Background

- Traditional psychometric models such as item response theory (IRT) models and cognitive diagnosis models (CDM) are static models.

- However, it is important to understand students’ learning trajectories:
  - Periodic tests during the school year
  - Interaction with intelligent tutors on a daily basis
  - Pre-post tests to evaluate educational interventions

- By understanding students’ learning over time,
  - Educators/intelligent tutors can adjust their instruction
  - Students can focus on improving the skills they lack
Background

Longitudinal psychometric models to understand students’ growth over time

- **IRT-based Longitudinal Models**
  - Multidimensional Rasch models (Andersen, 1985; Embretson, 1991)
  - Longitudinal IRT model with a growth curve (Pastor & Beretvas, 2006)
  - Longitudinal extension of mixture IRT models (Cho et al., 2010)

- **CDMs for assessing change in mastery of latent skills**
  - Latent Transition Analysis CDMs (Li et al., 2016; Kaya & Leite, 2016)
  - Higher-order hidden Markov CDMs (Wang et al., 2017)
  - **Growth curve CDMs (Lee & Rabe-Hesketh, today’s talk)**

- **Dynamic Bayesian Networks for assessing change in knowledge states**
  - Knowledge tracing models (Corbett & Anderson, 1994)
  - Markov decision process (Almond, 2007; LaMar, 2017)
Higher-order latent trait CDMs (de la Torre & Douglas, 2004)

Example: an assessment with 20 items measuring 4 skills.

The skills, $\alpha_{kj}$, are related to one or more broadly defined constructs of general intelligence or aptitude, $\theta_j$

$$\text{logit}[P(\alpha_{kj} = 1|\theta_j)] = \lambda_{0k} + \lambda_{1k}\theta_j$$

- The measurement part is defined by DINA (deterministic inputs, noisy “and” gate), or DINO, etc.
- The Q-matrix should be pre-defined.
Model

Growth curve cognitive diagnosis models (GC-CDM)

A unidimensional latent trait for person \( j \) at occasion \( t \), \( \theta_{jt} \), is modeled as

\[
\theta_{jt} = \zeta_{1j} + (\beta + \zeta_{2j}) \times \text{time}_{jt} + \epsilon_{jt}
\]

- \( \text{time}_{jt} \): time associated with occasion \( t \) for person \( j \)
- \( \beta \): mean slope of time (average growth)
- \( \zeta_{1j} \): random intercept for person \( j \)
- \( \zeta_{2j} \): random slope of time for person \( j \)
- \( \epsilon_{jt} \): time-specific error

\[
\begin{pmatrix}
\zeta_{1j} \\
\zeta_{2j}
\end{pmatrix}
\sim N\left[
\begin{pmatrix}
0 \\
0
\end{pmatrix},
\begin{pmatrix}
\psi_{11} & \psi_{12} \\
\psi_{21} & \psi_{22}
\end{pmatrix}
\right],
\epsilon_{jt} \sim N(0, \sigma_t^2)
\]
Model

Growth curve cognitive diagnosis models (GC-CDM)

\[ P(\alpha_{jt} | \theta_{jt}) = \prod_{k=1}^{K} P(\alpha_{jkt} | \theta_{jt}) \]

where \( P(\alpha_{jkt} = 1 | \theta_{jt}) = \frac{\exp(\lambda_{0k} + \lambda_{1k} \theta_{jt})}{1 + \exp(\lambda_{0k} + \lambda_{1k} \theta_{jt})} \)

With the DINA model, \( \pi_{ijt} = P(Y_{ijt} = 1 | \alpha_{jt}, \theta_{jt}) = (1 - s_{it})^{\xi_{ijt}} g_{it}^{1 - \xi_{ijt}} \)

where \( s_{it} \) and \( g_{it} \) are slipping and guessing parameters of item \( i \) at occasion \( t \)
\( \xi_{ijt} \) is the indicator whether respondent \( j \) possesses all required skills
Model

