A Bayesian Perspective on Structured Mixtures of IRT Models

Robert Mislevy, Roy Levy,
Marc Kroopnick, and Daisy Wise
University of Maryland

Presented at the Conference "Mixture Models in Latent Variable Research," May 18-19, 2006, Center for Integrated Latent Variable Research, University of Maryland

May 18, 2006 CILVR 2006 Slide 1

Probability is not really about numbers;
it is about the structure of reasoning.

Glen Shafer, quoted in Pearl, 1988, p. 77

May 18, 2006 CILVR 2006 Slide 2

Where we are going

- The structure of assessment arguments
- Probability-based reasoning in assessment
- Increasingly complex psychological narratives entail…
  - » Extended view of "data"
  - » More encompassing probability models,
    from Classical test theory to mixtures of structured Item response theory (IRT) models

May 18, 2006 CILVR 2006 Slide 3

The structure of assessment arguments

A construct-centered approach would begin by asking what complex of knowledge, skills, or other attribute should be assessed, …

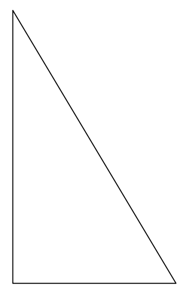
Next, what behaviors or performances should reveal those constructs, and

what tasks or situations should elicit those behaviors?

Messick, 1992, p. 17

May 18, 2006 CILVR 2006 Slide 4

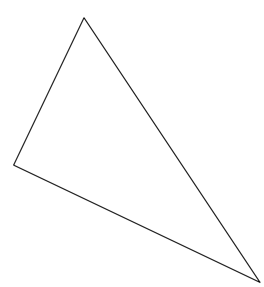
An Example: Mental Rotation Tasks



Stimulus

May 18, 2006 CILVR 2006 Slide 5

An Example: Mental Rotation Tasks



Target

May 18, 2006 CILVR 2006 Slide 6

Total Scores

- Counts of multiple observations which all signify proficiency--More is better.
- Everyone takes same tasks.
- Comparisons/decisions based on total scores X.
- Nature of tasks outside the model.
- No distinction between observation (X) and target of inference (proficiency at mental rotation);
- No probability-based model for characterizing evidence.
- No notion of “measurement error” (except when your score is lower than you think it ought to be)

May 18, 2006 CILVR 2006 Slide 7

Enter probability-based reasoning

A properly-structured statistical model overlays a substantive model for the situation with a model for our knowledge of the situation,

so that we may characterize and communicate what we come to believe—as to both content and conviction—and why we believe it—as to our assumptions, our conjectures, our evidence, and the structure of our reasoning.

Mislevy & Gitomer, 1996

May 18, 2006 CILVR 2006 Slide 8

Defining variables

- A frame of discernment is all the distinctions one can make within a particular model (Shafer, 1976).
  - » "To discern" means "to become aware of" and "to make distinctions among."
- In assessment, the variables relate to the claims we would like to make about students and the observations we need to make.
- All are framed and understood in terms appropriate to the purpose, the context, and psychological perspective that ground the application.

May 18, 2006 CILVR 2006 Slide 9

Conditional independence

- In assessment, the statistical concept of *conditional independence* formalizes the working assumption that if the values of the student model variables were known, there would have no further information in the details.
- We use a model at a given grainsize or with certain kinds of variables not because we think that is somehow "true", but rather because it adequately expresses patterns in the data in light of our perspective on knowledge/skill and the purpose of the assessment.

May 18, 2006 CILVR 2006 Slide 10

Classical Test Theory (CTT)

- Still total scores, but with the idea of replication—multiple "parallel" tests X_j that may differ, but are all "noisy" versions of the same "true score" θ:

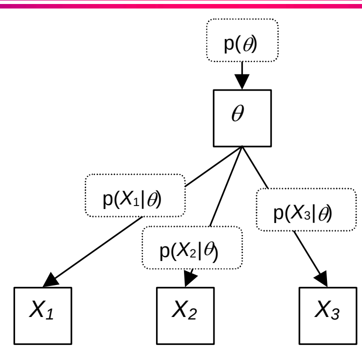
$$X_{ij} = \theta_i + e_{ij} ,$$

where $e_{ij} \sim N(0, \sigma_e^2)$.

- Details of cognition, observation task by task, and content of tasks lie outside the probability model.

May 18, 2006 CILVR 2006 Slide 11

Classical Test Theory (CTT)



- Directed graph representation of Bayesian probability model for multiple parallel tests.
- Note direction of conditional probability in model; contrast w. inference about θ once Xs are observed.

