Advances in Longitudinal Methods in the Social and Behavioral Sciences

Finite Mixtures of Nonlinear Mixed-Effects Models

Jeff Harring

Department of Measurement, Statistics and Evaluation
The Center for Integrated Latent Variable Research
University of Maryland, College Park
harring@umd.edu

Overview

- Mixture modeling
  - univariate and multivariate applications
- Characteristics of repeated measures learning data
- Nonlinear mixed-effects models
  - model description and analysis
- Nonlinear mixed-effects mixture (NLMM) model
  - model description

Overview

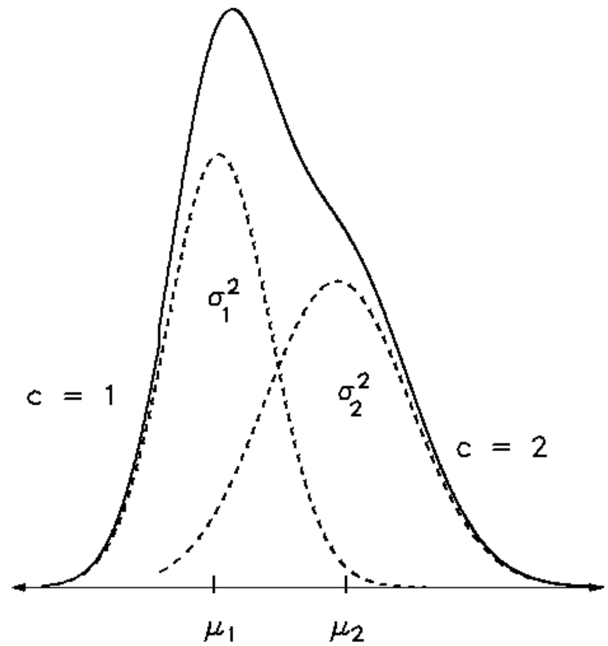
- **Learning data example revisited**
  - **analytic decision points**
- **Issues, challenges & considerations**

Finite Mixture Models

- **Karl Pearson (1894)**
- **Primary purposes…**
  - **model the density of complex distributions**
  - **model population heterogeneity**
- **Modeling heterogeneity $\Rightarrow$ mixture of distributions from the same parametric family**
- **Inferential goals**

Applications of Finite Mixture Models

- **Univariate mixtures**

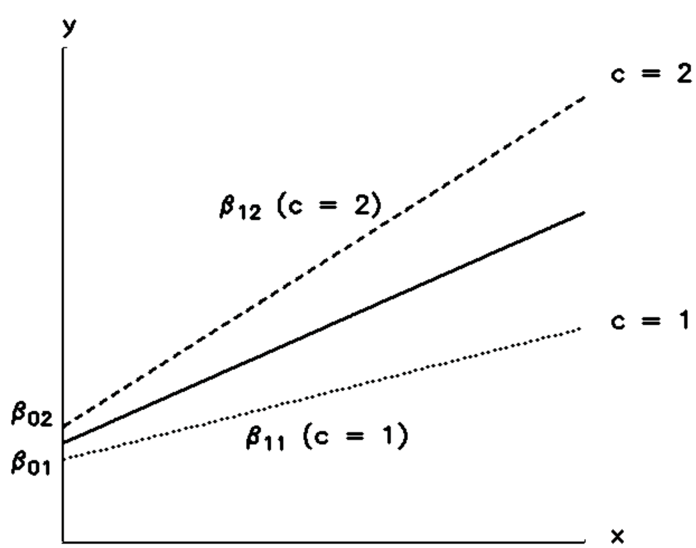


Applications of Finite Mixture Models

- **Regression mixtures**

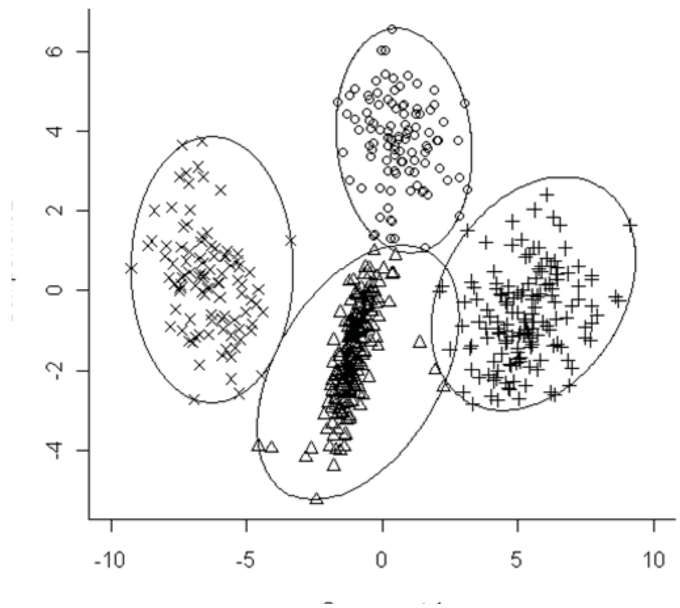
$$y_i = \beta_{0k} + \beta_{1k} x_i + e_{ik}$$

