Advancing Psychometric Methods for Assessing Cognitive Theory and Tracking Learning

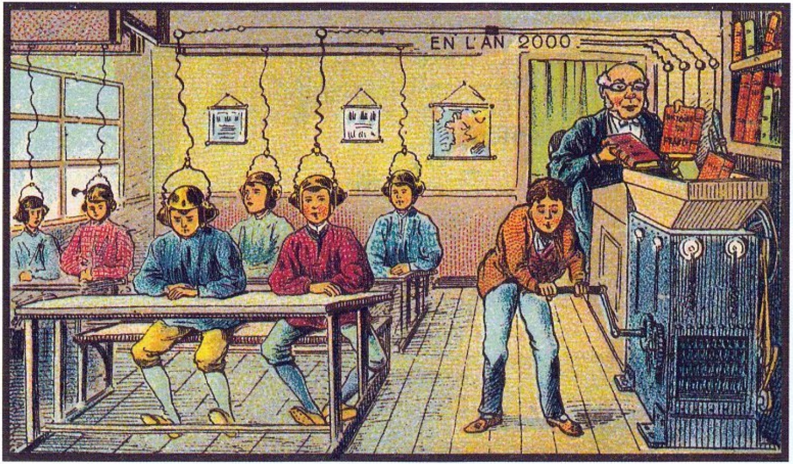
Steven Andrew Culpepper
Department of Statistics
University of Illinois at Urbana-Champaign

This research was partially supported by National Science Foundation Methodology, Measurement, and Statistics program grants #1632023 and #1758631 and Spencer Foundation grant #201700062.

The 18^{th} Annual Maryland Assessment Conference
University of Maryland College Park
Thursday, November 1, 2018, 8:45 - 9:30

School in the Year 2000, Jean-Marc Côté

Postcard from the 1910 World Exhibition in Paris



At School

https://publicdomainreview.org/collections/france-in-the-year-2000-1899-1910/

Jean-Marc Côté's Vision Realized...

Learning Technology

- "Machine-assisted" instruction has been a goal for over a century for some educators and researchers.
- Machine-assisted instruction requires:
  - Theory about the underlying cognitive attributes and processes (i.e., how the attributes interact) needed to master content.
  - Statistical models that support decision-making to advance development.
  - Knowledge about the learning process and how skills are acquired over time.

Overview

My research focuses on statistical methods for diagnostic and learning assessment:

- *Uncovering latent structure*: Developing exploratory methods to infer latent skills and processes.

- *Incorporating expert knowledge*: Creating algorithms that leverage expert knowledge to inform theory development and testing. We need to incorporate expert knowledge to make the results from algorithms more interpretable and useful.

- *Tracking learning trajectories*: Tracking skill acquisition in a dynamic, longitudinal manner.

Collaboration

Colleagues:

- Yuguo Chen
- Jeff Douglas
- Feng Liang

Students:

- James Balamuta
- Yinghan Chen, Univ. Nevada, Reno
- Yinyin Chen
- Ying Liu

Uncovering Latent Structure

- Cognitive diagnosis models (CDMs) specify the underlying structure with a $\boldsymbol{Q}$ matrix.
- Let K be the number of attributes and J the number of items.
- The $J \times K$ binary $\boldsymbol{Q} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_J)'$ matrix indicates which of the K skills are needed for each of the J items.

Example of Fraction-Subtraction Skills

- Experts identified $K = 8$ skills or operations to successfully subtract fractions.

I. Convert a whole number to fraction
II. Separate a whole number from fraction
III. Simplify before subtraction
IV. Find a common denominator
V. Borrow from the whole number part
VI. Column borrow to subtract the 2nd numerator from the 1st
VII. Subtract numerators
VIII. Reduce answers to simplest form

Fraction-Subtraction Example

- $\frac{3}{4} - \frac{3}{8}$ requires the following operations:

  IV. Find a common denominator
  VII. Subtract numerators

- We code the required operations as $\boldsymbol{q} = (0, 0, 0, 1, 0, 0, 1, 0)'$.

