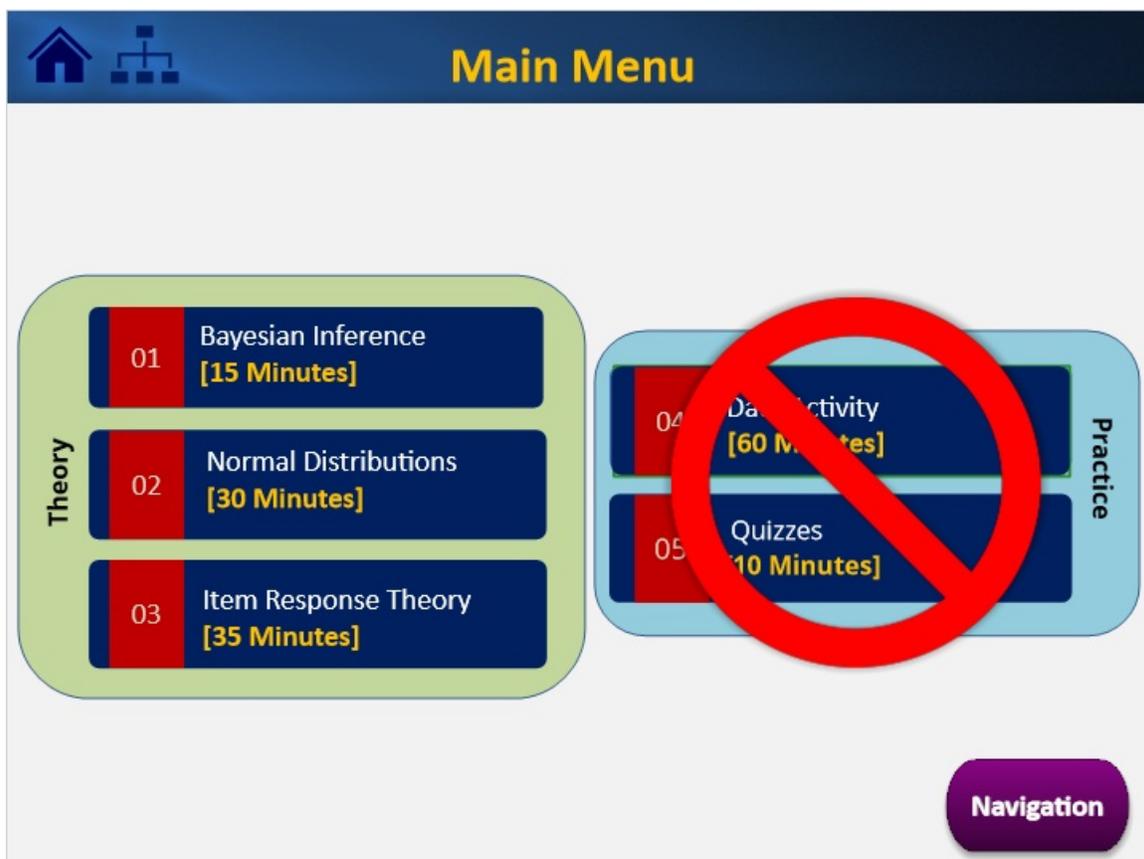


ITEMS Digital Module 11: Bayesian Psychometric Modeling

This document contains only the core content slides from the module. In the digital module all slides can be accessed individually.

Module Organization

The module starts with an introductory section that leads to the main menu from which learners can select individual theory and application sections; this slide deck contains the slides for the three content sections only:



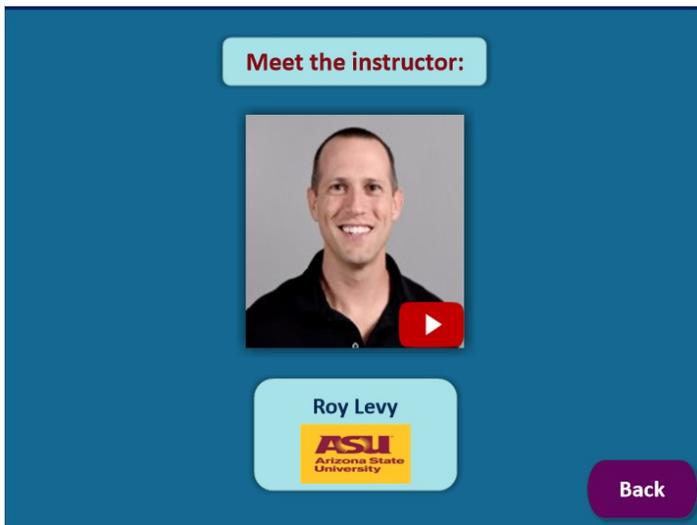
DM11 SLIDES (Bayesian Psychometrics, Version 1.0)

1. Module Overview

1.1 Module Cover



1.2 Instructor



1.3 Designers

Meet the instructional designer:



André A. Rupp
ETS



ETS
NCME
national council on
measurement
in education

Back

1.4 Welcome



Welcome to the
ITEMS Module!

The man to the left is Alberto!

Along with the instructor
he will be guiding you through
the module content.

Tell us your name here:

Untitled Layer 1 (Slide Layer)



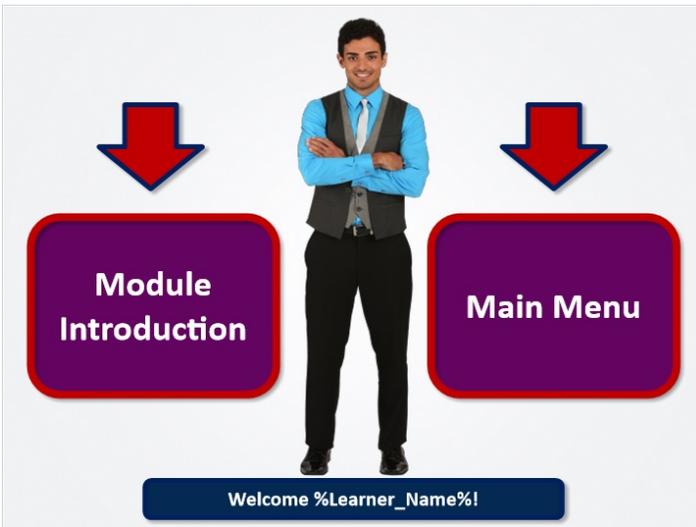
**Welcome to the
ITEMS Module!**

The man to the left is Alberto!

Along with the instructor
he will be guiding you through
the module content.

Tell us your name here:

1.5 Path Choice



Module
Introduction

Main Menu

Welcome %Learner_Name%!

1.6 Overview



1.7 Target Audience

Target Audience

Anyone who would like a **gentle statistical introduction** to this topic:

- graduate students and faculty in Master's, Ph.D., or certificate programs
- psychometricians and other measurement professionals
- data scientists / analysts
- research assistants or research scientists
- technical project directors
- assessment developers



However, we hope that you find the information in this module **useful no matter what your official title or role** in an organization is!

1.8 Expectations (I)



1.9 Expectations (II)

ITEMS Modules in Context

The image shows two examples of items in context. On the left is the cover of the book 'Bayesian Psychometric Modeling' by Roy Levy and Robert J. Mislevy. On the right is a webpage from the College of Education for the 'Measurement, Statistics and Evaluation, Master of Arts (M.A.)' program. Both images have a large red X over them, indicating that these items are not to be used in the current context.

1.11 Resources

Resources

Levy, R. (2020). *Bayesian psychometrics* (ITEMS Digital Module 11). Educational Measurement: Issues and Practice, 40, XX-XX. Available online at <https://ncme.elevate.commpartners.com>

Module Citation



Additional Resources

Resources 1 (Slide Layer)

Resources (1)

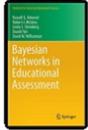
Levy, R., & Mislevy, R. J. (2016). Bayesian psychometric modeling.



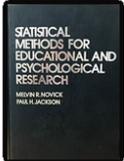
Fox, J.-P. (2010). Bayesian item response modeling: Theory and applications.



Almond, R. G., Mislevy, R. J., Steinberg, L. S., Yan, D., & Williamson, D. M. (2015). Bayesian networks in educational assessment.



Novick, M. R., & Jackson, P. (1974). Statistical methods for educational and psychological research.



Resources
1

Resources
2

Resources
3

Back

Resources 2 (Slide Layer)

Resources (2)

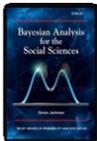
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.).



Gill, J. (2014). *Bayesian methods: A social and behavioral sciences approach* (3rd ed.).



Jackman, S. (2009). *Bayesian analysis for the social sciences*.



Kaplan, D. (2014). *Bayesian statistics for the social sciences*.



Resources 1 **Resources 2** **Resources 3** **Back**

Resources 3 (Slide Layer)

Resources (3)

Kim, J.-S., & Bolt, D. M. (2007). Estimating item response theory models using Markov chain Monte Carlo methods.

An NCME Instructional Module on
Estimating Item Response Theory Models Using Markov Chain Monte Carlo Methods
Jee-Seon Kim and Daniel M. Bolt, *University of Wisconsin, Madison*



Ames, A.J. & Myers, A.J. (2019). Digital Module 06: Bayesian psychometrics - Posterior predictive model checking



Resources 1 **Resources 2** **Resources 3** **Back**

1.12 Learning Objectives

Learning Objectives



1. Explain advantages of using Bayesian methods in assessment
2. Articulate general principles of specifying Bayesian models
3. Set up, estimate, and interpret models for normal distributions
4. Set up, estimate, and interpret unidimensional item response models

1.13 Main Menu

Main Menu



Theory

- 01 Bayesian Inference [15 Minutes]
- 02 Normal Distributions [30 Minutes]
- 03 Item Response Theory [35 Minutes]

Practice

- 04 Data Activity [60 Minutes]
- 05 Quizzes [10 Minutes]

Navigation

Navigation Help (Slide Layer)



2. Section 1: Bayesian Inference

2.1 Cover: Section 1 Why Bayes



2.2 Learning Objectives

  **Learning Objectives**



1. Understand the basic principles around Bayes' Theorem

2. List the distributional components for a Bayesian model

3. Express the key steps for parametric inference

4. Articulate what is produced by a Bayesian analysis

2.3 Big Picture (I)

  **Big Picture (I)**

Students engage in interactive activities / tasks or respond to items
(e.g., answer content questions, write essays, manipulate objects)

Reason from those actions about their capabilities

Reason from those actions about the activities

Reasoning is accomplished through Bayesian psychometric models

2.4 Big Picture (II)



Big Picture (II)

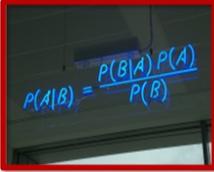


Bayes' Theorem

- **Facilitate inference** through parameter estimation
- **Expresses uncertainty** through probability
- **Propagates uncertainty** properly throughout the analysis
- **Allows us to work with models** that are very complex