- **Regression mixture modeling involves estimating separate regression coefficients and error for each latent class**



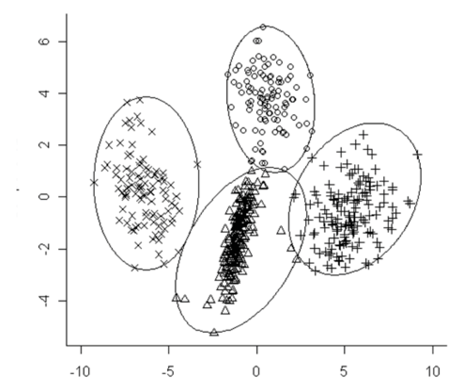
Applications of Finite Mixture Models

- **Multivariate normal mixtures**
  - **MVN mixtures $\Rightarrow$ cluster analysis**



Applications of Finite Mixture Models

- **Multivariate normal mixtures**

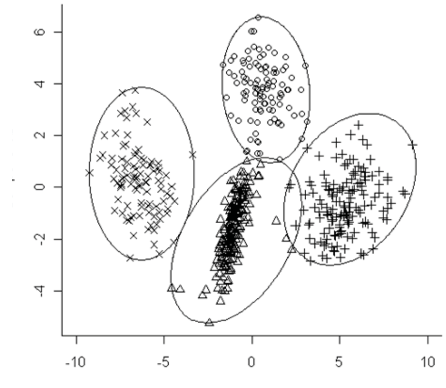


$$f(\mathbf{y}_i \mid \boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{y}_i \mid \boldsymbol{\theta}_k)$$

$$f_k(\mathbf{y}_i \mid \boldsymbol{\theta}_k) = (2\pi)^{-p/2} \left|\boldsymbol{\Sigma}_k\right|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_k)' \boldsymbol{\Sigma}_k^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_k)\right\}$$

Applications of Finite Mixture Models

- **Multivariate normal mixtures**

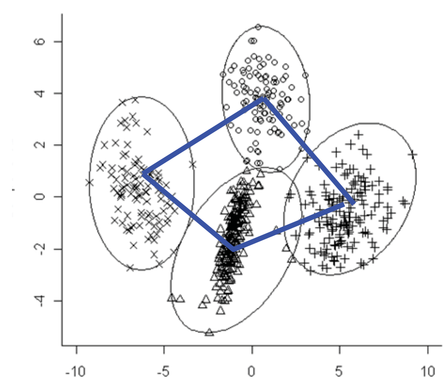


$$f(\mathbf{y}_i \mid \boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{y}_i \mid \boldsymbol{\theta}_k)$$

$$f_k(\mathbf{y}_i \mid \boldsymbol{\theta}_k) = (2\pi)^{-p/2} \left|\boldsymbol{\Sigma}_k\right|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_k)' \boldsymbol{\Sigma}_k^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_k)\right\}$$

Applications of Finite Mixture Models

- **Multivariate normal mixtures**

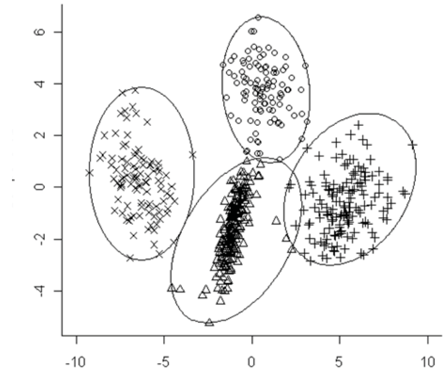


$$f(\mathbf{y}_i \mid \boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{y}_i \mid \boldsymbol{\theta}_k)$$

$$f_k(\mathbf{y}_i \mid \boldsymbol{\theta}_k) = (2\pi)^{-p/2} \left|\boldsymbol{\Sigma}_k\right|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_k)' \boldsymbol{\Sigma}_k^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_k)\right\}$$

Applications of Finite Mixture Models

- **Multivariate normal mixtures**

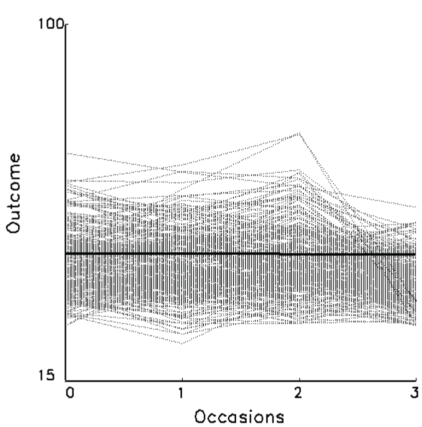


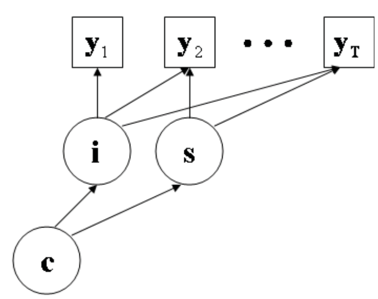
$$f(\mathbf{y}_i \mid \boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{y}_i \mid \boldsymbol{\theta}_k)$$

