Generalized Mixed Linear Models

Presentation at Fair Isaac

Jan de Leeuw

UCLA Statistics

07/01/04

Mixed Linear Models

$$\underline{y} = X\beta + Z\underline{\delta} + \underline{\epsilon}$$

Assume w.l.g. that the disturbances have zero expectation. Then

$$E(\underline{y}) = X\beta$$

as in ordinary linear models.

Assume the two vectors of disturbances are uncorrelated. Then

$$V(\underline{y}) = Z\Omega Z' + \Sigma,$$

where we often also assume homoscedasticity, i.e.

$$\Sigma = \sigma^2 I.$$

Thus MLM's can be seen as linear models with correlated disturbances, where the covariance matrix of the disturbances has a factor analytic structure.

Multilevel Models or HLM's

We have m groups of objects. For each group there is a random coefficient regression model of the form

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

where the disturbances are centered and homoscedastic.

This is the first level model. There is a second level model for the first level regression coefficients which says

$$\underline{b}_j = Z_j \gamma + \underline{\delta}_j.$$

This can be interpreted as "slopes as outcomes".

Alternatively we can rewrite the HLM as a LMM.

$$\underline{y}_j = X_j Z_j \gamma + X_j \underline{\delta}_j + \underline{\epsilon}_j,$$

where we see that the predictors with fixed coefficients are interactions between first and second level predictors. So an HLM can be thought of as an MLM with cross-level interactions.

For the Fair Isaac data the first level could be calls, and X has predictors describing calls, while the second level could be subscribers, where Z has subscriber characteristics.

In the usual HLM's, with say p first level and q second level predictors, the second level design matrices have the form

$$Z_j = \begin{bmatrix} z_j' & 0 & \cdots & 0 \\ 0 & z_j' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_j' \end{bmatrix}$$

In that case the pq cross level interactions are precisely the products of one first level (call) and one second level (subscriber) predictor. This can quickly become rather big (and rather ill-conditioned).

Generalized MLM's

We get a GLM by generalizing an LM. We get a GLML by generalizing an MLM.

$$\underline{y} \mid \underline{\delta} \sim_{id} \phi(y \mid \underline{\delta})$$

$$E(\underline{y} \mid \underline{\delta}) = \mu,$$

$$g(\mu_i) = x_i'\beta + z_i'\delta,$$

$$\underline{\delta} \sim \phi.$$

The usual GLM notions of link function and canonical link apply. As an example we use the mixed linear logit model, with likelihood

$$\mathcal{L} = \int \exp\{\sum_{i=1}^{n} y_i(x_i'\beta + z_i'\delta)\} \prod_{i=1}^{n}(1 + \exp(x_i'\beta + z_i'\delta))^{-1} d\phi(\delta)$$

This is generally difficult to evaluate, let alone optimize, because of the integral which usually can be written in a closed form.

Observe that this idea can be easily adapted to GHLM's, and that other link functions and exponential distributions can be used. For the Fair Isaac data, the logistic seems to be the preferred one.

Computation

Computationally there are a large number of possible approaches to avoid computing the integral.

- Quadrature (for instance Hermite)
- Expanding the likelihood around a fixed delta (PQL, MQL, linear, quadratic)
- MCMC
- Laplace approximation (number of terms)
- Nonparametric (point) distribution for delta

Software

- HLM
- MLWin
- Bugs
- R (various packages)
- MIXOR/MIXREG
- GLAMM in Stata
- Others (MLM, Mplus)