares for special parameterizations of marginal effects in the presence of higher-order interactions or other estimable hypotheses of low rank is then facilitated by the availability of the sparse factored model matrix after fitting a model.

The earliest references to the idea of exploiting sparsity of an ANOVA problem seem to have been made by Gentleman(1973, 1975), who discussed orthogonal given transformations for solving general sparse least-squares problems and noted their application to matrices arising in analysis-of-variance. Recently, Fellner(1987) discusses the use of sparse matrix methodology in estimation of variance components by likelihood methods. Givens rotations have been shown to be particularly suitable for the sparse least-squares problem by Gentleman(1973, 1975) and Heath(1984). Aside from the usual stability and precision advantages of operating directly on the X matrix, the matrix is more sparse than X^TX in large models. The model matrix is also highly rank deficient and this can be exploited to gain more sparsity. George and Heath(1980) have developed an algorithm for large sparse least-squares problems that uses orthogonal Givens rotations to process the matrix one row at a time. The method was later extended in Heath(1982) to rank-deficient problems. However, general sparse methods do not exploit the rank deficiency of a model matrix. Further substantial improvements can be made by recognizing the rank deficiency and special structure of a ANOVA model matrix. In this paper we develop a new method for the ANOVA problem that exploits the special structure and rank deficiency of the model matrix and is driven only by the model specification. In section 2, model specification and notation are used to describe the model matrix. In section 3, a procedure for discarding columns to obtain a full rank problem is discussed, and in section 4, testing of estimable hypotheses are discussed.

2. Model Specification and Notation

The usual model specification for a k-factor experiment with no interactions is

$$y = 1_N \mu + X_1 \beta_1 + \cdots + X_k \beta_k + \varepsilon, \tag{2.1}$$

where each factor i has m_i levels, and

y : an $N \times 1$ vector of observations

1_N : an N-dimensional column vector of ones

X_i : an $N \times m_i$ incidence matrix of i-th factor

β_i : an $m_i \times 1$ vector of parameters of i-th factor

ε : an $N \times 1$ vector of random errors

N : the total number of observations.

Let m be the total number of cells of the design, that is, $m = \prod_{i=1}^{k} m_i$, and $n_j (>0)$ be the number of observations in j-th cell. The model matrix X arising from a given model (2.1) is an $N \times p$ sparse matrix containing only zeros and ones, where $N = \sum_{j=1}^{m} n_j$ and $p = \sum_{i=0}^{k} m_i$, $m_0 = 1$. This matrix can be written as a product $X = TZ$, where T is a replication matrix and Z is a model matrix with one observation per cell. The matrix Z has m rows, where each row represents a distinct cell of the design. The matrix T has n rows and each row is a unit vector of dimension m that simply defines the cell sampled in that observation, and $T^T T$ is a diagonal matrix with diagonal entries n_j.

By fitting cell means, any least-squares ANOVA problem with one observation per cell. We illustrate this with the assumption of no missing cells. There exists an orthogonal matrix U such that $U^T T = \begin{bmatrix} D^{1/2} \\ O \end{bmatrix}$ and $U^T y = \begin{bmatrix} d \\ e \end{bmatrix}$, where $D^{1/2}$ is a diagonal matrix of order m with diagonal entries $\sqrt{n_j}$ giving the weights. If in addition we let $c = D^{-1/2} d$, the square of (1.1) becomes

$$\min_b \| D[\ c - Zb] \|_2^2 + \|e\|_2^2, \tag{2.2}$$

where the elements of c are the cell means, $c = D^{-1} T^T y$. The quantity $\|e\|_2^2$ is the residual sum of squares after fitting the cell means.

In view of the above discussion, the nonzero structure of problem (1.1) can be interpreted in terms of the matrix Z and the weighted problem (2.2).