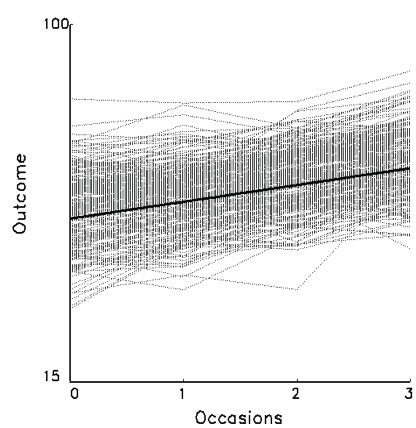
$$f_k(\mathbf{y}_i \mid \boldsymbol{\theta}_k) = (2\pi)^{-p/2} \left|\boldsymbol{\Sigma}_k\right|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_k)' \boldsymbol{\Sigma}_k^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_k)\right\}$$

Applications of Finite Mixture Models

- **Growth mixture modeling**

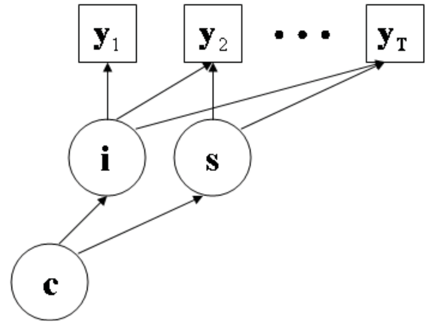






Applications of Finite Mixture Models

- **Growth mixture modeling**



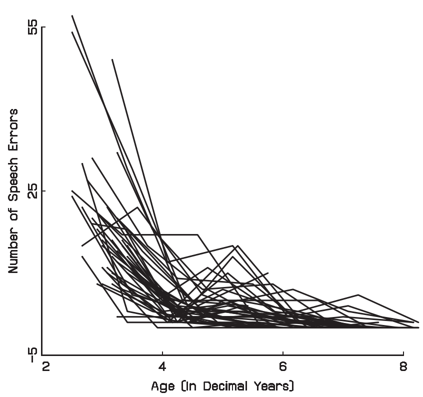
$$\mathbf{y}_i = \mathbf{\Lambda}\boldsymbol{\eta}_i + \mathbf{e}_i$$

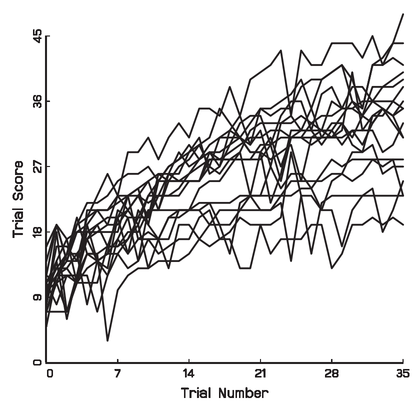
$$\boldsymbol{\eta}_i = \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \kappa_{\alpha k} \\ \kappa_{\beta k} \end{pmatrix} + \begin{pmatrix} \varsigma_{ik}^{\alpha} \\ \varsigma_{ik}^{\beta} \end{pmatrix}$$

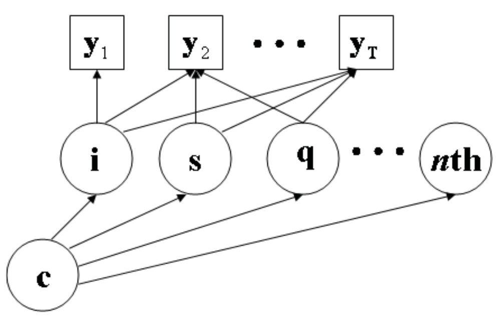
$$f_k(\mathbf{y}_i \mid \boldsymbol{\theta}_k) = (2\pi)^{-p/2} \left|\mathbf{\Sigma}_k\right|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_k)' \mathbf{\Sigma}_k^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_k)\right\}$$

$$\boldsymbol{\mu}_k = \mathbf{\Lambda}\boldsymbol{\kappa}_k \qquad\qquad \mathbf{\Sigma}_k = \mathbf{\Lambda}\mathbf{\Phi}_k\mathbf{\Lambda}' + \mathbf{\Theta}_k$$

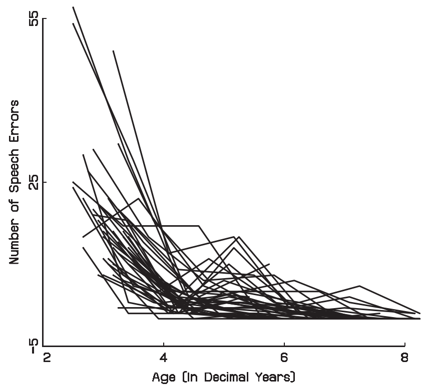
Applications of Finite Mixture Models

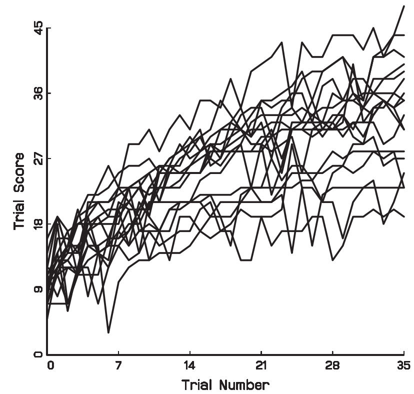






Applications of Finite Mixture Models





- **Handle nonlinear data with a nonlinear function**
- **Modeling potential population heterogeneity**

