Using Mixture Latent Markov Models for Analyzing Change with Longitudinal Data

Jay Magidson, Ph.D.
President, Statistical Innovations Inc.
Belmont, MA., U.S.

Presented at Modern Modeling Methods (M3) 2013, University of Connecticut
Abstract

Mixture latent Markov (MLM) models are latent class models containing both time-constant and time-varying discrete latent variables. We introduce and illustrate the utility of latent Markov models in 3 real world examples using the new GUI in Latent GOLD® 5.0.

MLM models often fit longitudinal data better than comparable mixture latent growth models because of the inclusion of transition probabilities as model parameters which directly account for first-order autocorrelation in the data. Lack of fit can be localized (e.g., first-order, second-order autocorrelation) using new longitudinal bivariate residuals (L-BVRs) to suggest model modification.
Outline of Presentation

• Introduction to latent Markov (LM) and mixture latent Markov (MLM) Models*

• Example 1: LM model with time-homogeneous transition probabilities

• Example 2: MLM model with time-heterogeneous transitions

• Example 3: MLM model with time-heterogeneous transitions and covariates

In each of these examples latent Markov outperformed the corresponding latent growth model, the largest improvement being in the Lag1 BVRs (i.e., first-order autocorrelation).

* (Mixture) latent Markov models are also known as (mixture) hidden Markov models
Latent (Hidden) Markov Models are Defined by
3 Sets of Equations

- **Initial State probs:** $P(X_0 = s)$ — categories of $X$ are called *latent states* (rather than *latent classes*) since a person may change from one state to another over time.

  **Logit model** may include covariates $Z$ (e.g., AGE, SEX) :
  \[
  \log \frac{P(x_0 = s)}{P(x_0 = 1)} = \alpha_{0s}
  \]

- **Transition probs:** $P(X_t = r | X_{t-1} = s)$

  **Logit** may include time-varying and fixed predictors :
  \[
  \log \frac{P(x_t = r | x_{t-1} = s)}{P(x_t = 1 | x_{t-1} = s)} = \gamma_{0r} + \gamma_{1rs} + \gamma_{2rt} + \gamma_{3rst}
  \]

- **Measurement model probs:** $P(y_t = j | X_t = s)$

  One or more dependent variables $Y$, of possibly different scale types (e.g., continuous, count, dichotomous, ordinal, nominal) :
  \[
  \log \frac{P(y_{it} = \ell | x_i = s)}{P(y_{it} = 1 | x_i = s)} = \beta_{0\ell} + \beta_{1\ell s}
  \]

* Equations can be customized using Latent GOLD 5.0 GUI and/or syntax.
Equations:

Latent Markov (LM):

\[
P(y_i \mid z_i) = \sum_{x_0=1}^{K} \sum_{x_1=1}^{K} \cdots \sum_{x_T=1}^{K} P(x_0, x_1, \ldots, x_T \mid z_i) P(y_i \mid x_0, x_1, \ldots, x_T, z_i)
\]

Initial latent state & Transition sub-models

\[
P(x_0, x_1, \ldots, x_T \mid z_i) = P(x_0 \mid z_{i0}) \prod_{t=1}^{T} P(x_t \mid x_{t-1}, z_{it})
\]

Measurement sub-model

\[
P(y_i \mid x_0, x_1, \ldots, x_T, z_i) = \prod_{t=0}^{T} P(y_{it} \mid x_t, z_{it}) = \prod_{t=0}^{T} \prod_{j=1}^{J} P(y_{ij} \mid x_t, z_{it})
\]

Mixture Latent Markov (MLM):

\[
P(y_i \mid z_i) = \sum_{w=1}^{L} \sum_{x_0=1}^{K} \sum_{x_1=1}^{K} \cdots \sum_{x_T=1}^{K} P(w, x_0, x_1, \ldots, x_T \mid z_i) P(y_i \mid w, x_0, x_1, \ldots, x_T, z_i)
\]

\[
P(w, x_0, x_1, \ldots, x_T \mid z_i) = P(w \mid z_i) P(x_0 \mid w, z_{i0}) \prod_{t=1}^{T} P(x_t \mid x_{t-1}, w, z_{it})
\]

\[
P(y_i \mid w, x_0, x_1, \ldots, x_T, z_i) = \prod_{t=0}^{T} P(y_{it} \mid x_t, w, z_{it}) = \prod_{t=0}^{T} \prod_{j=1}^{J} P(y_{ij} \mid x_t, w, z_{it})
\]
## Relationship between Mixture Latent Markov and Mixture Latent Growth Models

### Classification of latent class models for longitudinal research

<table>
<thead>
<tr>
<th>Model name</th>
<th>Transition structure</th>
<th>Unobserved heterogeneity</th>
<th>Measurement error</th>
</tr>
</thead>
<tbody>
<tr>
<td>I Mixture latent Markov</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>II Mixture Markov</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>III Latent Markov</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>IV Standard Markov*</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>V Mixture latent growth</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>VI Mixture growth</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>VII Standard latent class</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>VIII Independence*</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

*This model is not a latent class model.

