Bayesian IRT Modeling

Jim Albert

Bowling Green State University

October 12, 2009

Outline

18th IRT Workshop (7 years ago)

18th IRT Workshop (7 years ago)

- Why Bayes?

18th IRT Workshop (7 years ago)

- Why Bayes?
- Ease of fitting IRT models by MCMC.

18th IRT Workshop (7 years ago)

- Why Bayes?
- Ease of fitting IRT models by MCMC.
- Future of Bayes IRT?

18th IRT Workshop (7 years ago)

- Why Bayes?
- Ease of fitting IRT models by MCMC.
- Future of Bayes IRT?
    - Software to simply MCMC fitting

18th IRT Workshop (7 years ago)

- Why Bayes?
- Ease of fitting IRT models by MCMC.
- Future of Bayes IRT?
  - Software to simply MCMC fitting
  - Developments in hierarchical/multilevel modeling

18th IRT Workshop (7 years ago)

- Why Bayes?
- Ease of fitting IRT models by MCMC.
- Future of Bayes IRT?
  - Software to simply MCMC fitting
  - Developments in hierarchical/multilevel modeling
  - Development in Bayesian model checking

What's Happened in 7 Years?

What's Happened in 7 Years?

- R – the standard for the development of statistical software

What's Happened in 7 Years?

- R – the standard for the development of statistical software
- R packages are available for fitting Bayesian IRT models

What's Happened in 7 Years?

- R – the standard for the development of statistical software
- R packages are available for fitting Bayesian IRT models
- Developments in Bayesian IRT modeling

What's Happened in 7 Years?

- R – the standard for the development of statistical software
- R packages are available for fitting Bayesian IRT models
- Developments in Bayesian IRT modeling
- Developments in Bayesian model checking

Bayesian IRT Literature

Bayesian IRT Literature

- Bayesian modeling by MCMC

Bayesian IRT Literature

- Bayesian modeling by MCMC
- Data augmentation and Gibbs sampling (Albert, Patz and Junker)

Bayesian IRT Literature

- Bayesian modeling by MCMC
- Data augmentation and Gibbs sampling (Albert, Patz and Junker)
- Provides a framework for more sophisticated Bayesian modeling.

Bayesian Multilevel Modeling

Bayesian Multilevel Modeling

- Modeling of ability parameters (Fox and Glas).

Bayesian Multilevel Modeling

- Modeling of ability parameters (Fox and Glas).
- Combining IRT models across groups.

Bayesian Multilevel Modeling

- Modeling of ability parameters (Fox and Glas).
- Combining IRT models across groups.
- Modeling clustering patterns such as tests composed of testlets.

Bayesian Multilevel Modeling

- Modeling of ability parameters (Fox and Glas).
- Combining IRT models across groups.
- Modeling clustering patterns such as tests composed of testlets.
- Natural to think of multilevel models from a Bayesian perspective.

Bayesian Modeling Checking

Bayesian Modeling Checking

- Posterior predictive checking (Sinharray)

Bayesian Modeling Checking

- Posterior predictive checking (Sinharray)
- Comparing IRT models by Bayes factors?

Bayesian Modeling Checking

- Posterior predictive checking (Sinharray)
- Comparing IRT models by Bayes factors?
- Seems like there is a need for further work in this area.

Is Bayesian MCMC Popular?

Is Bayesian MCMC Popular?

- Ayala, *The Theory and Practice of Item Response Theory*, 2009.

Is Bayesian MCMC Popular?

- Ayala, *The Theory and Practice of Item Response Theory*, 2009.
- Mentions MCMC as a footnote on page 97.

Is Bayesian MCMC Popular?

- Ayala, *The Theory and Practice of Item Response Theory*, 2009.
- Mentions MCMC as a footnote on page 97.
- "Estimation of models ... is comparatively easier with MCMC than with either JMLE or marginal mle's"

Is Bayesian MCMC Popular?

- Ayala, *The Theory and Practice of Item Response Theory*, 2009.
- Mentions MCMC as a footnote on page 97.
- "Estimation of models ... is comparatively easier with MCMC than with either JMLE or marginal mle's"
- "However, MCMC implementation software is not as user-friendly"

Is Bayesian MCMC Popular?

- Ayala, *The Theory and Practice of Item Response Theory*, 2009.
- Mentions MCMC as a footnote on page 97.
- "Estimation of models ... is comparatively easier with MCMC than with either JMLE or marginal mle's"
- "However, MCMC implementation software is not as user-friendly"
- "Given this book's orientation, MCMC methods are not covered"

Is Bayesian MCMC Dangerous?

Is Bayesian MCMC Dangerous?

