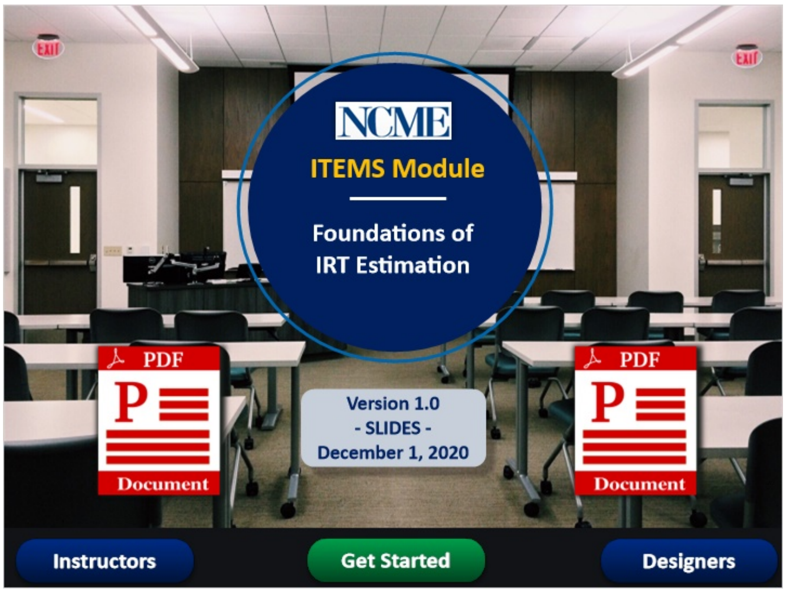
DM19 SLIDES (IRT Estimation, Version 1.0)

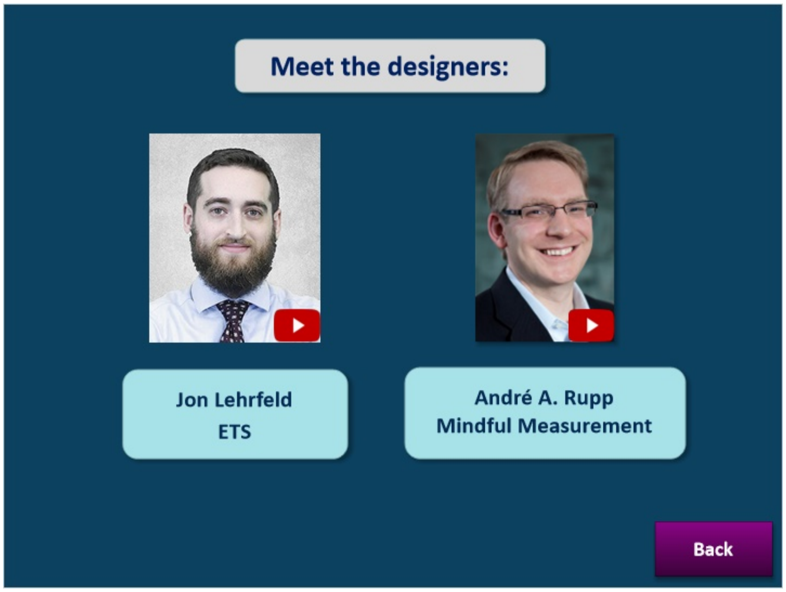
1. Module Overview

1.1 Module Cover (START)

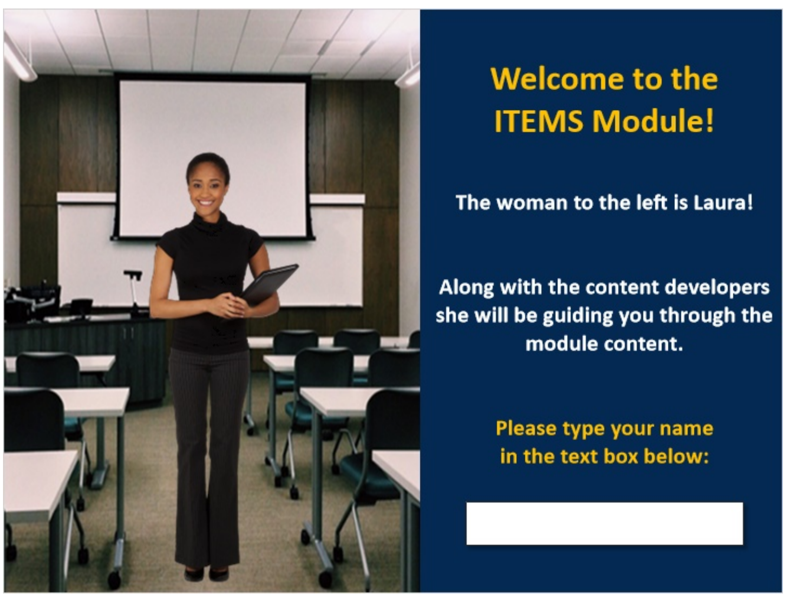


1.2 Instructors

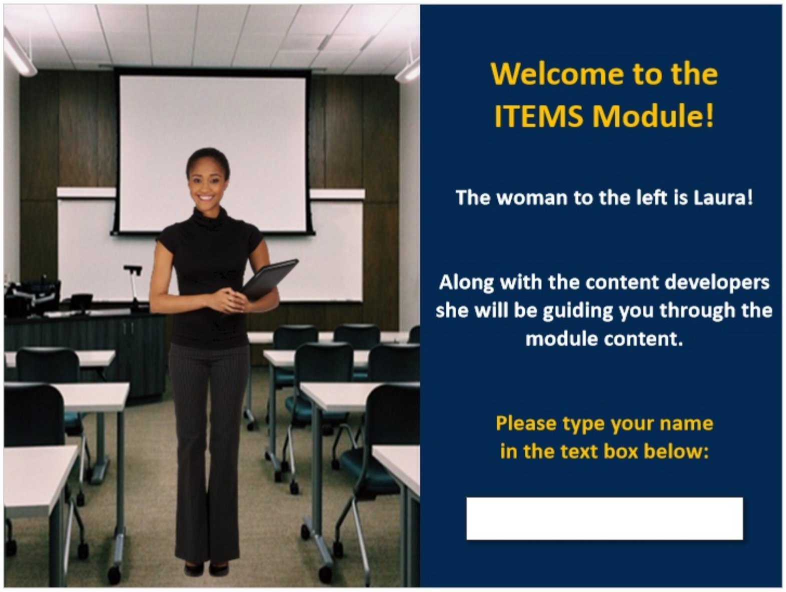
1.3 Designers



1.4 Welcome

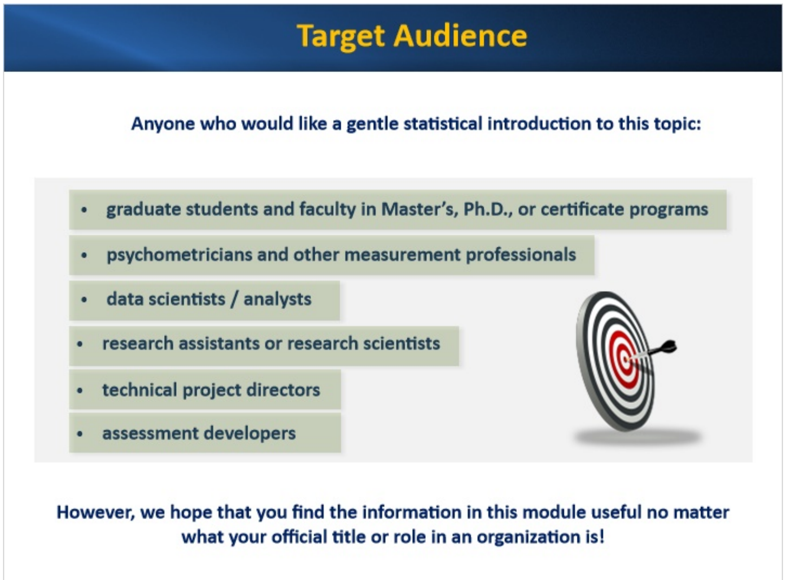


Untitled Layer 1 (Slide Layer)



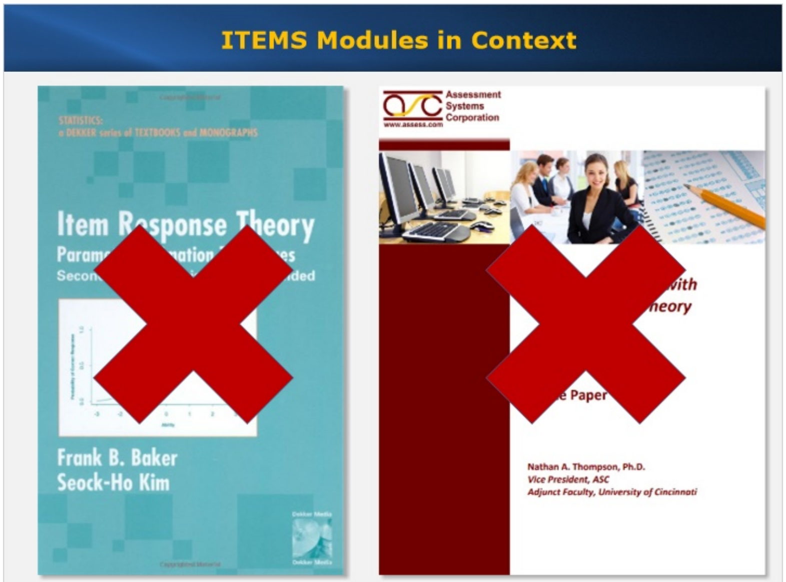
1.5 Overview

1.6 Target Audience



1.7 Expecations (I)

1.8 Expectations (II)



1.9 Learning Objectives

1.10 Prerequisites

Prerequisites

- **Foundational IRT concepts:**
  - ✓ Item response functions for 1/2/3PL models
  - ✓ Goal and logic of item calibration and examinee scoring
  - ✓ Parameter recovery using response matrix
- **Probability**
  - ✓ Prior and posterior distributions
- **R statistical software**
  - ✓ Useful for Section 3: Factors Affecting Estimation Accuracy

