AN AUGMENTATION METHOD FOR MULTILEVEL ANALYSIS

JAN DE LEEUW

Abstract. Dodo

1. Introduction

2. Model

The micro-model is

$$\underline{y}_j = X_j \underline{b}_j + \underline{\epsilon}_j, \tag{1a}$$

and the macro-model is

$$\underline{b}_j = Z_j \gamma + \underline{\delta}_j, \tag{1b}$$

which implies

$$\underline{y}_j = X_j Z_j \gamma + X_j \underline{\delta}_j + \underline{\epsilon}_j. \tag{2}$$

Here $\underline{y}_j$ has n_j elements, X_j is $n_j \times p$, and Z_j is $p \times q$. We suppose, in addition, that the disturbances $(\underline{\epsilon}_j, \underline{\delta}_j)$ are i.i.d. and satisfy

$$\begin{bmatrix} \underline{\epsilon}_j \\ \underline{\delta}_j \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 \mathcal{I} & 0 \\ 0 & \Omega \end{bmatrix} \right). \tag{3}$$

It follows that

$$\underline{y}_j \sim \mathcal{N}(X_j Z_j \gamma, X_j \Omega X_j' + \sigma^2 \mathcal{I}), \tag{4}$$

and the $\underline{y}_j$ are independent.

Date: December 31, 2005.

3. Likelihood

The negative log-likelihood, except for irrelevant constants, is

$$\mathcal{L} = \sum_{j=1}^{m} \log |X_j \Omega X_j' + \sigma^2 \mathcal{I}| + \sum_{j=1}^{m} (y_j - X_j Z_j \gamma)' [X_j \Omega X_j' + \sigma^2 \mathcal{I}]^{-1} (y_j - X_j Z_j \gamma). \tag{5}$$

This can be simplified considerably.

First, we write y_j in the form

$$y_j = X_j b_j + r_j, \tag{6}$$

where $r_j' X_j = 0$. This is always possible by taking b_j equal to a set of least-squares regression coefficients in the regression of y_j on X_j. If X_j is of full column rank p, then this decomposition is actually unique. Let us assume full rank for the time being.

Second, we can also write

$$X_j \Omega X_j' + \sigma^2 \mathcal{I} = X_j (\Omega + \sigma^2 (X_j' X_j)^{-1}) X_j' + \sigma^2 Q_j, \tag{7}$$

with $Q_j = \mathcal{I} - X_j (X_j' X_j)^{-1} X_j'$.

It follows from (6) and (7) that

$$(y_j - X_j Z_j \gamma)' [X_j \Omega X_j' + \sigma^2 \mathcal{I}]^{-1} (y_j - X_j Z_j \gamma) = (b_j - Z_j \gamma)' (\Omega + \sigma^2 (X_j' X_j)^{-1})^{-1} (b_j - Z_j \gamma) + \sigma^{-2} r_j' r_j, \tag{8a}$$

and

$$\log |X_j \Omega X_j' + \sigma^2 \mathcal{I}| = \log |X_j' X_j| + \log |\Omega + \sigma^2 (X_j' X_j)^{-1}| + (n_j - p) \log \sigma^2. \tag{8b}$$

Letting

$$W_j \triangleq \Omega + \sigma^2 (X_j' X_j)^{-1} \tag{9}$$

and

$$s_j^2 \triangleq \frac{1}{n_j - p} r_j' r_j \tag{10}$$

we see that

$$
\begin{aligned}
\text{(11)}\quad \mathcal{L} = \sum_{j=1}^{m} \log|X_j'X_j| + \\
\sum_{j=1}^{m}\{\log|W_j| + (b_j - Z_j\gamma)'W_j^{-1}(b_j - Z_j\gamma)\} + \\
\sum_{j=1}^{m}(n_j - p)\{\log\sigma^2 + \frac{s_j^2}{\sigma^2}\}
\end{aligned}
$$

4. The Balanced Case

The multilevel design is balanced if the n_j are the same for all j and if the $X_j'X_j$ are the same for all j as well. Balanced designs occur typically in repeated measures or growth curve experiments, in which there are no missing data. Then the columns of the X_j are simple functions (for instance monomials) of time, and thus if the time-points for all individuals are the same, the X_j are the same.

Suppose n is the common value of the n_j, $X'X$ is the common value of $X_j'X_j$, and W is the common value of the W_j. It becomes easy to give simple expressions for the maximum likelihood estimates of the parameters.

$$
\hat{\sigma}^2 = \frac{1}{m}\sum_{j=1}^{m} s_j^2, \tag{12a}
$$

$$
\hat{W} = \frac{1}{m}\sum_{j=1}^{m}(b_j - Z_j\hat{\gamma})(b_j - Z_j\hat{\gamma})', \tag{12b}
$$

$$
\hat{\gamma} = (\sum_{j=1}^{m} Z_j'\hat{W}^{-1}Z_j)^{-1}\sum_{j=1}^{m} Z_j'\hat{W}^{-1}b_j. \tag{12c}
$$

The maximum likelihood estimate of Ω is then simply

$$
\hat{\Omega} = \hat{W} - \hat{\sigma}^2 X'X. \tag{12d}
$$

Observe that it is possible that $\hat{\Omega}$ is not positive semi-definite. We could, of course, impose restrictions to exclude this unhappy event, we could also take it as an important diagnostic, because it indicates failure of the model.