- WinBUGS software User Manual states in User Manual:

Is Bayesian MCMC Dangerous?

- WinBUGS software User Manual states in User Manual:
- "Beware: MCMC sampling can be dangerous!"

Is Bayesian MCMC Dangerous?

- WinBUGS software User Manual states in User Manual:
- "Beware: MCMC sampling can be dangerous!"
- Some truth to this.

Is Bayesian MCMC Dangerous?

- WinBUGS software User Manual states in User Manual:
- "Beware: MCMC sampling can be dangerous!"
- Some truth to this.
- Have to think about starting values, monitor convergence, etc.

Is Bayesian MCMC Dangerous?

- WinBUGS software User Manual states in User Manual:
- "Beware: MCMC sampling can be dangerous!"
- Some truth to this.
- Have to think about starting values, monitor convergence, etc.
- But MCMC fitting works well for many applications.

Example: BGSU Math Placement Test

- Most freshmen take a math placement test.

Example: BGSU Math Placement Test

- Most freshmen take a math placement test.
- Focus on Form B of the test (students have had two years of high school algebra).

Example: BGSU Math Placement Test

- Most freshmen take a math placement test.
- Focus on Form B of the test (students have had two years of high school algebra).
- Use test results together with other info (ACT math score, high school GPA) to determine proper placement.

Example: BGSU Math Placement Test

- Most freshmen take a math placement test.
- Focus on Form B of the test (students have had two years of high school algebra).
- Use test results together with other info (ACT math score, high school GPA) to determine proper placement.
- Procedure seems to be generally effective in placing students in the "right" math course.

Questions

- What factors are most helpful in predicting student success in math courses?

Questions

- What factors are most helpful in predicting student success in math courses?
- Can we improve the math placement test?

Questions

- What factors are most helpful in predicting student success in math courses?
- Can we improve the math placement test?
- Are there redundant questions that can be removed?

Questions

- What factors are most helpful in predicting student success in math courses?
- Can we improve the math placement test?
- Are there redundant questions that can be removed?
- Are there particular questions that are helpful in discriminating between good and poor students?

General Two-Parameter IRT Model

- Have n students and k items, observe data matrix $\{y_{ij}\}$.

General Two-Parameter IRT Model

- Have n students and k items, observe data matrix $\{y_{ij}\}$.
- The probability that ith student gets jth question correct is given by

$$p_{ij} = \Phi\left(\beta_j \theta_i - \alpha_j\right).$$

General Two-Parameter IRT Model

- Have n students and k items, observe data matrix $\{y_{ij}\}$.
- The probability that ith student gets jth question correct is given by

$$p_{ij} = \Phi\left(\beta_j \theta_i - \alpha_j\right).$$

- Likelihood function of all unknowns is

$$L = \prod_i \prod_j p_{ij}^{y_{ij}} (1 - p_{ij})^{1 - y_{ij}}.$$

MCMCpack package in R

MCMCpack package in R

- Performs MCMC fitting for a variety of common Bayesian models.

MCMCpack package in R

- Performs MCMC fitting for a variety of common Bayesian models.
- Function `MCMCirt1d` implements Gibbs sampling for the 2-parameter model

MCMCpack package in R

- Performs MCMC fitting for a variety of common Bayesian models.
- Function `MCMCirt1d` implements Gibbs sampling for the 2-parameter model
- Inputs:

MCMCpack package in R

- Performs MCMC fitting for a variety of common Bayesian models.
- Function `MCMCirt1d` implements Gibbs sampling for the 2-parameter model
- Inputs:
  - data matrix `y`

MCMCpack package in R

- Performs MCMC fitting for a variety of common Bayesian models.
- Function `MCMCirt1d` implements Gibbs sampling for the 2-parameter model
- Inputs:
  - data matrix `y`
  - parameters of priors placed on item parameters

MCMCpack package in R

- Performs MCMC fitting for a variety of common Bayesian models.
- Function `MCMCirt1d` implements Gibbs sampling for the 2-parameter model
- Inputs:
  - data matrix `y`
  - parameters of priors placed on item parameters
  - aspects of MCMC fitting

MCMCpack package in R

- Performs MCMC fitting for a variety of common Bayesian models.
- Function `MCMCirt1d` implements Gibbs sampling for the 2-parameter model
- Inputs:
  - data matrix `y`
  - parameters of priors placed on item parameters
  - aspects of MCMC fitting
  - starting values

MCMCpack package in R

- Performs MCMC fitting for a variety of common Bayesian models.
- Function `MCMCirt1d` implements Gibbs sampling for the 2-parameter model
- Inputs:
  - data matrix `y`
  - parameters of priors placed on item parameters
  - aspects of MCMC fitting
  - starting values
  - number of MCMC iterations

What Priors Does it Allow?

