Estimation of Diagnostic Models with Mplus

Jonathan Templin
University of Georgia
National Council on Measurement in Education
Diagnostic Modeling
Training Session
April 13, 2009

Talk Overview

- Introduction of the log-linear cognitive diagnosis model (LCDM)
- Description of Mplus
  - Capabilities Mplus for LCDM estimation
- Brief overview of syntax
  - Introduction of SAS macro
    - Write, run, parse Mplus output
- Example analysis and results

THE LOG-LINEAR COGNITIVE DIAGNOSIS MODEL

Log-linear Cognitive Diagnosis Model

- The LCDM specifies the probability of a correct response as a function of a set of attributes and a Q-matrix:

$$P(X_{ij}=1|\boldsymbol{\alpha}_i)=\frac{e^{\boldsymbol{\lambda}_j^T\mathbf{h}(\mathbf{q}_j,\boldsymbol{\alpha}_i)}}{1+e^{\boldsymbol{\lambda}_j^T\mathbf{h}(\mathbf{q}_j,\boldsymbol{\alpha}_i)}}$$

- For an item, the LCDM has ANOVA-like parameters:

$$\boldsymbol{\lambda}_j^T\mathbf{h}(\mathbf{q}_j,\boldsymbol{\alpha}_i)=\lambda_{j,0}+\sum_{u=1}^{K}\lambda_{j,1,(u)}\left(\alpha_u q_{ju}\right)+\sum_{u=1}^{K}\sum_{v>u}\lambda_{j,2,(u,v)}\left(\alpha_{iu}\alpha_{iv}q_{ju}q_{jv}\right)+\ldots$$

Intercepts

Main Effects

Two-Way Interactions

Higher Interactions

LDCM Structural Model: Attribute Association

- The LCDM assumes attributes are distributed as multivariate Bernoulli
  - Estimate the probability of each attribute pattern in a sample
  - Incorporates the marginal frequency for each attribute and attribute correlation
  - For K attributes, 2^K - 1 parameters are estimated

- Reduced forms of the parameter space are often used
  - Log-linear structural model (Henson & Templin, 2005; Xu and von Davier, 2007)

- Mplus will estimate full and log-linear structural models

MPLUS ESTIMATION

The Mplus Statistical Package

- Mplus provides a general latent variable modeling framework that allows for combinations of:
  - Continuous or categorical latent variables (i.e. IRT or DCM)
  - Continuous, categorical, count, nominal or censored data

- Mplus is commercial software that is available without any research restrictions
  - DCM estimation requires base package + combination add-on
  - $350 (student); $895 (academic); $1,095 (non academic)
  - Available at http://www.statmodel.com

DCMs Able to Be Estimated via Mplus

- Although the LCDM is the main topic of this presentation, many DCMs can be estimated with Mplus:

  - LCDM — Implemented with SAS Macro
  - DINA
  - NIDA
  - RUM
  - DINO
  - NIDO
  - C-RUM

  Freeware program (CDM) available to aid in Mplus syntax building (contact me)

  - GDM — Easy modification to Mplus syntax

Key Mplus Syntax Features

- Estimation of LCDM enacted in specification of two Mplus syntax sections:

- Model
  - Defines sum of LCDM and attribute parameters

- Model Constraint
  - NEW command specifies LCDM parameters
    - Specify item parameters (i.e., intercepts, main effects, etc…)
    - Specify log-linear structural parameters (of any order)

Model Command Syntax

- For each attribute pattern and item, value of a threshold is given
  - Threshold represents one-minus the sum of the LCDM parameters
  - Each threshold gets a label

```
MODEL:

%c#1%  ! Model for Class 1 - Profile [0,0,0]
!----{code for other items omitted}----!
    [x2$1]  (T2_1);          ! Threshold 1 for item 2
    [x20$1] (T20_1);         ! Threshold 1 for item 20

%c#2%  ! Model for Class 2 - Profile [0,0,1]
!----{code for other items omitted}----!
    [x2$1]  (T2_1);          ! Threshold 1 for item 2
    [x20$1] (T20_2);         ! Threshold 2 for item 20

!----{code for other classes omitted}----!

%c#8%  ! Model for Class 8 - Profile [1,1,1]
    [x2$1]  (T2_2);          ! Threshold 2 for item 2
    [x20$1] (T20_4);         ! Threshold 4 for item 20
```

- Code is very lengthy and tedious to construct by hand
  - Done via SAS Macro
  - Code sample shown for 3 attribute (8 class) model