Fraction-Subtraction $\boldsymbol{Q}$, Items 1-7 of 20

	Item	I	II	III	IV	V	VI	VII	VIII
1.	$\frac{5}{3} - \frac{3}{4}$	0	0	0	1	0	1	1	0
2.	$\frac{3}{4} - \frac{3}{8}$	0	0	0	1	0	0	1	0
3.	$\frac{5}{6} - \frac{1}{9}$	0	0	0	1	0	0	1	0
4.	$3\frac{1}{2} - 2\frac{3}{2}$	0	1	1	0	1	0	1	0
5.	$4\frac{3}{5} - 3\frac{4}{10}$	0	1	0	1	0	0	1	1
6.	$\frac{6}{7} - \frac{4}{7}$	0	0	0	0	0	0	1	0
7.	$3 - 2\frac{1}{5}$	1	1	0	0	0	0	1	0

Specifying $\boldsymbol{Q}$

- Previously, expert knowledge directed $\boldsymbol{Q}$ specification; however, cognitive theory may be too underdeveloped to form a consensus among experts.
- Correctly specifying $\boldsymbol{Q}$ is fundamental for accurate diagnoses (Henson & Templin, 2007; Rupp & Templin, 2008).
- The unavailability of $\boldsymbol{Q}$ for many content areas poses a barrier to advancing learning technology.
- Exploratory methods are available to estimate $\boldsymbol{Q}$ and infer cognitive processes.

Binary Item Response Data

- We observe binary, correct/incorrect responses, $Y \in \{0, 1\}$.
- In a J item test, $\boldsymbol{Y} \in \{0, 1\}^J$ and there are 2^J possible response patterns.
- In practice, we approximate the high-dimensional space with more parsimonious models.

Cognitive Diagnosis Models (CDMs)

- CDMs classify students into pedagogically meaningful skill profiles.
- CDMs approximate the item response distribution with a fine-grained collection of binary attributes, $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{iK})' \in \{0, 1\}^K$.
- For $K = 3$, $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \alpha_{i3})'$ and there are $2^3 = 8$ profiles:

$$(0,0,0),\ (0,0,1),\ (0,1,0),\ (0,1,1)$$

$$(1,0,0),\ (1,0,1),\ (1,1,0),\ (1,1,1)$$

A General Diagnostic Model (GDM) for Binary Data

- General CDMs include: GDM (von Davier, 2008), GDINA (de la Torre, 2011), and LCDM (Henson et al., 2009).
- Chen, Culpepper, & Liang (2018) discuss a sparse latent class model,

$$P\left(Y_{ij} = 1 \,|\, \boldsymbol{\alpha}_i, \boldsymbol{\beta}_j\right) = \Phi\left(\boldsymbol{a}_i' \boldsymbol{\beta}_j\right)$$

  where $\Phi(\cdot)$ denotes the standard normal CDF.
- $\boldsymbol{\beta}_j$ is a 2^K vector of regression coefficients.
- For $K = 3$,

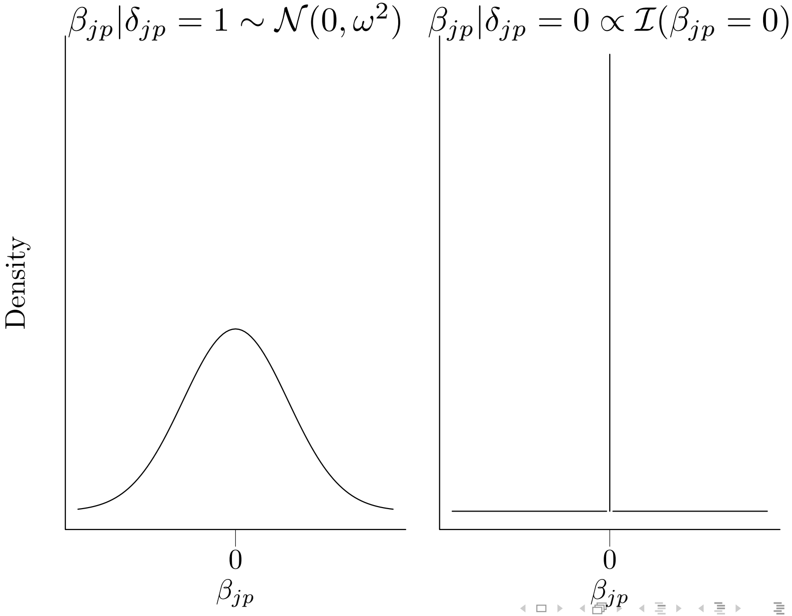
$$\boldsymbol{a}_i = (1, \alpha_1, \alpha_2, \alpha_3, \alpha_1\alpha_2, \alpha_1\alpha_3, \alpha_2\alpha_3, \alpha_1\alpha_2\alpha_3)'$$
$$\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \beta_{j2}, \beta_{j3}, \beta_{j12}, \beta_{j13}, \beta_{j23}, \beta_{j123})'$$