What Priors Does it Allow?

- $\theta_1, \ldots, \theta_N$ $\mathsf{N}(0, s_\theta)$

What Priors Does it Allow?

- $\theta_1, ..., \theta_N$ N$(0, s_\theta)$
- (α_j, β_j) iid N(μ, Σ)

What Priors Does it Allow?

- $\theta_1, ..., \theta_N$ N$(0, s_\theta)$
- (α_j, β_j) iid N(μ, Σ)
- Default values $s_\theta = 1, \mu = 0, \Sigma = 2I$.

Run of this R function

Run of this R function

- s=apply(y,1,mean)
- theta.start=(s-mean(s))/sd(s)

Run of this R function

- s=apply(y,1,mean)
- theta.start=(s-mean(s))/sd(s)
- fit=MCMCirt1d(y, store.item = TRUE, store.ability = TRUE, mcmc=10000, burnin=0, theta.start=theta.start)

Run of this R function

- s=apply(y,1,mean)
- theta.start=(s-mean(s))/sd(s)
- fit=MCMCirt1d(y, store.item = TRUE, store.ability = TRUE, mcmc=10000, burnin=0, theta.start=theta.start)
- (76 seconds on my laptop)

MCMC Diagnostics

MCMC Diagnostics

- Have the simulated draws converged to the posterior distribution?

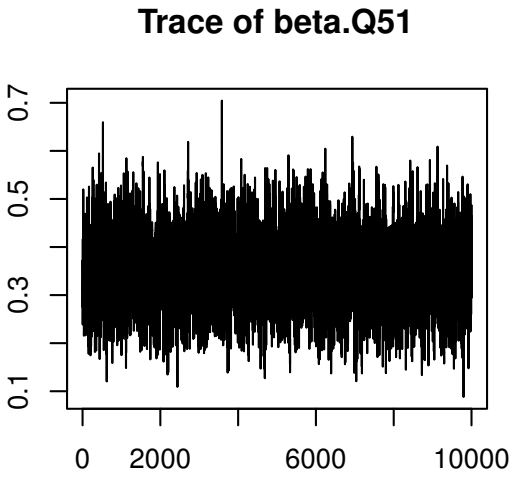
MCMC Diagnostics

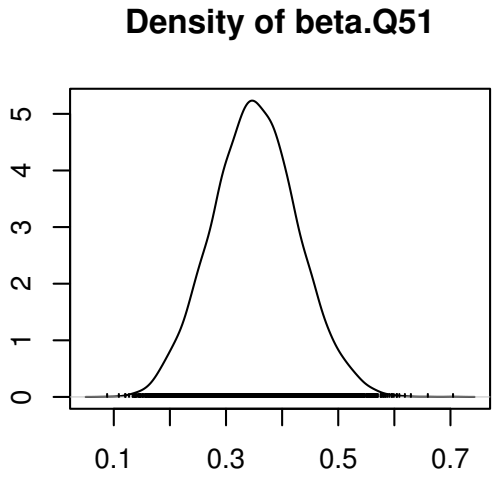
- Have the simulated draws converged to the posterior distribution?
- Use a R package `coda` that contains a number of diagnostic functions.

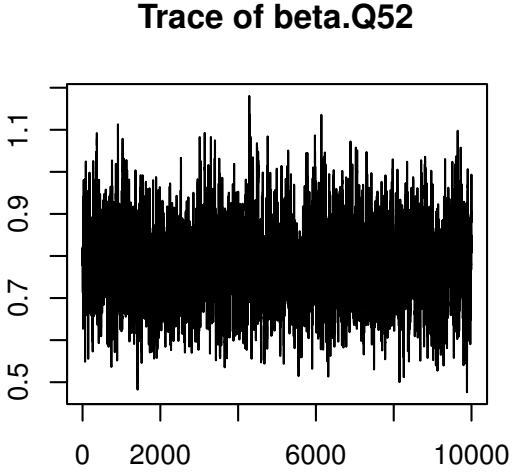
MCMC Diagnostics

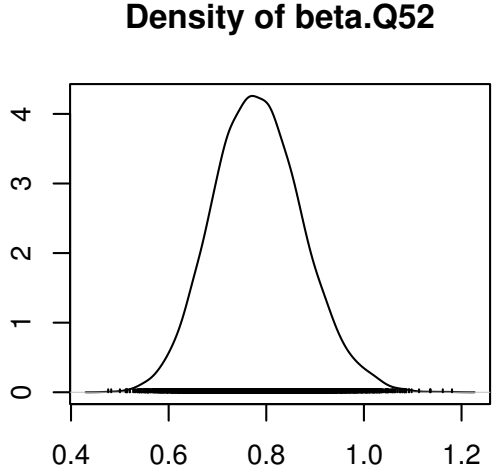
- Have the simulated draws converged to the posterior distribution?
- Use a R package `coda` that contains a number of diagnostic functions.
- Typically, this particular algorithm has rapid convergence.

