Searching for Latent Patterns in Longitudinal Data: How Well Do Latent Growth Mixture Models Work?

Kevin L. Delucchi, PhD
University of California, San Francisco
Acknowledgements

- Alan Bostrom, Ph.D.
- Connie Weisner, DPH
- Alcohol Research Group
- Sharon Hall, Ph.D.
- SF Treatment Research Center
Latent Constructs

- Common in social sciences
- Depression, affiliation, social pressure, big-five personality dimensions
- Factor analysis and structural equation modeling
- Growth mixture modeling
Individual Trajectories
Known Group Trajectories
Individual Trajectories

![Graph showing individual trajectories over time](image-url)
Latent Trajectories

Outcome

Time

1 2 3 4

0 5 10 15 20 25 30 35
Part 1:
Motivating Example

Trajectories of Alcohol Consumption over 5 Years

Research Goals

- Understanding the long-term course of problematic drinking
- Common patterns of drinking?
- Covariates related to those patterns?
Study Design

- Longitudinal survey of dependent and problem alcohol drinkers
- Data from 5 years (4 assessments)
- N = 1094 (complete data)
- Outcome: N drinks per year
First 100 Cases
Known Groups

- Mixed-effects models
  - Hierarchical linear models
  - Random effects

- Estimate effects for sex, treatment condition, etc.
Growth Mixture Models
Latent Class Growth Models

- Random effects models in K subgroups
- Finite mixture modeling
- Set of observed trajectories → smaller set of latent trajectories
- Improvement on “classify-then-analyze”
- Measures of model fit for model selection
- Software: Proc Traj, MPlus
Outcome Modeled

- Number of drinks of alcohol in prior year
- Log-10 transformed
- Fit models with 2 to 6 latent groupings
- Added covariates
<table>
<thead>
<tr>
<th>Number of Classes</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-6558.3</td>
</tr>
<tr>
<td>3</td>
<td>-6432.5</td>
</tr>
<tr>
<td>4</td>
<td>-6328.7</td>
</tr>
<tr>
<td>5</td>
<td>-6261.7</td>
</tr>
<tr>
<td>6</td>
<td>-6496.1</td>
</tr>
</tbody>
</table>
Results

Five class model produced best fit

- Early Quitters (N=88)
- Light/Non-drinkers (N=76)
- Gradual Improvers (N=129)
- Moderate Drinkers (N=229)
- Heavy Drinkers (N=572)
Early Quitters
3.5
Gradual Improvers
2.5
Heavy Drinkers
1.5
Non-Drinkers
0.5
Moderate Drinkers
0.5
Mean Log10 Volume Drinks

Early Quitters
Gradual Improvers
Heavy Drinkers
Non-Drinkers
Moderate Drinkers

Yr 0  Yr 1  Yr 3  Yr 5
# Mean Posterior Probabilities

<table>
<thead>
<tr>
<th>Group</th>
<th>Early Quit</th>
<th>Non-Drinkers</th>
<th>Gradual Improve</th>
<th>Mod. Drinkers</th>
<th>Heavy Drinkers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>.96</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td>Group 2</td>
<td>.00</td>
<td>.96</td>
<td>.00</td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td>Group 3</td>
<td>.01</td>
<td>.00</td>
<td>.89</td>
<td>.02</td>
<td>.00</td>
</tr>
<tr>
<td>Group 4</td>
<td>.03</td>
<td>.03</td>
<td>.07</td>
<td>.82</td>
<td>.10</td>
</tr>
<tr>
<td>Group 5</td>
<td>.00</td>
<td>.00</td>
<td>.03</td>
<td>.15</td>
<td>.90</td>
</tr>
</tbody>
</table>
## Percent No Drinks Prior Year

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Year 1</th>
<th>Year 3</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Early Quit</strong></td>
<td>0</td>
<td>97</td>
<td>100</td>
<td>77</td>
</tr>
<tr>
<td><strong>Non-Drink</strong></td>
<td>67</td>
<td>76</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td><strong>Improve</strong></td>
<td>0</td>
<td>13</td>
<td>48</td>
<td>74</td>
</tr>
<tr>
<td><strong>Moderate</strong></td>
<td>1</td>
<td>16</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td><strong>Heavy</strong></td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Part 2:

How well do growth mixture models work?
Three Normal Distributions
Growth Mixture Modeling

- Mixture modeling; Newcomb, 1886, Pearson, 1884
- Muthén & Shedden (1999)
- Nagin: latent class mixture modeling
Prior Research

Eggleston, et al. (2004) – length of follow-up effecting trajectory shapes

Jackson & Sher (2006) – effects of number of assessments

Henson, et al. (2007) fit statistics not accurate for small N (< 500)

Nylund, et al. (2007) - determining number of classes

Lubke & Muthén (2007) - FMM poor capture of classes without covariates
Current Project

- How well do growth mixture models capture group membership and model parameters?
- Simulations
- Two-group, longitudinal design
- Percent correctly classified, estimates of intercepts and slopes
Initial Conditions Simulated

112 cell-design (2x2x2x2x7)
- 2 sample sizes (300 and 900)
- 2 intercept effect sizes
- 2 slope effect sizes
- 2 levels of residual variance
- 7 levels of sample imbalance
- **Intercept means** -- 2 levels: \( \mu_{I1}, \mu_{I2} = (0,0.5) \) or \((0,1)\)

- **Slope means** -- 2 levels: \( \mu_{S1}, \mu_{S2} = (0.25, -0.25) \) or \((0.5, -0.5)\)

- **Residual variance** -- 2 levels: \( \sigma^2_R = 0.1 \) or \(0.5\).

