Two nonparametric approaches to cognitive dianosis modeling

Brian W. Junker
Carnegie Mellon University
brian@stat.cmu.edu

25th Workshop on Item Response Theory. Oct 12-15, 2009.
Twente University, Enschede, The Netherlands

Outline

- Cognitive Diagnosis Models (CDM's)
- The ASSISTMENTS Project
- Two Nonparametric Approaches in CDM's
  - Nonparametric Skills Inference
  - Nonparametric Skills Discovery
- Some Thoughts

Cognitive Diagnosis Models (CDM's)

- Basic ingredients

X_{ij} = 1 if examinee i performs task j correctly; 0 else

q_{jk} = 1 if skill k is related to task j; 0 else

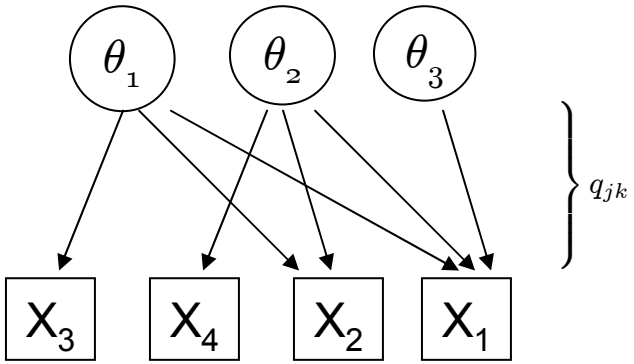
θ_{ik} = 1 if examinee i possesses skill k; 0 else

  - Additional parameters β_j for the structure of $P(\theta_i,\beta_j,Q) = P[X_{ij}=1|\theta_i,\beta_j,Q]$
  - θ_{ik} is latent (knowledge state component of examinee)

    [$\theta_i = (\theta_{i1}, \ldots, \theta_{iK})$ is a latent class label]
  - **Skill = attribute = knowledge component = …**

CDM's

- The Q matrix is the incidence matrix of a bipartite graph (Bayes network; conj/disj/otherwise)



$$Q = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

- Generative model looks much like IRT model:

$$P[X_{i1} = x_{i1}, \ldots, X_{iJ} = x_{iJ} | \theta_i, \beta_j, Q]$$

$$= \prod_{j=1}^{J} P(\theta_i, \beta_j, Q)^{x_{ij}} (1 - P(\theta_i, \beta_j, Q))^{1 - x_{ij}}$$

A simple parametric, conjunctive CDM: Deterministic Input, Noisy "And": DINA

$$\xi_{ij} = \prod_{k=1}^{K} \theta_{ik}^{q_{jk}} = \begin{cases} 1 \text{ if } i \text{ has all } \theta\text{'s for } j \\ 0 \text{ else} \end{cases}$$

$$P_j(\theta_i) = (1 - s_j)^{\xi_{ij}} g_j^{1-\xi_{ij}}, \; \beta_j = (\text{slip } s_j, \text{ guess } g_j)$$

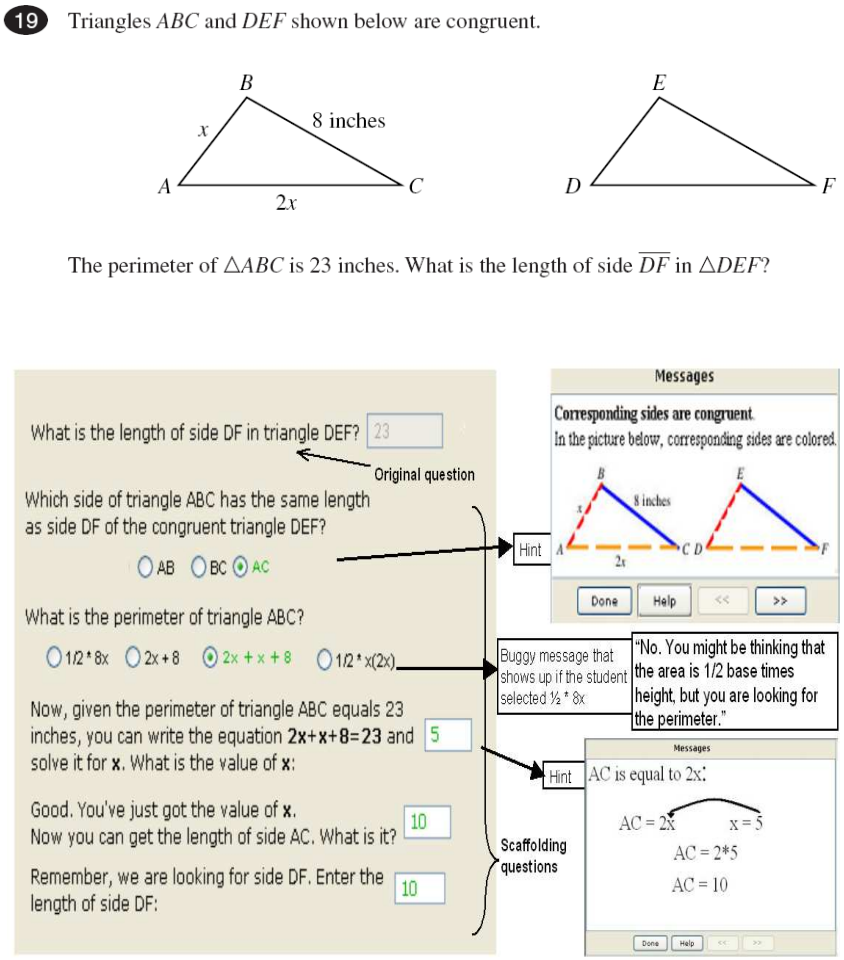
- Origins with Macready & Dayton (1977 *JES*), Tatsuoka (1983 *JEM*),Embretson (1984 *Pmka*), Haertel (1989 *JEM*)
- Reviewed/named by Junker & Sijtsma (2001, *APM*)
- Rediscovered often (e.g. educational data mining)
  - Barnes' (2006 AAAI EDM wkshp) Q-matrix work is based on a model like this with s=g=0
  - Pardos et al (2006 AAAI EDM wkshp) applied to tutoring data