Some coda Output









Inferences?

Inferences?

- Have matrix of simulated draws of parameters.

Inferences?

- Have matrix of simulated draws of parameters.
- Say you have particular parameter $h(\alpha, \beta)$ of interest.

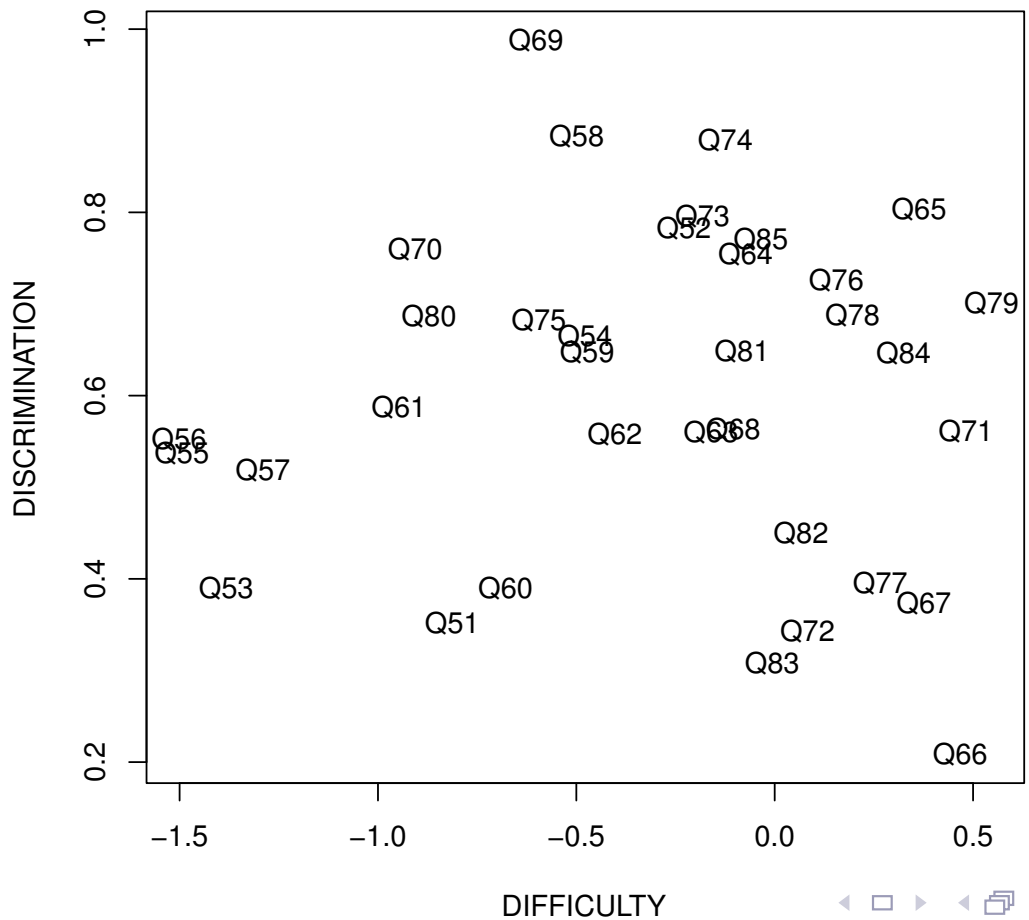
Inferences?

- Have matrix of simulated draws of parameters.
- Say you have particular parameter $h(\alpha, \beta)$ of interest.
- Apply this function to the simulated matrix – get sample from marginal posterior of h.

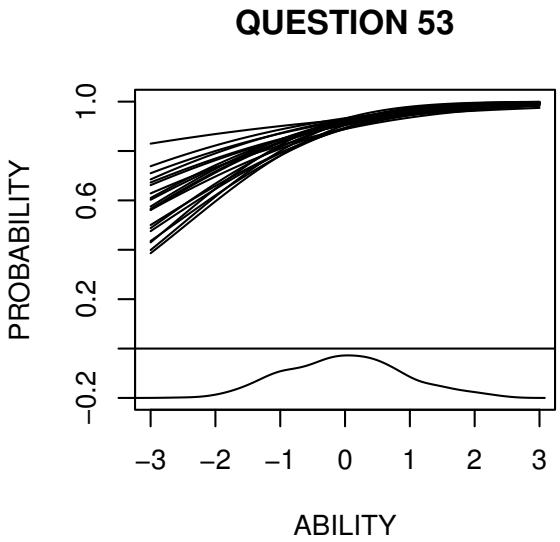
Inferences?