1.11 Resources

Resources

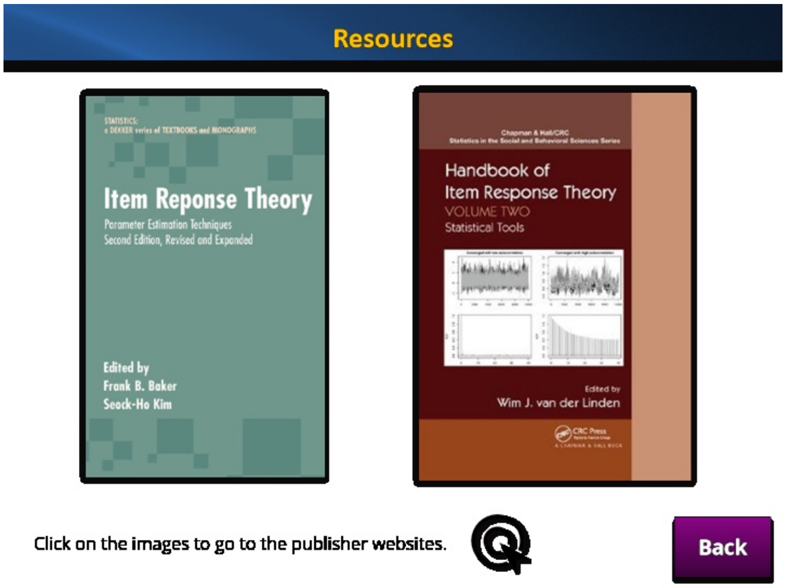
Module Citation

Wang, Z., & Thompson, N. (2020). Foundations of IRT estimation (Digital ITEMS Module 19). *Educational Measurement: Issues and Practice, 39*(4), 134-135.

Sample R Code

Additional Resources

References (Slide Layer)

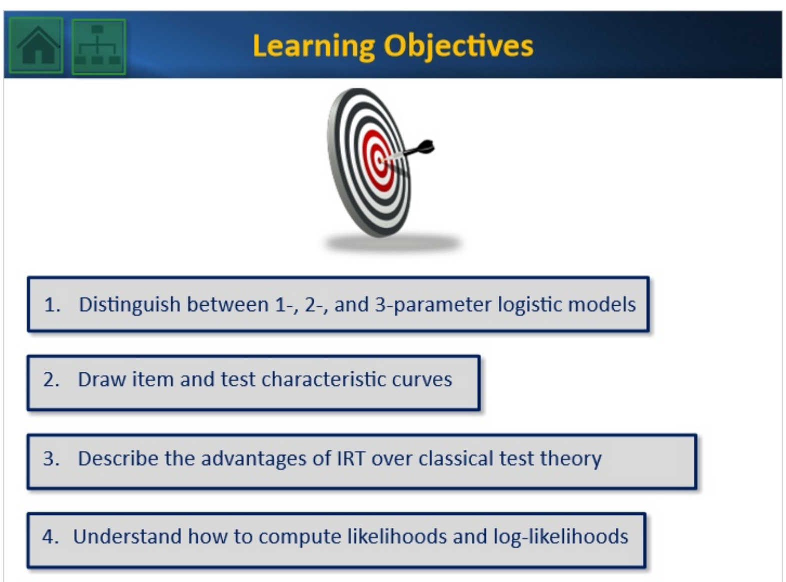


1.12 Main Menu

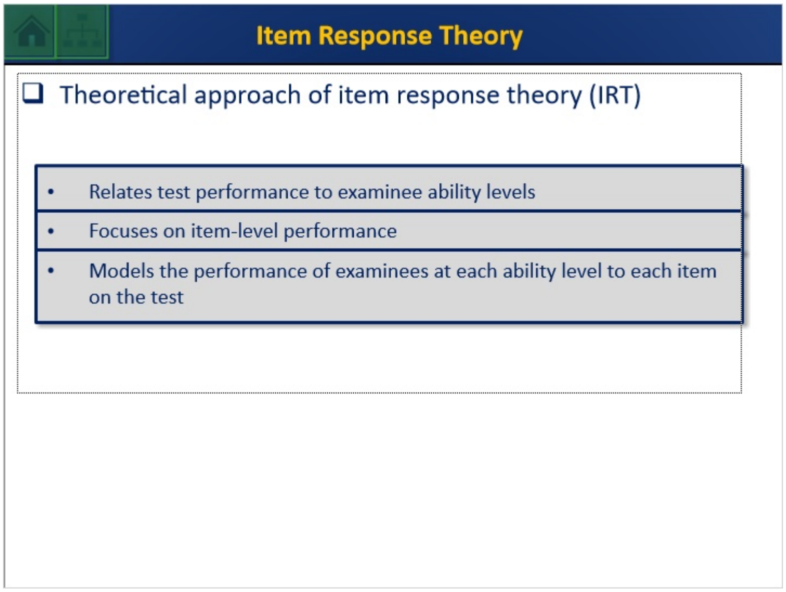
2. Section 1: IRT Basics

2.1 Cover: IRT Basics

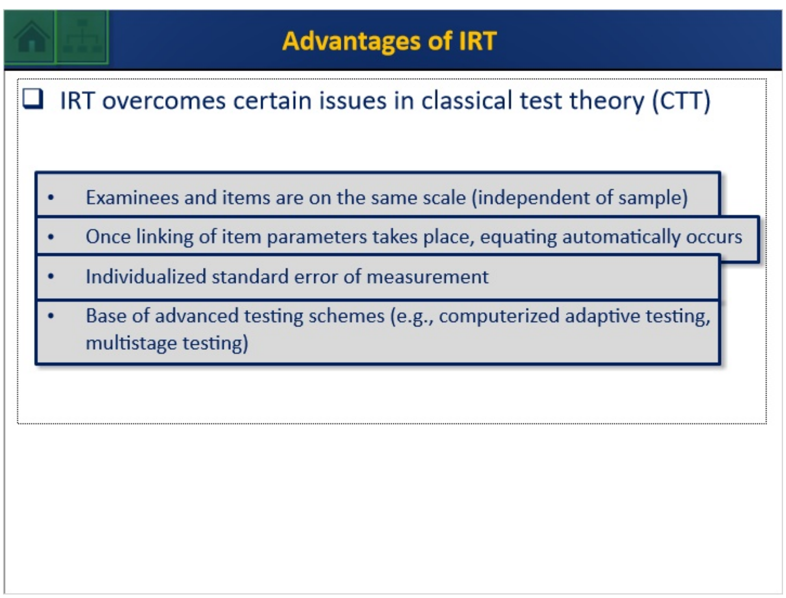
2.2 IRT Basics: Learning Objectives



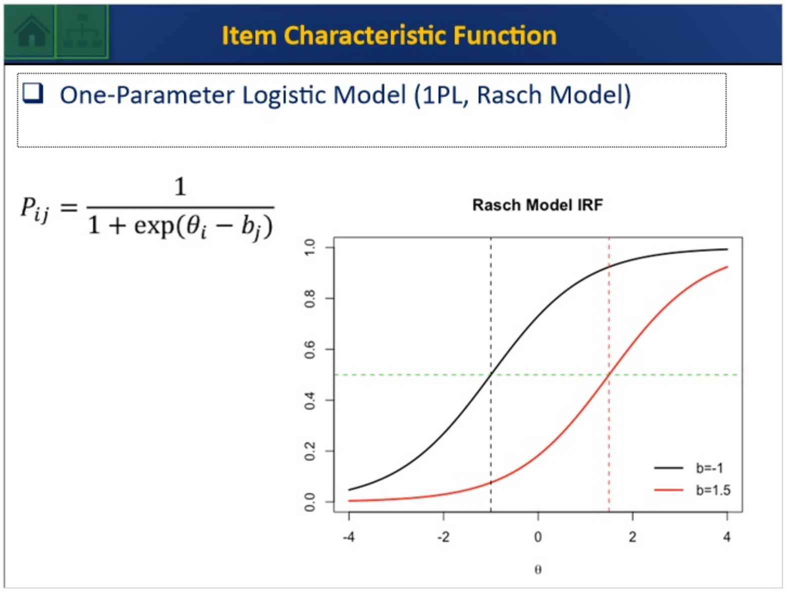
2.3 Item Response Theory



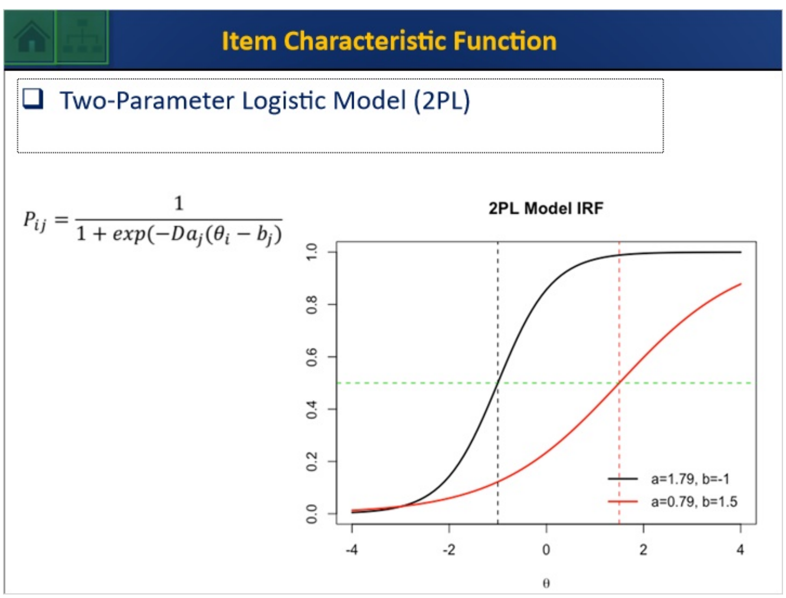
2.4 Advantages of IRT



2.5 Item Characteristic Functions I



2.6 Item Characteristic Functions II

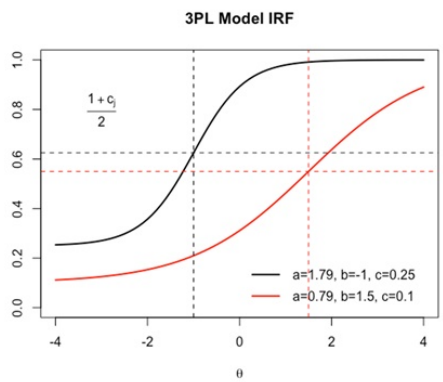


2.7 Item Characteristic Functions III

Item Characteristic Function

- Three-Parameter Logistic Model (3PL)

