Growth Mixture Models
-or-
Specific Cases of Mixtures of Mixed Models

Lecture 16

May 2, 2006

Clustering and Classification

Today's Lecture

- General Linear Model.
- General Linear Mixed-Effects Model.
  - Growth models are mixed-effects models.
- Mixtures of the General Linear Mixed-Effects Model.
  - Specifically, Growth Mixture Models.

General Linear Model

- Recall the "multiple regression" equation (for the i^{th} observation, prediction of Y_i by k variables X_{ik}):

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_2 + \ldots + \beta_k X_k + \epsilon_i$$

- The equation above can be expressed more compactly by a set of matrices:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

  - $\mathbf{y}$ is of size $(N \times 1)$.
  - $\mathbf{X}$ is of size $(N \times (1 + k))$.
  - $\boldsymbol{\beta}$ is of size $(k \times 1)$.
  - $\boldsymbol{\epsilon}$ is of size $(N \times 1)$.

GLM Illustrated

- In matrices, the regression equation from the previous slide gives:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \ldots & X_{1k} \\ 1 & X_{21} & X_{22} & \ldots & X_{2k} \\ 1 & X_{31} & X_{32} & \ldots & X_{3k} \\ 1 & X_{41} & X_{42} & \ldots & X_{4k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{N1} & X_{N2} & \ldots & X_{Nk} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \vdots \\ \epsilon_N \end{bmatrix}$$

Additional Notes

- Note that most everything is really straightforward in terms of matrix algebra.

- The matrix of predictors, **X**, has the first column containing all ones.
  - This represents the intercept parameter α.
  - This is also an introduction to setting columns of the **X** matrix to represent design and or group controls (as in ANOVA).

Distribution of Errors

- We often place distributional assumptions on our error terms, allowing for the development of hypothesis tests.
- With matrices, the distributional assumptions are no different, except for things are approached in a multivariate fashion:

$$\boldsymbol{\epsilon} \sim N_N(\mathbf{0}, \sigma^2_\epsilon \mathbf{I}_N)$$

- Having a multivariate normal distribution with independent (uncorrelated) variables (from $\mathbf{I}_N$) is identical to saying:

$$\epsilon_i \sim N(0, \sigma^2_\epsilon)$$

for all i observations.

Error Covariance Matrix

- Therefore, the fixed-effects GLM assumes the following covariance matrix for the residuals:

$$\sigma^2_\epsilon \mathbf{I}_N = \begin{bmatrix} \sigma^2_\epsilon & 0 & 0 & \dots & 0 \\ 0 & \sigma^2_\epsilon & 0 & \dots & 0 \\ 0 & 0 & \sigma^2_\epsilon & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma^2_\epsilon \end{bmatrix}$$

- In repeated measures or growth modeling, this assumption is not valid.

- The observations are not independent, so we must model the dependency into the error covariance matrix.

Mixed Models

- The general linear mixed model is given by:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\gamma + \epsilon$$

  - $\mathbf{y}$ is of size $(N \times 1)$.
  - $\mathbf{X}$ is of size $(N \times (1 + k))$ - observations from fixed variables.
  - β is of size $(k \times 1)$ - fixed effects.
  - $\mathbf{Z}$ is of size $(N \times g)$ - containing g variables for random effects.
  - γ is of size $(g \times 1)$ - random effects.
  - ϵ is of size $(N \times 1)$.

Mixed Model Assumptions

- The main assumption in the mixed model analysis is that γ and ϵ are both normally distributed with:

$$E\begin{bmatrix}\gamma\\ \epsilon\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}$$

and

$$\text{Var}\begin{bmatrix}\gamma\\ \epsilon\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}$$

- The resulting model estimated variance of $\mathbf{y}$ given $\mathbf{X}$ and $\mathbf{Z}$ is:

$$\text{Var}\left(\mathbf{y}\right) = \mathbf{ZGZ}' + \mathbf{R}$$

Growth Models

- Growth models are mixed models with very specific features:
  - Random intercept parameter.
  - Random slope (although this is sometimes omitted).
- Because of how they fit into the Linear Mixed Effects (LME) modeling framework, such growth models can be conceptualized by latent variable modeling methods such as structural equation models.
- Additionally, such growth models can be conceptualized as two-level hierarchical linear models.
  - First level are the observations.
  - Second level are the subjects.
    - Observations nested within people.