May 18, 2006 CILVR 2006 Slide 12

Classical Test Theory (CTT)

The full probability model:

$$p\left(\mathbf{x},\theta,\mu,\sigma_\theta^2,\sigma_e^2\right)=$$

$$\prod_i\prod_j p\left(x_{ij}\,|\,\theta_i,\sigma_e^2\right)\times p\left(\theta_i\,|\,\mu,\sigma_\theta^2\right)\times p\left(\mu\right)\times p\left(\sigma_\theta^2\right)\times p\left(\sigma_e^2\right).$$

Note what's there, conditional independence, relationships, what's not there.

Posterior inference, via Bayes theorem:

$$p\left(\theta,\mu,\sigma_\theta^2,\sigma_e^2\,\middle|\,\mathbf{x}^*\right)\propto$$

$$\prod_i\prod_j p\left(x_{ij}^*\,|\,\theta_i,\sigma_e^2\right)\times p\left(\theta_i\,|\,\mu,\sigma_\theta^2\right)\times p\left(\mu\right)\times p\left(\sigma_\theta^2\right)\times p\left(\sigma_e^2\right).$$

May 18, 2006 CILVR 2006 Slide 13

Classical Test Theory (CTT)

- For Student *i*, posterior inference is probability distribution for θ_i; that is,
  - » Expected value for test score, along with standard deviation of distribution for θ_i.
  - » People can be seen as differing only as to propensity to make correct responses.
- Inference bound to particular test form.
- Also posterior distribution for mean & variance of θs, error variance, and reliability, or

$$\rho=\sigma_\theta^2\Big/\left(\sigma_\theta^2+\sigma_e^2\right).$$

May 18, 2006 CILVR 2006 Slide 14

Item Response Theory (IRT)

- Modeling now at level of items
  - » Adaptive testing, matrix sampling, test assembly
- Multiple non-parallel item responses X_{ij} that may differ, but all depend on θ_i:

$$\Pr\left(x_{i1},\ldots,x_{iJ}\left|\theta_i,\beta_1,\ldots,\beta_J\right.\right)=\prod_j \Pr\left(x_{ij}\left|\theta_i,\beta_j\right.\right),$$

where β_j is the possibly vector-valued parameter of Item j.

- Conditional independence of item responses given θ and parameter(s) for each item.

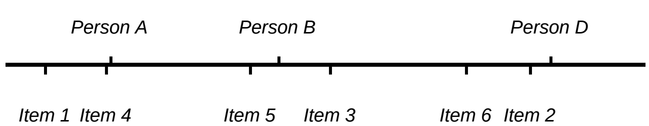
May 18, 2006 CILVR 2006 Slide 15

Item Response Theory (IRT)

- The Rasch IRT model for 0/1 items:

$\text{Prob}(X_{ij}=1|\theta_i,\beta_j) = \Psi(\theta_i - \beta_j)$, where $\Psi(x) = \exp(x)/[1+\exp(x)]$.

←less able θ more able→



←easier β harder→

- Same item ordering for all people.
- Content of tasks still outside the probability model.

May 18, 2006 CILVR 2006 Slide 16

A full Bayesian model for IRT

$$p\left(X,\theta,\beta,\tau,\eta\right)=\prod_i\prod_j p\left(X_{ij}\mid\theta_i,\beta_j\right)p\left(\theta_i\mid\eta\right)p\left(\beta_j\mid\tau\right)p\left(\tau\right)p\left(\eta\right)$$

The measurement model:
Item responses conditionally independent
given person & item parameters

Distributions for person parameters

Distributions for item parameters

Distribution for parameter(s) of distributions
for item parameters

Distribution for parameter(s) of distributions for person parameters

May 18, 2006 CILVR 2006 Slide 17

A full Bayesian model for IRT

Posterior inference:

$$p\left(\theta,\beta,\tau,\eta\middle|\boldsymbol{x}^*\right)\propto$$

$$\prod_i\prod_j p\left(x_{ij}^*\mid\theta_i,\beta_j\right)p\left(\theta_i\mid\eta\right)p\left(\beta_j\mid\tau\right)p\left(\tau\right)p\left(\eta\right)$$

- For people, posterior distributions for θs, or
  propensity to make correct responses.
- How/why outside model.
- For items, posterior distributions for βs.
- Some harder, some easier; How/why outside model.
- Presumed to be the same for all people.

May 18, 2006 CILVR 2006 Slide 18

The “cognitive revolution”

Summary test scores … have often been though of as “signs” indicating the presence of underlying, latent traits.