$$P_{ij} = c + \frac{1 - c}{1 + \exp(-Da_j(\theta_i - b_j))}$$

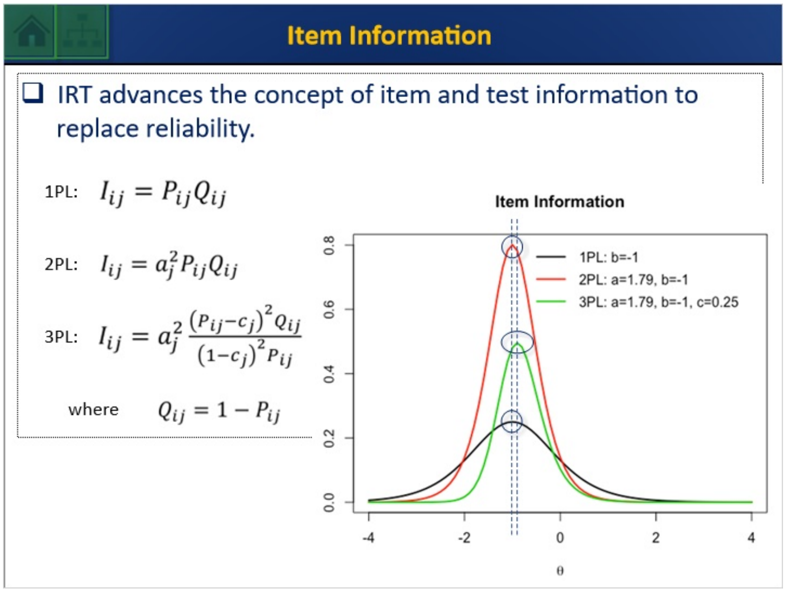


2.8 Simulation Example

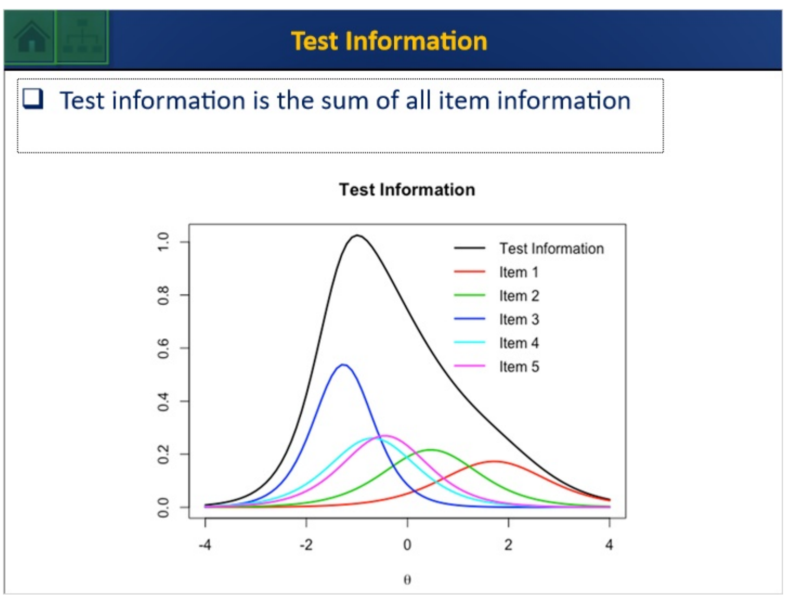
Simulation Example

- 2PL Example
  - 5 items
  - a~Lognormal(0,1)
  - b~Normal(0,1)
  - θ~Normal(0,1)
  - Response for examinee #1 is (0, 1, 0, 0, 1) [“Mixed” response]