- **Imbalance** -- 7 levels: \((N1, N2) = (480,420), (540,360), (600,300), (660,240), (720,180), (780,120), (840,60)\)
Grp 1 large effects
Grp 2 large effects

Time 0  Time 1  Time 2  Time 3
Data Model

Random slope and intercept selected from $N(\mu_i, \Sigma)$

$$\mu_i = \begin{bmatrix} \mu_{iI} \\ \mu_{iS} \end{bmatrix} \sum \begin{bmatrix} \sigma_{II}^2 & \sigma_{IS} \\ \sigma_{IS} & \sigma_{SS}^2 \end{bmatrix}$$
After sampling, 4 data points generated;

\[ Y_i = I + (i-1)S + R_i \text{ for } i = 1,\ldots,4 \]

where \( R_i \) is \( \sigma_R \varepsilon_i \) with \( s^2_R \) being a constant residual variance and \( \varepsilon_i \sim N(0,1) \).
- Data generated in SAS
- Analysis using Mplus (v4.2)
- Data summarized using SAS
TITLE: ANALYSIS 1;
DATA: FILE is C:\Delucchi\MPlus\Data\Sample1.txt;
VARIABLE: NAMES are Group Subject Y1-Y4;
   CLASSES = C(2);
   USEVARIABLES = Y1-Y4;
ANALYSIS: TYPE = MIXTURE;
   STARTS = 100, 10;
MODEL: %OVERALL%
   i s|y1@0 y2@1 y3@2 Y4@3;
   [c#1] (alpha);
   %c#2%
SAVEDATA: FILE = C:\Delucchi\MPlus\Data\File1.txt;
   RESULTS = C:\Delucchi\MPlus\Data\Analysis1.txt;
   SAVE = CPROBABILITIES;
OUTPUT: model constraint: alpha>0;
Results from 14 Conditions

- Residual var = 0.5
- N=900
- Largest and smallest effects
- All 7 levels of imbalance
All Four Conditions

% Classified Correctly

Percent of Sample in Larger Group (Total N = 900)

- Larger Group (large effects)
- Larger Group (smaller effects)
- Smaller Group (large effects)
- Smaller Group (small effects)
Mean Estimated Intercept by Class and Effect Size

![Graph showing mean estimated intercept by class and effect size.]

- Red line: Large ES, class 1 (0)
- Yellow line: Small ES, class 1 (0)
- Green line: Large ES, class 2 (1)
- Blue line: Small ES, class 2 (0.5)

Data points are plotted at 53%, 60%, 67%, 73%, 80%, 87%, and 93%. The graph illustrates how the estimated intercept changes across different effect sizes and classes.
Mean Estimated Slope by Class and Effect Size

- Large effect size, class 1 (0.5)
- Large effect size, class 2 (-0.5)
- Small effect size, class 1 (0.25)
- Small effect size, class 2 (-0.25)
Revised Simulations

- $N_s = 900 \ (300, 600) \ and \ 1800 \ (600, 1200)$
- Residual variance set to 0.1
- Intercepts at 0 and 1
- Slopes at -0.5 and 0.5
- Effect size from 0.2 to 2.2
Percent of Group Correctly Classified by Intercept, Slope Effect Size

- Group 1 (N = 600)
- Group 2 (N = 300)
- Group 1 (N = 1200)
- Group 2 (N = 600)
Mean Estimate of Intercept Means by Intercept, Slope Effect Size
Mean Estimate of Slope Means by Intercept, Slope Effect Size

Group 1 (N = 600)
Group 2 (N = 300)
Group 1 (N = 1200)
Group 2 (N = 600)
Percent of Group Correctly Classified by Intercept, Slope Effect Size

Residual variance = 0.1, with continuous covariate N(5,5) in Group 1, N(15,5) in Group 2.

Percent of Group Correctly Classified

Intercept and Slope Effect Size

- Group 1 (N = 600)
- Group 2 (N = 300)
- Group 1 (N = 1200)
- Group 2 (N = 600)
Mean Estimate of Intercept Means by Intercept, Slope Effect Size

Residual variance = 0.1, with continuous covariate N(5,5) in Group 1, N(15,5) in Group 2
Mean Estimate of Slope Means by Intercept, Slope Effect Size

Residual variance = 0.1, with continuous covariate N(5,5) in Group 1, N(15,5) in Group 2

CAPS 20 November 2009
Conclusions

- GMMs – potentially very informative
- Currently seen mainly in drug/alcohol and developmental studies – esp. criminology
- Simulation results raise concerns
  - Poor estimation of group membership
  - Model parameters require large effect sizes
Estimated Intercept Means by Intercept and Slope Variance

- Group 1 (N = 1800)
- Group 2 (N = 900)
- Group 3 (N = 450)