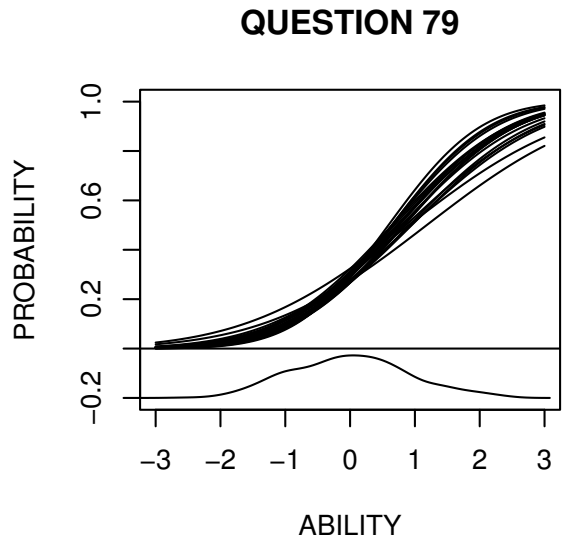
- Have matrix of simulated draws of parameters.
- Say you have particular parameter $h(\alpha, \beta)$ of interest.
- Apply this function to the simulated matrix – get sample from marginal posterior of h.
- In particular, can estimate item response curve $F(\beta_j\theta - \alpha_j)$.

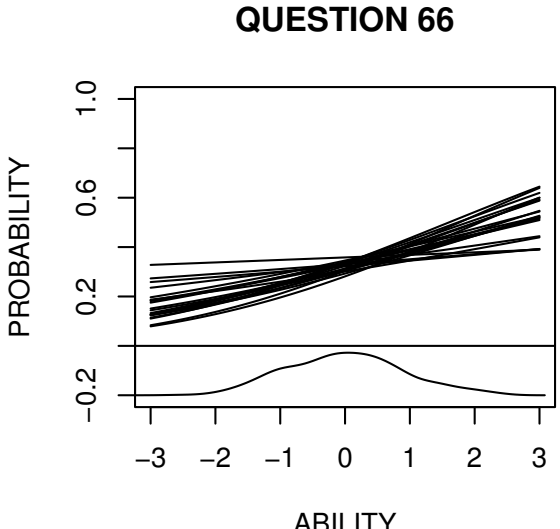
Posterior Means of Difficulty and Discrimination

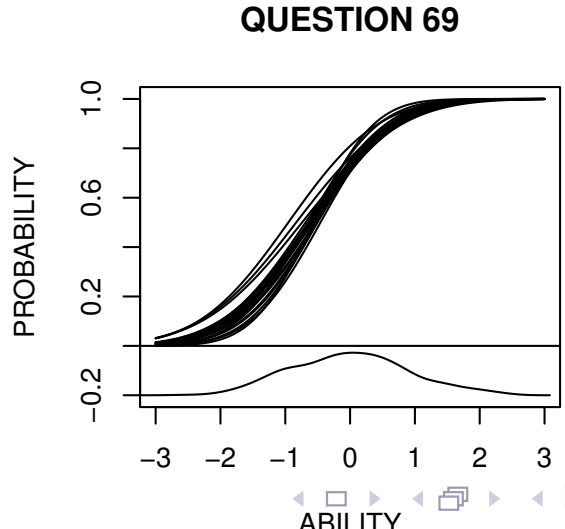


Some Fitted Item Response Curves









Two Questions

- Q 53, Low Difficulty, Low Discrimination
  QUESTION: If $n \times n \times n = 30$, which of the following best approximates n?
- Q 66, High Difficulty, Low Discrimination
  QUESTION: If $x < 2$, then $|2 - x| =$

Two Questions

- Q 69, Moderate Difficulty, High Discrimination
  QUESTION: The length L of a spring is given by $L = \frac{5}{6}F + 17$, where F is the applied force. What force F will produce a length L of 31?

- Q 79, High Difficulty, Moderate Discrimination
  QUESTION: In the system of equations $\{2x + 6y = 5, x - 3y = 8\}$, find a solution x.

Checking the Model

Checking the Model

- General idea: is the observed data consistent with predictions from the model?

Checking the Model

- General idea: is the observed data consistent with predictions from the model?
- Construct a discrepancy function $T(y^{obs})$ that measures a particular aspect of the model fit that you'd like to check.