2.5 Topic Selection





Bayesian
Inference: Big
Picture

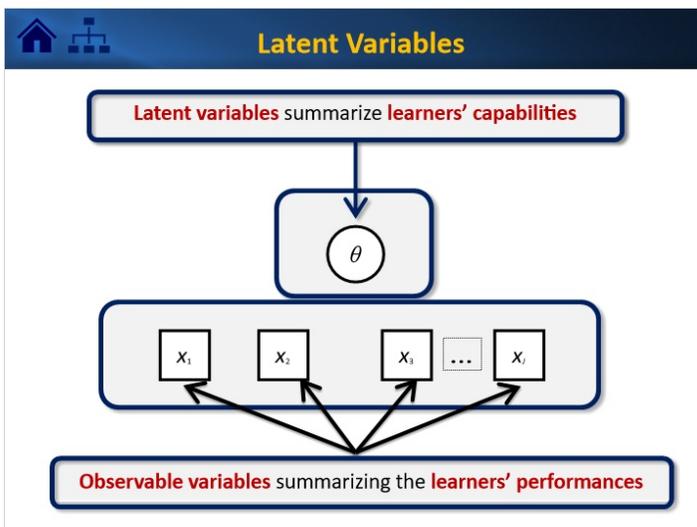
Inferences for
Learners & Tasks

Summary

2.6 Bookmark: Bayes' Theorem



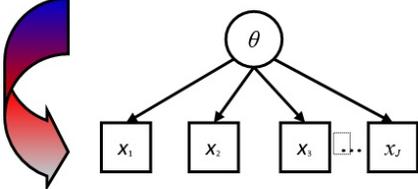
2.7 Latent Variables



2.8 Model Construction

  **Model Construction**

Model the **observable variables** as being dependent on the **latent variable**

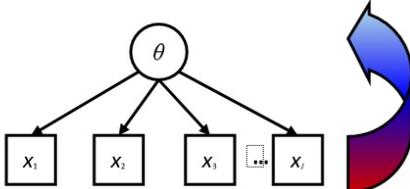


Model: What we would think about the x s if we knew θ
Goal: What we should think about θ if we knew x s

2.9 Inference about learners

  **Inference about Learners**

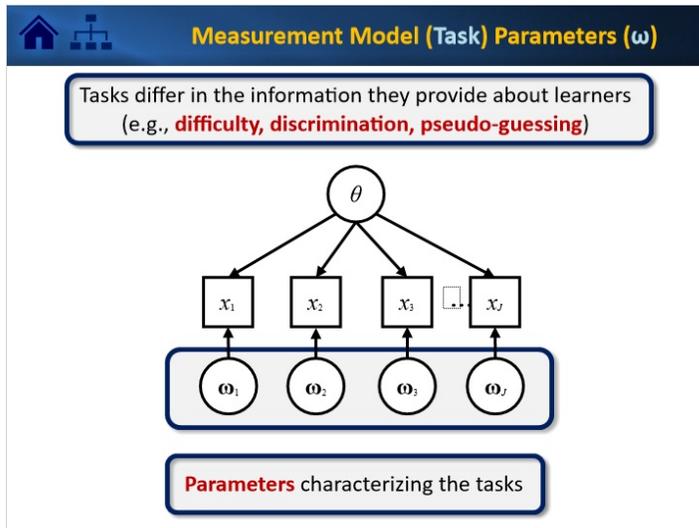
Bayes' Theorem:
A machinery that allows to reason back through the model



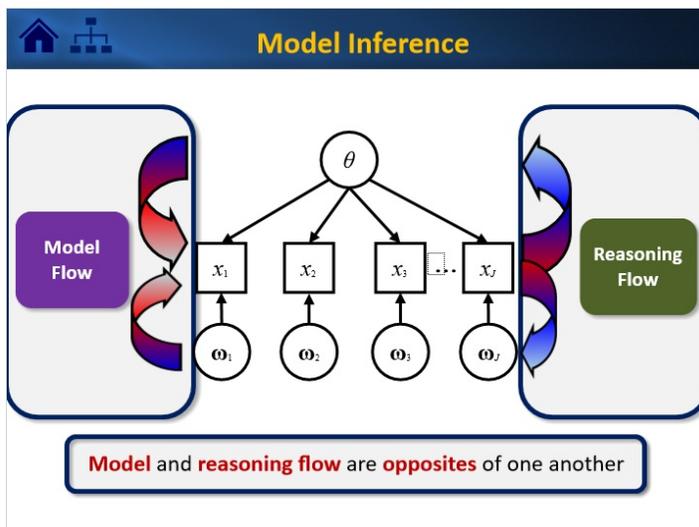
Reason "back through" the model

- Want to know what we should think about θ if we knew x s
- Opposite direction of model setup

2.10 Task Parameters

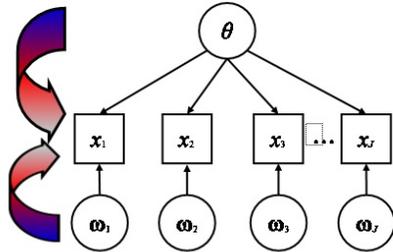


2.11 Model Inference



Model flow (Slide Layer)

Model Flow



Model Setup

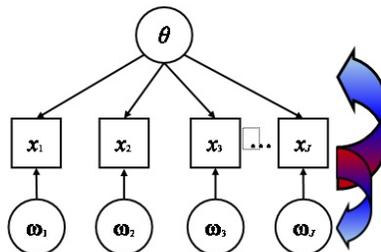
Observable variables modeled as dependent on:

- unobservable (latent) learner variables
- measurement model parameters

Back

Reasoning flow (Slide Layer)

Reasoning Flow



Reason "back through" the model

Want to know what we should think about θ and ω s, if we knew x s

Back

2.12 Summary



Summary

- **Goal:** Reason from learner performances (x_s) to make inferences about their capabilities (θ) and the features of the tasks (ω)
- **Strategy:** Set up the model with a particular flow from learner capabilities (θ) and the features of the tasks (ω) to performances (x_s)
- **Result:** Accomplish our goal by reversing the flow via Bayes' theorem



2.13 Bookend: Bayes' Theorem



This is the end of this part.

Topics

2.14 Bookmark: Learners & Tasks



2.15 Bayes' Theorem (I)

Notation	
θ	A parameter (e.g., a population mean)
$p(x \theta)$	Conditional probability of data (x) given the parameter (θ) Likelihood of the parameter (θ) given data (x)
$p(\theta)$	The <i>prior</i> distribution for θ
$p(\theta x)$	The <i>posterior</i> distribution for θ given x

2.16 Bayes' Theorem (I)

  **Purpose of Bayes' Theorem**

- Two entities:
 - x (data)
 - θ (parameter)
- We know the conditional probabilities $p(x | \theta)$, which tell us what to believe about x if we knew the value of θ
- When we learn value of x , what should we believe about θ ?



2.17 Bayes Theorem

  **Bayes' Theorem**

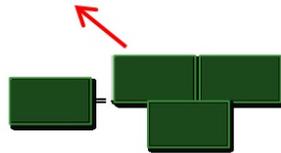

$$p(\theta | x) = \frac{p(x | \theta) p(\theta)}{p(x)}$$

Click on each component to learn more.

Likelihood (Slide Layer)

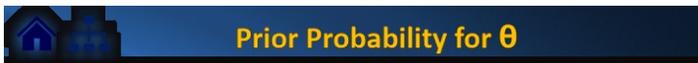


Conditional probability of x given θ

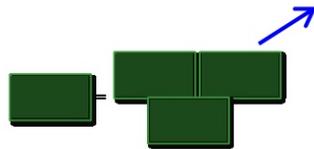


Also known as the **likelihood function**

Prior (Slide Layer)



Prior probability distribution
for unknown θ

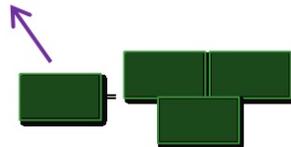


Captures everything we know about θ **before** we observe x

Posterior (Slide Layer)

Posterior Distribution for θ

Posterior probability distribution for unknown θ

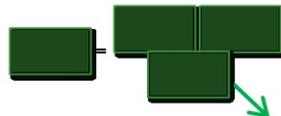


Captures what we think about θ **after** we have incorporated x

Constant (Slide Layer)

Marginal Probability

Serves to **normalize the distribution** and does **not** vary with θ



Marginal probability of x (over θ) $p(x) = \sum_{\theta} p(x|\theta)p(\theta)$

2.18 Posterior (II)

  **Proportionality Comment**

Discarding yields **proportionality**

$$p(\theta | x) = \frac{p(x | \theta) p(\theta)}{p(x)} \propto p(x | \theta) p(\theta)$$

Marginal probability of x (over θ)
Does not change for different values of θ

2.19 Topic Selection

Learner Parameters

Learner & Task Parameters

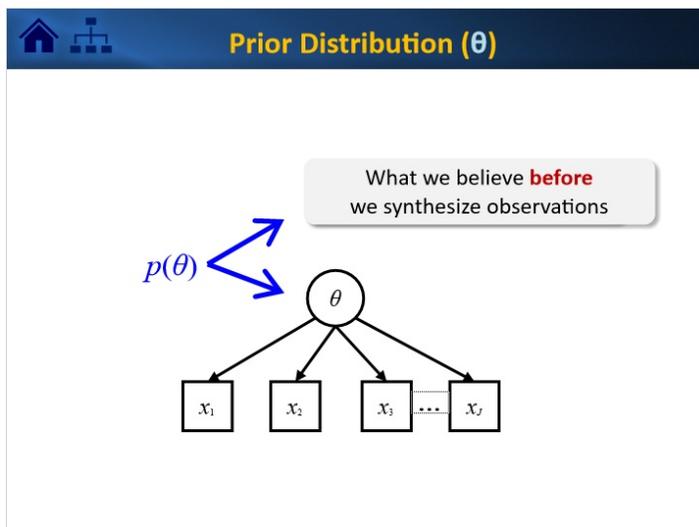
Part End

2.20 Bayes' Theorem: Model for ϑ

  **Bayes' Theorem: Model for θ**

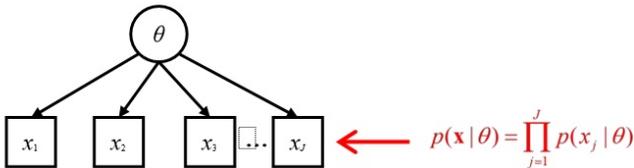

$$p(\theta | x) = \frac{p(x | \theta) p(\theta)}{p(x)} \propto p(x | \theta) p(\theta)$$


2.21 Prior Distribution (θ)



2.22 Conditional distribution

  Likelihood


$$p(\mathbf{x} | \theta) = \prod_{j=1}^J p(x_j | \theta)$$

Captures the **evidentiary value** of the observations

2.23 Bookend: Learner Parameters



This is the end of this part.