Applications of Finite Mixture Models

Intrinsically Nonlinear Functions → **Nonlinear Mixed-Effects Model**

Modeling Population Heterogeneity → **Finite Mixture Model**

Applications of Finite Mixture Models

Intrinsically Nonlinear Functions → Nonlinear Mixed-Effects Model

+

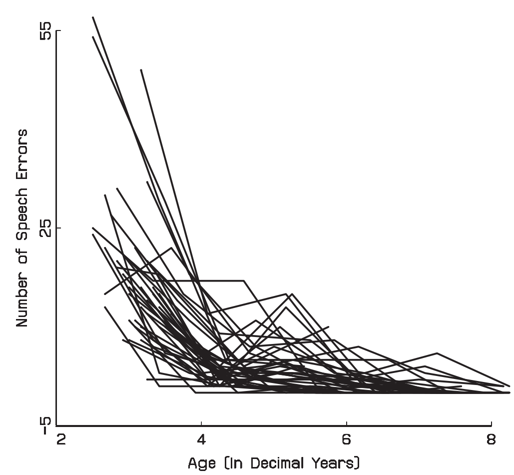
Modeling Population Heterogeneity → Finite Mixture Model

⇓

Nonlinear Mixed-Effects Mixture Model

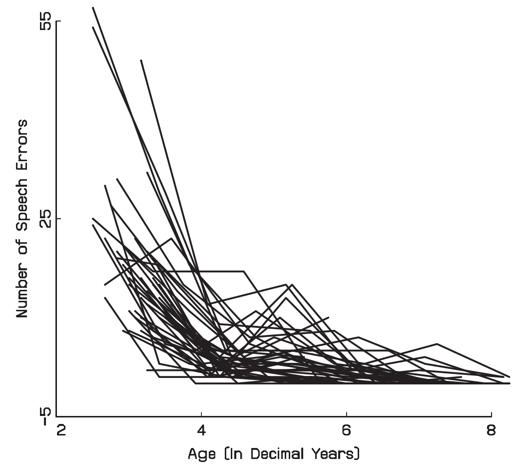
Nonlinear Learning Data – Speech Errors

- **Burchinal & Appelbaum (1991)**
  - Repeated measures data are recorded speech errors on a standard passage of text from an instrument of language proficiency
  - A sample of 43 young children ranging in ages from 2 to 8 years



Nonlinear Learning Data – Speech Errors

- **Burchinal & Appelbaum (1991)**
  - A maximum number of 6 repeated measures were taken
  - Ages at time of testings differed for each child
  - incomplete cases were observed
  - independent rating of overall speech intelligibility used as a covariate



Model for Nonlinear Learning Data

- The nonlinear mixed-effects (NLME) model is well- suited to handle both intrinsically nonlinear functions as well as key data and design characteristics
  - continuous response
  - compellingly strong individual difference in trajectories
  - substantial variability in time-response across subjects
  - distinct measurement occasions for each subject
  - interesting nonlinear change

$$y_{ij} = f(\beta_{1i}, \ldots, \beta_{pi}, x_{ij}) + e_{ij}$$

$$j = 1, \ldots, n_i; \quad i = 1, \ldots, m$$

NLME Model for Burchinal & Appelbaum Data

- **Following Davidian & Giltinan (2003) : two-stage hierarchy**
- **A subject-specific decelerating, decreasing exponential function is proposed**
- **Stage 1: *Individual-Level Model***

$$y_{ij} = \beta_{1i} \exp\left\{\beta_{2i}(x_{ij} - 3)\right\} + e_{ij}$$

  - β_{1i} **: the average number of speech errors at age 3**
  - β_{2i} **: rate parameter that governs functional decline**

NLME Model for Burchinal & Appelbaum Data

- **Stage 2: *Population-Level Model***

$$\boldsymbol{\beta}_i = \begin{pmatrix} \beta_{1i} \\ \beta_{2i} \end{pmatrix} = \begin{pmatrix} \beta_1 + b_{1i} \\ \beta_2 + b_{2i} \end{pmatrix} \qquad \textbf{unconditional}$$

$$\boldsymbol{\beta}_i = \begin{pmatrix} \beta_{1i} \\ \beta_{2i} \end{pmatrix} = \begin{pmatrix} \beta_1 + \gamma_1 z_i + b_{1i} \\ \beta_2 + \gamma_2 z_i + b_{2i} \end{pmatrix} \qquad \textbf{conditional}$$