Growth Model Example

Overview

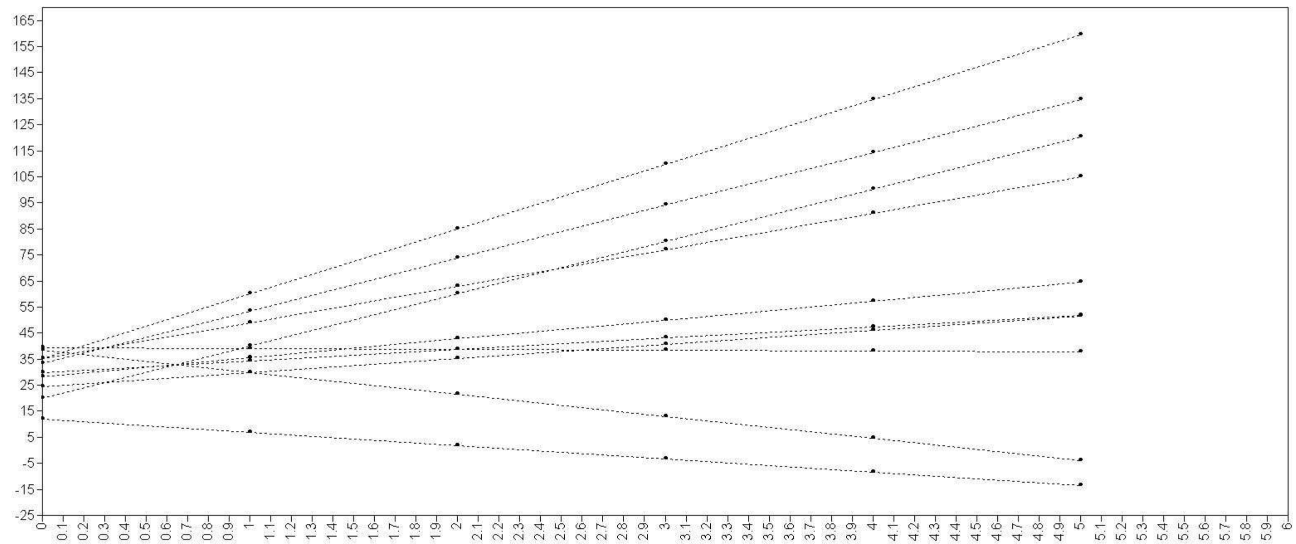
General Linear Model

General Linear Mixed Effects Model
- ➤ Assumptions
- ➤ Growth Models
- ➤ Example
- ➤ Parameterization
- ➤ Covariance Structure
- ➤ Estimates

Growth Mixture Models

Wrapping Up

- To demonstrate a growth model consider an example from a study of reading ability in a midwestern public school.
- The dependent variable, Nonsense Word Fluency (NWF) was measured across six consecutive time points (each separated by three month intervals).
- So, for each student, we have six observations of NWF (ignore issues of missing data for the moment).



Growth Model Example

- To parameterize the model as a LME model, we go back to our original LME model parameterization of $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\gamma + \epsilon$:

$$
\begin{bmatrix} \text{NWF}_{11} \\ \text{NWF}_{12} \\ \text{NWF}_{13} \\ \text{NWF}_{14} \\ \text{NWF}_{15} \\ \text{NWF}_{16} \\ \text{NWF}_{21} \\ \vdots \end{bmatrix}
=
\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 1 \\ \vdots & \vdots \end{bmatrix}
\begin{bmatrix} \alpha \\ \beta_1 \end{bmatrix}
+
\begin{bmatrix} 1 & 1 & 0 & 0 & \ldots \\ 1 & 2 & 0 & 0 & \ldots \\ 1 & 3 & 0 & 0 & \ldots \\ 1 & 4 & 0 & 0 & \ldots \\ 1 & 5 & 0 & 0 & \ldots \\ 1 & 6 & 0 & 0 & \ldots \\ 0 & 0 & 1 & 1 & \ldots \\ 0 & 0 & 1 & 2 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}
\begin{bmatrix} \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{31} \\ \vdots \end{bmatrix}
+
\begin{bmatrix} \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ \vdots \end{bmatrix}
$$

Model Specified Covariance of Y

- The variance of $\mathbf{y}$ is a block-diagonal matrix formed by the matrix product:

$$\text{Var}(\mathbf{y}) = \mathbf{ZGZ}' + \mathbf{R}$$

- Here, the matrix $\mathbf{G}$ is a 2×2 covariance matrix of the random effects:

$$\mathbf{G} = \begin{bmatrix} \sigma^2_{\gamma_1} & \sigma_{\gamma_1,\gamma_2} & 0 & 0 & \cdots \\ \sigma_{\gamma_1,\gamma_2} & \sigma^2_{\gamma_2} & 0 & 0 & \cdots \\ 0 & 0 & \sigma^2_{\gamma_1} & \sigma_{\gamma_1,\gamma_2} & \cdots \\ 0 & 0 & \sigma_{\gamma_1,\gamma_2} & \sigma^2_{\gamma_2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Model Specified Covariance of Y

- Similar to the fixed effects model, the **R** matrix is a diagonal matrix containing the residual variance, or $\sigma^2_\epsilon \mathbf{I}_N$:

$$\mathbf{R} = \begin{bmatrix} \sigma^2_\epsilon & 0 & \dots & 0 \\ 0 & \sigma^2_\epsilon & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2_\epsilon \end{bmatrix}$$

- So, the covariance matrix of **y** is then block diagonal, with repeating elements representing the covariation between observations within a person.