Bayesian Variable Selection for Model Selection

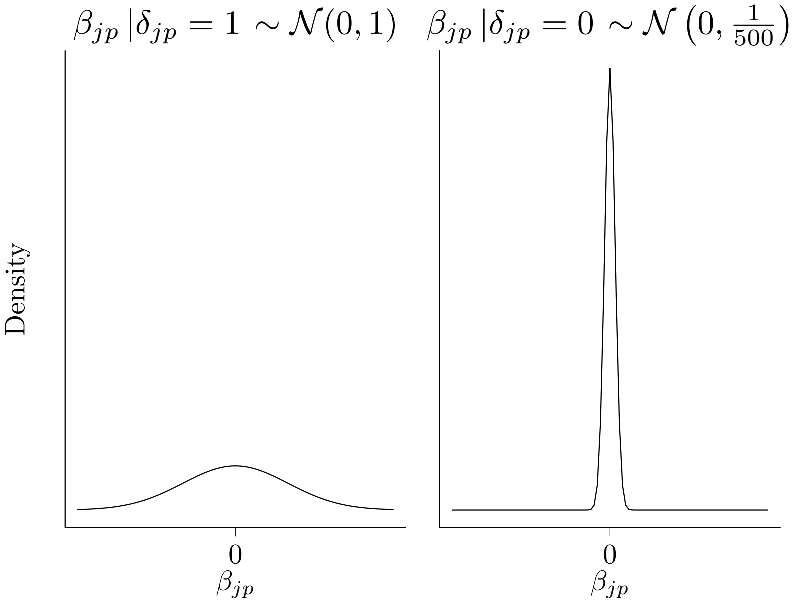
- CDMs are defined by which coefficients are active.
- Model selection is performed by determining which coefficients are nonzero.
- The standard strategy is to let δ_{jp} denote whether β_{jp} is active.
- Let $\delta_{jp} = 1$ indicate that β_{jp} is nonzero and $\delta_{jp} = 0$ otherwise.
- For $K = 3$,

$$\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \beta_{j2}, \beta_{j3}, \beta_{j12}, \beta_{j13}, \beta_{j23}, \beta_{j123})$$
$$\boldsymbol{\delta}_j = (\delta_{j0}, \delta_{j1}, \delta_{j2}, \delta_{j3}, \delta_{j12}, \delta_{j13}, \delta_{j23}, \delta_{j123})'$$

"Spike-Slab" Prior, $\beta_{jp}|\delta_{jp}$



"Spike-Slab" Prior, $\beta_{jp}|\delta_{jp}$



Extending the Standard Approach

- The model selection approach using δ_{jp} does not provide direct inference about $\boldsymbol{q}_j$.
- We use a new prior that specifies a prior for β_{jp} conditioned upon $\boldsymbol{q}_j$.