3. Model Matrix Column Structure

Let $r (< p)$ be the rank of Z and P be a permutation matrix so that the first r columns of ZP are linearly independent. Partition ZP into $[\ Z_1 \mid Z_2]$ such that Z_1 has full column

rank. The solution of (2.2) via Gram-Schmidt orthogonalization computes a factorization of $D^{1/2}Z_1$ that can be partitioned as

$$\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} D^{1/2}Z_1 = \begin{bmatrix} R \\ O \end{bmatrix}, \tag{3.1}$$

where $R = V_1^T D^{1/2} Z_1$ is upper triangular of order r, and $V = [\ V_1 \mid V_2]$ is a nonsingular matrix whose columns are orthogonal. A solution to (1.1) is a vector b given by

$$P^T b = \begin{bmatrix} R^{-1} V_1^T d \\ O \end{bmatrix} \tag{3.2}$$

and the residual sum of squares is given by

$$\left\| \begin{bmatrix} (I - V_1 D_1^{-1} V_1^T) d \\ e \end{bmatrix} \right\|_2^2, \tag{3.3}$$

where $D_1 = V_1^T V_1$ is a diagonal matrix.

The solution chosen in (3.2) is that of a parametrization where all parameters associated with columns of Z_2 are set to zero. Expressions (3.2) and (3.3) show that the calculation of a solution and the residual sum of squares requires only a set of r linearly independent columns of Z rather than the entire matrix. The matrix Z can be partitioned as $[\ Z(\mu) \mid Z(\beta_1) \mid \cdots \mid Z(\beta_k)]$ according to the model terms. Then the submatrix can be expressed using Kronecker products of identity matrix and of matrices with all entries equal to 1:

$$\begin{aligned} Z(\mu) &= 1_m \\ Z(\beta_i) &= 1_{m_1} \otimes \cdots \otimes I_{m_i} \otimes \cdots \otimes 1_{m_k}, \quad i = 1, \cdots k. \end{aligned} \tag{3.4}$$

A standard procedure for discarding columns to obtain a full rank problem is to drop any column associated with level one of any factor. This has the desired effect of leaving only a set of linearly independent columns; however, it is not the best set in terms of sparsity. Our proposed procedure for discarding columns to obtain better sparsity is to drop the column associated with the last level of the others except only one factor with the largest number of

levels. By the above procedure the partitioned matrix Z_1 can be described as $Z_1=[\ G_1 \mid G_2 \mid \cdots \mid G_k]$, where $G_1=Z(\beta_i)$, the index i is the one so that m_i is the largest number of levels, and the remaining submatrices are the matrices containing only the first m_j-1 columns of $Z(\beta_j)$ for all $j\neq i$.

Partition $D^{1/2}Z_1$ into $D^{1/2}Z_1=[\ W_1 \mid W_2 \mid \cdots \mid W_k]$, where $W_i=D^{1/2}G_i,\quad i=1,\cdots,k$. Since all the columns in each W_i are orthogonal, we can construct the matrix V_1 satisfying $V_1^TD^{1/2}Z_1=R$. Let V_1 be partitioned as $[\ \widehat{W_1} \mid \widehat{W_2} \mid \cdots \mid \widehat{W_k}]$. Given a model (2.1), Algorithm 1 computes the columns in V_1 by Gram-Schmidt orthogonalization.

Algorithm 1 (Orthogonal Columns)

1. $\widehat{W_1} \leftarrow W_1$.
2. For $i=2$ to k, compute $\widehat{W_i}$ from $\widehat{W_1}, \widehat{W_2}, \cdots, \widehat{W_{i-1}}, W_i$.

This algorithm shows that the calculation of orthogonal columns required is for only a set of $r-\max_i m_i$ linearly independent columns of V_1.

4. Testing of Estimable Hypothesis

In an F-test, where the denominator is obtained from a single fit, cancellation in the numerator will lead to a nonsignificant result, as if should. However, in cases where the denominator is a difference of two fits, cancellation can produce serious roundoff errors. When cancellation in the denominator occurs, we recommend the procedure below for computing the required sums of squares. This problem can occur in testing marginal variance components. In this case, the sums of squares produced by the $R(\ )$-notation may not be the ones of interest anyway and the procedure below is needed to produce the required sums of squares.