Checking the Model

- General idea: is the observed data consistent with predictions from the model?
- Construct a discrepancy function $T(y^{obs})$ that measures a particular aspect of the model fit that you'd like to check.
- Simulate draws from the posterior predictive distribution $T(y^*)$.

Checking the Model

- General idea: is the observed data consistent with predictions from the model?
- Construct a discrepancy function $T(y^{obs})$ that measures a particular aspect of the model fit that you'd like to check.
- Simulate draws from the posterior predictive distribution $T(y^*)$.
- Compute the probability that $T(y)$ is at least as extreme as $T(y^{obs})$.

Checking the Model

- General idea: is the observed data consistent with predictions from the model?
- Construct a discrepancy function $T(y^{obs})$ that measures a particular aspect of the model fit that you'd like to check.
- Simulate draws from the posterior predictive distribution $T(y^*)$.
- Compute the probability that $T(y)$ is at least as extreme as $T(y^{obs})$.
- If this probability is small, casts doubt on the model.

Posterior Predictive Checks on R

Posterior Predictive Checks on R

- Have matrix of simulated draws of item and ability parameters from posterior.

Posterior Predictive Checks on R

- Have matrix of simulated draws of item and ability parameters from posterior.
- From a simulated draws of θ, α, β, simulate probability matrix p.

Posterior Predictive Checks on R

- Have matrix of simulated draws of item and ability parameters from posterior.
- From a simulated draws of θ, α, β, simulate probability matrix p.
- Simulate future data y^* from Bernoulli distributions with probabilities p.

Posterior Predictive Checks on R

- Have matrix of simulated draws of item and ability parameters from posterior.
- From a simulated draws of θ, α, β, simulate probability matrix p.
- Simulate future data y^* from Bernoulli distributions with probabilities p.
- Compute discrepancy function $T(y^*)$.

Posterior Predictive Checks on R

- Have matrix of simulated draws of item and ability parameters from posterior.
- From a simulated draws of θ, α, β, simulate probability matrix p.
- Simulate future data y^* from Bernoulli distributions with probabilities p.
- Compute discrepancy function $T(y^*)$.
- Repeat process – get sample $\{T(y^*)\}$.

Check Conditional Independence Assumption

Check Conditional Independence Assumption

- Given θ, assume responses are independent.

Check Conditional Independence Assumption

- Given θ, assume responses are independent.
- Suggests using item correlation matrix as a discrepancy function.

Check Conditional Independence Assumption

- Given θ, assume responses are independent.
- Suggests using item correlation matrix as a discrepancy function.
- Use our method to single out pairs of items that have unusually small or large correlations.

Check Conditional Independence Assumption

- Given θ, assume responses are independent.
- Suggests using item correlation matrix as a discrepancy function.
- Use our method to single out pairs of items that have unusually small or large correlations.
- Use criterion Tail Probability < 0.02.

Find Questions Pairs that are Similar or Dissimilar

Find Questions Pairs that are Similar or Dissimilar

- Question Q83 that asks for the value of k when $f(x) = x^2 - kx - 3, f(2) = 0$, similar to question Q82 that asks for one of the solutions of the equation $x^2 + 4x = -13$.

Find Questions Pairs that are Similar or Dissimilar

- Question Q83 that asks for the value of k when $f(x) = x^2 - kx - 3, f(2) = 0$, similar to question Q82 that asks for one of the solutions of the equation $x^2 + 4x = -13$.
- Question Q83 that solves a equation is different from question Q53 that has the student choose the best approximation to n when $n \times n \times n = 30$.

Are Discrimination Parameter Estimates Suitable?

Are Discrimination Parameter Estimates Suitable?

- An estimate of a item's discrimination ability is by the item total correlation.

Are Discrimination Parameter Estimates Suitable?

- An estimate of a item's discrimination ability is by the item total correlation.
- Measure of variability of item total correlation is the standard deviation of these correlations.

Are Discrimination Parameter Estimates Suitable?

- An estimate of a item's discrimination ability is by the item total correlation.
- Measure of variability of item total correlation is the standard deviation of these correlations.
- Use this standard deviation as a discrepancy function.