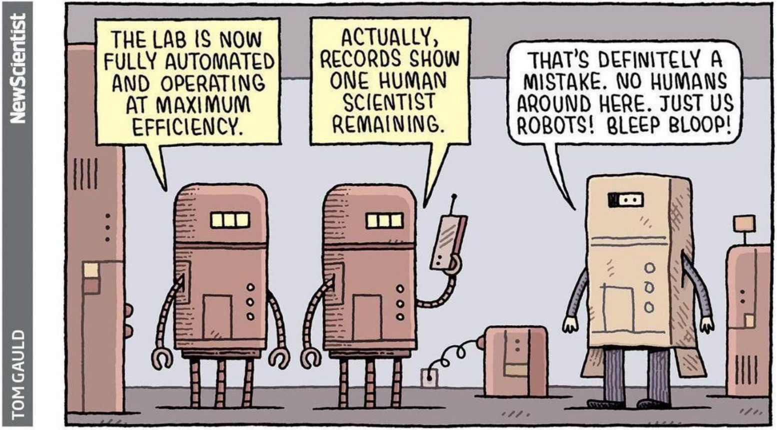
$\beta_{jp}|\boldsymbol{q}_j$

- The "activeness" of each coefficient is denoted by the 2^K vector $\tilde{\boldsymbol{q}}_j$.
- For $K = 3$, $\boldsymbol{q}_j = (q_{j1}, q_{j2}, q_{j3})'$ and

$$\tilde{\boldsymbol{q}}_j = (1, q_{j1}, q_{j2}, q_{j3}, q_{j1}q_{j2}, q_{j1}q_{j3}, q_{j2}q_{j3}, q_{j1}q_{j2}q_{j3})'$$
$$\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \beta_{j2}, \beta_{j3}, \beta_{j12}, \beta_{j13}, \beta_{j23}, \beta_{j123})'$$

- $\tilde{q}_{jp} = 1$ indicates β_{jp} is active and $\tilde{q}_{jp} = 0$ for inactive coefficients.

Leveraging Expert Knowledge



Estimating $\boldsymbol{Q}$ with Expert Knowledge

- Neglecting to use expert knowledge may be sub-optimal given there are 2^{JK} $\boldsymbol{Q}$'s.
  - For $J = 20$ and $K = 7$ there are 1.39×10^{42} different $\boldsymbol{Q}$ matrices.
- Exploratory methods do not always offer a clear interpretation of the uncovered skills.
- From a practical standpoint, we need a clear interpretation of $\boldsymbol{Q}$ to:
  - Design instructional interventions.
  - Develop item banks.

The Bayesian Strategy

- Formulate a prior for $\boldsymbol{Q}$ that incorporates expert knowledge.
- Use variable selection procedures to infer expert variables that explain the underlying structure.
- Use a fully exploratory method to identify *residual*, or unexplained, attributes that are not predicted by cognitive theory.

Expert Knowledge as Predictors

- Incorporate expert knowledge as predictors in a multivariate regression model.

  e.g., A provisional $\boldsymbol{Q}$ summarizes expert knowledge.

- Let x_{jv} be the value of "expert-predictor" v for item j.

- The expert-predictors for item j is $\boldsymbol{x}_j = (1, x_{j1}, \ldots, x_{jV})'$.

- $\boldsymbol{\gamma}_k = (\gamma_{0k}, \gamma_{1k}, \ldots, \gamma_{Vk})'$ relates column k of $\boldsymbol{Q}$ to $\boldsymbol{x}_j$.

Expert Knowledge Prior for $\boldsymbol{Q}$

- We relate $\boldsymbol{Q}$ to expert-predictors in the prior.
- The prior for q_{jk} is,

$$q_{jk}|\boldsymbol{\gamma}_k \overset{\text{ind.}}{\sim} Bernoulli\left[\Phi\left(\boldsymbol{x}_j' \boldsymbol{\gamma}_k\right)\right]$$

- We incorporate sparsity using a spike-slab prior for γ_{vk}.

Application: Fraction-Subtraction Dataset

- Tatsuoka's Fraction-Subtraction data includes responses to $J = 20$ items from $N = 536$ middle school students.
- There are 8 expert-predictors (de la Torre & Douglas, 2004):

I. Convert a whole number to fraction
II. Separate a whole number from fraction
III. Simplify before subtraction
IV. Find a common denominator
V. Borrow from the whole number part
VI. Column borrow to subtract the 2nd numerator from the 1st
VII. Subtract numerators
VIII. Reduce answers to simplest form