Generalizations Within and Beyond Conjunctive Modeling…

- RUM (Reduced, reparametrized, unified model);
  - Dibello et al (1995, *Cognitively Diagnostic Assessment*, Chipman et al, eds);
  - Roussos et al (2009; *CDA for Education: Theory and Practice*, Leighton & Gierl, eds)
- Expansions…
  - LCDM (Templin, IMPS09; Henson et al, *Pmka* 2009)
  - GDM (von Davier, 2005 ETS TR; 2008 *BJMSP*)
  - GDINA (de la Torre, IMPS09 & later in this workshop)
- Excellent reviews (Rupp & Templin, 2008, *Measurement* ; Fu & Li, 2007, NCME)

The ASSISTMENT Online Tutor

- *Assist* students toward success on Massachusetts Comprehensive System of Assessments (MCAS) in 8th Grade Mathematics
- *Assess* students through time, provides teacher reports per student & per class.
  - ***Predict MCAS Scores***
  - ***KC Feedback: learned/not-learned, etc.***
- Main Items: Released MCAS or "morphs"
- Incorrect Main→ "Scaffold" Items
  - "One-step" breakdowns of main task
  - Buggy feedback, hints on request, etc.
- *All items coded by*
  - *transfer model (Q-matrix) for*
  - *knowledge components (KC's; skills)*
- Led by ***Ken Koedinger*** (CMU Team), ***Neil Heffernan*** (WPI and WPS Team)



2004-2005 Data

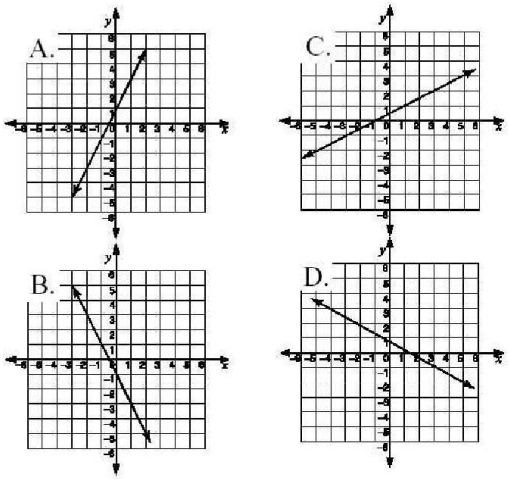
- Tutoring tasks
  - 493 main items (multiple skills)
  - 1216 scaffold items (single skills)
- Students
  - 912 eighth-graders in two middle schools
- Skills Models (Transfer Models / Q Matrices)
  - *1 "Proficiency":* Unidimensional IRT
  - *5 MCAS "strands"*: Number/Operations, Algebra, Geometry, Measurement, Data/Probability
  - *39 MCAS learning standards*: nested in the strands
  - *77 active skills* (~ 120 potential) in a task analysis performed by Neil Heffernan and his staff
- Assistment data available at http://pslcdatashop.org/

Parametric Fit/Calibration: Challenges

- Sheer size of data (tasks, skills, students) can be large
- Published examples tend to be small: 5-10 skills (e.g., de la Torre, 2008, *JEBS*; Templin et al, 2008, *APM;* von Davier, 2008, *BJMSP*)
- E-M good on small K, but combin. explosion as number of skills grow (2^K latent classes…)
- MCMC for DINA using Assistment data
  - Anozie (2007, NCME): ~ 300 tasks, 100 skills, 600 students (1-3 skills/task, 20-40 tasks/student)

Example: Guess and Slip Parameters

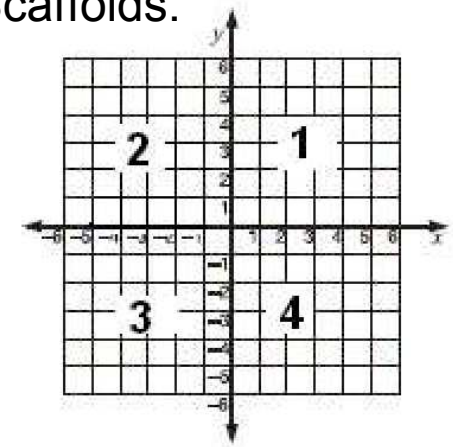
Main Item:



Which graph contains the points in the table?