To compute sums of squares for special estimable functions or specific parameterizations of marginal model terms, relatively little additional computation need to be done since the sparse R factor is available after fitting a model. Suppose we wish to compute the sum of squares for testing $H\beta=d$, where H is a $q\times p$ matrix of rank $q(<<p)$ that is in the rowspace of X, β is the vector of model parameters, and d is a $q\times 1$ vector. This is given in Kennedy and Gentle(1980) as

$$(Hb-d)^T[\ H(X^TX)^-H^T]^{-1}(Hb-d). \tag{4.1}$$

Since the solution vector in (3.1) contains only r nonzeros, let H_1 be the columns of H corresponding to these nonzeros and let b_1 be these nonzeros. That is, partition $HP=[\ H_1 \mid H_2]$ and $b^TP=[\ b_1^T \mid 0]$, where P is an appropriate permutation matrix. Expression (4.1) can be computed with the following algorithm.

Algorithm 2 (Sum of Squares due to Hyphothesis)

1. Compute $h=H_1b_1-d$.
2. Backsolve $R_1^TM=H_1^T$, where $R_1^T=R^TD_1^{-1/2}$.
3. Factor $M=U\begin{bmatrix} V \\ O \end{bmatrix}$.
4. Backsolve $V^Ts=h$.
5. Compute sum of squares s^Ts.

Because all computations can be done in place, the only additional storage required is for the $q\times r$ matrix H_1 and the $q\times 1$ vector d. The most expensive step of this algorithm is the orthogonal factorization of step 3 of a $q\times r$ matrix. Generally $q\ll r$, so the amount of additional time and storage is small.

In general, it is not easy to write down an estimable H, particularly for a large model. It is also not possible to anticipate all estimable hypotheses of interest and generate them automatically. However, for the case of no missing cells, the hypotheses that are normally tested in the balanced case could be generated automatically. More work is needed for the unbalanced case, particularly on the use of a set of linearly independent columns for the automatic generation of such hyphotheses.

References

[1] Feller, W.H. (1987). Sparse matrices and the estimation of variance components by likelihood methods, *Commnications in Statistics, Simulation*, Vol. 16, 439-463.

[2] Gentleman, W.M. (1973). Least squares computations by Givens transformations without squre roots, *Journal of institute of Mathematical Applications*, Vol. 12, 329-336.

[3] Gentleman, W.M. (1975). Error analysis of QR decompositions by Givens transformations, *Linear Algebra and its Applications*, Vol. 10, 189-197.

[4] George, J.A. and Heath, M.T. (1980). Solution of sparse least squares problems using Givens rotations, *Linear Algebra and its Applications*, Vol. 34, 69-83.

[5] Golub, G.H. and Nash, S.G. (1982). Nonorthogonal analysis of variance using a generalized

congugate-gradient algorithm, *Journal of the American Statistical Association*, Vol. 77, 109-116.

[6] Heath, M.T. (1982). Some extensions of an algorithm for sparse linear least squares problems, *SIAM Journal of Science and Statistical Computing*, Vol. 3, 223-237.

[7] Heath, M.T. (1984). Numerical methods for large sparse linear least squares problems, *SIAM Journal of Science and Statistical Computing*, Vol. 5, 497-513.

[8] Hemmerle, W.J. (1974). Algebraic specification of statistical models for analysis of variance computations, *Journal of ACM*, Vol. 11, 234-241.

[9] Jamshidian, M. and Jennrich, R.I. (1988). Nonorthogonal analysis variance using gradient methods, *Journal of the American Statistical Association*, Vol. 83, 483-489.

[10] Kennedy, W.J., JR. and Gentle, J.E. (1980). *Statistical Computing*, Marcel Dekker, New York.

[11] Kim, B.C., Marasinghe, M.G. and Kennedy, W.J., JR. (1984). A new congugate gradient algorithm for analysis of variance computations, *Proceeding of Statistical Computing Section, American Statistical Association*, 150-154.

[12] Speed, F.M. and Hocking, R.R. (1976). The use of R()-notation with unbalanced data, *American Statistician*, Vol. 30, 30-33.