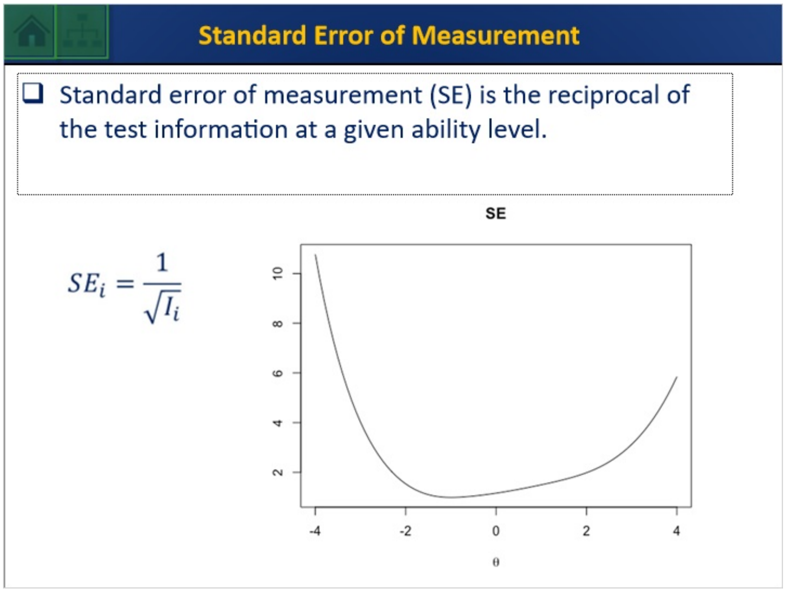
2.9 Item Information



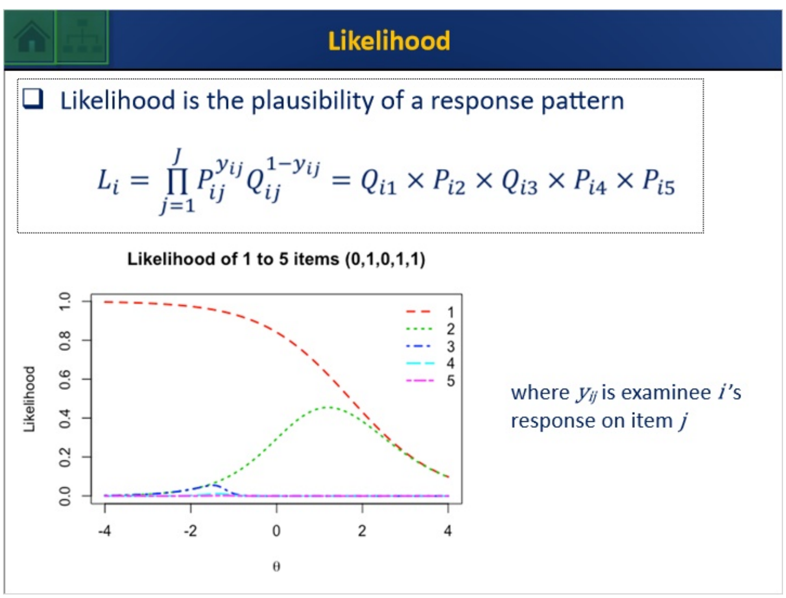
2.10 Test Information



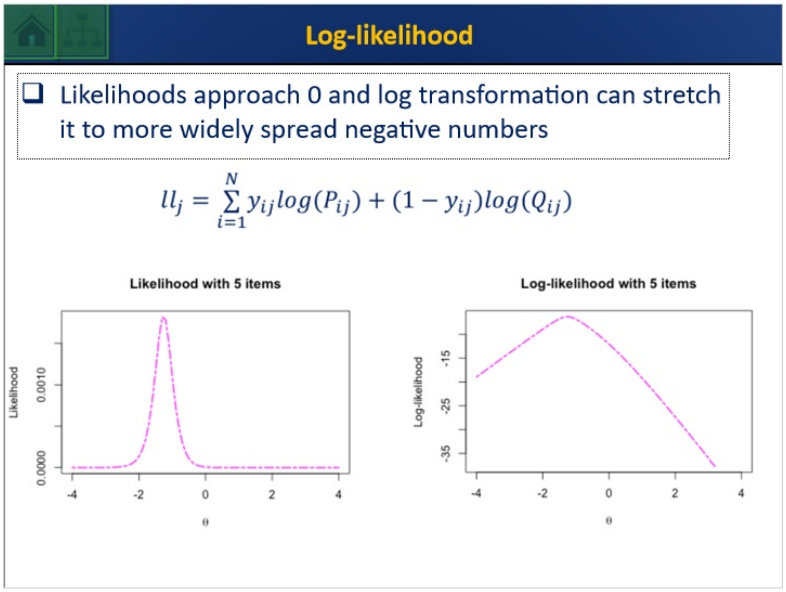
2.11 Standard Error of Measurement



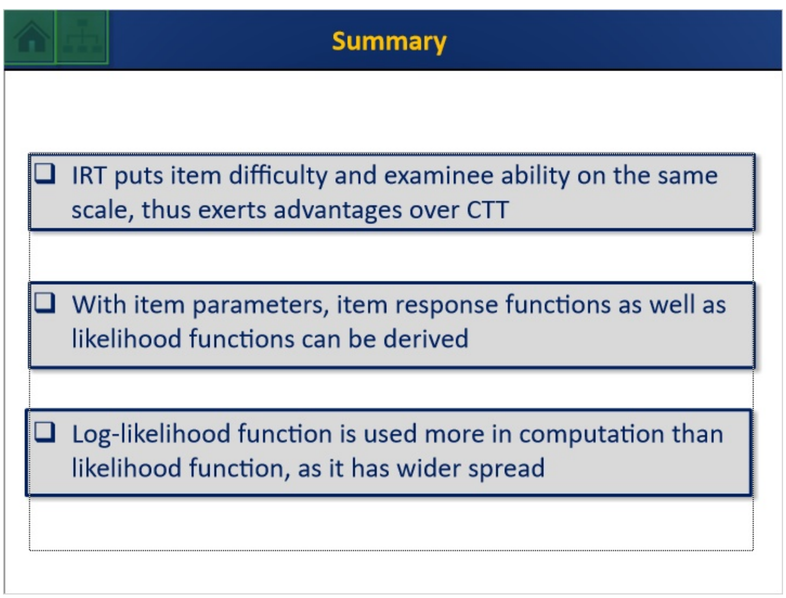
2.12 Likelihood



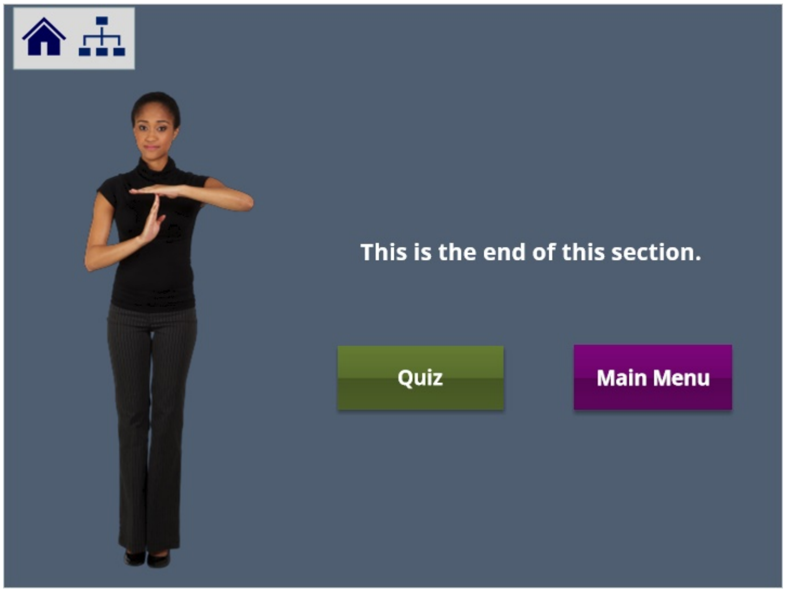
2.13 Log-Likelihood



2.14 Summary

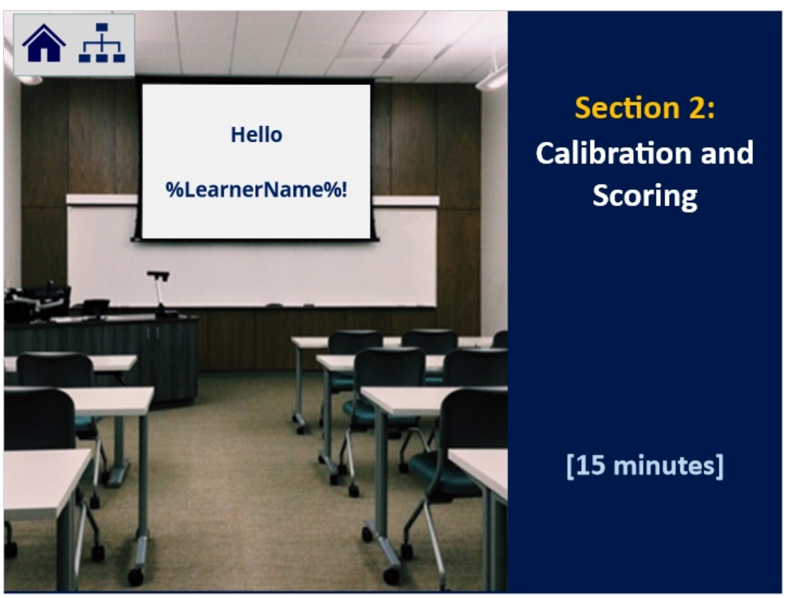


2.15 Bookend: IRT Basics

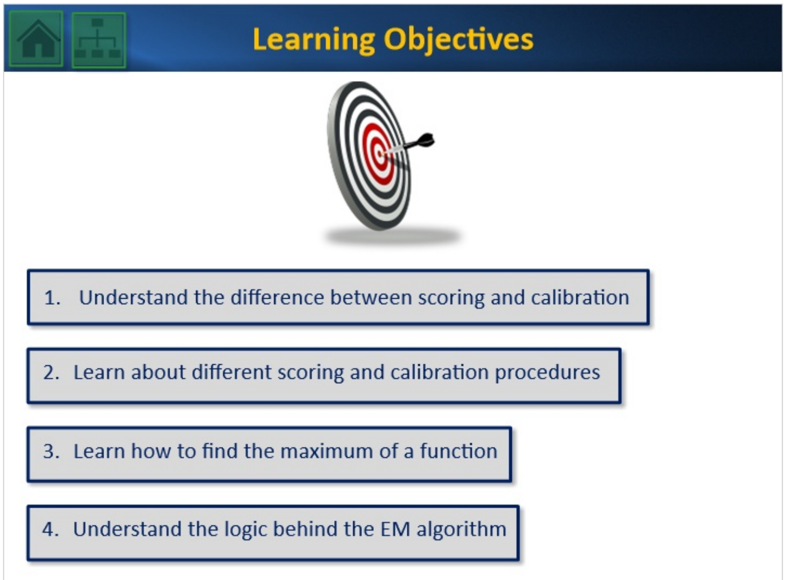


3. Section 2: Calibration and Scoring

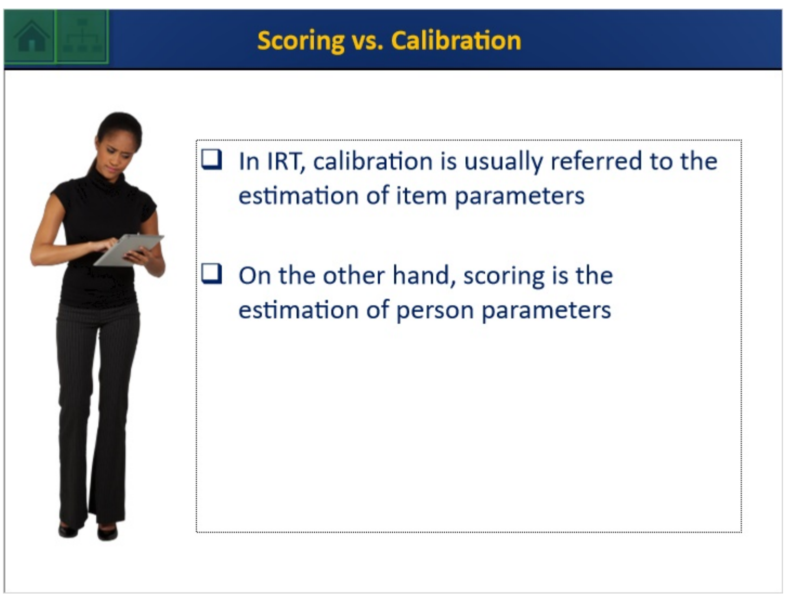
3.1 Cover: Calibration and Scoring



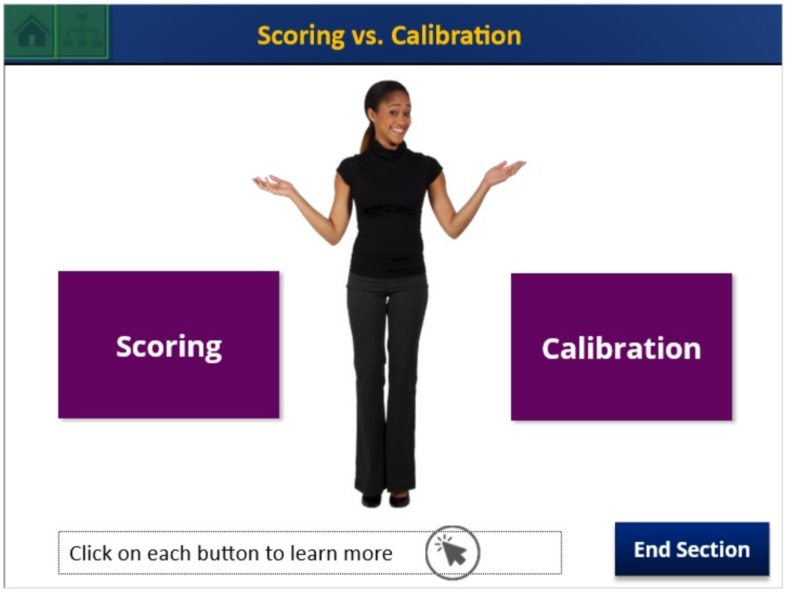
3.2 Estimation Algorithms: Learning Objectives