NLME Model for Burchinal & Appelbaum Data

- **Distributional assumptions**

$$\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Phi}) \qquad \mathbf{e}_i \sim N(\mathbf{0}, \boldsymbol{\Delta}_i(\boldsymbol{\delta})) \qquad \begin{aligned} \operatorname{cov}(\mathbf{e}_i, \mathbf{b}_i') &= \mathbf{0} \\ \operatorname{cov}(\mathbf{e}_i, \mathbf{e}_{i'}') &= \mathbf{0} \\ \operatorname{cov}(\mathbf{b}_i, \mathbf{b}_{i'}') &= \mathbf{0} \end{aligned}$$

$$\operatorname{cov}(\mathbf{b}_i) = \boldsymbol{\Phi} = \begin{pmatrix} \operatorname{var}(b_{1i}) & \\ \operatorname{cov}(b_{2i}, b_{1i}) & \operatorname{var}(b_{2i}) \end{pmatrix}$$

NLME Model for Burchinal & Appelbaum Data

- **Distributional assumptions**

$$\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Phi}) \qquad \mathbf{e}_i \sim N(\mathbf{0}, \boldsymbol{\Delta}_i(\boldsymbol{\delta})) \qquad \begin{aligned} \operatorname{cov}(\mathbf{e}_i, \mathbf{b}_i') &= \mathbf{0} \\ \operatorname{cov}(\mathbf{e}_i, \mathbf{e}_{i'}') &= \mathbf{0} \\ \operatorname{cov}(\mathbf{b}_i, \mathbf{b}_{i'}') &= \mathbf{0} \end{aligned}$$

$$\boldsymbol{\Delta}_i = \sigma^2 \mathbf{I}_{n_i}$$

Inference NLME Model : Maximum Likelihood

- **The marginal distribution of $\mathbf{y}_i$**

$$h(\mathbf{y}_i) = \int p(\mathbf{y}_i, \mathbf{b}_i) d\mathbf{b}_i = \int p(\mathbf{y}_i \mid \mathbf{b}_i) p(\mathbf{b}_i) d\mathbf{b}_i$$

- **Let $\boldsymbol{\theta} = \left(\boldsymbol{\beta}', \boldsymbol{\delta}', vech(\boldsymbol{\Phi})'\right)'$**

- **Parameter estimation is carried out by maximizing the log-likelihood**

AGHQ (Pinheiro & Bates),
GHQ (Davidian & Gallant),
Linearization (Lindstrom & Bates),
GTS (Davidian & Giltinan),
Bayesian…

$$l(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta} \mid \mathbf{y}) = \sum_{i=1}^{m} \ln\{h(\mathbf{y}_i)\}$$

Analysis

- **The unconditional model was fitted using SAS PROC NLMIXED (Gaussian Quadrature – 30 points)**

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{18.48}^* \\ \mathbf{-0.98}^* \end{pmatrix}$$

$$\hat{\boldsymbol{\Phi}} = \begin{pmatrix} \mathbf{77.02}^* & \\ \mathbf{0.15} & \mathbf{0.05} \end{pmatrix} \qquad \hat{\sigma}^2 = \mathbf{9.73}^*$$

$$-2\ln L = 1227.0 \qquad BIC = 1249.6$$

Analysis

- **The conditional model was fitted with mean-centered covariate – SAS PROC NLMIXED (GH - 30 points)**

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{18.22}^* \\ \mathbf{-0.97}^* \end{pmatrix} \qquad \hat{\boldsymbol{\gamma}} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{-2.76}^* \\ \mathbf{-0.06} \end{pmatrix}$$

$$\hat{\boldsymbol{\Phi}} = \begin{pmatrix} \mathbf{76.84}^* & \\ \mathbf{-0.29} & \mathbf{0.04} \end{pmatrix} \qquad \hat{\sigma}^2 = \mathbf{9.73}^*$$

$$-2\ln L = 1217.2 \qquad BIC = 1246.3$$

Potentially Clustered Data

- **In subject-specific models, like the NLME model, regression parameters are allowed to vary across individuals resulting in differing within-subject profiles**
- $E(\mathbf{b}_i) = \mathbf{0}$ & $E(\mathbf{e}_i) = \mathbf{0} \Rightarrow$ **assumption all subjects were sampled from a single populations with common parameters**
- **Finite mixture models relax the single population distributional assumption of the random effects and the conditional distribution for the data to allow for parameter differences across unobserved populations**

Verbeke & Lesaffre, 1996, Verbeke & Molenberghs, 2000
Muthén & Shedden, 1999; Muthén, 2001, 2003, 2004
Hall & Wang, 2005

NLMM Model

- **The exponential NLME model can be extended to a NLMM model for *K* latent classes where in class *k* (*k = 1, 2, …, K*)**

$$\mathbf{y}_i = \beta_{1i} \exp\{\beta_{2i}(x_{ij} - 3)\} + \mathbf{e}_i \qquad \mathbf{e}_i \sim N(\mathbf{0}, \mathbf{\Delta}_k)$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}_k + \boldsymbol{\Gamma}_k \mathbf{z}_i + \mathbf{b}_i \qquad \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{\Phi}_k)$$

γ_{1k} **and** γ_{2k}

- π_k **is the probability of belonging to class *k*** $\Rightarrow \sum_k \pi_k = 1$

$$\Rightarrow \quad 0 \le \pi_k \le 1$$

NLMM Model – Getting Started

- **Fit a series of models suppressing random effects :** $\mathbf{\Phi}_k = \mathbf{0}$ **&** $\mathbf{\Delta}_k = \mathbf{\Delta}$