Estimates of Growth Model Parameters

- So, to show how this works, I went and estimated a growth model for the example data.
- Because we will use Mplus for our growth mixture models, I used Mplus to fit the model.
- Nearly any software package can be used to fit the single-class model, including SAS, SPSS, LISREL, and so forth.
- The estimates:

Parameter	Estimate
α	24.011 (2.613)
β_1	0.605 (1.216)
$\sigma^2_{\gamma_1}$	675.386 (104.564)
$\sigma^2_{\gamma_2}$	157.171 (23.034)
$\sigma^2_{\gamma_1,\gamma_2}$	-162.651 (37.564)

Growth Mixture Models

- In the LME models, our assumptions were that conditional on **X** and **Z**, **y** was distributed multivariate normal with:

$$\mathrm{E}(\mathbf{y}|\mathbf{X},\mathbf{Z}) = \mathbf{X}\beta$$

$$\mathrm{Var}\left(\mathbf{y}\right) = \mathbf{ZGZ}' + \mathbf{R}$$

- In the growth model case, we have a very specific structure of, **Z**, **G**, and **R** - but this methodology is not limited to just growth models.
- Plugging these results into the MVN density function:

$$f\left(\mathbf{y}|\mathbf{X},\mathbf{Z}\right) = \frac{1}{(2\pi)^{p/2}|\mathbf{ZGZ}' + \mathbf{R}|^{1/2}} e^{-(\mathbf{y}-\mathbf{X}\beta)\left[\mathbf{Z}'\mathbf{GZ}+\mathbf{R}\right]^{-1}(\mathbf{y}-\mathbf{X}\beta)'/2}$$

Linear Mixed Effects Mixture Models

- Recall that we stated that a finite mixture model expresses the distribution of **X** as a function of the sum of weighted distribution likelihoods:

$$f(\mathbf{X}) = \sum_{g=1}^{G} \eta_g f(\mathbf{X}|g)$$

- We are now ready to construct the GMM or (LMEMM) likelihood.

- Here, we say that the conditional distribution of **X** given g is a sequence of multivariate normal variables.

LMEMM/GMM as FMM

Using some notation of Bartholomew and Knott, a LMEMM for the response vector of p variables ($i = 1, \ldots, p$) with K classes ($j = 1, \ldots, K$):

$$f(\mathbf{y}|\mathbf{X}, \mathbf{Z}) = \sum_{j=1}^{K} \eta_j \frac{1}{(2\pi)^{p/2}|\mathbf{Z}\mathbf{G}_j\mathbf{Z}' + \mathbf{R}_j|^{1/2}} e^{-(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_j)\left[\mathbf{Z}\mathbf{G}_j\mathbf{Z}'+\mathbf{R}_j\right]^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_j)'/2}$$

- η_j is the probability that any individual is a member of class j (must sum to one).

- $\boldsymbol{\beta}_j$ is the set of fixed effects for class j.

- $\mathbf{G}_j$ is the covariance matrix of the random effects $\boldsymbol{\gamma}_j$ for class j.

- $\mathbf{R}_j$ is the covariance matrix of ϵ for class j.

GMM Example

- To demonstrate what we get from GMM, we consider trying to extract a number of latent classes from the original data used for the mixed model.
- As usual, a common method for determining the number of classes to extract is the BIC (common $\neq$ good).
- The fit results for a set of classes are:

Classes	BIC
1	6492.360
2	6452.427
3	6423.997
4	6364.526*
5	6375.993

- The four-class solution is selected by us of the BIC.