3.3 Scoring vs. Calibration I

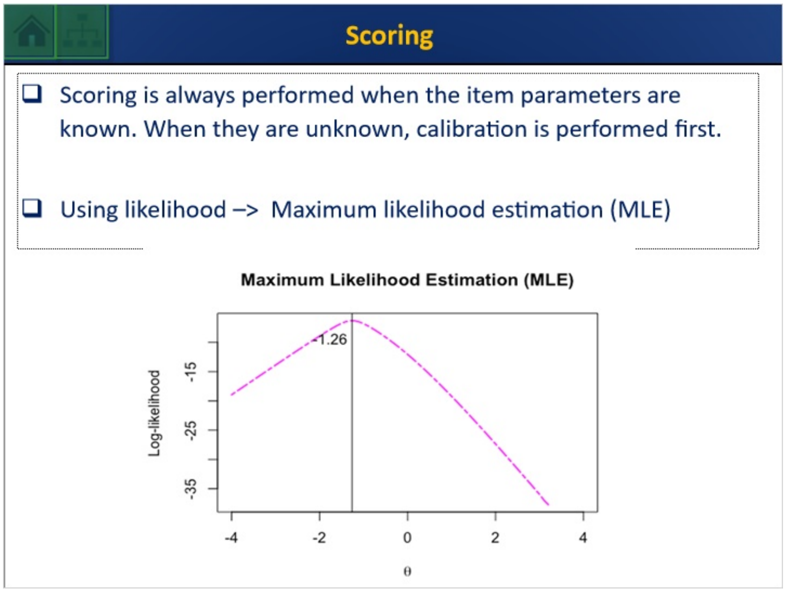


3.4 Topic Selection

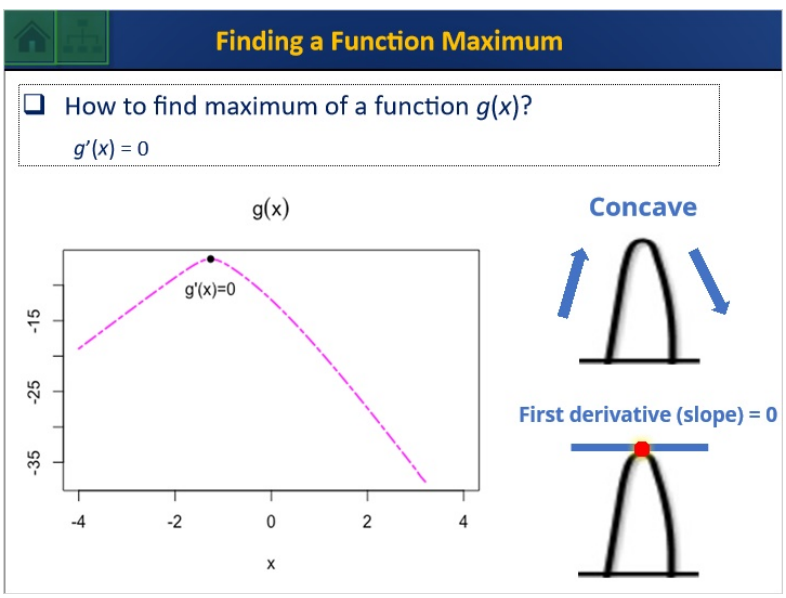


3.5 Bookmark: Scoring

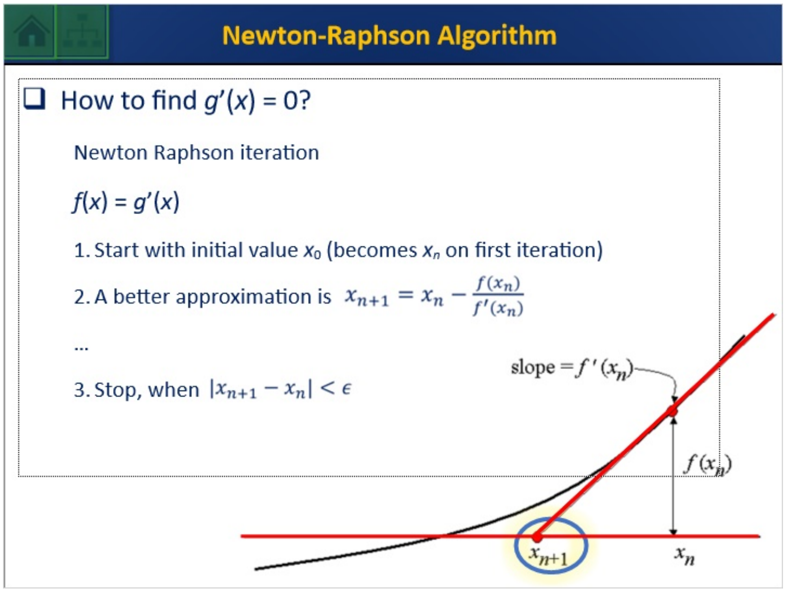
3.6 Scoring: MLE



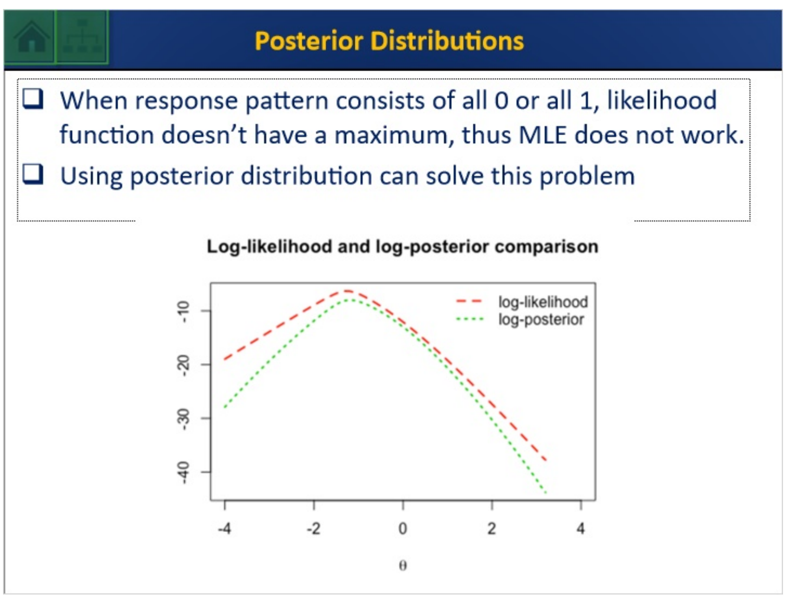
3.7 Finding a Function Maximum



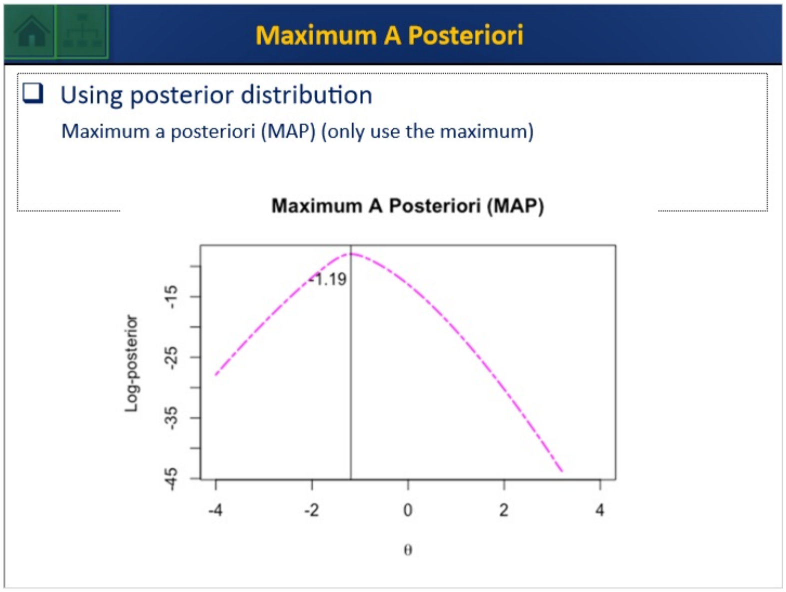
3.8 Newton-Raphson Algorithm



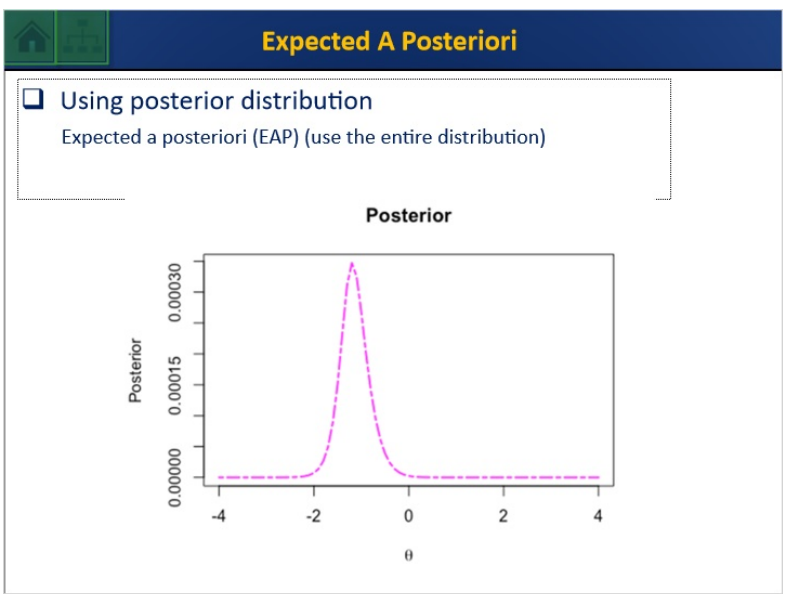
3.9 Posterior Distributions



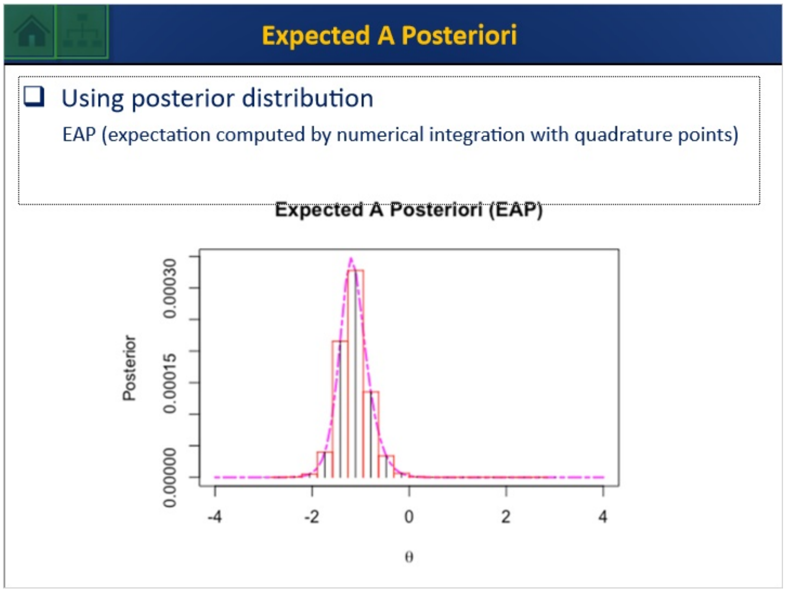
3.10 Maximum A Posteriori



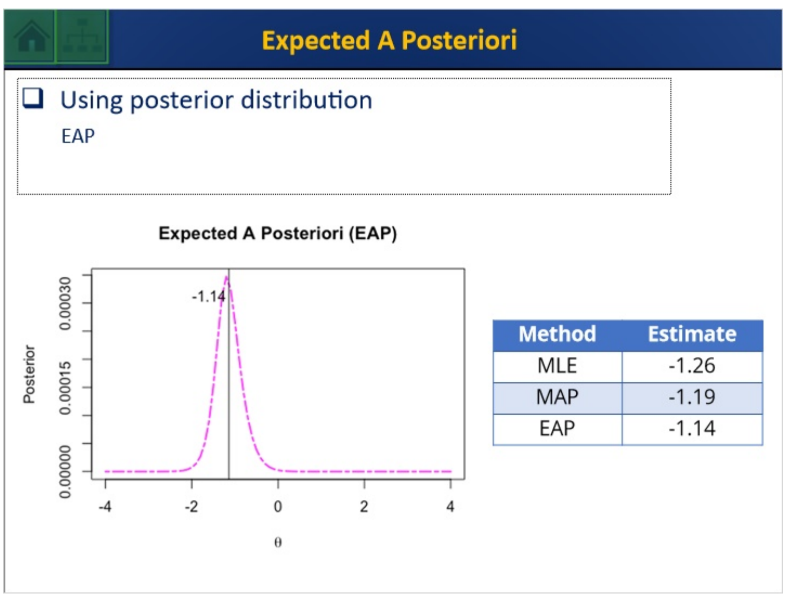
3.11 Expected A Posteriori I



3.12 Expected A Posteriori II



3.13 Expected A Posteriori III



Method	Estimate
MLE	-1.26
MAP	-1.19
EAP	-1.14

3.14 Bookend: Scoring

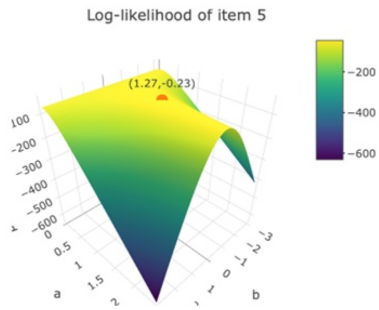
3.15 Bookmark: Calibration

3.16 Calibration with Unknown Item Parameters

Calibration with Known Examinee Parameters

- If examinee parameters are known, MLE is used

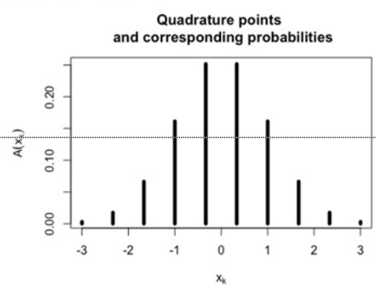
$$ll_j = \sum_{i=1}^{N} y_{ij} log(P_{ij}) + (1 - y_{ij}) log(Q_{ij})$$



3.17 Calibration with Unknown Item and Person Parameters

Calibration with Unknown Item and Person Parameters

- Marginal maximum likelihood estimation (MMLE)
  - Although examinee parameters are unknown, we can assume a known population distribution and replace the person parameters with quadrature points.
  - x_k are quadrature points and $A(x_k)$ are corresponding quadrature weights/ probabilities under the population distribution
  - The quadrature weights sum to 1



3.18 MMLE Log-Likelihood