  (LCGM – Nagin (1999))

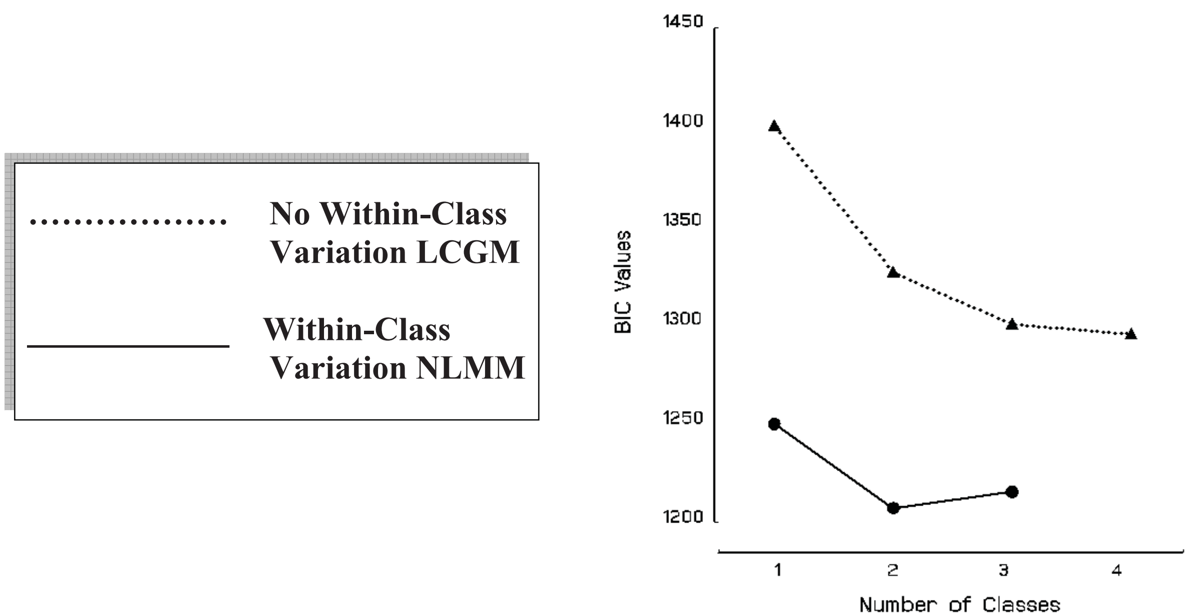
- **Method of deriving starting values for the mean structure of the NLMM model**

- **Use NLME output to suggest starting values for Φ and Δ**

NLMM Model – How Many Latent Classes?

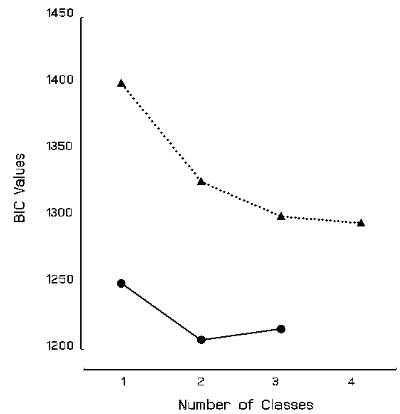
- Several statistics have been proposed and recommended in practice: AIC, BIC, SBIC, CLC, LMR-LRT (Lo, Mendell & Rubin, 2001), BLRT, Multivariate skewness and kurtosis indices
- Compute and plot BIC or SBIC values against number of classes

NLMM Model – How Many Latent Classes?



NLMM Model – How Many Latent Classes?

- **Decision based on…**
  - theoretical defensibility
  - model fit
  - parsimony
  - class separation and class incidence



- **Expertise of the substantive researcher**
- **Do the classes represent substantively meaningful groups?**

NLMM Model – Analysis of Learning Data

- **Parameter estimates for the two-class model**

NLME Estimates $\hat{\boldsymbol{\beta}}$	Class Means $\hat{\boldsymbol{\beta}}_k$		Class-Specific Covariate Estimates $\hat{\boldsymbol{\gamma}}_k$
	$\begin{pmatrix} 13.24^* \\ -0.66^* \end{pmatrix}$	$\hat{\pi}_1 = 0.33^*$	$\hat{\gamma}_{11} = -0.04 \quad \hat{\gamma}_{12} = 0.10^*$
$\begin{pmatrix} 18.22^* \\ -0.97^* \end{pmatrix}$			$\hat{\boldsymbol{\gamma}} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = \begin{pmatrix} -2.76^* \\ -0.08 \end{pmatrix}$
	$\begin{pmatrix} 19.45^* \\ -1.07^* \end{pmatrix}$	$\hat{\pi}_2 = 0.67$	$\hat{\gamma}_{21} = -2.72^* \quad \hat{\gamma}_{22} = -0.25^*$

NLMM Model – Analysis of Learning Data

- **Parameter estimates for the two-class model**

NLME Estimate Covariance Matrix $\hat{\Phi}$	Class-Specific Covariance Matrix $\hat{\Phi}_k = \hat{\Phi}$	NLME Estimate Residual Variance $\hat{\sigma}^2$	Class-Specific Residual Variance $\hat{\sigma}_k^2 = \hat{\sigma}^2$
$\begin{pmatrix} 77.02^* & \\ 0.15 & 0.05 \end{pmatrix}$	$\begin{pmatrix} 53.65^* & \\ 2.35^* & 0.62^* \end{pmatrix}$	9.73*	2.73*

NLMM Model – Assessing Model Fit?

