Multilevel Models for fMRI analysis

Presentation at IPAM

Jan de Leeuw
UCLA Statistics

07/19/04

http://gifi.stat.ucla.edu/fmri.pdf

What I wanted to do

- Present multilevel linear, nonlinear, and generalized linear models.
- Discuss how they have been applied in fMRI analysis so far.
- Use them to analyze a real (UCLA) example.

Unfortunately, I was overtaken by reality and I could only do the first two. The third will come later.

Data

fMRI data are often hierarchical. This means that we have *nested observation units*. Trials are nested within sessions, sessions within subjects, subjects within certain well-defined groups.

To fix our framework, let us consider the fMRI time series of hemodynamic signals at a single voxel, observed for each of m subjects. We look at single voxels because we want to use models with univariate outcomes. Ultimately, of course, multivariate extensions are needed.

Two-step Analysis

For each series we assume a linear model of the form

$$\underline{y}_j = X_j \beta_j + \underline{\epsilon}_j$$
$$\underline{\epsilon}_j \sim N\left[0, \Sigma_j(\theta)\right]$$

Parametric specification allows us to estimate the vector of regression coefficients and the matrix of error dispersions for each group. In the *second step* we then related the individual-level parameter estimates to individual-level regressors. In fMRI this is knows as the "summary statistics" approach, in multilevel analysis it is known as the "slopes as outcomes" approach.

Linear Multilevel Analysis

Multilevel analysis makes this two-step approach more precise and rigorous. We specify the first step as

$$\underline{y}_j = X_j \underline{\beta}_j + \underline{\epsilon}_j$$

and the second step as

$$\underline{\beta}_j = Z_j \gamma + \underline{\delta}_j$$

Generalization to more than two levels is obvious.

To make us statisticians happy, we also assume

$$\underline{\epsilon}_j \sim N\left[0, \Sigma_j(\theta)\right]$$

$$\underline{\delta}_j \sim N\left[0, \Omega_j(\xi)\right]$$

$$\underline{\delta}_j \perp \underline{\epsilon}_\ell \quad \forall j, \ell$$

$$\underline{\delta}_j \perp \underline{\delta}_\ell \quad \forall j \neq \ell$$

$$\underline{\epsilon}_j \perp \underline{\epsilon}_\ell \quad \forall j \neq \ell$$

The two specifications can be combined to

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

$$\underline{y}_j \sim N\left[X_j Z_j \gamma, X_j \Omega_j(\xi) X'_j + \Sigma_j(\theta)\right]$$

where we see how the resulting model is a mixed linear model, or a general linear model with a specific error dispersion structure that reflects information from both levels and with expectation structure that consists of *cross level interactions*.

In classical multilevel analysis we have some additional specifications

$$\Omega_j(\xi) = \Omega(\xi),$$
$$\Sigma_j(\theta) = \theta^2 I_j,$$

but especially the last one is usually inappropriate for time series data. Also

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

In that case the *cross level interactions* are precisely the products of one first level and one second level predictor.

$$X_j Z_j = [x_{j1} z_j' \;\; x_{j2} z_j' \;\; \cdots \;\; x_{jp} z_j']$$

If the regressions have an intercept, it means both design matrices have a column of ones, which means that predictors of both levels occur in the design matrix.

This design matrix can quickly become very big (and very ill-conditioned).

Estimation

- We distinguish between the maximum likelihood estimate (FIML) and the residual or restricted maximum likelihood estimate (REML).
- Algorithms use either scoring, or iterative generalized least squares, or EM, or MCMC.
- Random effects are not "estimated" in the usual sense but after we have parameter estimates we can compute the BLUP, i.e. the condional expectation of the random effect given the data. This gives "shrinkage estimates".

fMRI Applications

- Pan et al (Human Brain Mapping 2003). Simple two-level model.
- Friston et al (NeuroImage 2002ab). Theory (linear multilevel models, more tha two levels, EM estimation) and applications.
- Beckmann et al (NeuroImage, 2003). Two-level multi-subject/multisession model.
- Woolrich et al (NeuroImage 2004). Bayesian version of the two-level model.

Nonlinear MLM's

Work of statisticians such as Bates and Pinheiro (NLME in R) and of educational statisticians such as Goldstein (MLWIN) has generalized the basic multilevel model to

$$\underline{y}_{ij} = f(x_{ij}, \underline{\beta}_j) + \underline{\epsilon}_{ij}$$

Such models are fitted by using the same linearization techniques as in ordinary nonlinear least squares, leading to an approximate linear multilevel model.

Generalized MLM's

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

$$
\begin{aligned}
&\underline{y}_j \mid \delta_j \sim \phi(\bullet, \delta_j), \\
&E(\underline{y}_j \mid \delta_j) = \mu_j, \\
&g(\mu_j) = X_j Z_j \gamma + X_j \delta_j, \\
&\underline{\delta}_j \sim \psi.
\end{aligned}
$$

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

$$\mathcal{L} = \sum_{j=1}^{m} \log \int \prod_{i=1}^{n_j} \exp(y_{ij}(u_{ij}\gamma + x_{ij}\delta_j))[1+\exp(u_{ij}\gamma + x_{ij}\delta_j)]^{-1}\psi(\delta_j)d\delta_j$$

This is generally difficult to evaluate, let alone optimize, because of the integral which usually can not be evaluated in a closed form.

Observe that other link functions and exponential distributions can be used to accommodate outcomes that are counts or positive measurements.

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)

Discussion

- How Bayesian do you want to be ? And who cares ?
- Is MCMC really the computational tool of choice or just another sampling/optimization method?
- In multilevel analysis multivariate responses are incorporated as an additional "inner" level (variables nested in responses). Can this be used in fMRI ?
- Beckmann et al and Woolrich et al assume homoscedasticity of first-level errors. This seems optimistic for fMRI, certainly with multiple-time related voxels.