Longitudinal bivariate residuals quantify for each response variable $Y_k$ how well the overall trend as well as the first- and second-order autocorrelations are predicted by the model.

BVR.Time = BVR$_k$(time, $y_k$),

BVR.Lag1 = BVR$_k$(y$_k$[t-1], y$_k$[t])

and BVR.Lag2 = BVR$_k$(y$_k$[t-2], y$_k$[t])

Lag1:

$$BVR_k = \sum_r \sum_s \frac{(n_{krs} - m_{krs})^2}{m_{krs}}$$

where $m_{krs}$ is based on $P(Y_{kit-1} = r, Y_{kit} = s)$, that is

$$m_{krs} = \sum_i \sum_t P(Y_{kit-1} = r, Y_{kit} = s).$$

$$= P(Y_{kit-1} = r|X_{it-1} = x_{t-1})P(Y_{kit} = s|X_{it} = x_t)$$
Example 1: Loyalty Data in Long File Format

- N=631 respondents
- T= 5 time points
- Dichotomous Y:
  - Choose Brand A?
    - 1=Yes
    - 0=No
- \( Y=(Y_1, Y_2, Y_3, Y_4, Y_5) \)
- \( 2^5 = 32 \) response patterns id=1,2,…,32
- ‘freq’ is used as a case weight
Loyalty Model with Time-homogeneous Transitions

Measurement equivalence

Initial State Probability \((b_0)\)

Transition Probabilities \((b)\)

Measurement Model Probabilities \((a)\)
Example 1: LM Model Outperforms Latent Growth Model

2-state time-homogeneous latent Markov model fits well: (p=.77), small L-BVRs

2-class latent growth model (“2-class Regression”) is rejected (p=.0084)

<table>
<thead>
<tr>
<th>L-BVR</th>
<th>2-class Regression</th>
<th>LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>0.000</td>
<td>.0785</td>
</tr>
<tr>
<td>Lag1</td>
<td>8.739* (p = .003)</td>
<td>.0115</td>
</tr>
<tr>
<td>Lag2</td>
<td>0.065</td>
<td>.3992</td>
</tr>
</tbody>
</table>

Lag1 BVR pinpoints problem in LC growth model as failure to explain 1st order autocorrelation
Example 2: Life Satisfaction Model with Time-Heterogeneous Transitions

- N=5,147 respondents
- T= 5 time points
- Dichotomous Y:
  - Satisfied with life?
    - 1=No
    - 2=Yes
- \( Y=(Y_1, Y_2, Y_3, Y_4, Y_5) \)
- \( 2^5 = 32 \) response patterns id=1,2,...,32
- ‘weight’ is used as a case weight

Models:
- Null
- Time heterogeneous LM
- 2-class mixture LM
- Restricted (Mover-Stayer)
Separate transition probability parameters exist for each pair of adjacent time points.
### Example 2: Model Parameters Time-heterogeneous Model

Estimated values for 2-state LM model

<table>
<thead>
<tr>
<th>State[-0]</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.6045</td>
<td>0.3955</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transition Probabilities (b_t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>1987</td>
</tr>
<tr>
<td>1987</td>
</tr>
<tr>
<td>1988</td>
</tr>
<tr>
<td>1988</td>
</tr>
<tr>
<td>1989</td>
</tr>
<tr>
<td>1989</td>
</tr>
<tr>
<td>1990</td>
</tr>
<tr>
<td>1990</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measurement Model Probabilities (a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

Initial State Probability (b_0)
Ex. 2: Mover-Stayer Time-heterogeneous Latent Markov Model

Both the unrestricted and Mover-Stayer 2-class MLM models fit well ($p=.95$ and .71), the BIC statistic preferring the Mover-Stayer model.

<table>
<thead>
<tr>
<th>L-BVR</th>
<th>2-class Regression</th>
<th>2-class LM</th>
<th>2-class LM Mover-Stayer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>0.0</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>Lag1</td>
<td>55.1 ($p=1.1E-13$)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Lag2</td>
<td>1.7</td>
<td>0.10</td>
<td>1.97</td>
</tr>
</tbody>
</table>
Mover-Stayer Time-heterogeneous 2-class MLM Model