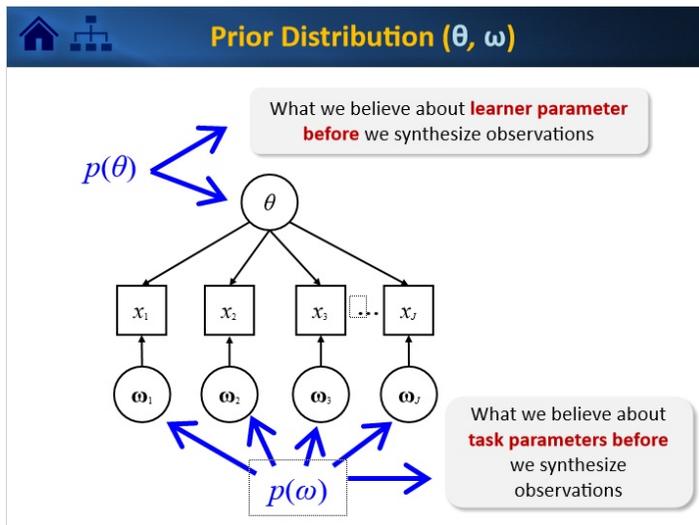
[Topics](#)

2.24 Bayes' Theorem: Model for ϑ and ω

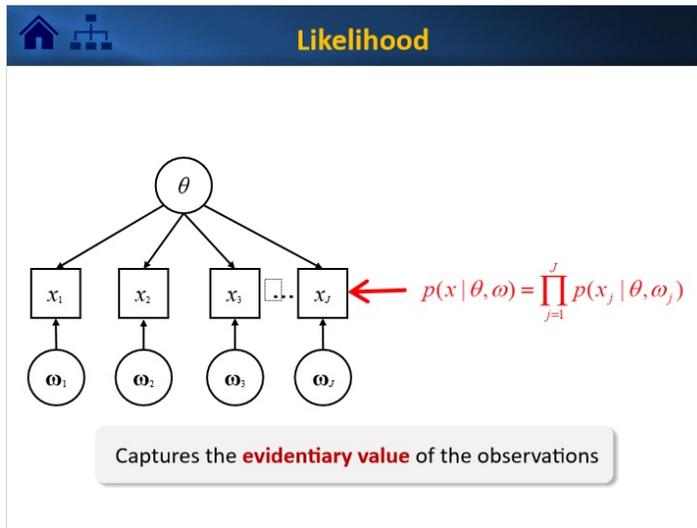
  **Bayes' Theorem: Model for θ and ω**


$$p(\theta, \omega | x) = \frac{p(x | \theta, \omega) p(\theta, \omega)}{p(x)} \propto p(x | \theta, \omega) p(\theta, \omega)$$

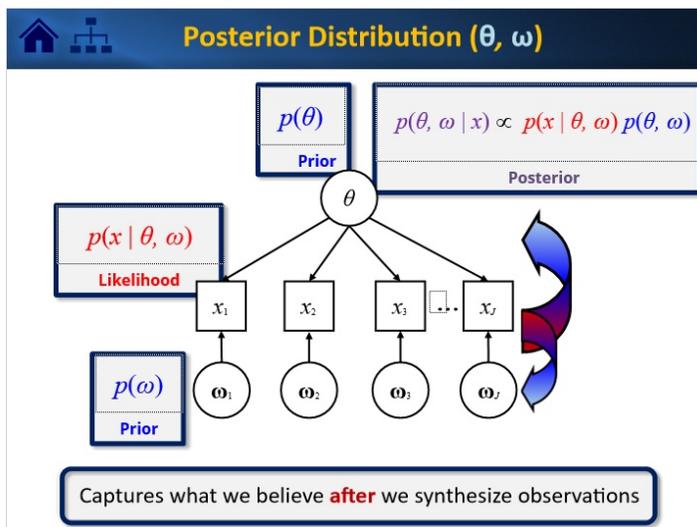

2.25 Prior Distribution (ϑ)



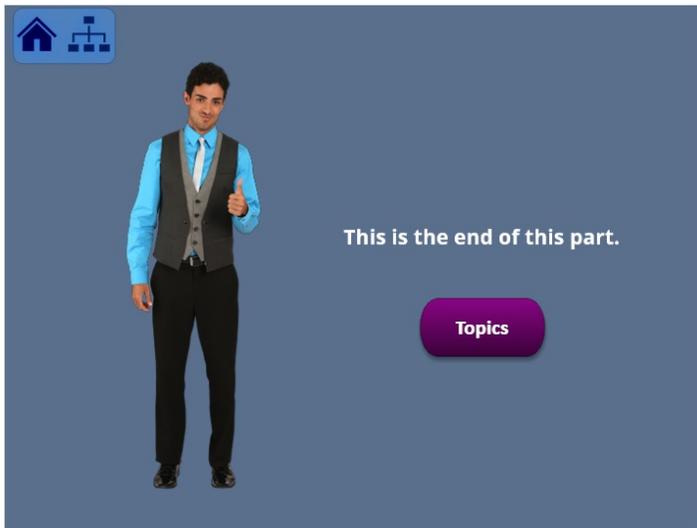
2.26 Conditional distribution



2.27 Posterior



2.28 Bookend: Learners & Items



2.29 Bookmark: Summary



2.30 Summary

Summary

- Set up the model as **flowing from** learner capabilities (θ) and the features of the tasks (ω) **to** performances (x s)
- Inference about learner capabilities (θ) and the features of the tasks (ω) based on observed performances (x s) by **reversing the flow via Bayes' theorem**:
$$p(\theta | x) = \frac{p(x | \theta) p(\theta)}{p(x)} \propto p(x | \theta) p(\theta)$$
- Applies to learner capabilities (θ) but can be extended to include parameters for tasks (ω s)

2.31 Bookend: Section 1



If you are interested in taking a **self-assessment** on this section click here:

If you want to return to the **main menu** click here:

Quiz

Main Menu

3. Section 2: Normal Distributions

3.1 Cover: Section 2 Bayesian Inference



Section 2:

Normal Distributions

[30 Minutes]

3.2 Learning Objectives

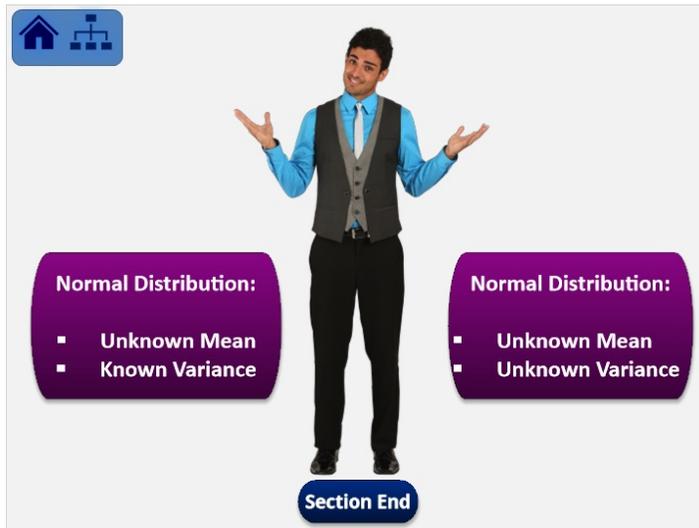


  **Learning Objectives**



1. Differentiate posterior distribution, prior distribution, and likelihood
2. Visualize the posterior as the synthesis of prior and likelihood
3. Introduce role of Markov-chain Monte Carlo (MCMC) estimation

3.3 Topic Selection



Normal Distribution:

- Unknown Mean
- Known Variance

Normal Distribution:

- Unknown Mean
- Unknown Variance

Section End

3.4 Bookmark: Setup 1



Setup 1:
Unknown Mean
Known Variance

3.5 Data Overview

  **Example: Data Overview**



- Sample size of learners: 10
- Number of items: 100
- Item scores: Dichotomous (0-1)
- Total score range: 0-100

Interested in the mean score for the population of learners

3.6 Modeling Setup

  **General Setup**

$$\mathbf{x} = (x_1, x_2, x_3, \dots, x_n)$$
$$p(\mathbf{x}) = p(x_1, x_2, x_3, \dots, x_n)$$
$$x_i | \mu, \sigma^2 \sim N(\mu, \sigma^2)$$

- μ **unknown**, σ^2 **known**
- desired **inference** for μ

$$p(\mu | \mathbf{x}, \sigma^2) = \frac{p(\mathbf{x} | \mu, \sigma^2) p(\mu)}{p(\mathbf{x})} \propto p(\mathbf{x} | \mu, \sigma^2) p(\mu)$$

3.7 Likelihood

  **Likelihood**

Conditional probability of \mathbf{x} given μ, σ^2 ; likelihood for μ

$$p(\mu | \mathbf{x}, \sigma^2) \propto p(\mathbf{x} | \mu, \sigma^2) p(\mu)$$

$$p(\mu | \mathbf{x}, \sigma^2) \propto \prod_{i=1}^n p(x_i | \mu, \sigma^2) p(\mu)$$

$$p(\mathbf{x} | \mu, \sigma^2) = \prod_{i=1}^n p(x_i | \mu, \sigma^2) = \prod_{i=1}^n N(x_i | \mu, \sigma^2)$$

$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right)$$

$$p(\bar{x} | \mu, \sigma^2) = N(\bar{x} | \mu, \sigma^2/n)$$

3.8 Prior

  **Prior Distribution for μ**

Prior probability distribution for unknown μ

$$p(\mu | \mathbf{x}, \sigma^2) \propto \prod_{i=1}^n p(x_i | \mu, \sigma^2) p(\mu)$$

A popular choice for a prior distribution in this context is a **normal distribution**

Normal

Normal (Slide Layer)

🏠
Prior Distribution for μ

Prior probability distribution
for unknown μ

↓

$$p(\mu | \mathbf{x}, \sigma^2) \propto \prod_{i=1}^n p(x_i | \mu, \sigma^2) p(\mu)$$

$$p(\mu | \mu_\mu, \sigma_\mu^2) = N(\mu_\mu, \sigma_\mu^2) = \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left(-\frac{1}{2\sigma_\mu^2}(\mu - \mu_\mu)^2\right)$$

Mean of the prior
distribution for μ

Variance of the prior
distribution for μ

↑ ↓

Hyperparameters

Back

3.9 Posterior (Mean)

🏠
Posterior Distribution for μ

$$p(\mu | \mathbf{x}, \sigma^2) \propto p(\mathbf{x} | \mu, \sigma^2) p(\mu)$$

$$\propto \prod_{i=1}^n N(x_i | \mu, \sigma^2) N(\mu | \mu_\mu, \sigma_\mu^2)$$

The **posterior distribution** is a **normal distribution**:

$$p(\mu | \mathbf{x}, \sigma^2) = N(\mu | \mu_{\mu|\mathbf{x}}, \sigma_{\mu|\mathbf{x}}^2)$$

The **normal prior** for the unknown mean is a **conjugate prior**

Mean

Variance/
Precision

Distribution

Distribution (Slide Layer)

Posterior Distribution

Prior: $N(\mu_\mu, \sigma_\mu^2)$ Likelihood: $\prod_i N(x_i | \mu, \sigma^2) = N(\bar{x} | \mu, \sigma^2/n)$

Posterior: $p(\mu | \mathbf{x}, \sigma^2) = N(\mu | \mu_{\mu|\mathbf{x}}, \sigma_{\mu|\mathbf{x}}^2)$

Posterior mean of μ

$$\mu_{\mu|\mathbf{x}} = \left(\frac{\mu_\mu}{\sigma_\mu^2} + \frac{n\bar{x}}{\sigma^2} \right) / \left(\frac{1}{\sigma_\mu^2} + \frac{n}{\sigma^2} \right)$$

Posterior variance of μ

$$\sigma_{\mu|\mathbf{x}}^2 = 1 / \left(\frac{1}{\sigma_\mu^2} + \frac{n}{\sigma^2} \right)$$