Model Constraint Item Syntax

- NEW command defines LCDM item parameters
- Threshold values are then set to LCDM-specified values for item parameters
  - Based on Q-matrix

```
MODEL CONSTRAINT:           ! Used to define LCDM parameters
                            ! Mplus uses P(X=0) so multiply by -1

!-------------------------------------------------------------------------------
! Item 2: Define LCDM parameters present for item 2
! Q-matrix entry: [0,1,0]

NEW(L2_0 L2_12);            ! Define two LCDM parameters for the item
T2_1=-(L2_0);               ! Item 2 Threshold 1
T2_2=-(L2_0+L2_12);         ! Item 2 Threshold 2

! Main effect order constraints
L2_12>0;
!-------------------------------------------------------------------------------
! Item 20: Define LCDM parameters present for item 20
! Q-matrix entry: [1,0,1]
NEW(L20_0 L20_11 L20_13 L20_213); ! Define four LCDM parameters for the item
T20_1=-(L20_0);                     ! Item 20 Threshold 1
T20_2=-(L20_0+L20_13);              ! Item 20 Threshold 2
T20_3=-(L20_0+L20_11);              ! Item 20 Threshold 3
T20_4=-(L20_0+L20_11+L20_13+L20_213); ! Item 20 Threshold 4

! Main effect order constraints
L20_11>0;
L20_13>0;

! Two-way interaction order constraints
L20_213>-L20_11;
L20_213>-L20_13;
!-------------------------------------------------------------------------------
```

Model Constraint Structural Syntax

- The structural model is optional.
  - Default: saturated model

```
MODEL CONSTRAINT:

!STRUCTURAL MODEL:
NEW(g_0 g_11 g_12 g_13 g_212 g_213 g_223 g_3123);
m1=-(g_11+g_12+g_13+g_212+g_213+g_223+g_3123);
m2=g_13-(g_11+g_12+g_13+g_212+g_213+g_223+g_3123);
m3=g_12-(g_11+g_12+g_13+g_212+g_213+g_223+g_3123);
m4=g_12+g_13+g_223-(g_11+g_12+g_13+g_212+g_213+g_223+g_3123);
m5=g_11-(g_11+g_12+g_13+g_212+g_213+g_223+g_3123);
m6=g_11+g_13+g_213-(g_11+g_12+g_13+g_212+g_213+g_223+g_3123);
m7=g_11+g_12+g_212-(g_11+g_12+g_13+g_212+g_213+g_223+g_3123);
g_0=-(g_11+g_12+g_13+g_212+g_213+g_223+g_3123);
```

- Log-linear syntax shown
  - Saturated model

- Lower-order structural achieved by deleting parameters from model

Data Size and Estimation Timing

- Mplus uses marginal maximum likelihood for estimation
  - Accelerated E-M algorithm
    - Combines EM, Quasi-Newton, and Fisher Scoring (for continuous LV)
- Size of model limited by memory capacity of machine
  - Mplus 64-bit versions have virtually limitless capacity
- Estimation time can be an issue
  - More than 6 attributes can take hours
    - Can be much faster for simpler DCMs (constraints slow estimation)

SAS MACRO

Aiding Mplus Estimation: SAS Macro

- Mplus syntax is tedious to construct
  - Human errors can occur very easily
- We developed a SAS macro to aid in Mplus estimation
- The macro:
  - Creates Mplus syntax
  - Calls Mplus to estimate model parameters
  - Parses Mplus output into SAS data sets
- Available to anyone without restrictions (contact me)

SAS Macro Syntax

```
* Location of original data files - CHANGE BOTH;
* Permanent SAS library;                LIBNAME folder "C:\DCM_Mplus";
* Path to import/export files from;     %LET filesave=  C:\DCM_Mplus;

* Name for files to be created;             %LET filename =     DCM_Mplus;
* Name of Q matrix file;                    %LET Qname=         Qmatrix;
* Name of dataset with extension;           %LET dataname=      DCM_data.dat;
* Name of person ID variable;               %LET IDname=        ID;
* List of variables in dataset;             %LET varlist=       ID x1-x28;
* List of items to be modeled;              %LET uselist=       x1-x28;

* Stem of item names;                       %LET itemstem=      x;
* First item number;                        %LET firstitem=     1;
* Last item number;                         %LET lastitem=      28;
* Total number of items;                    %LET numitem=       28;
* Variable for order of item model;         %LET ordervar=      itemorder;
* Max order of interaction in item model;   %LET maxitemorder=  2;

* Attribute stem in Q matrix;               %LET attstem=       attribute;
* First attribute number;                   %LET firstatt=      1;
* Last attribute number;                    %LET lastatt=       3;
* Total number of attributes;               %LET numatt=        3;
* Number of total classes (2^A);            %LET numclass=      8;
* Use structural model(0=n,1=y);            %LET structon=      1;
* Order of interaction in structural model; %LET structorder=   2;
*************************************************************************;
```

DCM ESTIMATION OUTPUT

LCDM Item Parameters

In Mplus

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
New/Additional Parameters				
L1_0	-0.913	0.053	-17.141	0.000
L1_11	1.872	0.071	26.297	0.000
L2_0	-1.006	0.069	-14.636	0.000
L2_12	2.045	0.082	24.960	0.000
L3_0	-0.961	0.062	-15.460	0.000
L3_13	2.014	0.079	25.619	0.000
L4_0	-2.437	0.157	-15.522	0.000
L4_11	2.047	0.205	10.007	0.000
L4_12	1.858	0.212	8.768	0.000
L4_212	1.034	0.338	3.057	0.002