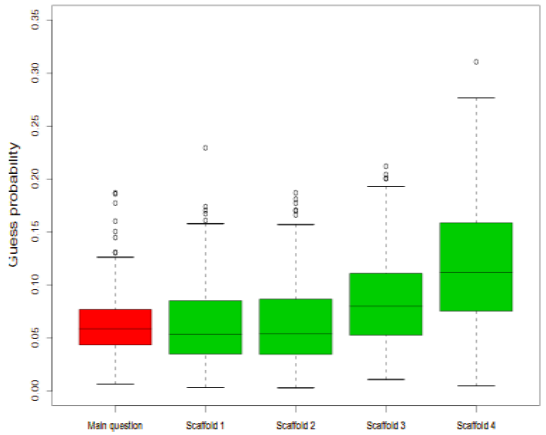
X	*Y*
-2	-3
-1	-1
1	3

Scaffolds:

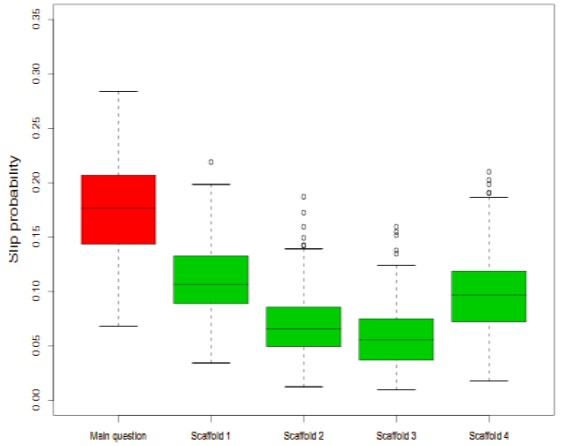


1. Quadrant of (-2,-3)?
2. Quadrant of (-1,-1)?
3. Quadrant of (1,3)?
4. [Repeat main]

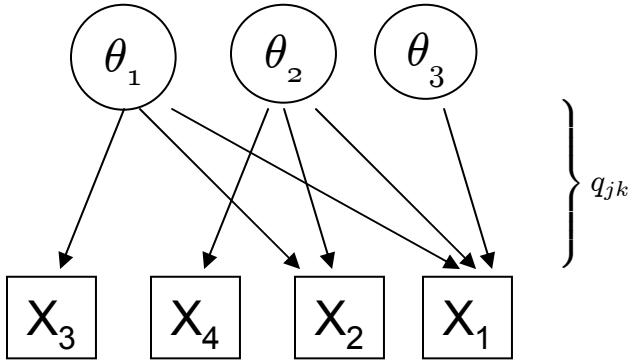
Guess (posterior boxplots)



Slip (posterior boxplots)



Nonparametric Skills Inference: Henson et al.



$$Q = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

- Henson,Templin & Douglas (2007, *JEM*): Given Q=[q_{jk}], compute *sum-scores*

$$W_i = (W_{i1}, \ldots, W_{iK}) \text{ where } W_{ik} = \sum_{j:i \text{ saw } j} x_{ij} q_{jk}$$

as a measure of mastery of individual skills

Nonparametric Skills Inference: Chiu et al.

- Chiu (2008 Dissertation)
  - Recover "perfect" θ_i's from W_i's: K-means, hierarchical agglomerative clustering
  - ***Theorem:*** If there is a non-vanishing proportion of single-skill items for each skill, then as J grows, all 2^K latent classes can be recovered
  - For K=3, 4 skills, N=500 subjects, recovery of all 2^K clusters moderate (ARI~0.50) for J=20 items to excellent (ARI~0.95) for J=80 items