[Back](#)

Precision Posterior (Slide Layer)

Precision: Posterior

Prior: $N(\mu_\mu, \tau_\mu)$ Likelihood: $\prod_i N(x_i | \mu, \tau) = N(\bar{x} | \mu, n\tau)$

Posterior: $p(\mu | \mathbf{x}, \tau) = N(\mu | \mu_{\mu|\mathbf{x}}, \tau_{\mu|\mathbf{x}})$

The posterior precision is the sum of the prior precision and the precision in the data

Posterior precision of μ

$$\tau_{\mu|\mathbf{x}} = \tau_\mu + n\tau$$

Precision of the prior

Precision of the all the data = $n\tau$

[Back](#)

Mean (Slide Layer)

Posterior Mean

$$\text{Prior: } N(\mu_\mu, \tau_\mu) \quad \text{Likelihood: } \prod_i N(x_i | \mu, \tau) = N(\bar{x} | \mu, n\tau)$$

$$p(\mu | \mathbf{x}, \tau) = N(\mu | \mu_{\mu|x}, \tau_{\mu|x})$$

Posterior mean of μ

$$\mu_{\mu|x} = \frac{\tau_\mu \mu_\mu + n\tau \bar{x}}{\tau_\mu + n\tau} = \frac{\tau_\mu}{\tau_\mu + n\tau} \mu_\mu + \frac{n\tau}{\tau_\mu + n\tau} \bar{x}$$

Posterior mean is a **weighted average** of the **prior mean** and the **mean of the data**, where the weights are **proportional to the precision**

Back

Precision Prior Data (Slide Layer)

Precision: Prior & Data

- $\tau = 1/\sigma^2$
- $x | \mu, \sigma^2 \sim N(\mu, \sigma^2) \rightarrow x | \mu, \tau \sim N(\mu, \tau)$
- Conditional probability of the data
 $x_i | \mu, \tau \sim N(x_i | \mu, \tau)$
Precision of the data
- Prior distribution
 $\mu | \mu_\mu, \tau_\mu \sim N(\mu | \mu_\mu, \tau_\mu)$
Precision of the prior

Precision: Posterior

Back

3.10 Data Values

  **Example: Data Values**



$\sigma^2 = 25$ **Known variance**
 $\mathbf{x} = (91, 85, 72, 87, 71, 77, 88, 94, 84, 92)$ **Observed test scores**
 $\bar{x} = 84.1$ **Sample mean**

3.11 Modeling Setup Revisited

  **Example: Modeling Setup**

x_i **Score on the test for student i**
 $x_i \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$ **Conditional test score distribution**

✓ μ : **Population mean of learners' scores**
✓ σ^2 : **Population variance of learners' scores**

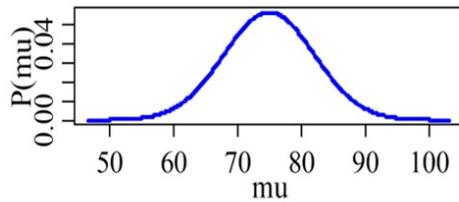
$\mu \sim N(\mu_\mu, \sigma_\mu^2)$ **Prior distribution for μ**

✓ $\mu_\mu = 75$ **Mean of prior distribution**
✓ $\sigma_\mu^2 = 50$ **Variance of prior distribution** [Graph](#)

Graph of Prior Mean (Slide Layer)

 **Prior for μ**

$$\mu \sim N(\mu_\mu, \sigma_\mu^2) \quad \text{with } \mu_\mu = 75 \text{ and } \sigma_\mu^2 = 50$$



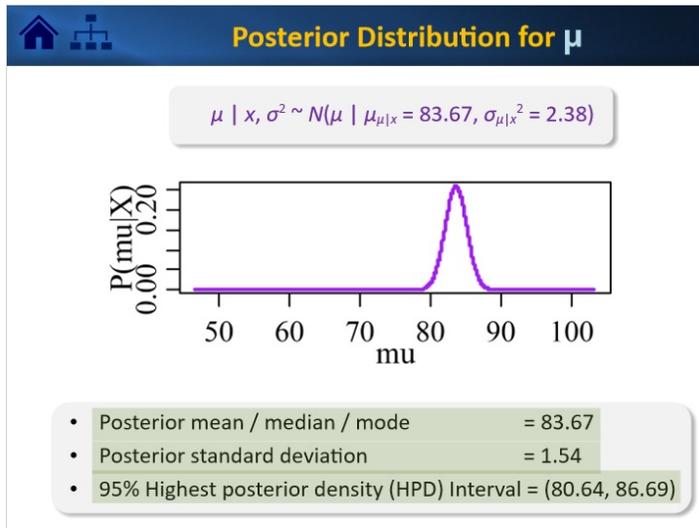
Back

3.12 Distributions

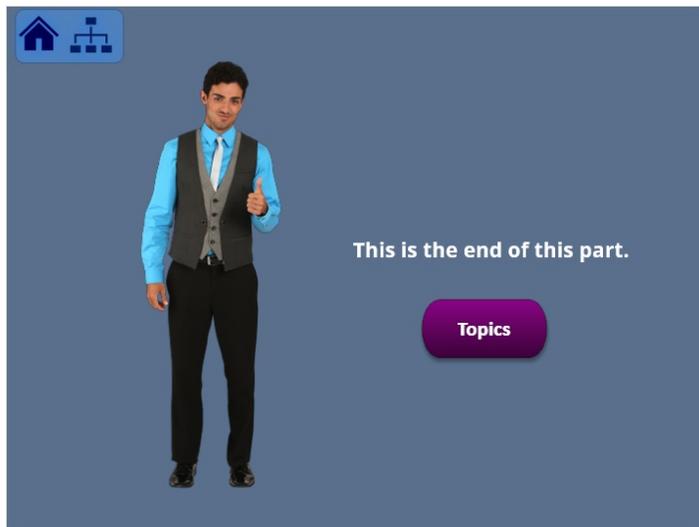
 **Distributions**

<p>Prior</p>	$\mu \sim N(\mu \mu_\mu = 75, \sigma_\mu^2 = 50)$
<p>Likelihood</p>	$L(\mu \mathbf{x}, \sigma^2) = p(\mathbf{x} \mu, \sigma^2)$ $= \prod_i p(x_i \mu, \sigma^2)$ $= \prod_i N(x_i \mu, \sigma^2 = 25)$
<p>Posterior</p>	$\mu \mathbf{x}, \sigma^2 \sim N(\mu \mu_{\mu \mathbf{x}}, \sigma_{\mu \mathbf{x}}^2)$ $\sim N(\mu \mu_{\mu \mathbf{x}} = 83.67, \sigma_{\mu \mathbf{x}}^2 = 2.38)$

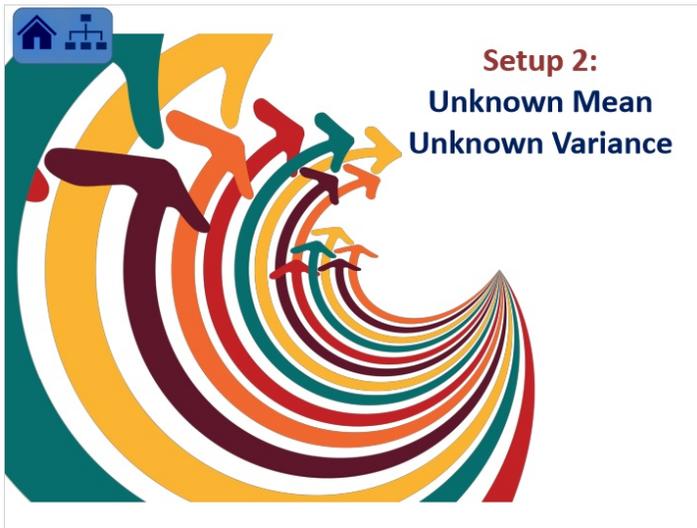
3.13 Posterior



3.14 Bookend: Setup 1



3.15 Bookmark: Setup 2



3.16 Data Overview

Example: Data Overview

- Sample size of learners: 10
- Number of items: 100
- Item scores: Dichotomous (0-1)
- Total score range: 0-100

Interested in the mean score and variance of scores for the population of learners

3.17 Modeling Setup

General Setup

$$\mathbf{x} = (x_1, x_2, x_3, \dots, x_n)$$

$$p(\mathbf{x}) = p(x_1, x_2, x_3, \dots, x_n)$$

$$x_i | \mu, \sigma^2 \sim N(\mu, \sigma^2)$$

- μ and σ^2 **unknown**
- desired **inference** for μ and σ^2

$$p(\mu, \sigma^2 | \mathbf{x}) = \frac{p(\mathbf{x} | \mu, \sigma^2) p(\mu, \sigma^2)}{p(\mathbf{x})} \propto p(\mathbf{x} | \mu, \sigma^2) p(\mu, \sigma^2)$$

3.18 Likelihood

Likelihood

Conditional probability of \mathbf{x}
 given μ, σ^2 ; likelihood for μ

↓

$$p(\mu, \sigma^2 | \mathbf{x}) \propto p(\mathbf{x} | \mu, \sigma^2) p(\mu, \sigma^2)$$

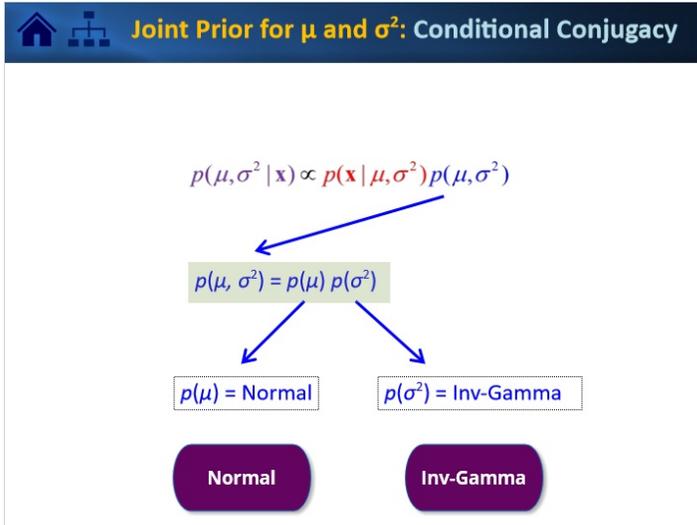
$$p(\mu, \sigma^2 | \mathbf{x}) \propto \prod_{i=1}^n p(x_i | \mu, \sigma^2) p(\mu, \sigma^2)$$

$$p(\mathbf{x} | \mu, \sigma^2) = \prod_{i=1}^n p(x_i | \mu, \sigma^2) = \prod_{i=1}^n N(x_i | \mu, \sigma^2)$$