- **The quality of the mixture based on the precision of the classification – classification is based on estimated posterior probabilities**

- **For $K \geq 2$, average posterior probabilities can be computed. A $K \times K$ matrix should have high diagonal and low off-diagonal values indicating good classification quality**

  – **Average Latent Class Probabilities for Most Likely Latent Class Membership (Row) by Latent Class (Column)**

	1	2
1	0.854	0.146
2	0.204	0.796

NLMM Model – Assessing Model Fit?

- **Entropy – a summary measure of the classification based on individuals' estimated posterior probabilities can be computed with values close to 1 indicating near perfect classification**

$$E_K = 1 - \frac{\sum_{i=1}^{m}\sum_{k=1}^{K}(-\hat{\pi}_{ik}\ln\hat{\pi}_{ik})}{m\ln K}$$

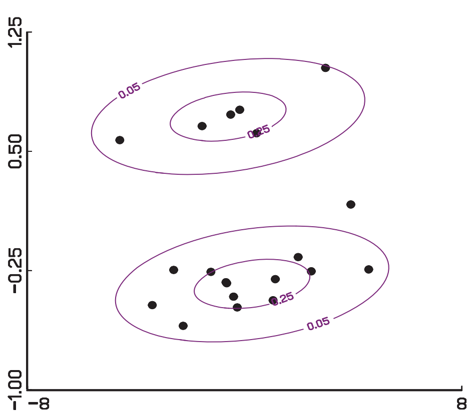
$$E_{K=2} = 0.73$$

Graphical Summaries

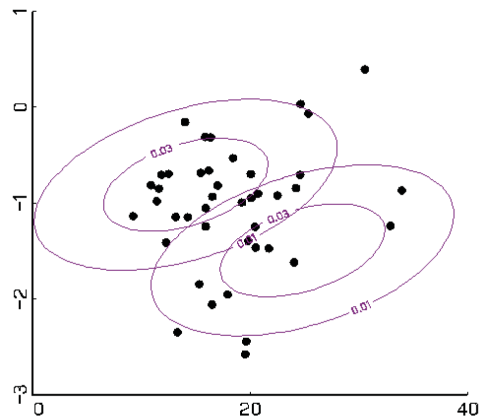
- **Plot predicted random coefficients on a contour plot or surface plot**