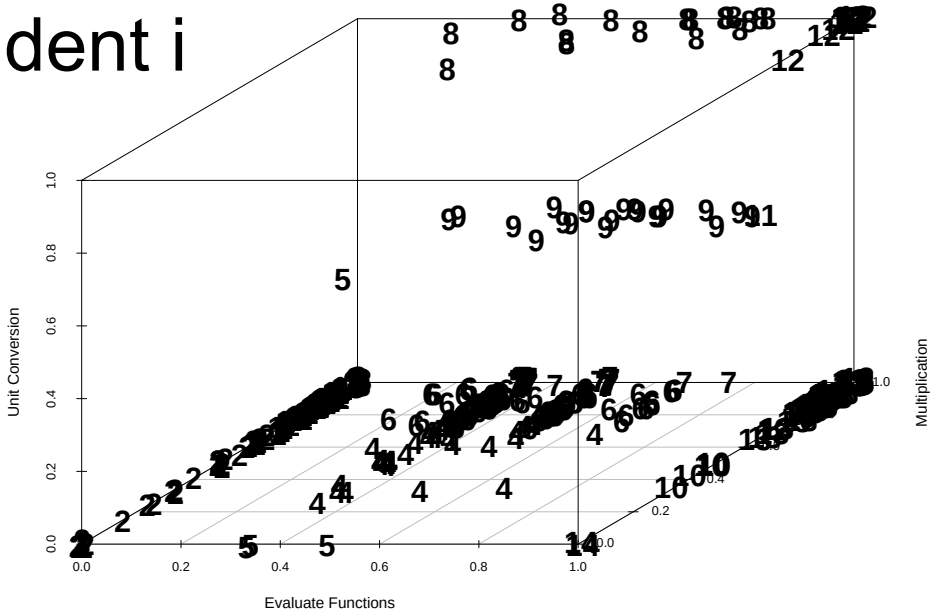
Nonparametric Skills Inference: Ayers

- Ayers, Nugent & Dean (2008, *First Int'l Conf on Educ. Data Mining;* 2009, *IMPS*) consider the *capability matrix* with rows

$$B_i=(B_{i1}, \ldots, B_{iK})$$

where $B_{ik}= W_{ik}/J_{ik}$, J_{ik} = number of tasks with attribute k seen by student i

- Maps sum scores directly into unit hypercube; corners are pure 0/1 attribute/KC profiles
- Accommodates MCAR data
- Use *model-based clustering* (MBC) to identify major skill groups

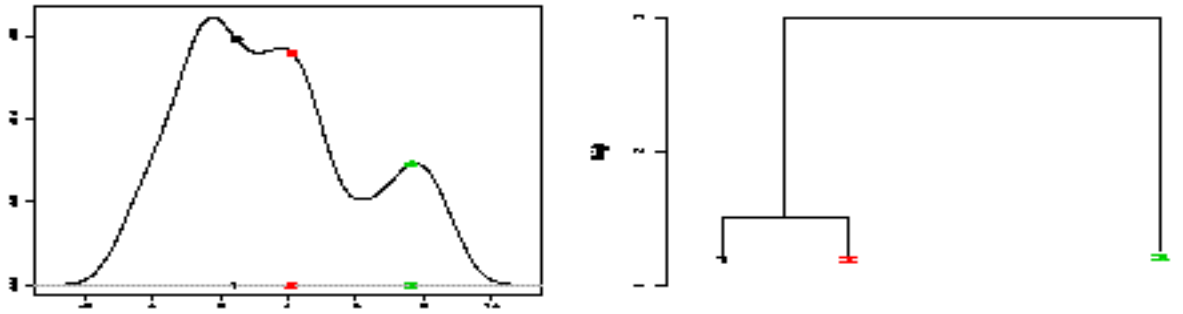


Nonparametric Skills Inference: Ayers

- Typical Result (N=250, J=30, K=3, unbalanced mixed design):

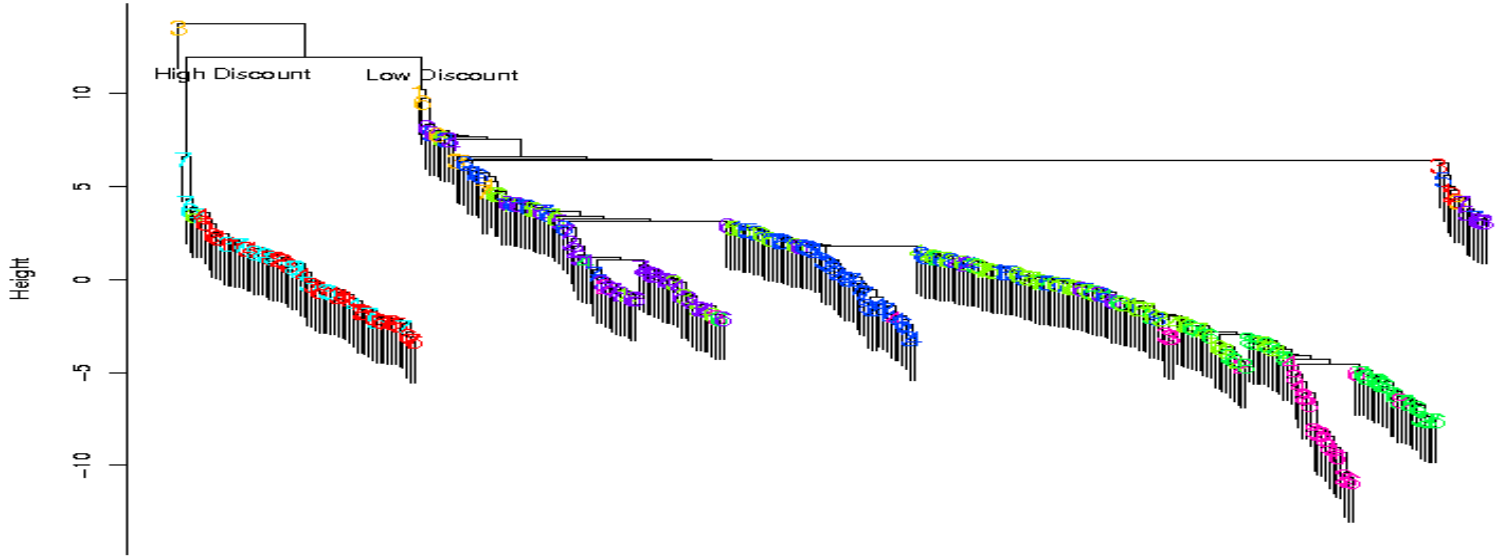
	% Missing				
ARI	0	10	20	30	40
DINA (MCMC)	0.92	0.91	0.88	0.82	0.77
Sum-Score (HAC)	0.65	0.57	0.48	0.37	0.33
Capability (MBC)	0.66	0.55	0.47	0.48	0.43