An alternative interpretation of test scores as samples of cognitive processes and contents, and of correlations as indicating the similarity or overlap of this sampling, is equally justifiable and could be theoretically more useful.

The evidence from cognitive psychology suggests that test performances are comprised of complex assemblies of component information-processing actions that are adapted to task requirements during performance.

(Snow & Lohman, 1989, p. 317)

Research on mental rotation tasks

- Roger Shepard’s studies in the early 70’s shows difficulty of mental rotation tasks depends mainly on how much it is rotated.

Shepard, R. N. & Meltzer, J (1971) Mental rotation of three-dimensional objects. *Science, 171,* 701-703.

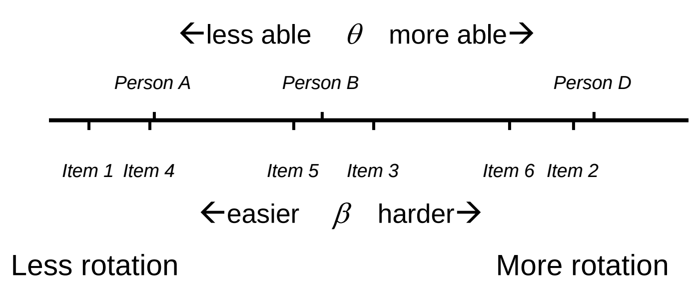
Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual Information Processing* (pp. 75–176). New York: Academic Press.

A structured IRT model: The LLTM

- The linear logistic test model (Fischer, 1973)
- Rasch model but item parameters conditional on item features q_j: $$\beta_j = \sum_k q_{jk}\tau_k = q_j'\tau,$$ where τ_k is a contribution to difficulty from feature *k.*
- Now difficulty modeled as function of task features, as correlated with demands for aspects of knowledge or processing. Conditional independence of item parameters given features: They "explain" item difficulty.
- Task features bring psychological theory into model.
- For mental rotation, can use degree of rotation for q_j .

May 18, 2006 CILVR 2006 Slide 21

A structured IRT model: The LLTM



- Content of tasks inside the probability model.

May 18, 2006 CILVR 2006 Slide 22

A full Bayesian model for LLTM

$$p\left(X,\theta,\beta,\tau,\eta,q\right)=$$
$$\prod_i\prod_j \underbrace{p\left(X_{ij}\mid\theta_i,\beta_j\right)}p\left(\theta_i\mid\eta\right)\underbrace{p\left(\beta_j\mid q_j,\tau\right)}p\left(\tau\right)p\left(\eta\right)$$

The measurement model:
Item responses conditionally independent given person & item parameters

Distributions for item parameters conditional on item features

Note on extensions:
•Rating scale, count, vector-valued observations.
•Multivariate student models for multiple ways people differ; conditional independence given vector θ.
•Different item features relevant to different components of θ.

May 18, 2006 CILVR 2006 Slide 23

A full Bayesian model for LLTM

Posterior inference:

$$p\left(\theta,\beta,\tau,\eta \middle| \mathbf{x}^*,q\right)\propto$$
$$\prod_i\prod_j p\left(x_{ij}^*\mid\theta_i,\beta_j\right)p\left(\theta_i\mid\eta\right)p\left(\beta_j\mid q_j,\tau\right)p\left(\tau\right)p\left(\eta\right).$$

•For people, posterior distributions for θs, or propensity to make correct responses.
•How/why interpreted through psychological model.
•For items, posterior distributions for βs.
•Still resumed to be the same for all people.
•Some harder, some easier; How/why inside model.
•Hypothesized patterns can be checked statistically.

May 18, 2006 CILVR 2006 Slide 24

Structured mixtures of IRT models

- What makes items hard may depend on solution strategy.
  - » John French (1965) on mental rotation.
  - » Siegler re balance beam--development.
  - » Gentner & Gentner on electrical circuits.
  - » Tatsuoka on mixed number subtraction
- Theory says what the relationship ought to be; the trick is putting it into a probability model.