Fraction-Subtraction Expert-Predictors

	Item	I	II	III	IV	V	VI	VII	VIII
1.	$\frac{5}{3} - \frac{3}{4}$	0	0	0	1	0	1	1	0
2.	$\frac{3}{4} - \frac{3}{8}$	0	0	0	1	0	0	1	0
3.	$\frac{5}{6} - \frac{1}{9}$	0	0	0	1	0	0	1	0
4.	$3\frac{1}{2} - 2\frac{3}{2}$	0	1	1	0	1	0	1	0
5.	$4\frac{3}{5} - 3\frac{4}{10}$	0	1	0	1	0	0	1	1
6.	$\frac{6}{7} - \frac{4}{7}$	0	0	0	0	0	0	1	0
7.	$3 - 2\frac{1}{5}$	1	1	0	0	0	0	1	0
8.	$\frac{2}{3} - \frac{2}{3}$	0	0	0	0	0	0	1	0
9.	$3\frac{7}{8} - 2$	0	1	0	0	0	0	0	0
10.	$4\frac{4}{12} - 2\frac{7}{12}$	0	1	0	0	1	0	1	1
11.	$4\frac{1}{3} - 2\frac{4}{3}$	0	1	0	0	1	0	1	0
12.	$\frac{11}{8} - \frac{1}{8}$	0	0	0	0	0	0	1	1
13.	$3\frac{3}{8} - 2\frac{5}{6}$	0	1	0	1	1	0	1	0
14.	$3\frac{4}{5} - 3\frac{2}{5}$	0	1	0	0	0	0	1	0
15.	$2 - \frac{1}{3}$	1	0	0	0	0	0	1	0
16.	$4\frac{5}{7} - 1\frac{4}{7}$	0	1	0	0	0	0	1	0
17.	$7\frac{3}{5} - \frac{4}{5}$	0	1	0	0	1	0	1	0
18.	$4\frac{1}{10} - 2\frac{8}{10}$	0	1	0	0	1	1	1	0
19.	$4 - 1\frac{4}{3}$	1	1	1	0	1	0	1	0
20.	$4\frac{1}{3} - 1\frac{5}{3}$	0	1	1	0	1	0	1	0

Relative Model Fit

- We estimated the GDM with $K = 5$, 6, 7, and 8 to ascertain the number of underlying attributes and assessed relative fit of the models using the marginal Deviance Information Criterion (DIC).

- The DIC for $K = 7$ provided the best fit; the DIC for GDMs with $K = 5$, 6, 7, and 8 were 8762, 8678, 8676, and 8740, respectively.

- The DIC for the DINA with $K = 5$ equaled 10804, which suggests the more flexible GDM is needed to describe the latent class probabilities.

Summary of Expert Predictor Results

- $\boldsymbol{Q}$ related to expert-predictors (I) and (IV):

  I. Convert a whole number to fraction
  IV. Find a common denominator

- There is some evidence attribute three relates to (V) "Borrow from the whole number part".

Summary of “Residual” Attributes

- $\widehat{\boldsymbol{Q}}_5$ loads onto items 6 and 12, which only require numerator subtraction.
- $\widehat{\boldsymbol{Q}}_6$ loads onto item 4 and 8 and may be interpreted as, “subtract equivalent fractions”.
- $\widehat{\boldsymbol{Q}}_7$ relates to five items (6, 8, 12, 14, 16), which are the items that could be solved with either only numerator subtraction or both numerator and whole number subtraction.

Benefits of the Expert Q Model

- We are able to validate expert knowledge, in addition to uncovering possibly different cognitive processes (e.g., conjunctive, compensatory, disjunctive, etc.) for the items.
- Using the MVN prior to predict elements of Q significantly improves recovery.
- Several expert-predictors were validated and four residual attributes were uncovered.
- The results could be shared with experts to determine whether there is evidence to refine the cognitive theory.

Next Steps

- Disseminate methods in the `ecdm` R package: `https://github.com/tmsalab/ecdm`.
- `ecdm` currently includes functions for estimating the exploratory DINA and rRUM.
- Develop methods for inferring K.

