THE UNIVERSITY OF KANSAS

A Bayesian Method For Incorporating Uncertainty into Q-matrix Estimation in Skills Assessment

Jonathan Templin
Department of Psychology
University of Kansas

Robert A. Henson
Department of Educational Research Methodology
The University of North Carolina at Greensboro

Skills Assessment Procedures

- Up to this point, we have seen a couple demonstrations of Q-matrix building with differing cognitive diagnosis models.
- In both cases, Q-matrix construction was accomplished by a sequence of analyses.
  - The Q-matrix went through successive revisions after each analysis.
- In this talk, I will present a Bayesian method that allows for Q-matrix entries to be evaluated algorithmically.
  - The aim of this method is to allow for empirical evidence to guide in Q-matrix development.

The Q-Matrix

- A Q-matrix is a matrix of binary indicators that specify the set of attributes measured by each item.
- The Q-matrix specifies the set of latent traits necessary for each item (like the factor pattern matrix input into a confirmatory factor analysis).
- As such, individuals experienced in specific content domains typically construct the Q-matrix.
  - Larger scale tests/applications may have a set of several experts to design Q-matrix specifications.

Q-matrix Example

- As an example, consider selected items from a test of fraction subtraction (Tatsuoka, 1990), where eight attributes were defined by de la Torre and Douglas (2004).
- Example Items:

4. $3\frac{1}{2} - 2\frac{3}{2}$
   - Skills #2, 3, 5, and 7 needed.

15. $2 - \frac{1}{3}$
   - Skills #1 and 8 needed.

Fraction Subtraction Skills

1. Convert a whole number to a fraction.
2. Separate a whole number from fraction.
3. Simplify before subtracting.
4. Find a common denominator.
5. Borrow from whole number part.
6. Column borrow to subtract the second numerator from the first.
7. Subtract numerators.
8. Reduce answers to simplest form.

Models for Skills Assessment

Cognitive diagnosis models are constrained latent class models, commonly used for skills assessment:

- For K specified skills, set of 2^K classes are defined.
- Choice of cognitive diagnosis model and Q-matrix determines the set of equality constraints placed on the latent class response probabilities.
- Models such as the DINA and RUM are non-compensatory, meaning that the absence of necessary traits cannot be compensated for by others.

Latent Class Models

A latent class model for the response vector of individual i (C classes, $j = 1, \ldots, J$ items):

Structural Measurement

$$P(\mathbf{X}_i = \mathbf{x}_i) = \sum_{c=1}^{C} \underbrace{P(c)}_{\text{Structural}} \prod_{j=1}^{J} \underbrace{P(X_{ij} = x_{ij} | c)}_{\text{Measurement}}$$

In cognitive diagnosis models:

- Structural portion represents probability of a skill pattern.
- Measurement portion represents "item response" function.
  - Where different cognitive diagnosis models are defined.

The DINA Model

Introduction
Latent Class Models
➤ Latent Class Model
➤ The DINA Model
Probabilistic Q-matrix
Algorithm Testing
Discussion

- Deterministic Input; Noisy "And" Gate (Macready and Dayton, 1977; Haertel, 1989; Junker and Sijtsma, 2001).

$$P(x_{ij} = 1|\xi_{ij}) = (1 - s_j)^{\xi_{ij}} g_j^{(1-\xi_{ij})}$$

where

$$\xi_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$$

$$1 - s_j = P(x_{ij} = 1|\xi_{ij} = 1)$$

$$g_j = P(x_{ij} = 1|\xi_{ij} = 0)$$

- Q-matrix entries act as "switches" where examinee response probabilities are forced into one of two possible values: $(1 - s_j)$ or g_j.

- No matter how many skills are indicated for an item, the DINA model has only two probability classes per item.

DINA Item Response Constraints

Total Attributes: 4

Total Classes: 16

Model: DINA

Q-matrix:

	α_1	α_2	α_3	α_4
Item	Add	Sub	Mult	Div
1.) $2+3-1$	1	1	0	0
2.) $4/2$	0	0	0	1
3.) $3 \times (4-2)$	0	1	1	0

Equivalence Classes Per Item: 2

Item	1	2	3
Classes	2	2	2

Equal $P(X_{ij} = 1|\boldsymbol{\alpha}_i)$ Denoted By Color

	Item		
$\boldsymbol{\alpha}$	1	2	3
[0000]	g_1	g_2	g_3
[0001]	g_1	$1-s_2$	g_3
[0010]	g_1	g_2	g_3
[0011]	g_1	$1-s_2$	g_3
[0100]	g_1	g_2	g_3
[0101]	g_1	$1-s_2$	g_3
[0110]	g_1	g_2	$1-s_3$
[0111]	g_1	$1-s_2$	$1-s_3$
[1000]	g_1	g_2	g_3
[1001]	g_1	$1-s_2$	g_3
[1010]	g_1	g_2	g_3
[1011]	g_1	$1-s_2$	g_3
[1100]	$1-s_1$	g_2	g_3
[1101]	$1-s_1$	$1-s_2$	g_3
[1110]	$1-s_1$	g_2	$1-s_3$
[1111]	$1-s_1$	$1-s_2$	$1-s_3$

Q-Matrix Background

- Preponderance of cognitive diagnosis models require Q-matrix elements to be fixed.
  - A method for estimating the magnitude of Q-matrix entries for polytomous-valued skills was developed by Karelitz (2004).
- Because of the Q-matrix certainty needed by most cognitive diagnosis models, it is difficult to determine *all* elements of the Q-matrix correctly.
  - Evidence of improper Q-matrix specification is manifested in model parameter estimates indicating “masters” often miss items or “non-masters” often answer correctly.
  - With the DINA model, this would be reflected by having large slip or guess parameters.