MMLE Log-Likelihood

- MMLE
  - Replace single response y_{ij} with $\overline{r}_{kj}$ which is number of correct answer in group *k*; replace $1-y_{ij}$ with $\overline{n}_k - \overline{r}_{kj}$, where $\overline{n}_k$ is the size of group *k*.
  - $\overline{n}_k$ and $\overline{r}_{kj}$ are called artificial data
  - The log-likelihood function for item *j* becomes:

$$ll_j = \sum_{i=1}^{N} \overline{r}_{kj} log(P_{kj}) + (\overline{n}_k - \overline{r}_{kj}) log(Q_{kj})$$

3.19 MMLE Log-Likelihood II

MMLE Log-Likelihood

- MMLE
  - $L_{ik} = \prod_{j=1}^{J} P_{kj}^{y_{ij}} Q_{kj}^{1-y_{ij}}$ is the likelihood of a examinee *i* with ability x_k
  - Compute $\overline{n}_k$ and $\overline{r}_{kj}$ based on known item parameters and responses

$$\overline{n}_k = \sum_{i=1}^{N} \frac{L_{ik} A(x_k)}{\sum_{k=1}^{K} L_{ik} A(x_k)}$$

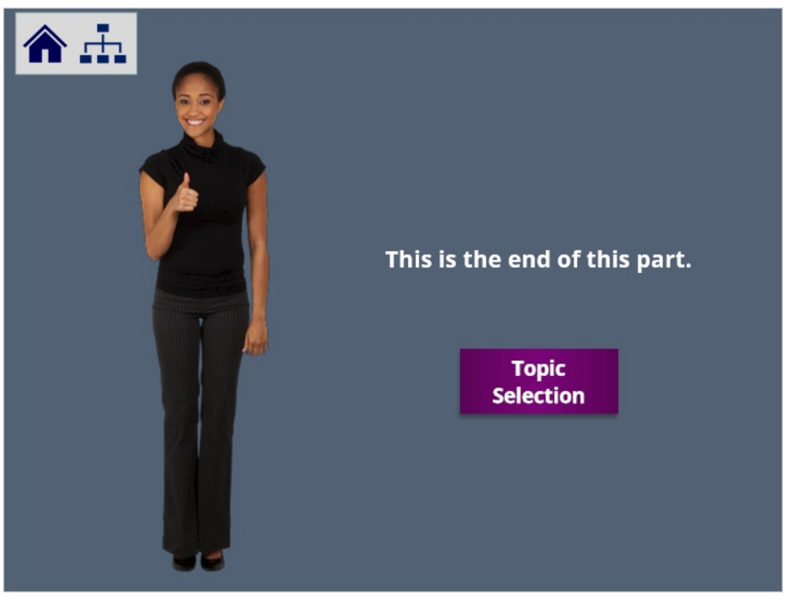
$$\overline{r}_{jk} = \sum_{i=1}^{N} \frac{y_{ij} L_{ik} A(x_k)}{\sum_{k=1}^{K} L_{ik} A(x_k)}$$

3.20 EM Algorithm

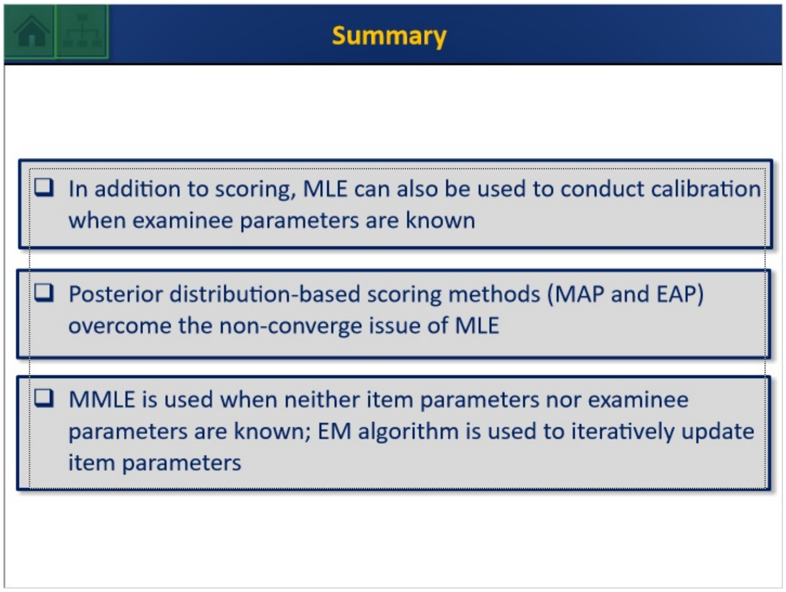
Expectation-Maximization (EM) Algorithm

- In reality, item parameters are usually unknown
- The EM algorithm iteratively estimates item parameters
- Begin by assigning starting values for item parameters, then cycle between E-steps and M-steps until the estimates converge
  - E-step: assume item parameters are known and compute $\overline{n}_k$ and $\overline{r}_{kj}$
  - M-step: using previously calculated values of $\overline{n}_k$ and $\overline{r}_{kj}$, and compute MLE of item parameters