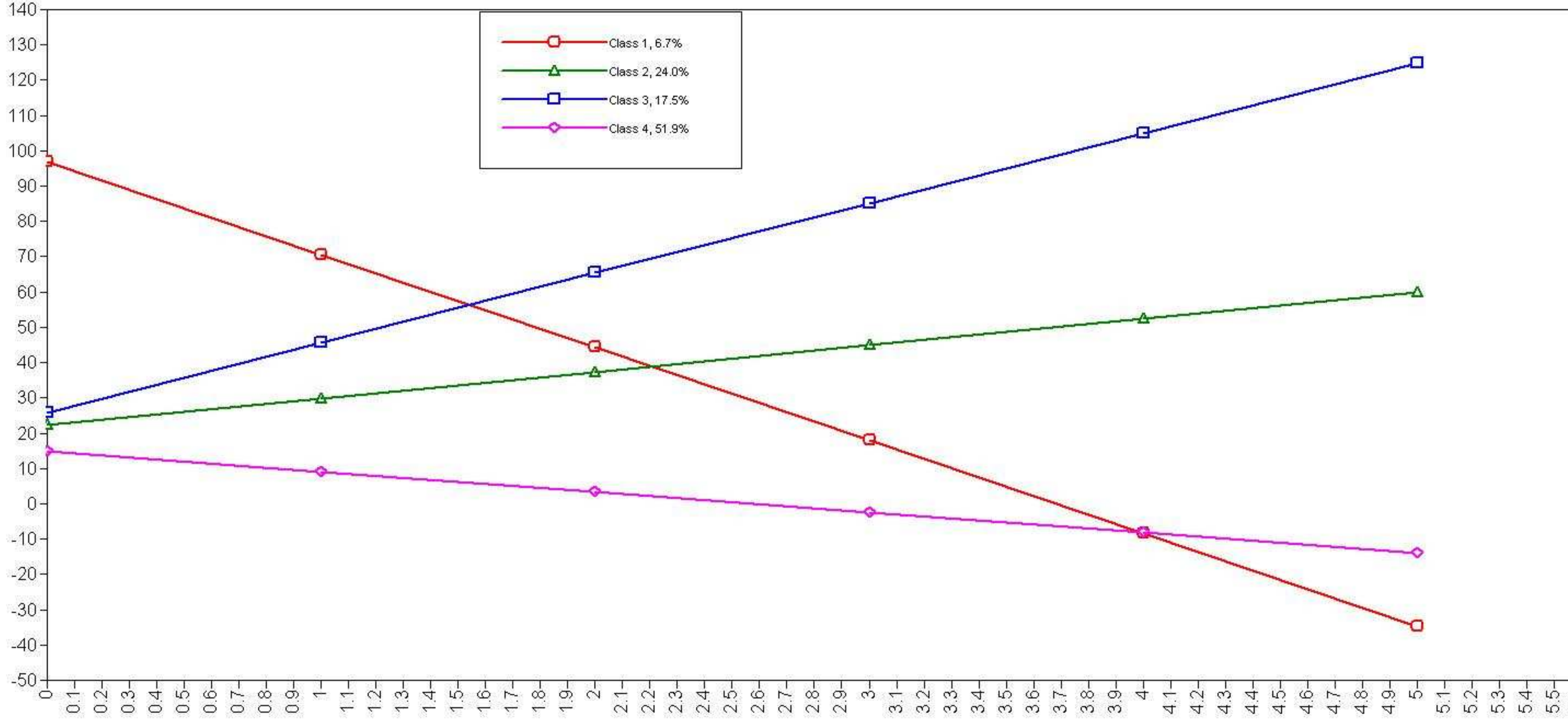
GMM Example

- The estimates:

Parameter	Class 1	Class 2	Class 3	Class 4
α	96.900	22.238	25.867	14.736
β_1	-26.323	7.550	19.778	-5.723
$\sigma^2_{\gamma_1}$	277.266	277.266	277.266	277.266
$\sigma^2_{\gamma_2}$	22.774	22.774	22.774	22.774
$\sigma^2_{\gamma_1,\gamma_2}$	-69.707	-69.707	-69.707	-69.707
η	0.07080	0.23894	0.17699	0.51327

- Notice anything strange?
- Can you find the Mplus default (which I didn't realize until after 2:00AM last night)?
- That's nice, but what do these look like?

GMM Example



Class 1, 6.7%
Class 2, 24.0%
Class 3, 17.5%
Class 4, 51.9%
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
0
1
2
3
4
5

Issues in GMM

- One of the biggest problems in GMM is the extraction of too many classes.
- This is not unlike what we witnessed with using LPA on Fisher’s Iris data.
- Like in Fisher’s Iris data, over extraction of classes occurs when violations of the assumptions are present.
- Clear need for model diagnostics, especially at one-class model.

Final Thought

- GMM is but one example of FMM applied to a specific example.
- The use of GMM has increased markedly in Psychology over the past few years.

- Running GMM without understanding the pitfalls can lead to seriously misleading results.
- I am going to bring out the “bad science” label for instances of reckless use of GMM.

Next Time

- How to fit GMM in Mplus.
- Empirical example (from Kevin).