Estimated Values output for 2-state time-heterogeneous MLM model with 2 classes

<table>
<thead>
<tr>
<th>Time</th>
<th>Class</th>
<th>State[-1]</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>1987</td>
<td>1</td>
<td>1</td>
<td>0.9591</td>
</tr>
<tr>
<td>1987</td>
<td>2</td>
<td>2</td>
<td>0.0409</td>
</tr>
<tr>
<td>1988</td>
<td>1</td>
<td>1</td>
<td>0.9539</td>
</tr>
<tr>
<td>1988</td>
<td>2</td>
<td>2</td>
<td>0.0461</td>
</tr>
<tr>
<td>1989</td>
<td>1</td>
<td>1</td>
<td>0.9580</td>
</tr>
<tr>
<td>1989</td>
<td>2</td>
<td>2</td>
<td>0.0420</td>
</tr>
<tr>
<td>1990</td>
<td>1</td>
<td>1</td>
<td>0.9572</td>
</tr>
<tr>
<td>1990</td>
<td>2</td>
<td>2</td>
<td>0.0428</td>
</tr>
</tbody>
</table>
Example 2: Mover-Stayer Time-heterogeneous Latent Markov Model

The Estimated Values output shows that 52.25% of respondents are in the Stayer class, who tend to be mostly Satisfied with their lives throughout this 5 year period -- 67.85% are in state 1 (‘Satisfied’ state) initially and remain in that state. In contrast, among respondents whose life satisfaction changed during this 5 year period (the ‘Mover’ class), fewer (54.82%) were in the Satisfied state during the initial year.

- Class Size
- Initial State
- Transition Probabilities – note that class 1 probability of staying in the same state has been restricted to 1.
- Measurement model
Example 2: Longitudinal Profile Plot for the Mover-Stayer LM model

- Stayer class showing 61.75% satisfied each year
- Mover class showing changes over time
Example 2: Longitudinal-Plot with Predicted Probability of Being Satisfied (‘Overall Prob’) Appended
Example 2: L-BVRs for the Mover-Stayer Model
Example 3: Latent GOLD Longitudinal Analysis of Sparse Data

- N=1725 pupils who were of age 11-17 at the initial measurement occasion (in 1976)
- Survey conducted annually from 1976 to 1980 and at three year intervals after 1980
- 23 time points (T+1=23), where t=0 corresponds to age 11 and the last time point to age 33.
- For each subject, data is observed for at most 9 time points (the average is 7.93) which means that responses for the other time points are treated as missing. (See Figure 2)
- Dichotomous dependent variable – ‘drugs’ indicating whether respondent used hard drugs during the past year (1=yes; 0=no).
- Time-varying predictors are ‘time’ (t) and ‘time_2’ (t^2); time-constant predictors are ‘male’ and ‘ethn4’ (ethnicity).
Example 3: Latent GOLD Longitudinal Analysis of Sparse Data

The plot on the left shows the overall trend in drug usage during this period is non-linear, with zero usage reported for 11 year olds, increasing to a peak in the early 20s and then declining through age 33. The plot on the right plots the results from a mixture latent Markov model suggesting that the population consists of 2 distinct segments with different growth rates, Class 2 consisting primarily of non-users.
Example 3: 2-class MLM Model

<table>
<thead>
<tr>
<th>Class</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6556</td>
<td>0.3444</td>
<td></td>
</tr>
</tbody>
</table>

Class size

Initial state probabilities by class

Transition probabilities by class

Measurement model probabilities
Example 3: Including Gender and Ethnicity as Covariates in Model
Example 3: Including Gender and Ethnicity as Covariates in Model

Model Summary output, showing that adding gender and ethnicity improves the fit.
Example 3: Including Gender and Ethnicity as Covariates in Model
For concreteness we will focus on 18 year olds (see highlighted cells in figure). We see that 18 year olds who were in the lower usage state (State 1) at age 17 have a probability of .1876 of switching to the higher usage state (State 2) if they are in Class 2 compared to a probability of only .0211 of switching if they were in Class 1. In addition, if they were in the higher use state (State 2) at age 17, they have a probability of .9589 of remaining in that state compared to only .3636 if they were in Class 1. Thus, based on these different transition probabilities we see that Class 2 is more likely to move to and remain in a higher drug usage state than Class 1.
Appendix: Equations

Latent Markov:

\[
P(y_i) = \sum_{x_0=1}^{K} \sum_{x_1=1}^{K} \cdots \sum_{x_T=1}^{K} P(x_0, x_1, \ldots, x_T) P(y_i \mid x_0, x_1, \ldots, x_T)
\]

\[
P(x_0, x_1, \ldots, x_T) = P(x_0) \prod_{t=1}^{T} P(x_t \mid x_{t-1})
\]

\[
P(y_i \mid x_0, x_1, \ldots, x_T) = \prod_{t=0}^{T} P(y_{it} \mid x_t) = \prod_{t=0}^{T} \prod_{j=1}^{J} P(y_{itj} \mid x_t)
\]

Latent Growth:

\[
P(y_i \mid z_i) = \sum_{w=1}^{L} P(w \mid z_i) \prod_{t=0}^{T} P(y_{it} \mid w, z_{it})
\]
References