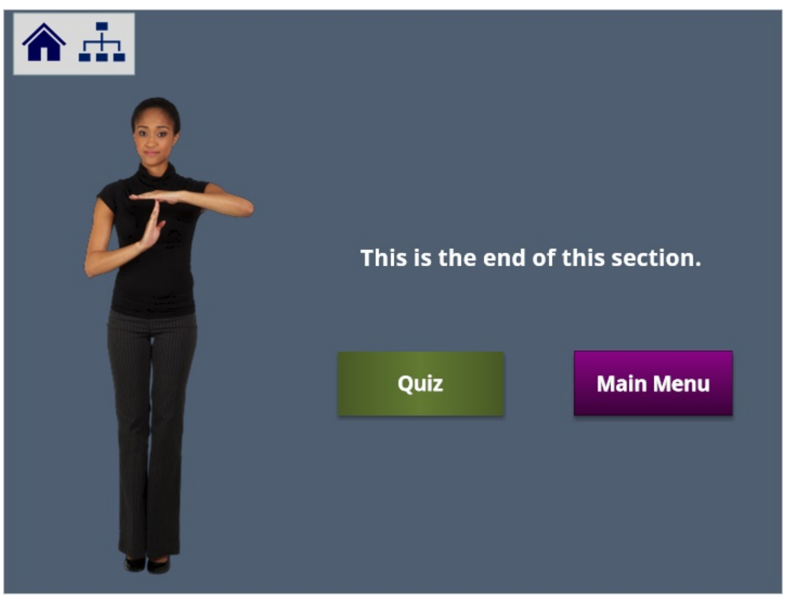
3.21 Bookend: Calibration



3.22 Summary

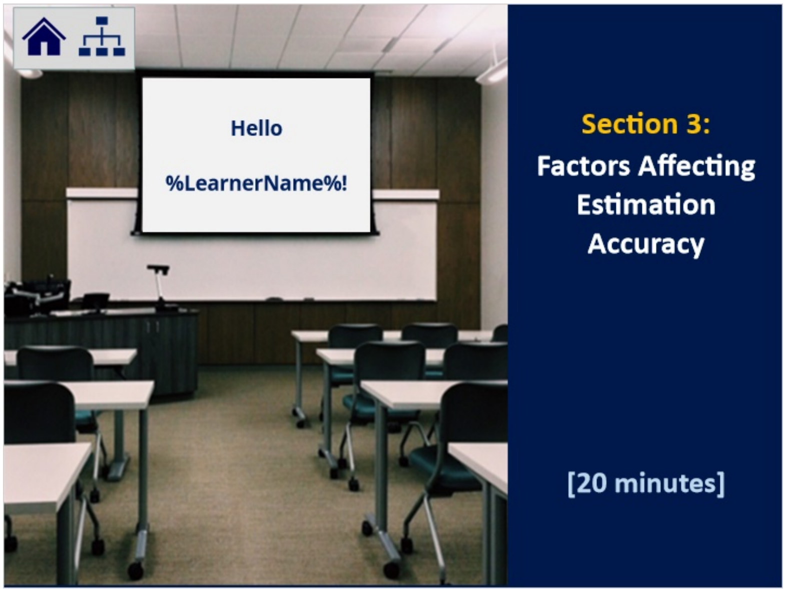


3.23 Bookend: IRT Basics

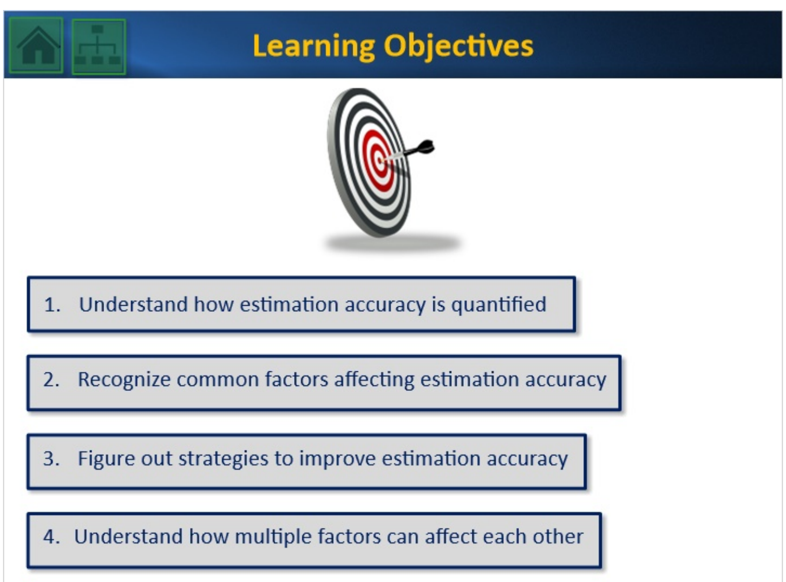


4. Section 3: Factors Affecting Estimation Accuracy

4.1 Cover: Factors Affecting Estimation Accuracy



4.2 Factors Affecting Estimation Accuracy: Learning Objectives



4.3 Estimation Accuracy Evaluation Criteria

Estimation Accuracy Evaluation Criteria

- ❑ Bias and root mean square error (RMSE) are used to evaluate the estimation accurac of calibration and scoring
- ❑ π is the true value of a parameter and $\hat{\pi}$ is the estimated value of this parameter

Bias:

$$\boxed{\frac{\sum_{i=1}^{n}(\hat{\pi}_i - \pi_i)}{n}}$$

RMSE:

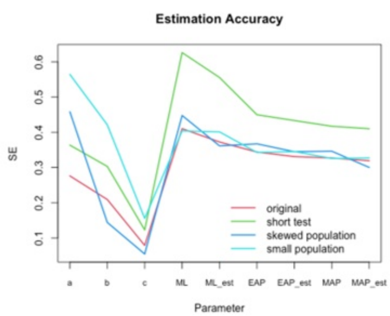
$$\boxed{\sqrt{\frac{\sum_{i=1}^{n}(\hat{\pi}_i - \pi_i)^2}{n}}}$$

- ❑ The *proportional* bias and RMSE are computed as the proportion of bias and RMSE of a parameter over the SD of that parameter.

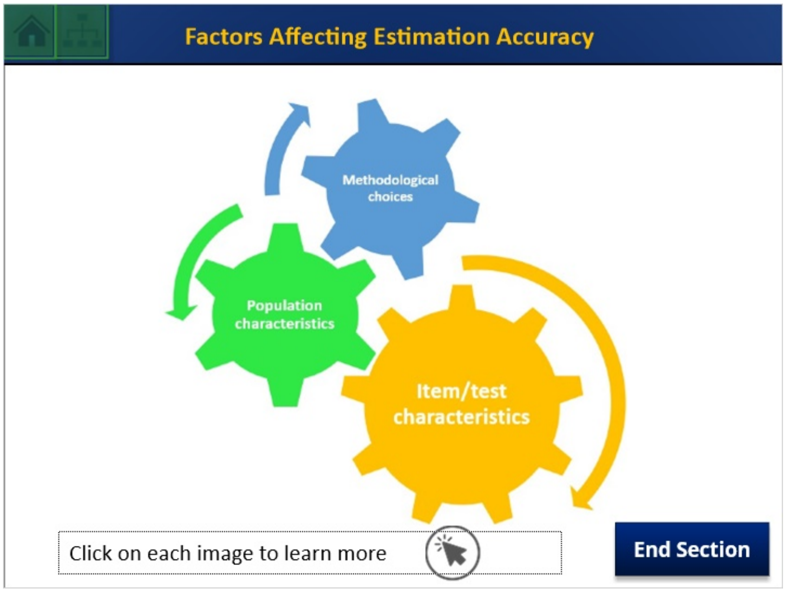
4.4 Identify the Need to Improve Estimation Accuracy

Identify the Need to Improve Estimation Accuracy

- ❑ How to identify the need of improving estimation accuracy?
  - Look at population size.
    - ➢ 500 for 2PL
    - ➢ 1000 for 3PL
  - Look at SE of parameter estimation (especially *a* parameter)