A GC-CDM for four skills and four time points
Estimation

Maximum Marginal Likelihood Estimation

\[ L(y) = \prod_{j=1}^{J} \int_{\theta_j} \left\{ \prod_{t=1}^{T} \left( P(\alpha_{jt} | \theta_{jt}) \prod_{i=1}^{I} P(y_{ijt} | \alpha_{jt}) \right) \right\} P(\theta_j) d\theta_j \]

where \( \theta_j = (\zeta_{1j}, \zeta_{2j}, \epsilon_{j1}, \ldots, \epsilon_{jT})^T \) and \( P(\theta_j) = P(\zeta_{1j}, \zeta_{2j}) \prod_{t=1}^{T} P(\epsilon_{jt}) \)

- The marginal likelihood of the GC-CDM can be computed by numerical integration techniques (e.g., Gaussian quadrature) - evaluation of this likelihood requires \((T + 2)\)-dimensional integration
- But, when the number of time point increases, the computational complexity increases exponentially
Estimation

Maximum Marginal Likelihood Estimation

\[ L(y) = \prod_{j=1}^{J} \int_{\zeta_{1j}, \zeta_{2j}} \left\{ \prod_{t=1}^{T} \int_{\epsilon_{jt}} \left( P(\alpha_{jt} | \zeta_{1j}, \zeta_{2j}, \epsilon_{jt}) \prod_{i=1}^{I} P(y_{ijt} | \alpha_{jt}) \right) P(\epsilon_{jt}) d\epsilon_{jt} \right\} P(\zeta_{1j}, \zeta_{2j}) d\zeta_{1j} d\zeta_{2j} \]

- Use factorized likelihood with nested integration, reflecting multilevel structure where occasions are nested within persons
  - Only 3-dimensional integration regardless of the number of time points
- The marginal likelihood is maximized using the Expectation Maximization (EM) algorithm
- Estimation was implemented using Mplus
Simulation Design (GC-DINA model)

- Three factors:
  - Number of respondents (1,000 vs 5,000)
  - Design of the Q-matrix (Simple vs Complex)
  - Number of time points (3 vs 4)

- Compared the estimates with,
  - the generating parameters
  - estimates when the skill mastery indicators are observed (mixed-effects logistic model; growth IRT model) - Benchmark 1
  - estimates when the higher-order latent traits are observed (linear growth curve model) - Benchmark 2
Simulation Design (GC-DINA model)

**Simple Q-matrix** of 20 items and skills for the simulation study

<table>
<thead>
<tr>
<th>Skill</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

**Complex Q-matrix** of 20 items and skills for the simulation study

<table>
<thead>
<tr>
<th>Skill</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Simulation Results

- Effect of design of the Q-matrix
  - Benchmark 1 comparison: worse performance with the complex Q-matrix, especially for point estimates

- Effect of sample size
  - Standard errors are larger with smaller sample size, also for benchmarks

- Effect of number of time points
  - No significant change between three time points and four time points

- Overall, good recovery across all conditions, especially the average growth (the parameter of interest)

  ➡ It appears to work reasonably even with the complex Q-matrix and small sample
Application

Study Design

- Two interventions called on Kim’s Koment (KK) and Fraction of the Cost (FOC) (Bottge et al., 2007)
- 109 students from six math classrooms
- 50 males and 59 females in the 7th grade
- Fraction of Cost (FOC) test: 23 items & 4 skills
  (1) “Number & Operation”, (2) Measurement, (3) Problem Solving and (4) Presentation
Application

Result

- The estimated average growth = 1.81 logits; improved over time on average
- The variance of the person-specific random intercept = 7.2
  : large variation between students in their overall higher-order latent traits

Predicted growth trajectories in the higher-order latent traits for 109 students

Proportion of students predicted to have mastered each skill at each occasion

<table>
<thead>
<tr>
<th></th>
<th>t=1</th>
<th>t=2</th>
<th>t=3</th>
<th>t=4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number &amp; Operation</td>
<td>0.49</td>
<td>0.72</td>
<td>0.90</td>
<td>0.99</td>
</tr>
<tr>
<td>Measurement</td>
<td>0.32</td>
<td>0.56</td>
<td>0.77</td>
<td>0.97</td>
</tr>
<tr>
<td>Problem Solving</td>
<td>0.01</td>
<td>0.03</td>
<td>0.07</td>
<td>0.32</td>
</tr>
<tr>
<td>Presentation</td>
<td>0.08</td>
<td>0.18</td>
<td>0.35</td>
<td>0.78</td>
</tr>
</tbody>
</table>
GC-CDM when T=2