May 18, 2006 CILVR 2006 Slide 25

Structured mixtures of IRT models

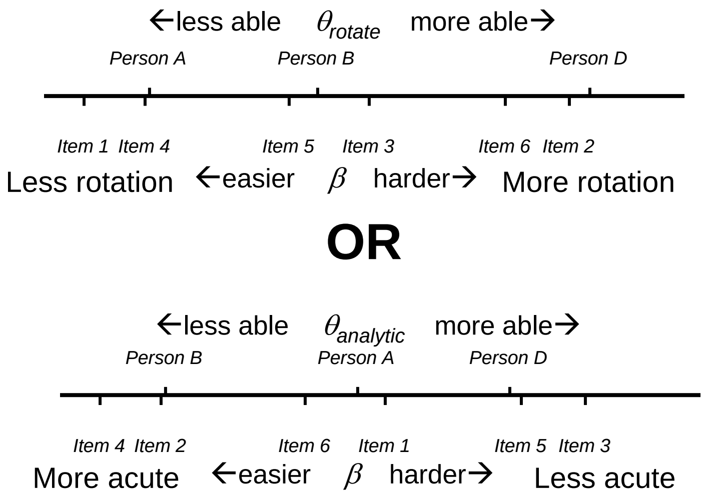
For mental rotation items:

- Difficulty depends on angle of rotation if mental rotation strategy.
- Difficulty depends on acuteness of angle if analytic strategy.
- Can make inferences about group membership using items that are relatively hard under one strategy, relatively easy under the other.

Mislevy, R.J., Wingersky, M.S., Irvine, S.H., & Dann, P.L. (1991). Resolving mixtures of strategies in spatial visualization tasks. *British Journal of Mathematical and Statistical Psychology, 44,* 265-288.

May 18, 2006 CILVR 2006 Slide 26

A structured IRT mixture



May 18, 2006 CILVR 2006 Slide 27

Structured mixtures of IRT models

- Groups of people distinguished by the way they solve tasks.
- ϕ_{ik} =1 if Person *i* is in Group *k*, 0 if not.
- People differ as to knowledge, skills, proficiencies within group, expressed by θ_{ik}'s.
- Items differ as to knowledge, skills, demands within group, expressed by q_{jk}'s.
- Thus, LLTM models within groups.
- Conditional independence of responses given person, item, and group parameters.

May 18, 2006 CILVR 2006 Slide 28

Structured mixtures of IRT models

- Consider M strategies; each person applies one of them to all items, and item difficulty under strategy m depends on features of the task that are relevant under this strategy in accordance with an LLTM structure.
- The difficulty of item j under strategy m is $b_{jm} = \sum_k q_{jmk}\eta_{mk}$
- The probability of a correct response is

$$\Pr\left(X_{ij}=1 \middle| \theta_i, \phi_i, q_j, \eta\right) = \prod_m \left[\Psi\left(\theta_{im} - \sum_k q_{jmk}\eta_{mk}\right)\right]^{\phi_i}.$$

May 18, 2006 CILVR 2006 Slide 29

Structured mixtures of IRT models

$$p\left(X,\theta,\phi,\beta,\tau,\eta,q\right) =$$

$$\prod_i \prod_j \prod_{k \ni \phi_{ik}=1} \underbrace{p\left(X_{ij} \mid \theta_{ik}, \beta_{jk}, \phi_i\right)} p\left(\theta_{ik} \mid \eta_k\right) p\left(\phi_i\right) \underbrace{p\left(\beta_{jk} \mid q_{jk}, \tau_k\right)} p\left(\tau_k\right) p\left(\eta_k\right)$$

Item responses conditionally independent given person's group and person & item parameters relevant to that group.

Distributions for item parameters conditional on item features and feature effects relevant to each group

May 18, 2006 CILVR 2006 Slide 30

Structured mixtures of IRT models

Posterior inference:

$$p\left(\theta,\phi,\beta,\tau,\eta,\middle|\mathbf{x}^{*},q\right)\propto$$

$$\prod_{i}\prod_{j}\prod_{k\ni\phi_{ik}=1} p\left(x_{ij}^{*}\mid\theta_{ik},\beta_{jk},\phi_{i}\right)p\left(\theta_{ik}\mid\eta_{k}\right)p\left(\phi_{i}\right)p\left(\beta_{jk}\mid q_{jk},\tau_{k}\right)p\left(\tau_{k}\right)p\left(\eta_{k}\right)$$

- •For people, posterior probs for ϕs, or group memberships.
- •Posterior distributions for θs within groups.
- •How/why interpreted through psychological model.
- •For items, posterior distributions for βs for each group.
- •Items differentially difficult for different people, based on theories about what makes items hard under different strategies.

May 18, 2006 CILVR 2006 Slide 31

Conclusion

- Evolution of psychometric models joint with psychology.
- Bayesian probability framework well suited for building models that correspond to "narratives"
- Can't just "throw data over the wall" like was done with CTT;
- Need to build coordinated observational model and probability model, from psychological foundation.

May 18, 2006 CILVR 2006 Slide 32