Probabilistic Q-Matrix Algorithm

- For these reasons, we introduce a Bayesian estimation procedure that estimates entries in the Q-matrix.

- In this procedure, users are allowed to specify Q-matrix entries in terms of the (subjective) *probability* an item "requires" a given attribute.

  - The common Q-matrix can be thought of as having all elements decided with absolute certainty (probability of zero or one).

- As a result of the estimation procedure, posterior probabilities of Q-matrix entries are obtained, indicating the likelihood a skill is "required" for a successful response to an item.

Estimation Algorithm

- MCMC estimation algorithm estimating (by use of Metropolis-Hastings with in Gibbs algorithm):
  - Q-matrix elements (all that specified with uncertainty).
    - Prior for each Q-matrix element is defined by specified probability of necessity of Q-matrix element.
  - DINA item parameters (s, g).
    - Uniform prior for all DINA item parameters (s, g).
  - Structural parameters describing the joint distribution of skill patterns (using Tetrachoric correlations).
    - Uniform prior for all mean and covariance structural parameters $(\kappa, \boldsymbol{\rho})$.
  - Latent skill patterns $(\boldsymbol{\alpha})$ modeled with empirical prior defined by structural model (underlying $N_K(0, \boldsymbol{\rho})$ variates bisected by cut-point parameters κ).

Simulation Study Design

- Under each condition, a simulated 30-item test was generated, with responses from 1000 generated examinees.
- The simulated item parameters were indicative of a quality diagnostic test.
  - All s and g parameters were less than 0.25.
- The experimental conditions randomly varied the type of entries in the Q-matrix in terms of the proportion of subjective probability entries in the Q-matrix, and the subjective probability specified for each entry.
- For each condition:
  - Chain length of 20,000 (first 10,000 used as burnin).
  - Chain convergence checked by Geweke test and by visual inspection of chains for each parameter.

Simulation Study Design

- Two crossed experimental conditions:

1. Proportion of Q-matrix entries specified probabilistically.
   - Each element of the true Q-matrix had chance to be replaced by subjective probability according to random process.
   - Three conditions for the probability an element was replaced: 0.25, 0.50, and 0.75.

2. Subjective probability assigned to replaced elements.
   - If an element of the Q-matrix was selected to be replaced by subjective probability, probability level was assigned randomly.
   - Three levels: subjective probability randomly chosen (with equal chance) from 0.25, 0.50, and 0.75.

Simulation Study Results

Average absolute difference between true Q-matrix entry and estimated posterior probability:

	Probability of replacement: 0.25			Probability of replacement: 0.50			Probability of replacement: 0.75		
	Subjective Probability			Subjective Probability			Subjective Probability		
True Q	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
0	0.0019	0.0019	0.0016	0.0024	0.0018	0.0029	0.0050	0.0070	0.0381
1	0.0000	0.0000	0.0004	0.0000	0.0008	0.0004	0.0000	0.0002	0.0000

- For all Q-matrix element replacement conditions, complete recovery (100%) of Q-matrix elements was obtained.
  - Q-matrix recovery was defined by rounding Q-matrix element posterior probabilities, and determining concordance rate of estimate with truth.

Real-World Application

- Fraction subtraction test (Tatsuoka, 1990).
  - 20 item math test given to 2,144 middle school students.
- All Q-matrix entries for Items 2, 10, and 19 were all converted to subjective probabilities of 0.50.

2.	$\frac{3}{4} - \frac{3}{8}$	(4, 7).
10.	$4\frac{4}{12} - 2\frac{7}{12}$	(2, 5, 7, 8).
19.	$7 - 1\frac{4}{3}$	(1, 2, 3, 5, 7).

Fraction Subtraction Skills

1. Convert a whole number to a fraction.
2. Separate a whole number from fraction.
3. Simplify before subtracting.
4. Find a common denominator.
5. Borrow from whole number part.
6. Column borrow to subtract the second numerator from the first.
7. Subtract numerators.
8. Reduce answers to simplest form.

Real-World Application Results

Q-matrix element posterior probabilities:

	Skill							
Item	1	2	3	4	5	6	7	8
2.	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00
10.	0.00	0.99	0.74	0.00	1.00	0.00	1.00	1.00
19.	1.00	0.98	1.00	0.00	1.00	0.05	1.00	0.00

- For each item, Q-matrix entry posterior probabilities indicated Q-matrix entries consistent with original definitions.
- For Item 10, one element (for Skill 3) was indicated by probabilistic method but not by de la Torre and Douglas.
  - The indicated attribute, “Simplify before subtracting,” seems reasonable for this item, $4\frac{4}{12} - 2\frac{7}{12}$.

Discussion

- Simulation studies indicate probabilistic Q-matrix specification can recover true structure when some elements of Q-matrix are not known with certainty.
  - Simulated data had item parameters indicative of a quality diagnostic test.
  - As diagnostic information decreases, prior Q-matrix probabilities are expected to provided greater influence on the results.
- Fraction subtraction results showed that the probabilistic method produced Q-matrix estimates consistent with those defined previously.

Future Directions

- Use of probabilistic Q-matrix augments traditional estimation algorithms in cognitive diagnosis by allowing for exploratory analyses.
- Further studies are needed to understand limitations of probabilistic Q-matrix method.
  - Not all elements can be missing.
  - Omitted skills are not currently considered.
  - Extend method to find total number of skills needed in a test.
- Q-matrix entries can be representative of degree of certainty a set of experts have in an item's ability to measure a given attribute.