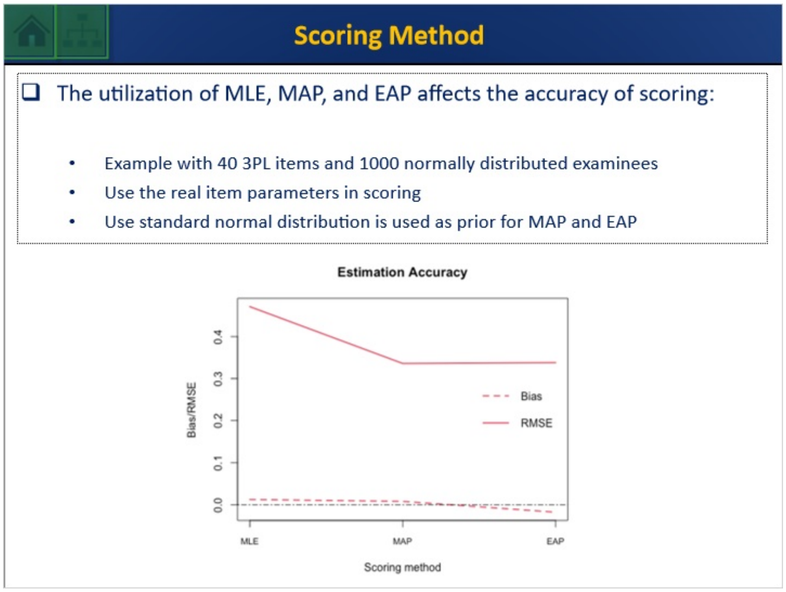


4.5 Topic Selection

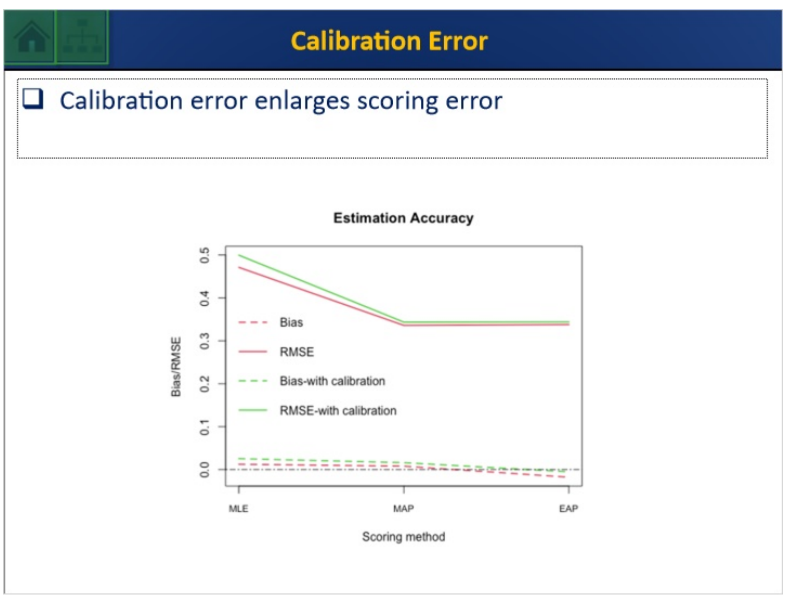


4.6 Bookmark: Methodological Choices

4.7 Scoring Method



4.8 Calibration Error



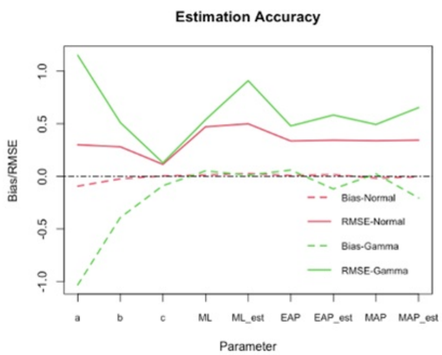
4.9 Bookend: Calibration

4.10 Bookmark: Population Characteristics

4.11 Population Distribution I

Population Distribution

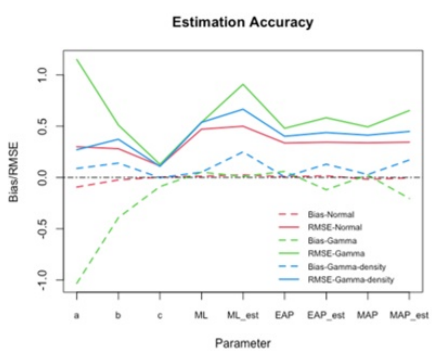
- Skewed ability distribution increases error in both item and examinee parameter estimation



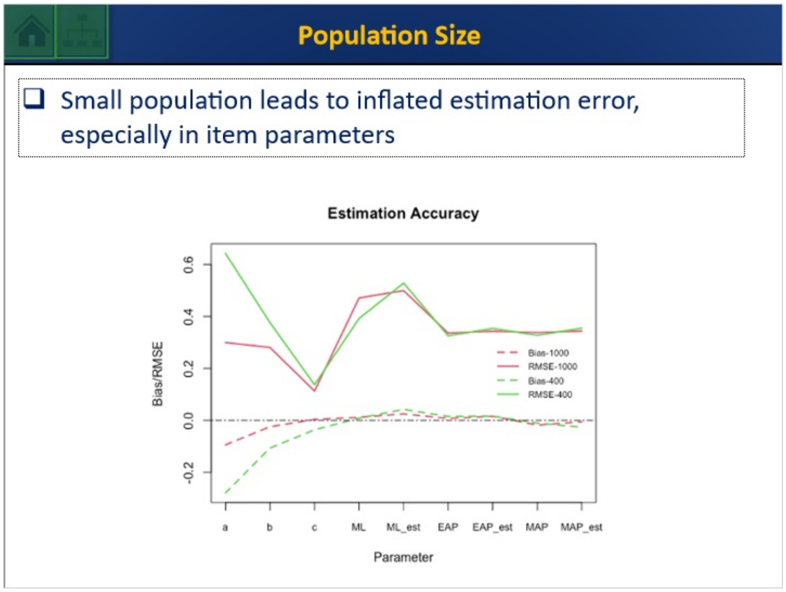
4.12 Population Distribution II

Population Distribution

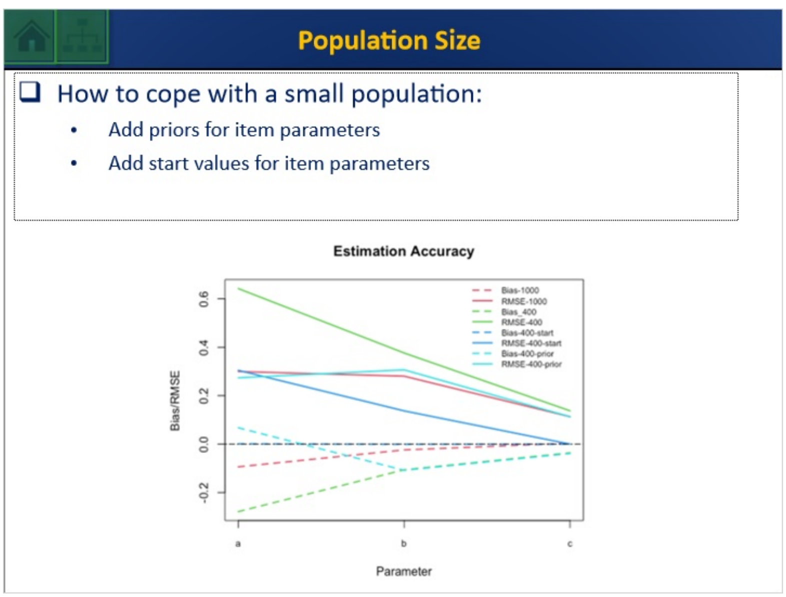
- Cope with skewed ability distribution:
  - Add population density
  - Use empirical histogram for latent trait distribution in MMLE



4.13 Population Size I



4.14 Population Size II



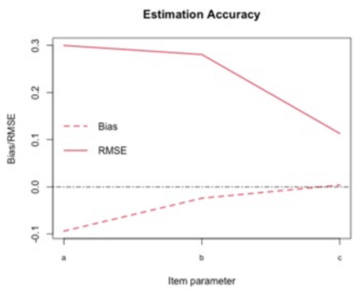
4.15 Bookend: Calibration

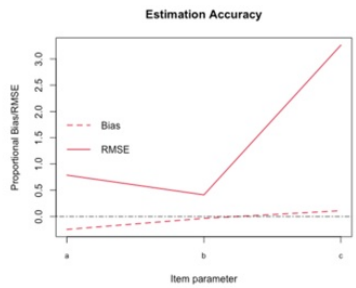
4.16 Bookmark: Item and Test Characteristics

4.17 Item Parameter Types

Item Parameter Type

- In 3PL, *b* parameters are easier to calibrate, thus have relatively smaller estimation error

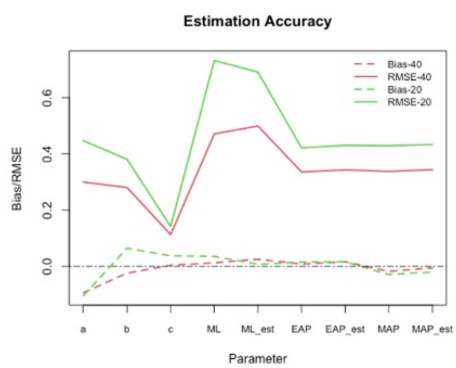




4.18 Test Length

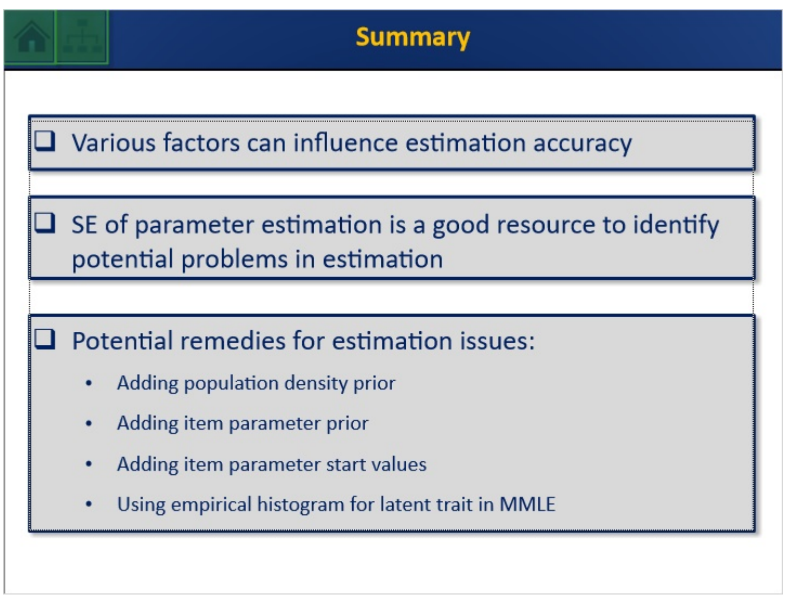
Test Length

- Short tests increase error in both item and person parameter estimation
- The advantage of MAP and EAP is greater for short tests

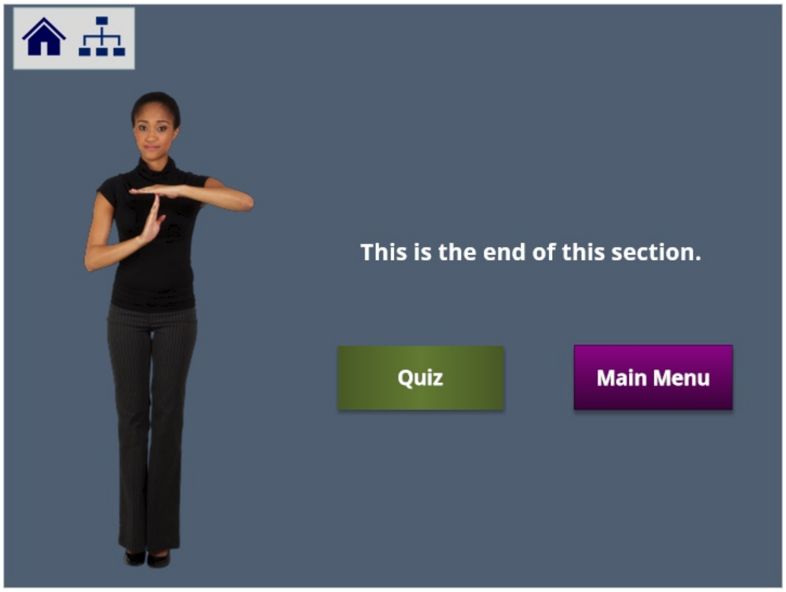


4.19 Bookend: Calibration

4.20 Summary



4.21 Bookend: IRT Basics



4.22 Module Cover (START)