- Speedups can be substantial. E.g. for K=10 attributes/KC's:
  - Winbugs estimation of DINA: 1 day
  - EM estimation of DINA: 15 min
  - Clustering: 2 sec
- Visualization using Nugent & Dean (2009) minimum-density linkage



Example from Assistments Project

- J = 135 items, K = 13 attributes, N = 344 students



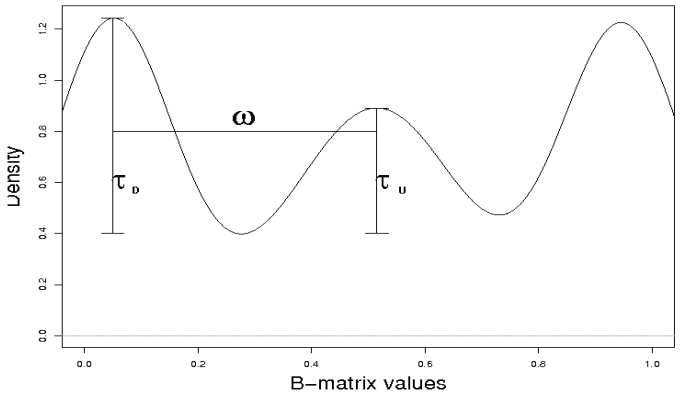
- One outlying student
- Higher dimensional information
  - Equation Solving, Reading a Graph, Substitution
- Really less than 2^{13} interesting groups here

Nonparametric Skills Inference: Ayers

- Ayers et al (2009, 2nd EDM Conference) explore *Conditional*

  Subspace (Pre)Clustering

  - Find dimensions with modes separated by well-def valleys
  - Cluster remaining dimensions, conditional on these modes

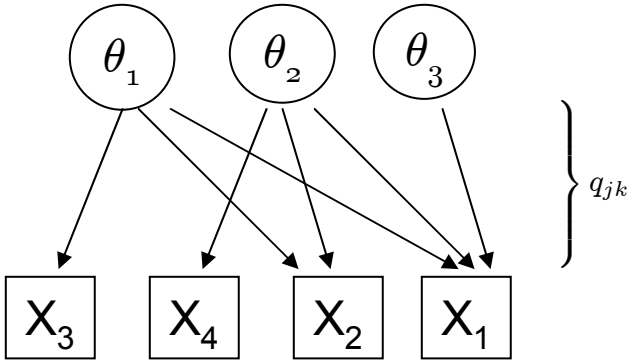


- Often comparable to straight Model-Based Clust (MBC)
- Observed computational savings (simulation):

	DINA-MCMC	DINA-EM	MBC
DINA-EM	96–840		
MBC	3–1900	1–450	
Conditional	85–12000	1–900	3–900

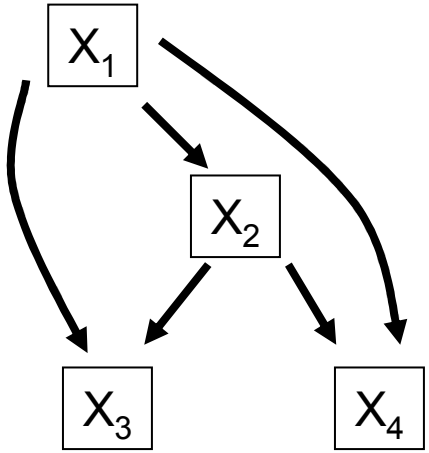
Nonparametric Skills Discovery

- Q-Matrix (Skills Graph):



$$Q = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

- R-Matrix (Surmise Graph):



$$R = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Nonparametric Skills Discovery

- Algebraically, $R = (Q^c \text{ o } Q^T)^c$
  - $Q^c = 1\text{-}Q$
  - o is matrix mult with “and” for * and “or” for +