When only two time points are available (T= 2) and the timing is identical across subjects, time_{jt} = time_{t}, the growth curve model is not identified

\[
\begin{pmatrix}
\theta_{j1} \\
\theta_{j2}
\end{pmatrix}
\sim N
\left[
\begin{pmatrix}
\mu_1 \\
\mu_2
\end{pmatrix},
\begin{pmatrix}
\sigma_1^2 & \sigma_{12} \\
\sigma_{21} & \sigma_2^2
\end{pmatrix}
\right]
\]

where \((\mu_1, \mu_2)\)' is the mean vector of the higher-order latent traits
\(\sigma_1^2\) and \(\sigma_2^2\) are the variances of \(\theta_{j1}\) and \(\theta_{j2}\) respectively
\(\sigma_{12}(= \sigma_{21})\) is the covariance between \(\theta_{j1}\) and \(\theta_{j2}\)
Thank you!
Questions?
Simulation

Data generation

- Generated response data for 20 items & 4 skills

- Generating parameter values
  - Guessing and slipping parameters $\sim$ uniform(0.1, 0.3)
  - The variance of the random intercept $\psi_{11} = 0.4$
  - The variance of the random slope of time $\psi_{22} = 0.02$
  - The covariance between the random intercept and random slope $\psi_{12} = \psi_{21} = 0.02$
  - The average growth $\beta = 0.3$
  - The variance of the occasion-specific error $\sigma^2 = 0.6$
  - $\lambda_0 = (\lambda_{01}, \lambda_{02}, \lambda_{03}, \lambda_{04}) = (1.51, -1.42, -0.66, 0.50)$
Application

Effects of Enhanced Anchored Instruction (EAI): Kim’s Koment (KK) and Fraction of the Cost (FOC) (Bottge et al., 2007)

**KK** includes video instruction depicting two girls competing in pentathlon events. Here, with instruction from the video anchor, students learn to identify the fastest cars in the race, based on times and distances and also learn to construct the “line of best fit” to predict the speed of the cars when released from various points on the ramp.

**FOC** depicts three middle school students trying to buy materials for a skateboard ramp. The aim is that students learn various concepts and skills and apply them holistically to solve a problem. The skills include (a) calculate the percent of money in a savings account and sales tax on a purchase, (b) read a tape measure, (c) convert feet to inches, (d) decipher building plans, (e) construct a table of materials, (f) compute whole numbers and mixed fractions, (g) estimate and compute combinations, and (h) calculate total cost.
Application

Result

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\psi}_{11}$</th>
<th>$\hat{\psi}_{22}$</th>
<th>$\hat{\psi}_{12}$</th>
<th>$\hat{\beta}$</th>
<th>$\hat{\sigma}^2$</th>
<th>$\hat{\lambda}_{01}$</th>
<th>$\hat{\lambda}_{02}$</th>
<th>$\hat{\lambda}_{03}$</th>
<th>$\hat{\lambda}_{04}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Est</strong></td>
<td>7.20</td>
<td>0.42</td>
<td>-1.74</td>
<td>1.81</td>
<td>1.41</td>
<td>-0.30</td>
<td>-1.61</td>
<td>-6.52</td>
<td>-4.20</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>2.85</td>
<td>0.25</td>
<td>0.84</td>
<td>0.25</td>
<td>0.91</td>
<td>0.39</td>
<td>0.49</td>
<td>0.98</td>
<td>0.69</td>
</tr>
</tbody>
</table>

- The variance of the person-specific random intercept = 7.2
  - large variation between students in their overall higher-order latent traits; the skill mastery is highly correlated across skills (the estimated intraclass correlation of the latent response for skill mastery is 0.6).

- The correlation between random intercept and the random slope of time is close to -1. The negative relationship corresponds to the idea that the EAI treatments were developed to be effective for students with LD. More benefit to the low achieving students.

- The estimated average growth=1.81
  - students' overall abilities improved over time on average; the corresponding odds-ratio of 6.1 means that, for a midian student, the odds of mastering each skill increases by a factor of six between testing occasions.