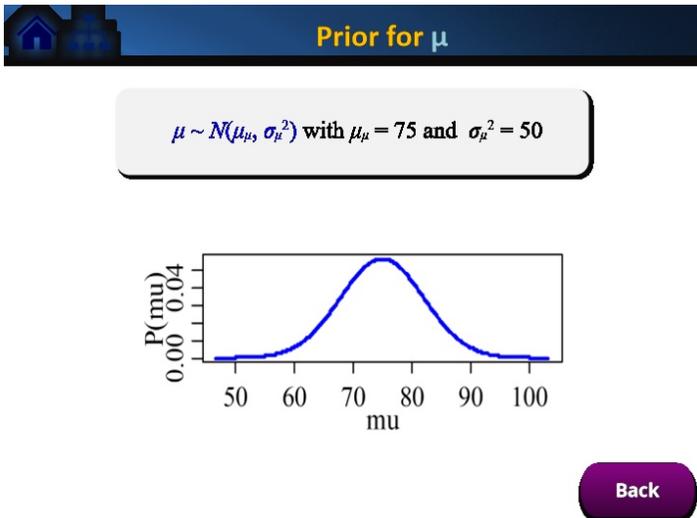
$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right)$$

$$p(\bar{x} | \mu, \sigma^2) = N(\bar{x} | \mu, \sigma^2/n)$$

3.19 Prior (Mean, Variance)



Prior Mean (Slide Layer)



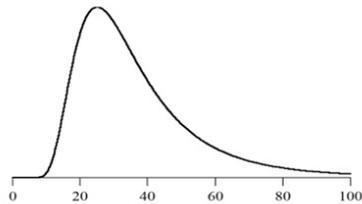
Prior Variance (Slide Layer)

Prior for σ^2

Informed by sample variance of 30 from 10 learners:

- ✓ $\sigma_0^2 = 30$
 - ✓ $\nu_0 = 10$
- $\sigma^2 \sim \text{Inv-Gamma}(\sigma^2 \mid \nu_0/2, \nu_0 * \sigma_0^2/2)$

$$\sigma^2 \sim \text{Inv-Gamma}(\sigma^2 \mid 10/2, 10*30/2) = \text{Inv-Gamma}(\sigma^2 \mid 5, 150)$$

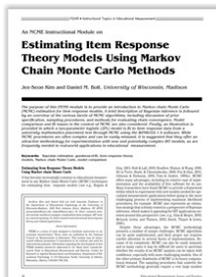


Back

3.20 Posterior (Mean, Variance)

Joint Posterior for μ and σ^2

- **Does not** have a **closed form**
- Can be obtained using **MCMC estimation**
- **MCMC = Markov chain Monte Carlo**



3.21 Data Values

  **Example: Setup and Data**



- x_i **Score on the test for student i**
- $x_i \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$ **Conditional score distribution**
- $\mathbf{x} = (91, 85, 72, 87, 71, 77, 88, 94, 84, 92)$ **Observed test scores**
- $\bar{x} = 84.1$ **Sample mean**
- $s^2 = 66.77$ **Sample variance**

3.22 JAGS Implementation

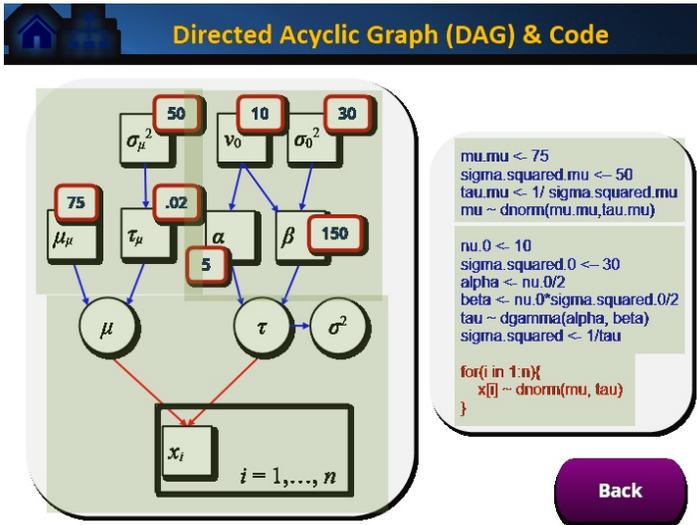
  **Running the Model in JAGS**



Equations

Graph & Code

Graph (Slide Layer)



Equation (Slide Layer)

Equations for Model

$$x_i \mid \mu, \tau \sim N(x_i \mid \mu, \tau)$$

$\mu \sim N(\mu \mid \mu_\mu, \tau_\mu)$	$\tau \sim \text{Gamma}(\tau \mid \nu_0, \sigma_0^2)$
$\tau_\mu = \sigma_\mu^{-2}$	$\sigma^2 = 1/\tau$

Precision **Inverse Gamma** **Back**

Gamma (Slide Layer)

Gamma and Inverse Gamma

- Typically specified with parameters α and β
$$\sigma^2 \sim \text{Inv-Gamma}(\alpha, \beta)$$
$$\tau \sim \text{Gamma}(\alpha, \beta)$$
- A convenient **reparameterization** has $\alpha = \nu_0/2$ and $\beta = \nu_0\sigma_0^2/2$
$$\sigma^2 \sim \text{Inv-Gamma}(\nu_0/2, \nu_0\sigma_0^2/2)$$
$$\tau \sim \text{Gamma}(\nu_0/2, \nu_0\sigma_0^2/2)$$
- σ_0^2 interpreted as the best estimate of the **variance**
- ν_0 interpreted as a **degrees of freedom** or **pseudo-sample size**

Model Back

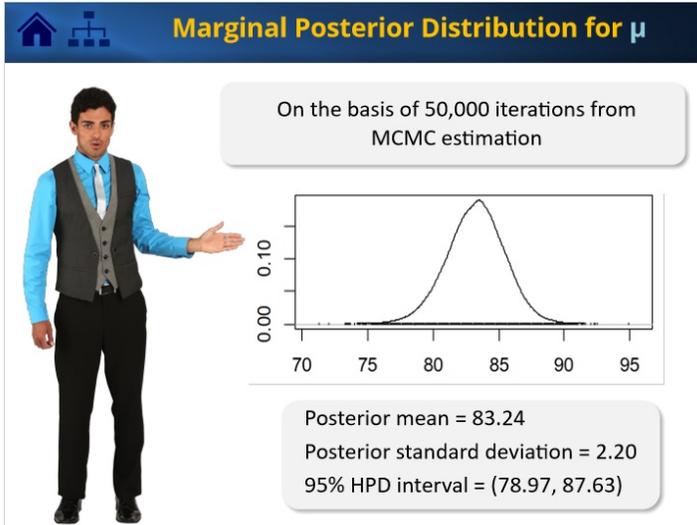
Precision (I) (Slide Layer)

Precision: Prior & Data

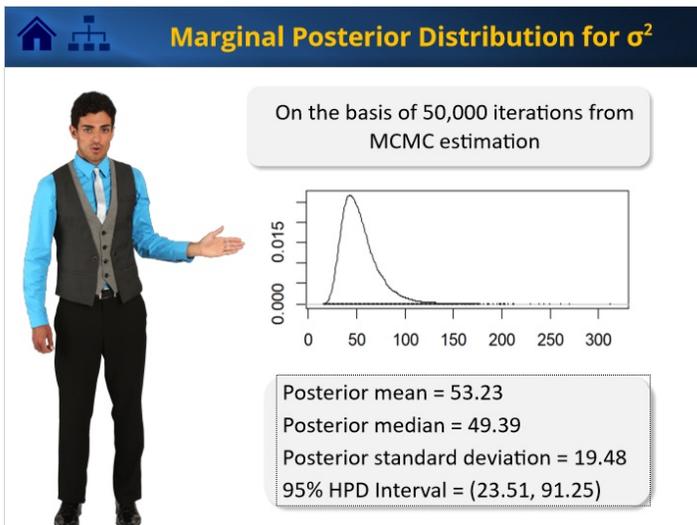
- $\tau = 1/\sigma^2$
- $x \mid \mu, \sigma^2 \sim N(\mu, \sigma^2) \rightarrow x \mid \mu, \tau \sim N(\mu, \tau)$
- Conditional probability of the data
 $x_i \mid \mu, \tau \sim N(x_i \mid \mu, \tau)$
Precision of the data
- Prior distribution
 $\mu \mid \mu_\mu, \tau_\mu \sim N(\mu \mid \mu_\mu, \tau_\mu)$
Precision of the prior

Model Back

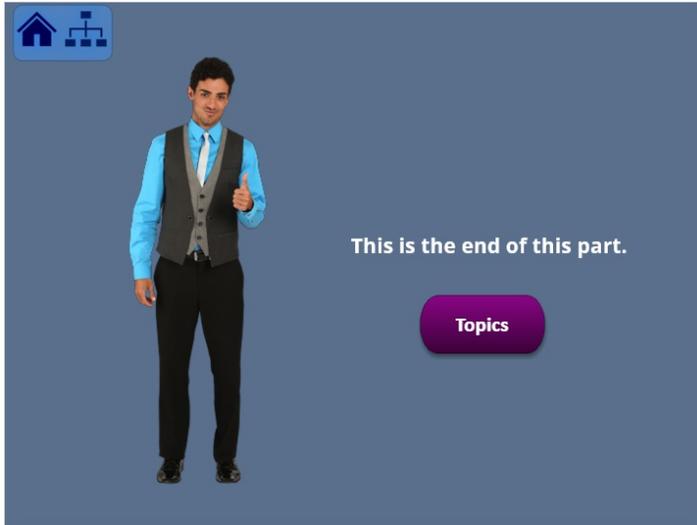
3.23 Posterior (Mean)



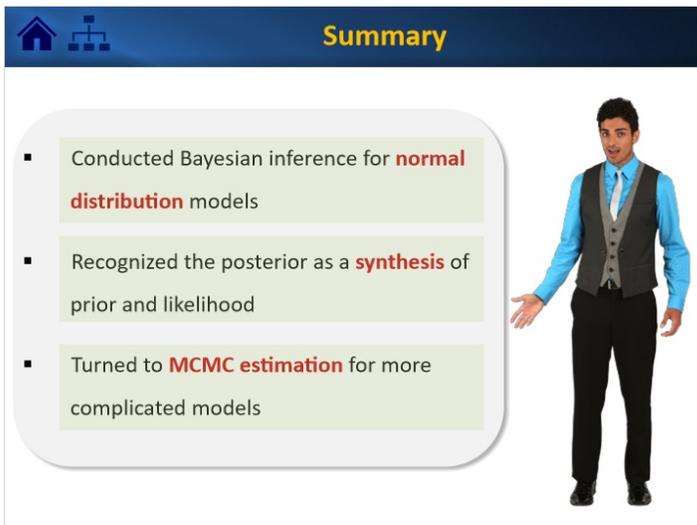
3.24 Posterior (Variance)



3.25 Bookend: Setup 2



3.26 Summary



A slide with a dark blue header. On the left, there is a navigation icon of a house and a tree. The word "Summary" is written in yellow on the right side of the header. Below the header, on the left, is a rounded rectangle containing a list of three items. On the right, a man in a blue shirt, grey vest, and black pants is gesturing towards the list.

- Conducted Bayesian inference for **normal distribution** models
- Recognized the posterior as a **synthesis** of prior and likelihood
- Turned to **MCMC estimation** for more complicated models

3.27 Bookend: Section 2



A navigation menu for Section 2. It features a blue icon with a house and a tree in the top left corner. On the left is a full-body image of a man in a blue shirt, grey vest, and black pants. To the right of the image are three text prompts, each followed by a colored button: a green button for 'Quiz', a blue button for 'Data Examples', and a purple button for 'Main Menu'.