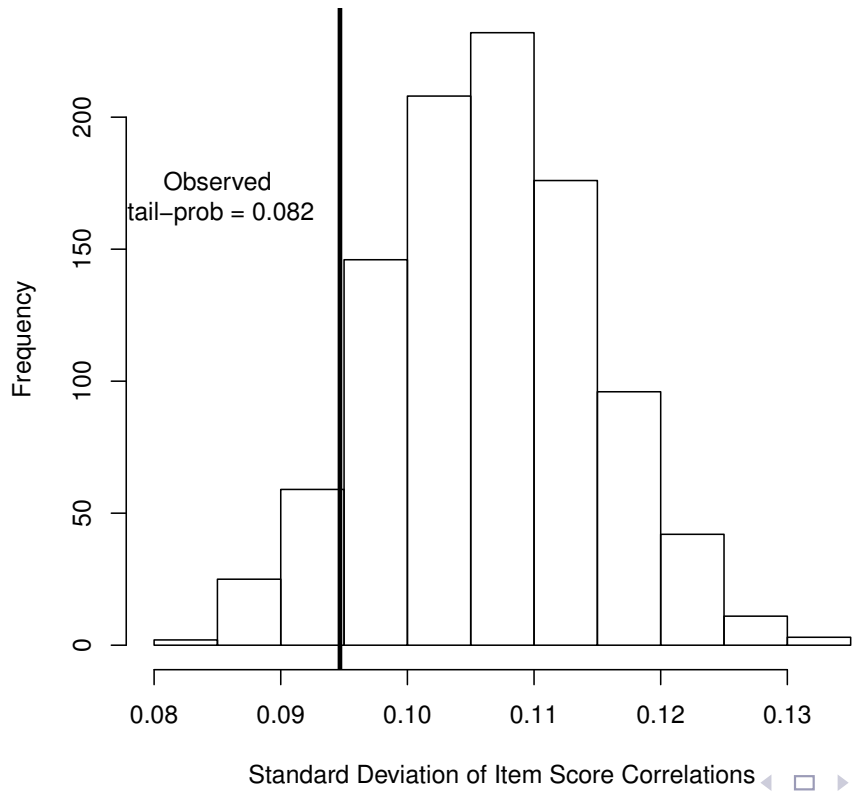
Are Discrimination Parameter Estimates Suitable?

- An estimate of a item's discrimination ability is by the item total correlation.
- Measure of variability of item total correlation is the standard deviation of these correlations.
- Use this standard deviation as a discrepancy function.
- We find that the item total correlations predicted from the model are more disperse than those from the observed data.

Are Discrimination Parameter Estimates Suitable?

- An estimate of a item's discrimination ability is by the item total correlation.
- Measure of variability of item total correlation is the standard deviation of these correlations.
- Use this standard deviation as a discrepancy function.
- We find that the item total correlations predicted from the model are more disperse than those from the observed data.
- This suggests our discrimination parameter estimates are too disperse.

Histogram of the Posterior Predictive Distribution of the Standard Deviation of the Item Total Correlations



1 or 2 Parameter Model?

1 or 2 Parameter Model?

- Do items have equal discrimination ability (one parameter model) or different discriminations (two parameter model)?

1 or 2 Parameter Model?

- Do items have equal discrimination ability (one parameter model) or different discriminations (two parameter model)?
- Actually I think both models are incorrect.

1 or 2 Parameter Model?

- Do items have equal discrimination ability (one parameter model) or different discriminations (two parameter model)?
- Actually I think both models are incorrect.
- I believe there are several clusters of questions and questions within each clusters have same discrimination.

1 or 2 Parameter Model?

- Do items have equal discrimination ability (one parameter model) or different discriminations (two parameter model)?
- Actually I think both models are incorrect.
- I believe there are several clusters of questions and questions within each clusters have same discrimination.
- Problem is to detect clusters.

Multilevel Model

Multilevel Model

- Assume discrimination parameters $\beta_1, ..., \beta_k$ are random sample from a common distribution $g(\beta)$.