In SAS

.New

	Item	Order	Atts	parameter	estimate	se	z	pvalue
1	1	0	0	L1_0	-0.913	0.053	-17.141	0
2	1	1	1	L1_11	1.872	0.071	26.297	0
3	2	0	0	L2_0	-1.006	0.069	-14.636	0
4	2	1	2	L2_12	2.045	0.082	24.96	0
5	3	0	0	L3_0	-0.961	0.062	-15.46	0
6	3	1	3	L3_13	2.014	0.079	25.619	0
7	4	0	0	L4_0	-2.437	0.157	-15.522	0
8	4	1	1	L4_11	2.047	0.205	10.007	0
9	4	1	2	L4_12	1.858	0.212	8.768	0
10	4	2	12	L4_212	1.034	0.338	3.057	0.002

LCDM Structural Parameters

In Mplus

Saturated Model

Pattern Proportions

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODEL

Latent Classes		
1	2645.23894	0.26452
2	799.04351	0.07990
3	748.57761	0.07486
4	836.23876	0.08362
5	785.83601	0.07858
6	875.16371	0.08752
7	863.55607	0.08636
8	2446.34540	0.24463

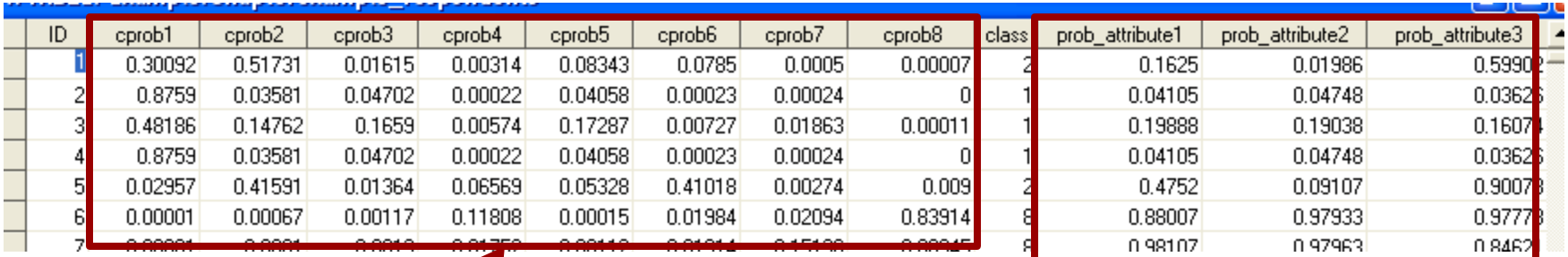
Log-Linear Model

ODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
M_11	-1.109	0.114	-9.701	0.000
M_12	-1.145	0.123	-9.299	0.000
M_13	-1.091	0.120	-9.075	0.000
M_212	1.142	0.147	7.771	0.000
M_213	1.104	0.143	7.701	0.000
M_223	1.085	0.140	7.748	0.000

LCDM Examinee Parameters

In SAS



ID	cprob1	cprob2	cprob3	cprob4	cprob5	cprob6	cprob7	cprob8	class	prob_attribute1	prob_attribute2	prob_attribute3
1	0.30092	0.51731	0.01615	0.00314	0.08343	0.0785	0.0005	0.00007	2	0.1625	0.01986	0.59902
2	0.8759	0.03581	0.04702	0.00022	0.04058	0.00023	0.00024	0	1	0.04105	0.04748	0.03626
3	0.48186	0.14762	0.1659	0.00574	0.17287	0.00727	0.01863	0.00011	1	0.19888	0.19038	0.16074
4	0.8759	0.03581	0.04702	0.00022	0.04058	0.00023	0.00024	0	1	0.04105	0.04748	0.03626
5	0.02957	0.41591	0.01364	0.06569	0.05328	0.41018	0.00274	0.009	2	0.4752	0.09107	0.90078
6	0.00001	0.00067	0.00117	0.11808	0.00015	0.01984	0.02094	0.83914	8	0.88007	0.97933	0.97778
7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	8	0.98107	0.97963	0.84626

Attribute ***Pattern*** Probability Estimates (from Mplus)

Marginal Attribute Probability Estimates (computed in SAS)

ADDITIONAL FEATURES

Additional Mplus Features for DCMs

- Mplus also features a set of other features that will aid in estimation of DCMs across virtually any psychometric context:
  - Monte Carlo simulation of examinees
  - Data types: Continuous, categorical, count, nominal, censored
  - Missing data: by design, at random
  - Sampling weights
  - Multiple group estimation
  - Stratification variables
  - Modeling covariate relations to attributes and structural models

Hybrid-Type Models

- Mplus syntax can easily be modified to include continuous latent variables with DCMs
  - Hybrid models
  - General diagnostic model
  - Mixture Rasch models
- Unlike the DCM syntax which is tedious, hybrid models take very few lines of code
  - Displayed for Bifactor DCM (Henson, Templin, & Willse, 2009)

```
MODEL:
%OVERALL%
theta by x1-x28@1 (1);
theta;
```

FUTURE IMPLEMENTATIONS

Future Directions of Mplus Macros

- Mplus macro syntax will be adapted to better incorporate all features of the Mplus package into an easy interface

- Additionally, the ability to conduct simulation studies will be included

- For any questions or requests for software, please contact me at jtemplin@uga.edu