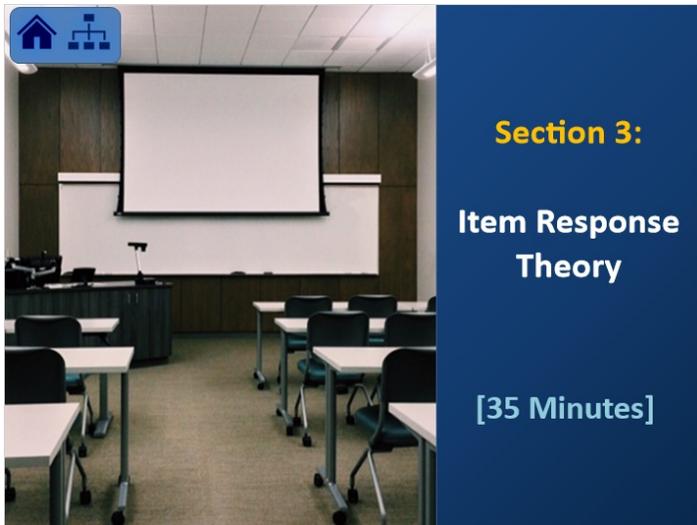
If you are interested in taking a **self-assessment** on this section click here: [Quiz](#)

If you are interested in seeing worked data examples of analyses in **R** click here: [Data Examples](#)

If you want to return to the **main menu** click here: [Main Menu](#)

4. Section 3: Item Response Theory

4.1 Cover: Section 3 Item Response Theory



A cover image for Section 3. The left side shows a photograph of a classroom with desks and a projector screen. The right side is a dark blue vertical panel with text. At the top left of the panel is a blue icon with a house and a tree. The text on the panel reads 'Section 3: Item Response Theory' in white, with '[35 Minutes]' below it in a lighter blue color.

Section 3:
Item Response Theory
[35 Minutes]

4.2 Learning Objectives



Learning Objectives



1. Articulate IRT notation
2. Describe Bayesian IRT models
3. Understand the results from fitting a Bayesian IRT model
4. Become aware of the extensions and applications of Bayesian IRT models

4.3 Extensions (II)

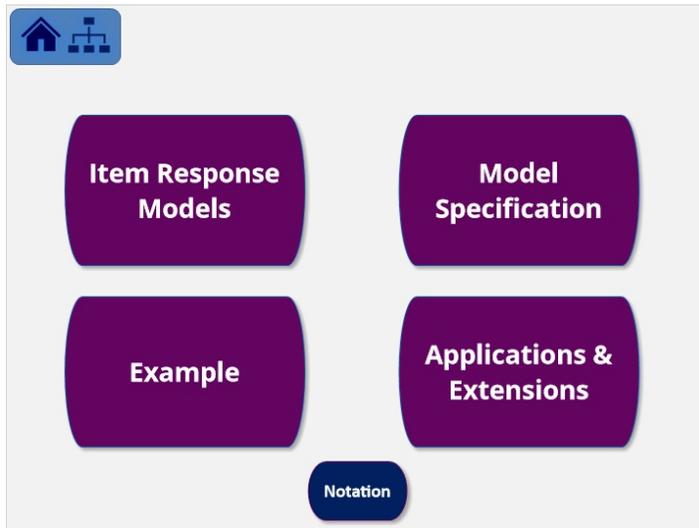


Applications & Extensions



- **Manage** inferences at different levels (groups vs. individuals)
- **Conduct** adaptive testing
- **Incorporate** different sources of information
- **Represent** substantive beliefs and collateral information
- **Integrate** assessment and learning systems
- **Represent** Bayesian thinking even without formal Bayesian statistics

4.4 Topic Selection

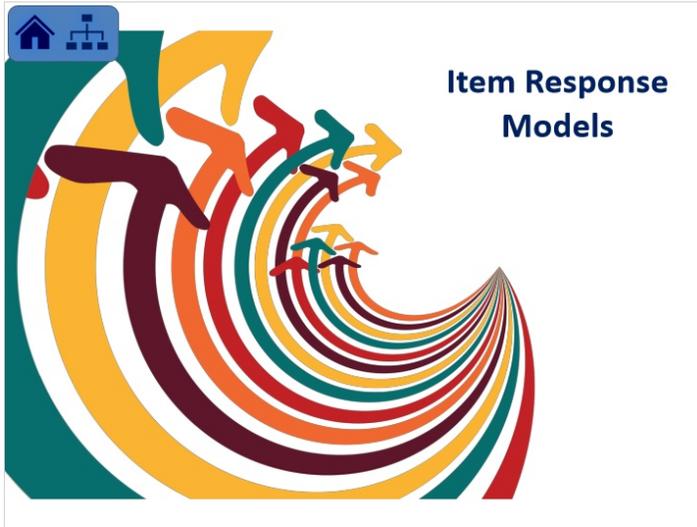


Notation (Slide Layer)

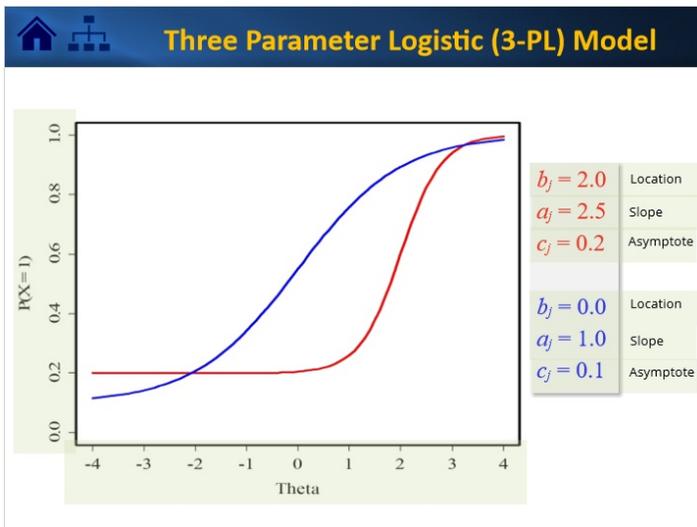
x_{ij}	Dichotomous observable score for examinee i on observable j ; $\in \{0,1\}$
θ_i	Value of latent variable for examinee i
b_j, d_j	Location (difficulty) parameter for of observable (item) j
a_j	Discrimination (loading) parameter for observable (item) j
c_j	Lower-asymptote (pseudo-guessing) param. for obs. (item) j
μ_θ	Mean of latent variable
σ_θ^2	Variance of the latent variable

Back

4.5 Bookmark: IRT Models



4.6 Three-parameter Model



4.7 Link Functions

  **Logistic and Normal IRT Models**

Logistic IRT models use cumulative *logistic* distribution

$$P(x_{ij} = 1 | \theta_i, b_j, a_j) = \Psi[a_j(\theta_i - b_j)] = \frac{\exp(a_j(\theta_i - b_j))}{1 + \exp(a_j(\theta_i - b_j))}$$

Normal-Ogive IRT models use the cumulative *normal* distribution

$$P(x_{ij} = 1 | \theta_i, b_j, a_j) = \Phi[a_j(\theta_i - b_j)] = \int_{-\infty}^{a_j(\theta_i - b_j)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

4.8 Alternative Parameterizations

  **Alternative Parameterizations**

$$P(x_{ij} = 1 | \theta_i, b_j, a_j, c_j) = c_j + (1 - c_j)F[a_j(\theta_i - b_j)]$$

- Focuses on examinee location relative to item
- $P = .5$ when $\theta = b$ for two-parameter models

$$P(x_{ij} = 1 | \theta_i, d_j, a_j, c_j) = c_j + (1 - c_j)F(a_j\theta_i + d_j)$$

- $d_j = -a_j b_j$
- Useful for viewing IRT similar to regression and factor analysis
- Natural extensions to multidimensional models

F denotes distribution function (logistic or normal)

4.9 Indeterminacies



Indeterminacies

For 2-parameter (and higher) models

- Location often resolved by fixing mean of θ to 0
- Metric often resolved by fixing variance of θ to 1
- Orientation often resolved by constraining the a s to be > 0

For 1-parameter / Rasch models

- Location indeterminacy
- Often resolved by constraints on location (b) parameter(s)

4.10 Bookend: IRT Models



This is the end of this part.

Topics

4.11 Bookmark: Model Specification



4.12 Model Specification

  **Model Specification**

Fixing $\mu_\theta, \sigma_\theta^2$ to **resolve indeterminacies**:

$$\boldsymbol{\theta} = \theta_1, \dots, \theta_n \quad \mathbf{d} = d_1, \dots, d_J \quad \mathbf{a} = a_1, \dots, a_J \quad \mathbf{c} = c_1, \dots, c_J$$

The **posterior distribution** is proportional **likelihood** times the **prior**:

$$p(\boldsymbol{\theta}, \mathbf{d}, \mathbf{a}, \mathbf{c} \mid \mathbf{x}) \propto p(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{d}, \mathbf{a}, \mathbf{c}) p(\boldsymbol{\theta}, \mathbf{d}, \mathbf{a}, \mathbf{c})$$

Via common **exchangeability** and **conditional independence assumptions**:

$$p(\boldsymbol{\theta}, \mathbf{d}, \mathbf{a}, \mathbf{c} \mid \mathbf{x}) \propto \prod_{i=1}^n \prod_{j=1}^J p(x_{ij} \mid \theta_i, d_j, a_j, c_j) p(\theta_i) p(d_j) p(a_j) p(c_j)$$

4.13 Bayes Theorem



Bayes' Theorem


$$p(\theta, \mathbf{d}, \mathbf{a}, \mathbf{c} \mid \mathbf{x}) \propto \prod_{i=1}^n \prod_{j=1}^J p(x_{ij} \mid \theta_i, d_j, a_j, c_j) p(\theta_i) p(d_j) p(a_j) p(c_j)$$

Posterior \propto Likelihood \times Prior

4.14 Likelihood



Likelihood

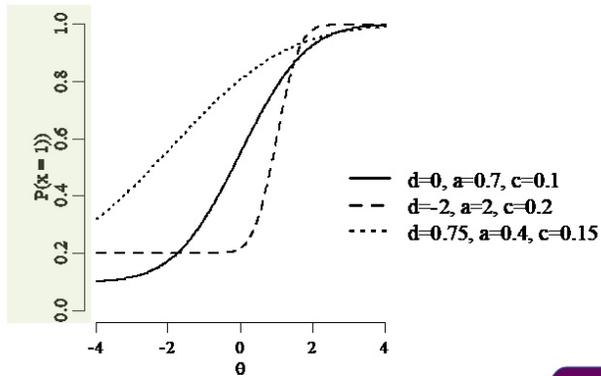
$$p(\theta, \mathbf{d}, \mathbf{a}, \mathbf{c} \mid \mathbf{x}) \propto \prod_{i=1}^n \prod_{j=1}^J p(x_{ij} \mid \theta_i, d_j, a_j, c_j) p(\theta_i) p(d_j) p(a_j) p(c_j)$$

$x_{ij} \mid \theta_i, d_j, a_j, c_j \sim \text{Bernoulli}[c_j + (1-c_j)F(a_j\theta_i + d_j)]$ [Graph](#)