- Discovering Q:
  - Part I. Estimate R from Data
  - Part II. Find a Q that factors R as above

Nonparametric Skills Discovery: Estimating R from Data

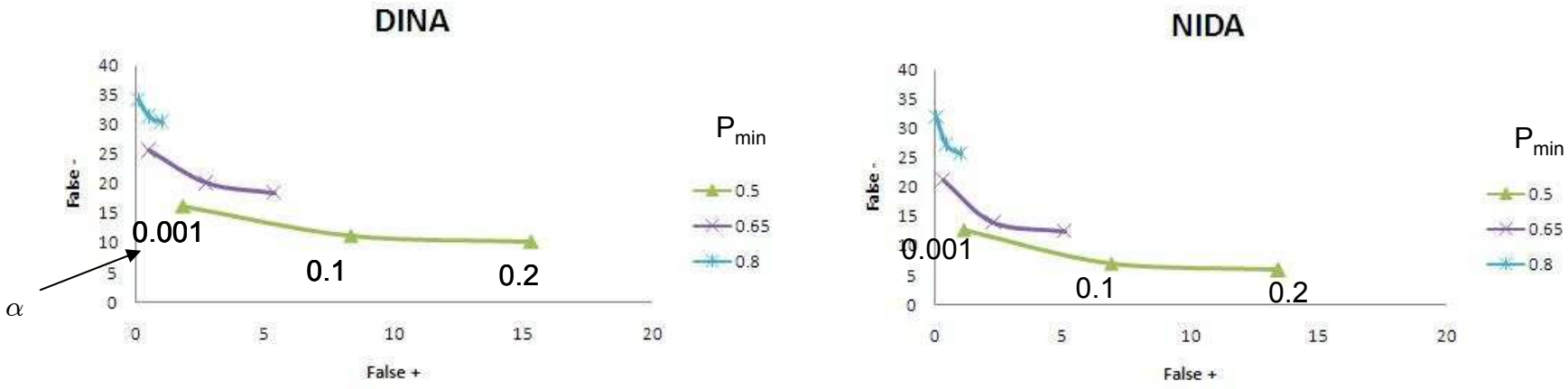
- $X_a \rightarrow X_b$ iff success on X_b can be surmised from success on X_a
  - $P[X_b=1|X_a=1]$ large and $P[X_a=0|X_b=0]$ large
  - Test $Odds(X_b|X_a) \approx n_{11}/n_{10}$ large,
    $Odds(\sim X_a|\sim X_b) \approx n_{00}/n_{10}$ large:

		X_b	
		n_{00}	n_{01}
X_a			
		n_{10}	n_{11}

  - Want n_{10} small; similar to Guttman condition
- Desmarais & Pu (2005, *IntJ AI in Ed)*

Nonparamtric Skills Discovery: Estimating R from Data

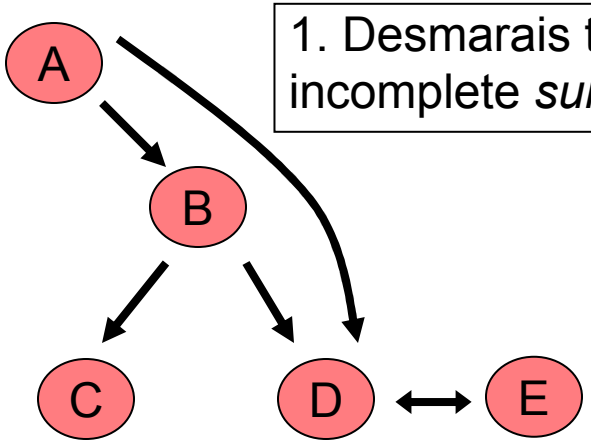
- Tucker (2009 Honors Thesis), Part I:
  - Simulated 45,000 data sets from DINA and NIDA
  - Conditional binomial tests to recover "true" R…



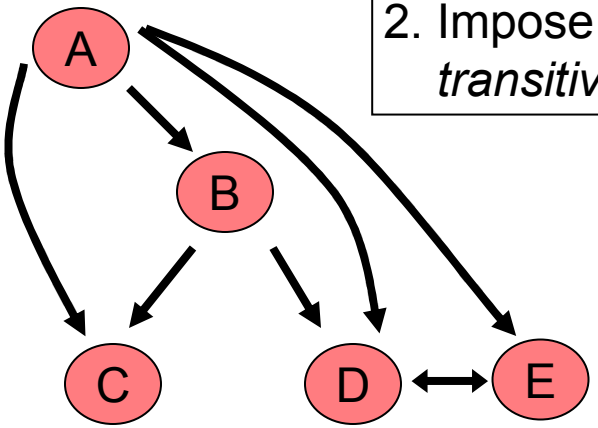
Nonparametric Skills Discovery: Finding Q to Factor R