Learning Trajectories in Cognitive Diagnosis Models

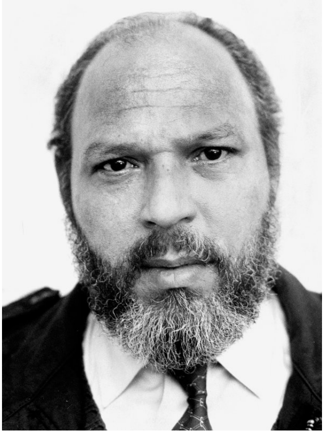
Collaboration with:
Yinghan Chen
Department of Mathematics and Statistics
University of Nevada, Reno

Jeffrey Douglas
Department of Statistics
University of Illinois at Urbana-Champaign

Shiyu Wang
Department of Educational Psychology
University of Georgia

Susu Zhang
Department of Statistics
Columbia University

Learning is a Discontinuous Process



"The universe stuttered, and everything fell into place."
–August Wilson, Playwright

Bessie Smith, *Nobody In Town Can Bake A Sweet Jelly Roll Like Mine*

Learning in Cognitive Diagnosis Models (CDMs)

- *Learning* is the process of transitioning from not knowing to knowing.
- CDMs are ideal for tracking the skill mastery process and for evaluating factors that promote learning.
- For CDMs, learning, or insight, is characterized by $\alpha_{t-1} = 0$ and $\alpha_t = 1$.

Learning Trajectories

- Recent psychometric research considers longitudinal CDMs where students may learn (Kaya & Leite, 2016; Li et al., 2016; Madison & Bradshaw, 2017)
- Online learning technology also uses change-point detection algorithms (Ye et al., 2016).
- This research is also related to “knowledge-tracing”.
- There are also IRT learning models (e.g., Saltus, Mislevy & Wilson, 1996)

Definitions

- Suppose tests are given in T different time points, each time point has J test items.
- Subject i's attribute profile over time (learning trajectories):

  $\boldsymbol{\alpha}_i = (\boldsymbol{\alpha}_{i1}, \ldots, \boldsymbol{\alpha}_{iT})'$, at time t: $\boldsymbol{\alpha}_{it} = (\alpha_{i1t}, \ldots, \alpha_{iKt})'$.

Modeling Learning Trajectories

- The learning trajectory space for binary attributes is high dimensional:
    - 2^{KT} unrestricted trajectories
    - $(T+1)^K$ non-decreasing trajectories
- We considered several approximations of the 2^{KT} learning trajectory space:

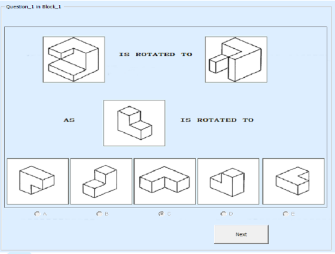
  1) first-order hidden Markov models

  2) latent multivariate growth curves

  3) higher-order factor model.

Ex. #1: Spatial Rotation Reasoning

- Wang et al. (2016) designed a training tool for learning of rotation tasks.
- After each test block, examinees are exposed to learning intervention. 5 test blocks with 10 test items per block under the balanced design.
- Four mental rotation skills are tested:



- 90^0 x-axis
- 90^0 y-axis
- 180^0 x-axis
- 180^0 y-axis

- Contains responses of 351 individuals to 50 test items.

First-Order Hidden Markov Model (FOHM)

Chen, Y., Culpepper, S. A., Wang, S., & Douglas, J. (2018). A Hidden Markov Model for Learning Trajectories in Cognitive Diagnosis with Application to Spatial Rotation Skills. Applied Psychological Measurement, 42, 1, 5 - 23.

- The changes at time t only depend on attribute class at time $t-1$
- Assume invariant transition probabilities over time.
- $\boldsymbol{\pi}_1$ are the baseline probabilities of attribute classes at time $t = 1$.
- The unrestricted and non-decreasing FOHM approximate the learning trajectory space with 4^K and 3^K parameters, respectively.