Expressed as a **Bernoulli distribution** for dichotomous outcomes with **probability of '1'** depending on **model parameters**

IRT Probabilities (Slide Layer)

IRT Response Probabilities



Back

4.15 Priors: Learner Parameters

Priors: Learner Parameters

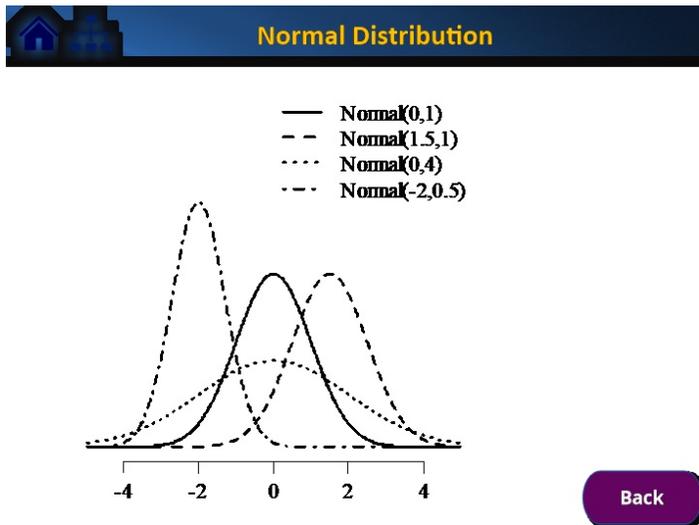
$$p(\theta, \mathbf{d}, \mathbf{a}, \mathbf{c} \mid \mathbf{x}) \propto \prod_{i=1}^n \prod_{j=1}^J p(x_{ij} \mid \theta_i, d_j, a_j, c_j) p(\theta_i) p(d_j) p(a_j) p(c_j)$$

$$\theta_i \sim N(\mu_\theta, \sigma_\theta^2)$$

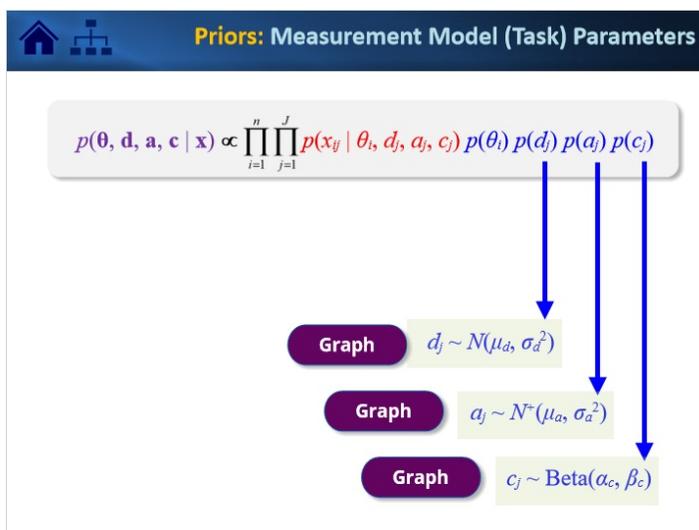
Graph

Chosen values for $\mu_\theta, \sigma_\theta^2$ serve to **resolve indeterminacies** in location and metric

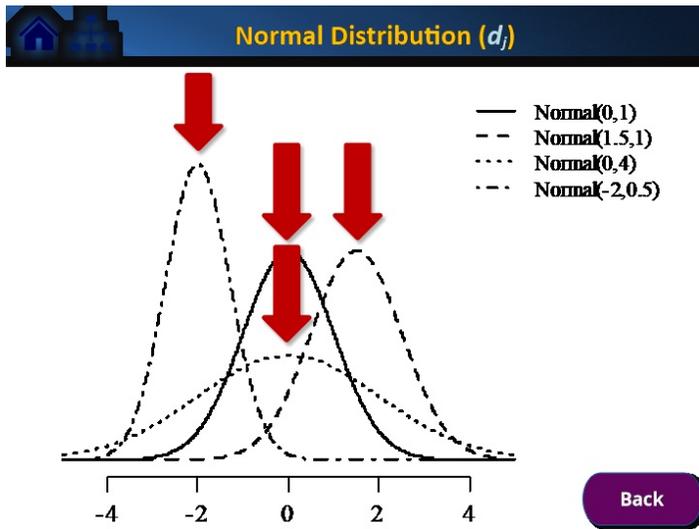
Graph Normal (Slide Layer)



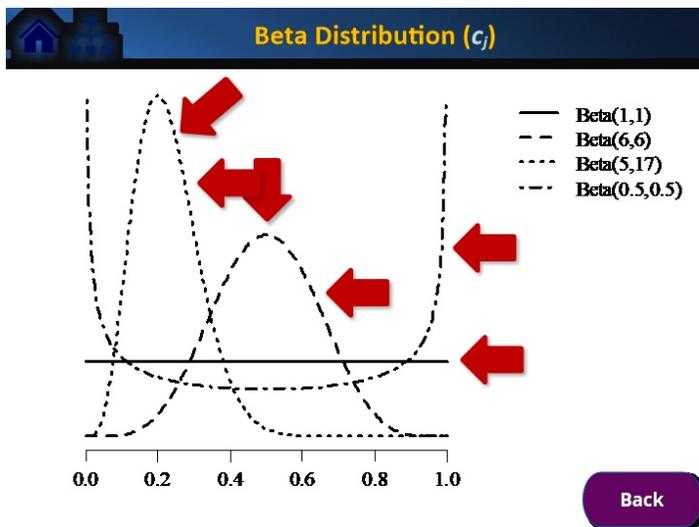
4.16 Priors: Item Parameters



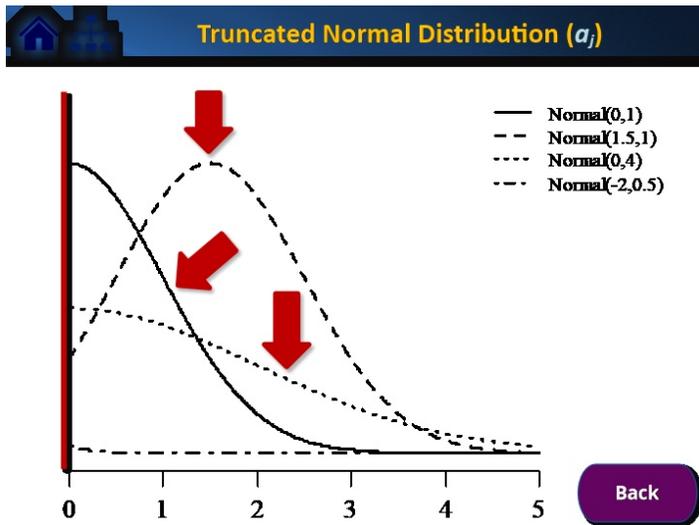
Graph Normal (Slide Layer)



Graph Beta (Slide Layer)



Graph Truncated Normal (Slide Layer)



4.17 Bookend: Model Specification



4.18 Bookmark: Example



4.19 Example (I)

  **Example: Law School Admissions Test**

Vector ID	Item 1	Item 2	Item 3	Item 4	Item 5	Frequency	Vector ID	Item 1	Item 2	Item 3	Item 4	Item 5	Frequency
1	0	0	0	0	0	3	17	1	0	0	0	0	10
2	0	0	0	0	1	6	18	1	0	0	0	1	29
3	0	0	0	1	0	2	19	1	0	0	1	0	14
4	0	0	0	1	1	11	20	1	0	0	1	1	81
5	0	0	1	0	0	1	21	1	0	1	0	0	3
6	0	0	1	0	1	1	22	1	0	1	0	1	28
7	0	0	1	1	0	3	23	1	0	1	1	0	15
8	0	0	1	1	1	4	24	1	0	1	1	1	80
9	0	1	0	0	0	1	25	1	1	0	0	0	16
10	0	1	0	0	1	8	26	1	1	0	0	1	56
11	0	1	0	1	0	0	27	1	1	0	1	0	21
12	0	1	0	1	1	16	28	1	1	0	1	1	173
13	0	1	1	0	0	0	29	1	1	1	0	0	11
14	0	1	1	0	1	3	30	1	1	1	0	1	61
15	0	1	1	1	0	2	31	1	1	1	1	0	28
16	0	1	1	1	1	15	32	1	1	1	1	1	298

Summary of Response Patterns (5 items, 1000 examinees)

[Reference](#)

Reference (Slide Layer)

Reference



Springer Link

Published: June 1970

Fitting a response model for dichotomously scored items

B. Darrell Bock & Marcus Lieberman

Psychometrika 35, 179–197(1970) | [Cite this article](#)

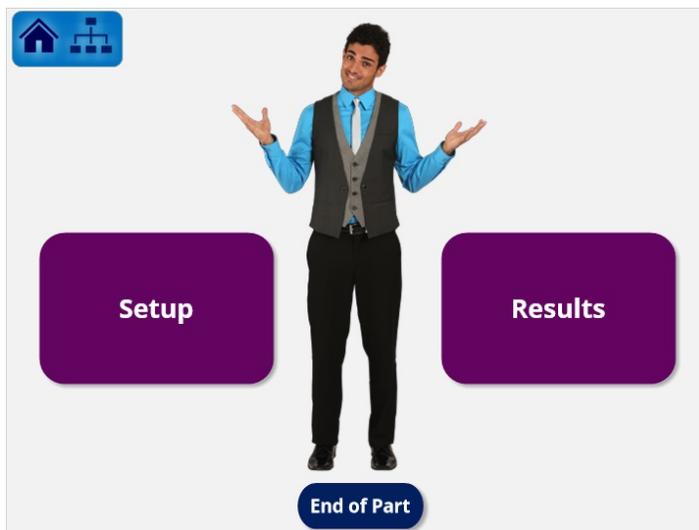
795 Accesses | 375 Citations | 6 Altmetric | [Metrics](#)

Abstract

A method of estimating the parameters of the normal ogive model for dichotomously scored item-responses by maximum likelihood is demonstrated. Although the procedure requires numerical integration in order to evaluate the likelihood equations, a computer implemented Newton-Raphson solution is shown to be straightforward in other respects. Empirical tests of the procedure show that the resulting estimates are very similar to those based on a conventional analysis of item "difficulties" and first factor loadings obtained from the matrix of tetrachoric correlation coefficients. Problems of testing the fit of the model, and of obtaining invariant parameters are discussed.