Graphical Summaries

- **Plot predicted random coefficients on a contour plot or surface plot**



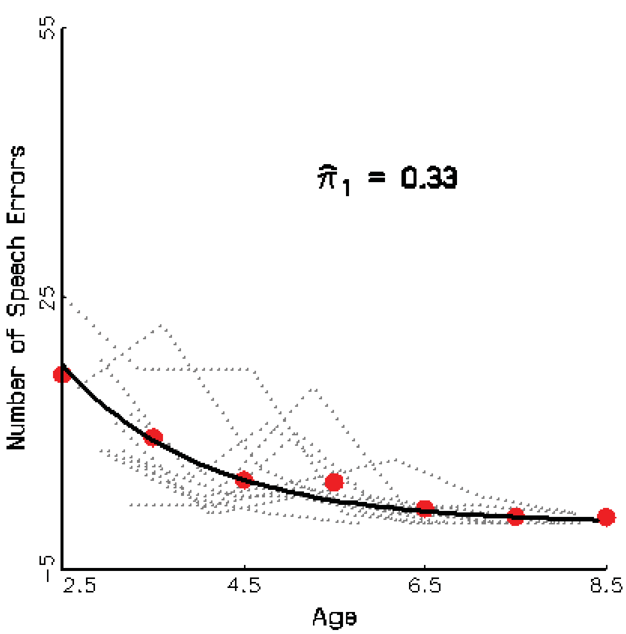
Good Separation

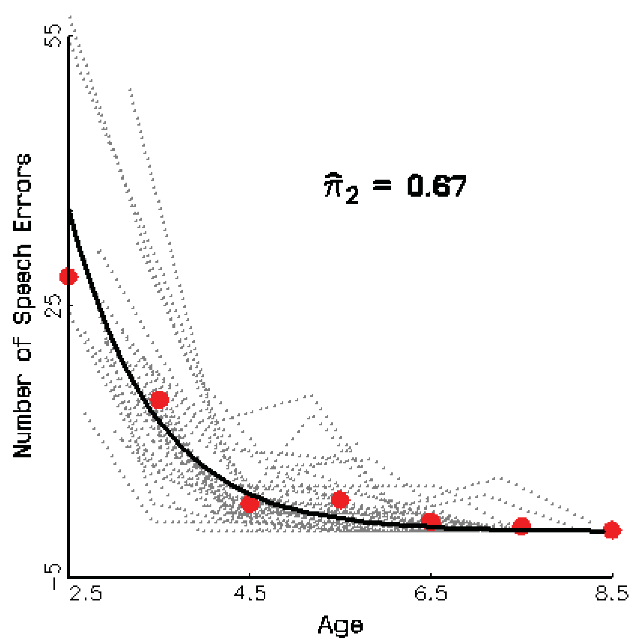


Modest Separation

Graphical Summaries

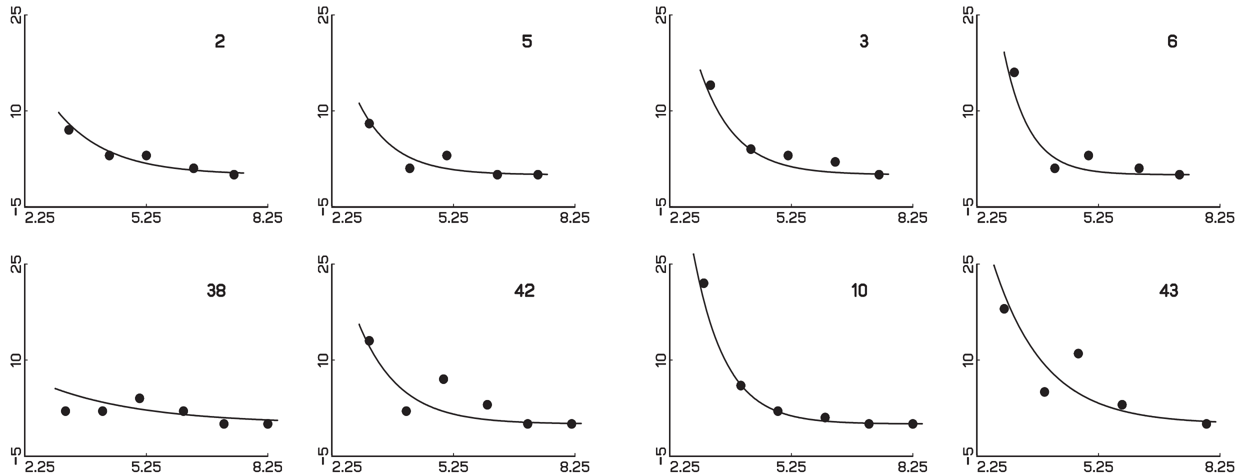
Two Latent Classes: Mean Trajectories





Graphical Summaries

Fitted Curves for 4 Individuals in Class 1 | **Fitted Curves for 4 Individuals in Class 2**



Practical Issues…

- **Estimation**
- **Computing standard errors**

Log-likelihood

- **Let** $\boldsymbol{\pi}' = (\pi_1, \ldots, \pi_{K-1})$
- **Let** $\boldsymbol{\xi}' = (\boldsymbol{\beta}'_k, vech(\boldsymbol{\Phi}_k), vech(\boldsymbol{\Delta}_k))$
- **Let** $\boldsymbol{\theta} = (\boldsymbol{\pi}', \boldsymbol{\xi}')'$ **all model parameters, then the log-likelihood**

$$
\begin{aligned}
l(\boldsymbol{\theta}) &= \ln L(\boldsymbol{\theta} \mid \mathbf{y}) \\
&= \ln\left(\prod_{i=1}^{m} \sum_{k=1}^{K} \pi_k h_k(\mathbf{y}_i)\right) \quad \textbf{where} \quad h_k(\mathbf{y}_i) = \int p_k(\mathbf{y}_i \mid \mathbf{b}_i) p_k(\mathbf{b}_i) d\mathbf{b}_i \\
&= \sum_{i=1}^{m} \ln\left(\sum_{k=1}^{K} \pi_k h_k(\mathbf{y}_i)\right)
\end{aligned}
$$

Estimation

- **<u>If</u> the nonlinear regression coefficients are fixed across individuals : M*plus* (however – does not handle unique measurement occasions)**

$$y_{ij} = \beta_{1i} \exp\left\{\beta_2 (x_{ij} - 3)\right\} + e_{ij}$$

- **Directly maximize the log-likelihood using gradient methods like N-R or Quasi-Newton**
- **Use EM (Expectation – Maximization) algorithm treating class membership as missing data**

Standard Errors

- Direct maximization uses the diagonal elements of the Hessian matrix at convergence for a model-based estimate of the standard errors.
- EM algorithm – no standard errors are computed as a by-product of the algorithm
  - At convergence, use a direct maximization step to produce SE

More Practical Issues and Future Considerations…

- Evidence of local extrema
- Covariates predicting individual coefficients & class membership
- Does the method of handling the intractable integration influence the number of latent classes?
- Normality of the random effects distribution: non-parametric alternatives?

Advances in Longitudinal Methods in the Social and Behavioral Sciences

Finite Mixtures of Nonlinear Mixed-Effects Models

Jeff Harring

Department of Measurement, Statistics and Evaluation
The Center for Integrated Latent Variable Research
University of Maryland, College Park
harring@umd.edu