Estimated Transition Matrix

	$\boldsymbol{\alpha}_c$	$\hat{\boldsymbol{\pi}}_1$	$\hat{\omega}_{1\|i}$	$\hat{\omega}_{2\|i}$	$\hat{\omega}_{3\|i}$	$\hat{\omega}_{4\|i}$	$\hat{\omega}_{5\|i}$	$\hat{\omega}_{6\|i}$	$\hat{\omega}_{7\|i}$	$\hat{\omega}_{8\|i}$	$\hat{\omega}_{9\|i}$	$\hat{\omega}_{10\|i}$	$\hat{\omega}_{11\|i}$	$\hat{\omega}_{12\|i}$	$\hat{\omega}_{13\|i}$	$\hat{\omega}_{14\|i}$	$\hat{\omega}_{15\|i}$	$\hat{\omega}_{16\|i}$
1	0000	0.31	0.84	0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.01	0.01
2	0001	0.02		0.13		0.11		0.18		0.10		0.11		0.12		0.12		0.13
3	0010	0.01			0.12	0.11			0.13	0.13			0.16	0.11			0.12	0.11
4	0011	0.01				0.31				0.25				0.21				0.23
5	0100	0.05					0.37	0.15	0.11	0.08					0.09	0.07	0.07	0.07
6	0101	0.02						0.72		0.12						0.09		0.06
7	0110	0.02							0.50	0.17							0.18	0.15
8	0111	0.01								0.78								0.22
9	1000	0.03									0.13	0.09	0.17	0.10	0.19	0.10	0.12	0.10
10	1001	0.01										0.22		0.25		0.29		0.24
11	1010	0.02											0.73	0.08			0.12	0.07
12	1011	0.02												0.43				0.57
13	1100	0.05													0.72	0.08	0.09	0.12
14	1101	0.02														0.47		0.53
15	1110	0.03															0.55	0.45
16	1111	0.38																1.00

Higher-order Hidden Markov Model

Wang, S, Yang, Y., Culpepper, S. A., & Douglas, J. (2018). Tracking skill acquisition with cognitive diagnosis models: Application to spatial rotation skills. Journal of Educational and Behavioral Statistics, 43, 57-87.

- An alternative is to model transition probabilities by conditioning on a learning factor, θ_i:

$$\omega_{1|0,ik} = p\left(\alpha_{ikt} = 1 | \alpha_{ik,t-1} = 0, \theta_i\right)$$
$$\omega_{1|1,ik} = p\left(\alpha_{ikt} = 1 | \alpha_{ik,t-1} = 1, \theta_i\right)$$

- Non-decreasing restriction implies $\omega_{1|1,ik} = 1$.
- One possibility: $\omega_{1|0,ik} = \Psi\left(\gamma_{0k} + \gamma_{1k}\theta_i + \text{covariates}\right)$ where $\Psi\left(\cdot\right)$ is a cumulative distribution function.

Ex. # 2: Adaptive Content with Evidence-Based Diagnosis (ACED) Evaluation Study

Collaboration with Shannon Sledz, Rahul Kalluri, Ben Olson
Thank you for sharing the data Professors Shute and Almond!

- Shute & Almond (2008) designed a pretest-treatment-posttest study ($N = 268$) to assess the impact of three interventions on students' performance on a geometric sequence test.

- Students completed counter-balanced, parallel forms of 25 items at pre- and post-test.

- Students were randomly assigned to one of four intervention groups.

ACED Study Intervention Groups

During the practice phase,

- Groups 1 ($N_1 = 71$) and 2 ($N_2 = 75$) received adaptive training where practice was based on their solution history.
- Group 3 ($N_3 = 67$) was presented tasks in a pre-designed order.
- Groups 1 and 3 were given feedback to verify the correctness of their solutions and were provided more detailed explanation.
- Group 2 only received feedback as to the correctness of answers.
- Group 4 ($N_4 = 55$) was the control and received content irrelevant to math.

Multivariate Latent Growth Curves

Collaboration with Yinghan Chen and Jeff Douglas

- We use the multivariate probit model for each time point,

$$\alpha_{ikt} = \mathcal{I}\left(\alpha^*_{ikt} > 0\right)$$
$$\boldsymbol{\alpha}^*_{it} \,|\boldsymbol{\beta}_t, \boldsymbol{R}_t \sim \mathcal{N}_K\left(\boldsymbol{x}'_{it}\boldsymbol{\beta}_t, \boldsymbol{R}_t\right)$$

  $\boldsymbol{x}_{it}$ is a vector of covariates, $\boldsymbol{\beta}_t$ are coefficients, and $\boldsymbol{R}_t$ is a correlation matrix.