Back

4.20 Topic Selection



Home icon

Setup

Results

End of Part

4.21 Example (II)

Joint Specification

$$p(\theta, \mathbf{d}, \mathbf{a}, \mathbf{c} \mid \mathbf{x}) \propto \prod_{i=1}^n \prod_{j=1}^J p(x_{ij} \mid \theta_i, d_j, a_j, c_j) p(\theta_i) p(d_j) p(a_j) p(c_j)$$

$x_{ij} \mid \theta_i, d_j, a_j, c_j \sim \text{Bernoulli}[c_j + (1-c_j)\Phi(a_j\theta_i + d_j)]$ **Graph** **Code**

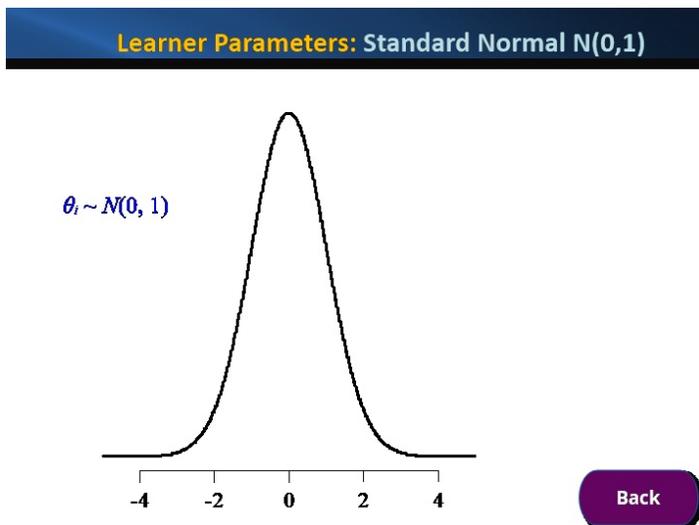
$\theta_i \sim N(0, 1)$ **Graph** **Code**

$d_j \sim N(0, 2)$ **Graph** **Code**

$a_j \sim N^+(1, 2)$ **Graph** **Code**

$c_j \sim \text{Beta}(5, 17)$ **Graph** **Code**

Theta Graph (Slide Layer)



Theta Code (Slide Layer)

Learner Parameters: Standard Normal $N(0,1)$

$$\theta_i \sim N(0, 1)$$

```
for(i in 1:n){  
  theta[i] ~ dnorm(0, 1)  
}
```

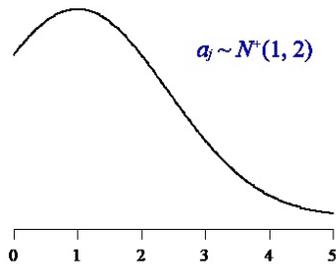
- Note JAGS software uses the **precision parameterization**
- Precision is the **inverse of variance**
- Precision = $1/\text{variance}$
- $1/1 = 1$

Precision
Parameterization

Back

Discrimination Graph (Slide Layer)

Discrimination Parameters: $N^+(1,2)$



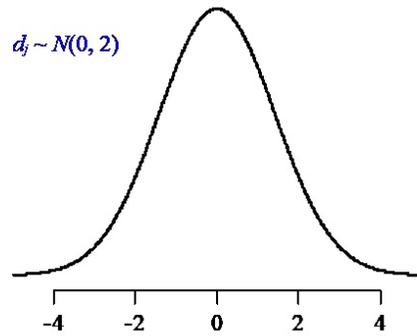
Normal distribution truncated to the positive real number line to:

- **resolve** rotational indeterminacy
- **reflect** desired interpretation of θ

Back

Difficulty Graph (Slide Layer)

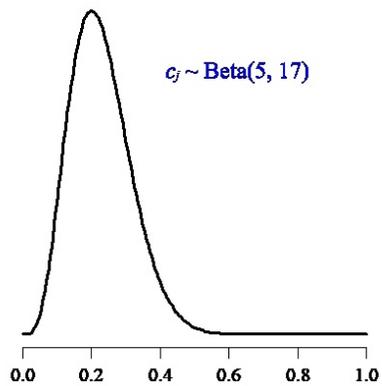
Difficulty Parameters: $N(0,2)$



Back

Guessing Graph (Slide Layer)

Lower Asymptote Parameters: $\text{Beta}(5,17)$



Back

Difficulty Code (Slide Layer)

Difficulty Parameters: $N(0,2)$

$$d_j \sim N(0, 2)$$

```
for(j in 1:J){  
  d[j] ~ dnorm(0, .5)  
}
```

- Note JAGS software uses the **precision parameterization**
- Precision is the **inverse of variance**
- Precision = 1/ variance
- $1/2 = 0.5$

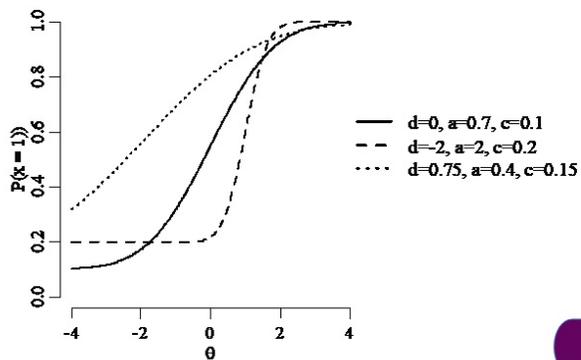
Precision
Parameterization

Back

IRT Curves Graph (Slide Layer)

IRT Response Probabilities

$$x_{ij} \mid \theta_i, d_j, a_j, c_j \sim \text{Bernoulli}[c_j + (1 - c_j)\Phi(a_j\theta_i + d_j)]$$



Back

IRT Model Code (Slide Layer)



$$x_{ij} \mid \theta_i, d_j, a_j, c_j \sim \text{Bernoulli}[c_j + (1 - c_j)\Phi(a_j\theta_i + d_j)]$$

```
for (i in 1:n){
  for (j in 1:J){
    P[i,j] <- c[j] + (1 - c[j]) * phi(a[j] * theta[i] + d[j])
    x[i,j] ~ dbern(P[i,j])
  }
}
```

Back

Discrimination Code (Slide Layer)



$$a_j \sim N^+(1, 2)$$

```
for (j in 1:J){
  a[j] ~ dnorm(1, .5) T(0, )
}
```

- Note JAGS software uses the **precision parameterization**
- Precision is the **inverse of variance**
- Precision = 1 / variance
- 1/2 = 0.5
- T(0,) imposes the truncation at 0

Precision
Parameterization

Back

Guessing Code (Slide Layer)



$$c_j \sim \text{Beta}(5, 17)$$

```
for(j in 1:J){  
  c[j] ~ dbeta(5, 17)  
}
```

Back

Precision (Slide Layer)



- $\tau = 1/\sigma^2$

- $x \mid \mu, \sigma^2 \sim N(\mu, \sigma^2) \rightarrow x \mid \mu, \tau \sim N(\mu, \tau)$

- Conditional probability of the data

$$x_i \mid \mu, \tau \sim N(x_i \mid \mu, \tau)$$

Precision of the data

- Prior distribution

$$\mu \mid \mu_\mu, \tau_\mu \sim N(\mu \mid \mu_\mu, \tau_\mu)$$

Precision of the prior

Back

4.22 Bookend: Setup



4.23 Posterior: Item Parameters

Home icon | **Posterior: Item Parameters**

Summary of Posterior (Burning in 11000, Saving 30,000 from 3 Chains)

Parameter	Mean	Median	SD	95% HPD lower	95% HPD Upper	Effective Size
d[1]	1.42	1.41	0.15	1.14	1.7	4109.02
d[2]	0.26	0.32	0.27	-0.22	0.62	822.88
d[3]	-0.57	-0.44	0.49	-1.6	0.1	559.73
d[4]	0.51	0.53	0.16	0.18	0.78	2973.71
d[5]	1.03	1.03	0.12	0.78	1.27	5722.96
a[1]						8112.45
a[2]						511.56
a[3]						475.5
a[4]						2394.03
a[5]						3169.05
c[1]	0.24	0.23	0.09	0.07	0.42	13572.3
c[2]	0.26	0.25	0.11	0.07	0.48	1515.2
c[3]	0.27	0.27	0.08	0.11	0.42	1092.04
c[4]	0.26	0.25	0.1	0.08	0.45	3473.51
c[5]	0.24	0.23	0.09	0.07	0.43	8034.22

Data not very informative about c_s , or at least not contradictory to the prior

- Recall the items are pretty easy
- Different amounts of evidence for different parameters

4.24 Extensions (I)



Extensions



- Complex models
- Polymtous observables
- Multiple latent variables
- Alternative priors
- Hierarchical specifications
- Covariates
- Other modeling families

4.25 Posterior (Learners)



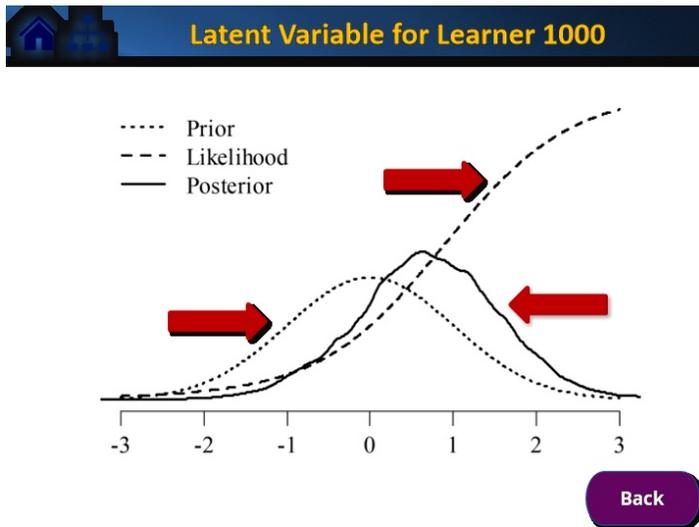
Posterior: Learner Parameters

Summary of Posterior (Burning in 11000, Saving 30,000 from 3 Chains)

Parameter	Mean	Median	SD	95% HPD lower	95% HPD Upper	Effective Size
theta[1]	-1.63	-1.61	0.77	-3.18	-0.17	50473.72
.
.
.
theta[1000]	0.70	0.72	0.88	-1.10	2.38	54701.95

[Example](#)

Example (Slide Layer)



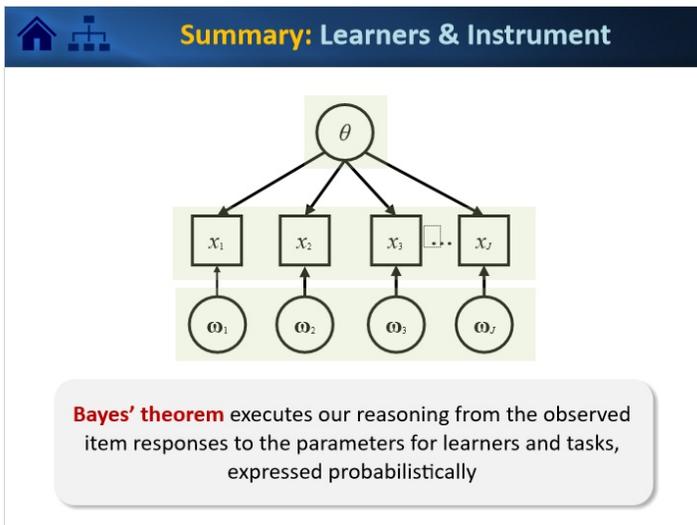
4.26 Bookend: Example



4.27 Bookmark: Applications and Extensions



4.28 Summary



4.29 Advantages



Advantages

- **Align with** theory and experience
- **Allow** estimation with “perfect” response patterns
- **Account for** and propagate uncertainty
- **Contrast with** conventional divide-and-conquer strategies



4.30 Bookend: Section 3



If you are interested in taking a **self-assessment** on this section click here: [Quiz](#)

If you are interested in seeing worked data examples of analyses in an **R package** click here: [Data Examples](#)

If you want to return to the **main menu** click here: [Main Menu](#)