- Tucker (2009), Part II:
  - Compute transitive closure of estimated R
    - Compensate partially for false-negative edges
    - $R = (Q^c \text{ o } Q^T)^c$ is transitively closed
  - Compute Hasse Diagram from estimated R
    - Minimal directed graph with same transitive closure
    - Combine mutually surmised items into equivalence classes
  - Q is not unique
    - $Q = R^T$ is always possible
    - Minimal Q is NP-hard (Leenen, van Mechelen & DeBoeck, 1999, *JMP*)
    - Find a labelling of Hasse Diagram that is “useful”

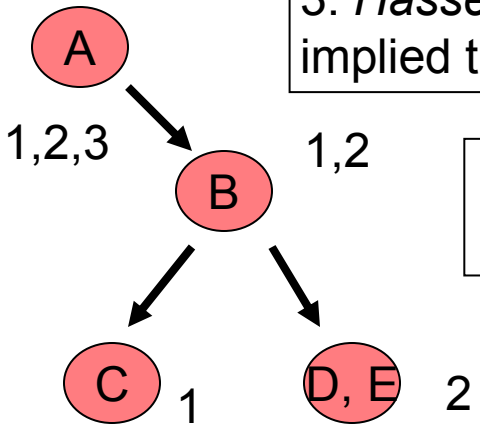
Nonparametric Skills Discovery (Tucker)



1. Desmarais tests produce incomplete *surmise network*



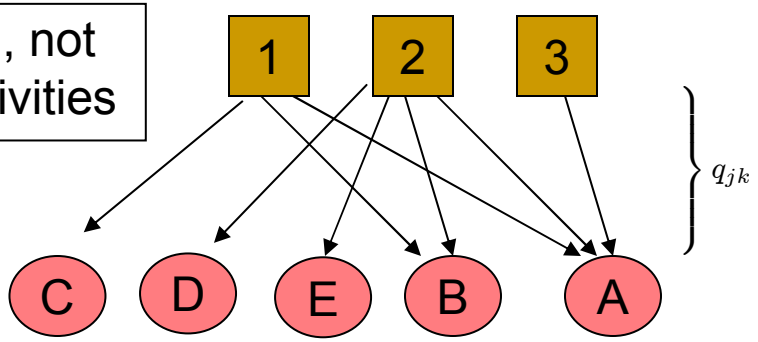
2. Impose constraint of *transitive closure*



3. *Hasse diagram* drops implied transitivities

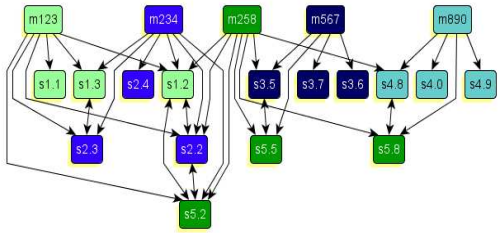
4. *Label* Hasse diagram, not introducing more transitivities

5. Read *Q-matrix* off labelled Hasse diagram

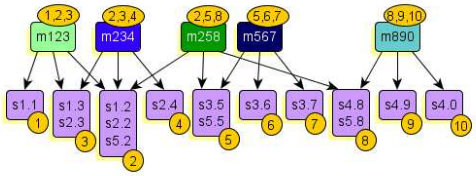


Nonparametric Skills Discovery Examples

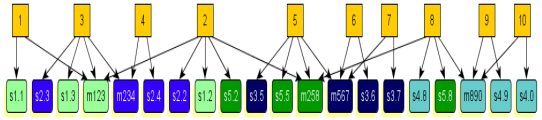
Target R



Target Hasse Diagram

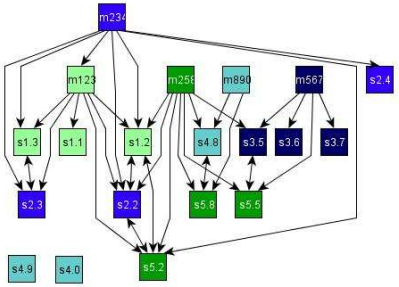


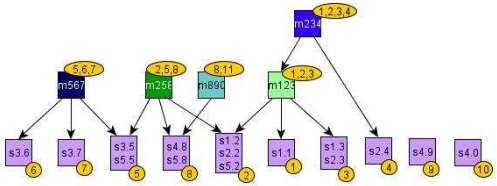
Target Q-Matrix



- 10 attribs in all

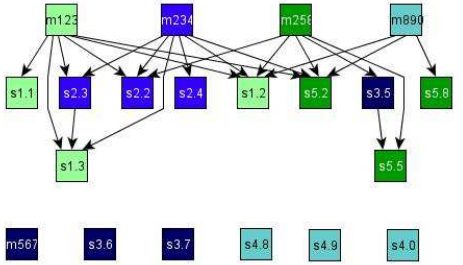
Liberal Discovery

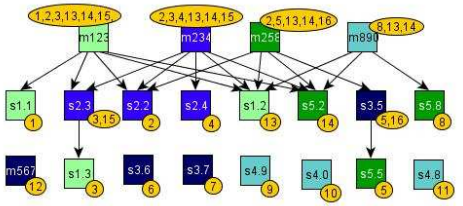




- All equivalence classes detected
- 1 additional skill required for M234
- 11 attribs in all