The equations (12) can be solved in a very straightforward way by *block relaxation*. We start with an estimate of γ, then use (12b) to find the optimal W for this γ, then use (12c) to find a new γ, optimal for the current W, and so on. The computations in each step are very simple, and the interpretation of the computations is simple as well. $\hat{\sigma}^2$ is just the variance of the first order residuals, $\hat{W}$ is the variance-covariance matrix of the second order residuals, and $\hat{\gamma}$ is the two step estimator of the fixed regression coefficients, suitably weighted.

5. Dealing with Imbalance

If the X_j are different, then the W_j are different, and solving the likelihood equations is no longer simple. But we can use a classical trick, sometimes called *augmentation*, to balance the data. Suppose C is a matrix of order p, satisfying $C \succcurlyeq X_j'X_j$ in the Loewner sense for all j. Thus $C - X_j'X_j$ is positive semi-definite. This means that we can find a $p \times p$ matrix H_j such that $H_j'H_j = C - X_j'X_j$. Augmentation works by adding the p rows of H_j as additional pseudo-observations to X_j. Clearly the augmented X_j, say $\tilde{X}_j$, now satisfies $\tilde{X}_j'\tilde{X}_j = C$.

Because we have augmented X_j we also need to augment y_j. The p additional elements of y_j, let's call them g_j, just become additional parameters we have to minimize over. We just use

$$\mathcal{L} = \sum_{j=1}^m \log|X_j\Omega X_j' + \sigma^2\mathcal{I}| + \sum_{j=1}^m (\tilde{y}_j - \tilde{X}_jZ_j\gamma)'[\tilde{X}_j\Omega\tilde{X}_j' + \sigma^2\mathcal{I}]^{-1}(\tilde{y}_j - \tilde{X}_jZ_j\gamma), \tag{13}$$

which now must be minimized over γ, σ^2, Ω, and the unknown elements of the $\tilde{y}_j$. Minimizing over these unknown elements is real easy, we just set

$$g_j = H_jZ_j\hat{\gamma} \tag{14}$$

in an additional substep of the iteration. The g_j are used to compute the corresponding $\tilde{b}_j$, with

$$\tilde{b}_j = C^{-1}\{(X_j'X_j)b_j + (C - X_j'X_j)Z_j\hat{\gamma}\}. \tag{15}$$

Since choosing the optimal g_j by (14) makes the corresponding residuals zero, augmenting does not change the sum of squares of the residuals, i.e. does not change s_j^2. Also, we see from (15) that we never really need H_j and g_j. All that is needed is C. Observe that the theory in this section can also be applied if one or more of the $X_j'X_j$ are singular, as long as we choose C to be nonsingular.

It seems clear that C chose be chosen as small as possible from those matrices satisfying the condition $C \succcurlyeq X_j'X_j$. A C which is very large leads to $b_j \approx Z_j\hat{\gamma}$, which means there will be almost no change from one iteration to the next. Convergence will be slow. The obvious choice

$$C = \sum_{j=1}^{m} X_j'X_j \tag{16}$$

may not be very good, and a more careful choice may be called for in many cases.

6. Balancing

Consider the problem of minimizing C under the condition $C \succcurlyeq X_j'X_j$.

7. Analysis of Boundary Cases

8. Relationship with EM

9. Relationship with EB

10. Restrictions on Ω

Our developments so far depend critically on balance in the design, or on balance introduced by augmentation, but also on the fact that Ω can be any positive semi-definite matrix. If Ω is restricted in some form or another, then we cannot find it simply by using (**??**), and we have to impose the restrictions directly.

Most commonly, we write Ω in the form

$$\Omega = \sum_{k=1}^{\ell} \omega_k A_k, \tag{17}$$

i.e. Ω must be a weighted sum of known matrices.

11. Multilevel Block Structure

In many classical multilevel situations, the matrices Z_j are of the form

$$Z_j = \begin{bmatrix} z_j' & 0 & \dots & 0 \\ 0 & z_j' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & z_j' \end{bmatrix}, \tag{18}$$

i.e. Z_j is the direct sum of p copies of a vector z_j, with, say, v elements. If we have this additional structure, then the likleihood eqautions can be written more compactly in matrix notation.

Clearly in this case $Z_j\gamma$ can be written as Γz_j, with Γ an $p\times v$ matrix. Collecting all the b_j and z_j in an $p \times m$ matrix B and a $v \times m$ matrix Z we find

$$\hat{W} = (B - \Gamma Z)(B - \Gamma Z)', \tag{19a}$$

and

$$\hat{\Gamma} = (W^{-1} \otimes ZZ')^{-1} W^{-1} BZ' \tag{19b}$$

Interdivisional Program in Statistics, University of California at Los Angeles