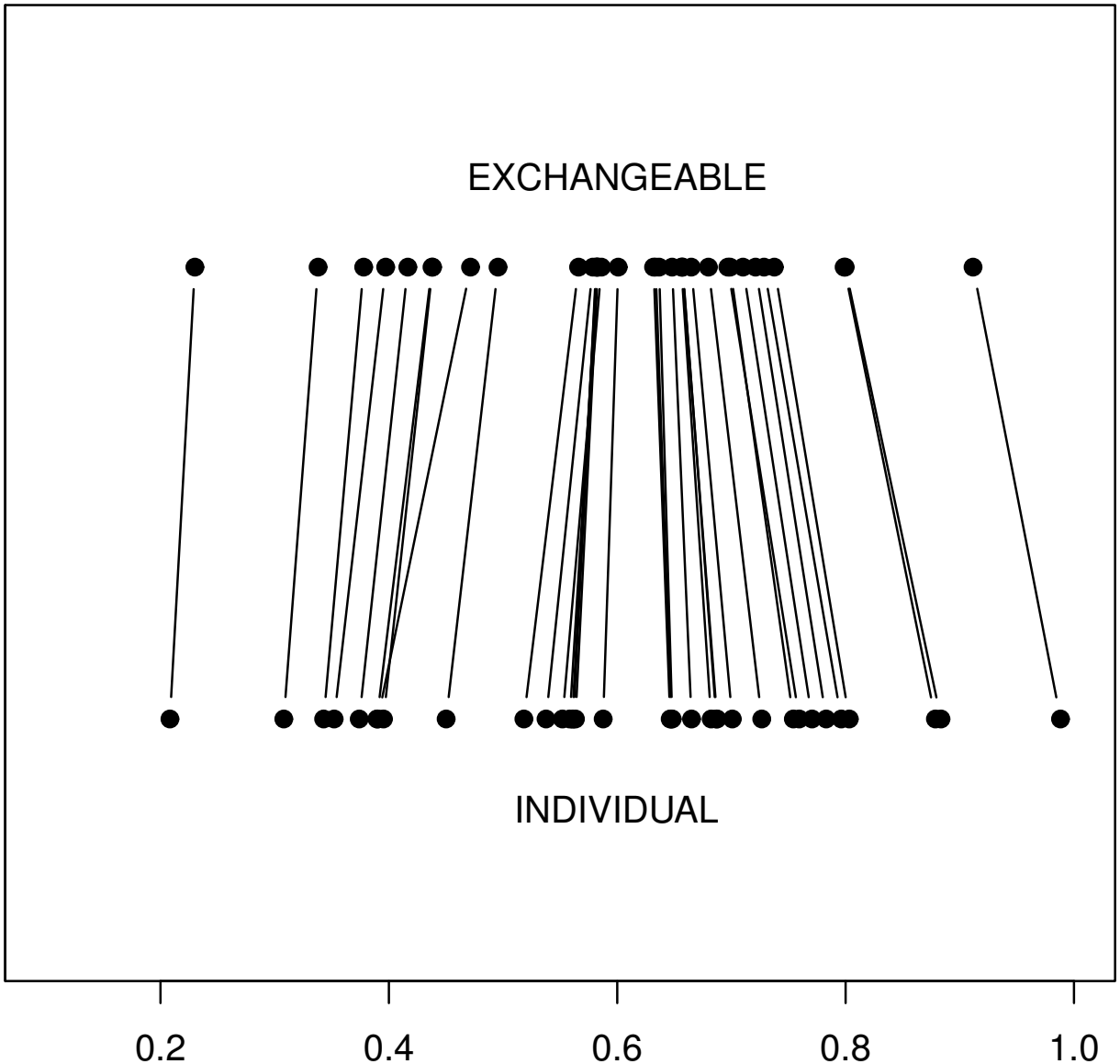
Multilevel Model

- Assume discrimination parameters $\beta_1, ..., \beta_k$ are random sample from a common distribution $g(\beta)$.
- If g is normal, posterior estimates of β_j will shrink towards a common value.

Multilevel Model

- Assume discrimination parameters $\beta_1, ..., \beta_k$ are random sample from a common distribution $g(\beta)$.
- If g is normal, posterior estimates of β_j will shrink towards a common value.
- If instead you choose a Cauchy density for g, get more adaptive shrinkage. (Seems more appropriate here.)

Multilevel Modeling with a Cauchy Density



Math Placement Data – Further Analysis

Math Placement Data – Further Analysis

- Multilevel model where θ_i are modeled in terms of student covariates (gender, ACT math, GPA).

Math Placement Data – Further Analysis

- Multilevel model where θ_i are modeled in terms of student covariates (gender, ACT math, GPA).
- Combine modeling across different groups of students, where "groups" are defined in terms of entering course, form of test, year, etc.

Math Placement Data – Further Analysis

- Multilevel model where θ_i are modeled in terms of student covariates (gender, ACT math, GPA).
- Combine modeling across different groups of students, where "groups" are defined in terms of entering course, form of test, year, etc.
- Test can be grouped by different sections (Arithmetic, Fractions, Graphing, Advanced, Solving) and this motives a "testlet" model where responses to questions within each testlet are assumed positively correlated.

The Future of Bayesian IRT Modeling

The Future of Bayesian IRT Modeling

- Model checking and model comparison.

The Future of Bayesian IRT Modeling

- Model checking and model comparison.
- Multilevel modeling. Effective way of combining similar sources of data.

The Future of Bayesian IRT Modeling

- Model checking and model comparison.
- Multilevel modeling. Effective way of combining similar sources of data.
- New book by Jean-Paul Fox on Bayesian IRT (with software).

The Future of Bayesian IRT Modeling

- Model checking and model comparison.
- Multilevel modeling. Effective way of combining similar sources of data.
- New book by Jean-Paul Fox on Bayesian IRT (with software).
- Break down the "wall of resistance" in using Bayesian methods in fitting and checking IRT models.