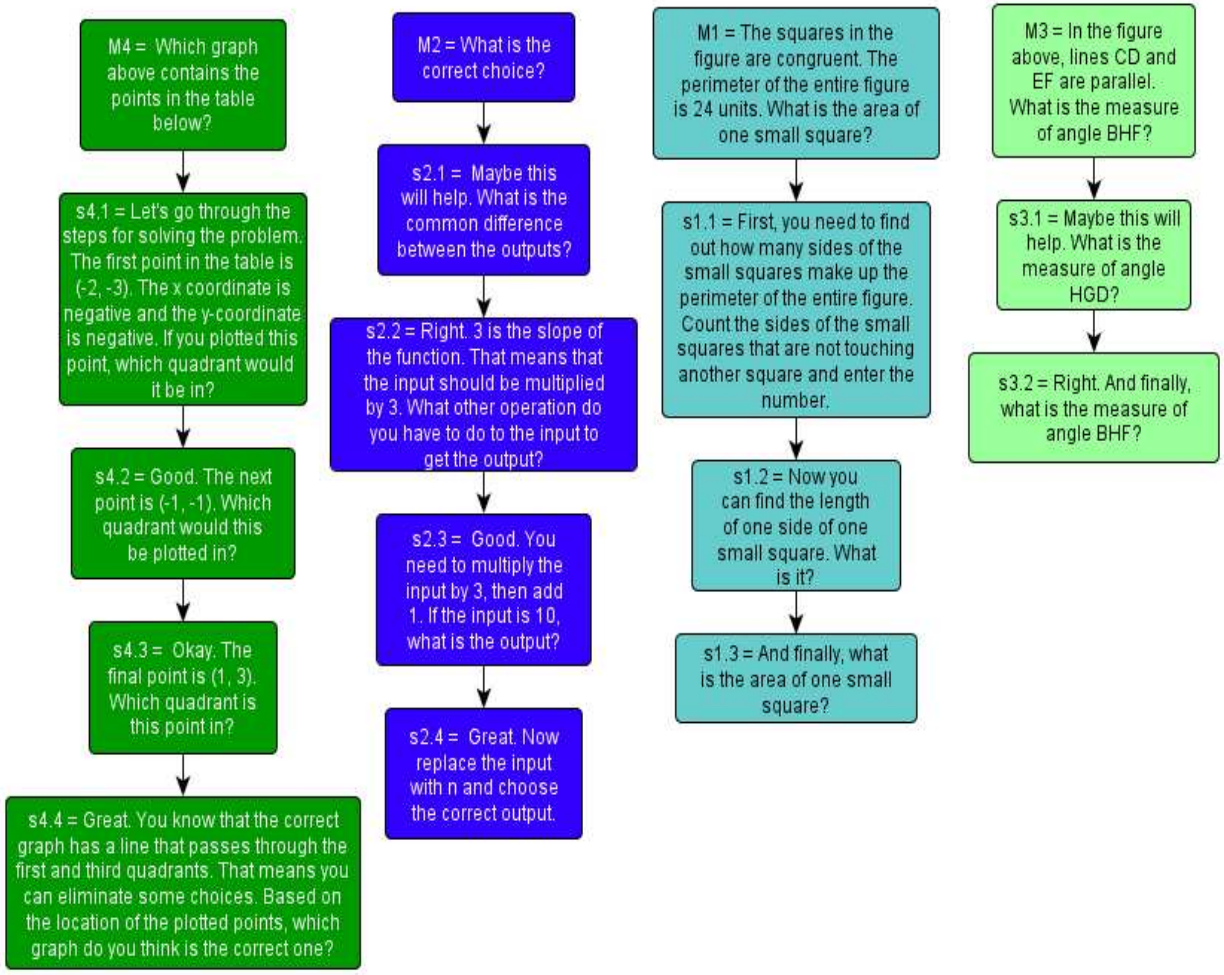
Conservative Discovery





- Many edges & equivalent nodes not detected
- Many more skills needed
- 16 attribs in all

Example from Assistments Project



Some Thoughts…

- Nonparametric approaches to CDM's are in their infancy
- Key motivating factor is computation that scales up as data size scales up
  - parametric model computation need not scale well
- As with Nonparametric IRT, success with
  - Partial sum scores, for scoring students
  - Manifest covariance structure, for discovering latent structure
  - Clustering methods

More Thoughts…

- Building a Nonparametric CDM theory may help understanding of the models as well as speed/scalability of inference
- Data sharing & data repositories matter. http://pslcdatashop.org …and what else?
- The educational data mining community has
  - interesting psychometric problems,
  - lots of data, and
  - competitors/collaborators.

  We should see what they're up to! http://www.educationaldatamining.org

END