- We provide a fine-grained assessment of educational interventions.

ACED Design

- There were $T = 2$ time points.
- They identified $K = 8$ skills and specified a $\boldsymbol{Q}$ matrix.
- Notice that prior learning models are less applicable with this smaller dataset.

  Unrestricted and non-decreasing FOHMs require 65,536 and 6,561 parameters to describe learning,respectively.
- Our model requires 68 parameters for this dataset.

Relative Model Fit

- We compared our method with a traditional two-parameter, longitudinal IRT model.
- We assess relative model fit using the marginal Deviance Information Criterion.
- Our method has a DIC value of 14996, which is smaller than the DIC value of 16625 for the longitudinal IRT model.

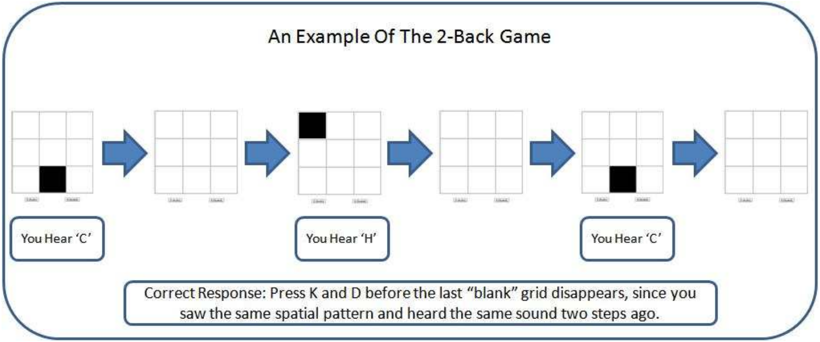
Summary of ACED Application

- Group 1 improved on all eight skills relative to the control group.
- Groups 2 and 3 improved relative to the control on a subset of skills.
- The application provides a fine-grained evaluation of an educational intervention.
- The results could be disseminated to practitioners to recommend which types of feedback promote learning of which skills in an effort to create student-tailored instructional interventions.

Ex. # 3: The n-Back Working Memory Task

Collaboration with Albert Man, Aron Barbey, Chris Zwilling, Evan Anderson, Tanveer Talukdar

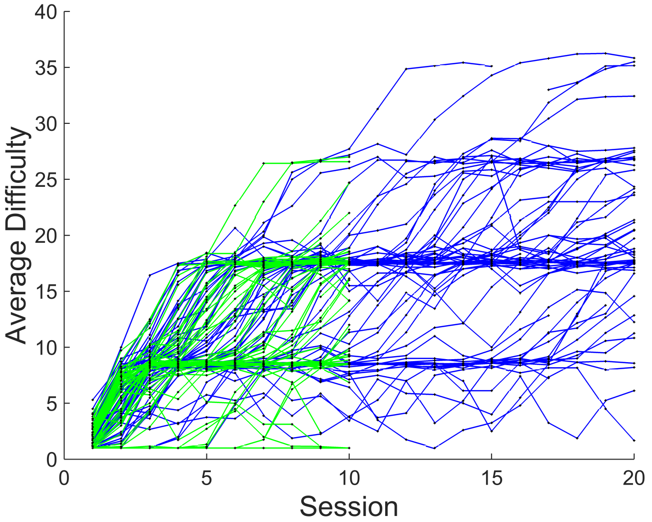
The N-back game tests working memory. Subjects are presented with a sequence of stimuli and asked if the current stimuli matches the one from n steps ago.



The n-Back Working Memory Task

- We modeled learning of the dual n-back working memory task.
- Each observed level of the game differs in the number of visual and audio stimuli.
- We use a polytomous CDM to relate observed and *latent* n-back levels.
- Changes in attributes are modeled with an exploratory factor model.

n-Back Summary Plot



Summary

- Longitudinal CDMs provide new opportunities for fine-grained evaluation of educational interventions.
- Advancing these methods are central to optimizing machine-assisted learning.