Mplus Short Courses
Topic 9

Bayesian Analysis Using Mplus

Bengt Muthén
Linda K. Muthén
Tihomir Asparouhov

www.statmodel.com
06/01/2011

1

Table Of Contents

2

Table Of Contents (Continued)

3

Mplus Background

- Inefficient dissemination of statistical methods:
  - Many good methods contributions from biostatistics, psychometrics, etc are underutilized in practice
- Fragmented presentation of methods:
  - Technical descriptions in many different journals
  - Many different pieces of limited software
- Mplus: Integration of methods in one framework
  - Easy to use: Simple, non-technical language, graphics
  - Powerful: General modeling capabilities

4

Mplus Background

- SBIR from NIH (NIAAA)
- Mplus versions
  - V1: November 1998
  - V2: February 2001
  - V3: March 2004
  - V4: February 2006
  - V5: November 2007
  - V5.21: May 2009
  - V6: April 2010
  - V6.11: April 2011
- Mplus team: Linda & Bengt Muthén, Thuy Nguyen, Tihomir Asparouhov, Michelle Conn, Jean Maninger

5

Statistical Analysis With Latent Variables A General Modeling Framework

Statistical Concepts Captured By Latent Variables

Continuous Latent Variables

- Measurement errors
- Factors
- Random effects
- Frailties, liabilities
- Variance components
- Missing data

Categorical Latent Variables

- Latent classes
- Clusters
- Finite mixtures
- Missing data

6

Statistical Analysis With Latent Variables
A General Modeling Framework (Continued)

Models That Use Latent Variables

Continuous Latent Variables

- Factor analysis models
- Structural equation models
- Growth curve models
- Multilevel models

Categorical Latent Variables

- Latent class models
- Mixture models
- Discrete-time survival models
- Missing data models

Mplus integrates the statistical concepts captured by latent variables into a general modeling framework that includes not only all of the models listed above but also combinations and extensions of these models.

7

Mplus

Several programs in one

- Exploratory factor analysis
- Structural equation modeling
- Item response theory analysis
- Growth modeling
- Mixture modeling (latent class analysis)
- Longitudinal mixture modeling (Hidden Markov, LTA, LCGA, GMM)
- Survival analysis (continuous- and discrete-time)
- Multilevel analysis
- Complex survey data analysis
- Bayesian analysis
- Monte Carlo simulation

Fully integrated in the general latent variable framework

8

Bayesian Analysis In Mplus

Mplus conceptualization:

- Mplus was envisioned 15 years ago as both a frequentist and a Bayesian program
- Bayesian analysis firmly established and its use growing in mainstream statistics
- Much less use of Bayes outside statistics
- Bayesian analysis not sufficiently accessible in other programs
- Bayes provides a broader platform for further Mplus development

9

Bayesian Analysis

Why do we have to learn about Bayes?

- More can be learned about parameter estimates and model fit
- Better small-sample performance, large-sample theory not needed
- Priors can better reflect substantive hypotheses
- Analyses can be made less computationally demanding
- New types of models can be analyzed

10

Writings On The Bayes Implementation In Mplus

- Asparouhov & Muthén (2010). Bayesian analysis using Mplus: Technical implementation. Technical Report. Version 3.
- Asparouhov & Muthén (2010). Bayesian analysis of latent variable models using Mplus. Technical Report. Version 4.
- Asparouhov & Muthén (2010). Multiple imputation with Mplus. Technical Report. Version 2.
- Asparouhov & Muthén (2010). Plausible values for latent variable using Mplus. Technical Report.
- Muthén (2010). Bayesian analysis in Mplus: A brief introduction. Technical Report. Version 3.
- Muthén & Asparouhov (2010). Bayesian SEM: A more flexible representation of substantive theory.

Posted under Papers, Bayesian Analysis

11

Overview Of Bayesian Features In Mplus

12

Bayesian Estimation In Mplus

- Single-level, multilevel, and mixture models
- Continuous and categorical outcomes (probit link)
- Default non-informative priors or user-specified informative priors (MODEL PRIORS)
- Multiple chains using parallel processing (CHAIN)
- Convergence assessment using Gelman-Rubin potential scale reduction factors
- Posterior parameter distributions with means, medians, modes, and credibility intervals (POINT)
- Posterior parameter trace plots
- Autocorrelation plots
- Posterior predictive checking plots

Multiple Imputation (DATA IMPUTATION)

- Carried out using Bayesian estimation to create several data sets where missing values have been imputed
- The multiple imputation data sets can be used for subsequent model estimation using ML or WLSMV
- The imputed data sets can be saved for subsequent analysis or analysis can be carried out in the same run
- Imputation can be done based on an unrestricted H1 model using three different algorithms including sequential regressions
- Imputation can also be done based on an H0 model specified in the MODEL command

Multiple Imputation (Continued)

- The set of variables used in the imputation of the data do not need to be the same as the set of variables used in the analysis
- Single-level and multilevel data imputation are available
- Multiple imputation data can be read using TYPE=IMPUTATION in the DATA command

15

Plausible Values (PLAUSIBLE)

- Plausible values are multiple imputations for missing values corresponding to a latent variable
- Plausible values used in IRT contexts such as the ETS NAEP, The Nation's Report Card (Mislevy et al., 1992)
- Available for both continuous and categorical latent variables (factors, random effects, latent classes)
- More informative than only an estimated factor score and its standard error or a class probability
- Plausible values are more accurate than factor scores
- Plausible values are given for each observation together with a summary over the imputed data sets for each observation and latent variable
- Multiple imputation and plausible values examples are given in the User's Guide, Chapter 11

16

Bayesian Analysis Using Mplus: An Ongoing Project

Features that are not yet implemented include:

- EFA and ESEM
- Logit link
- Censored, count, and nominal variables
- XWITH
- Weights
- c ON x in mixtures
- Mixture models with more than one categorical latent variable
- Two-level mixtures
- MODEL INDIRECT
- MODEL CONSTRAINT except for NEW parameters
- MODEL TEST

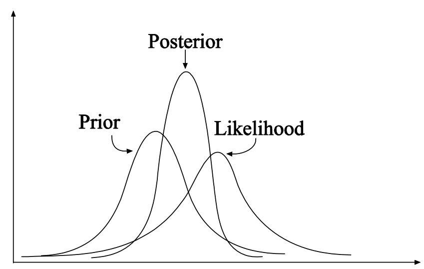
17

Bayesian Estimation

18

Prior, Likelihood, And Posterior

- Frequentist view: Parameters are fixed. ML estimates have an asymptotically-normal distribution
- Bayesian view: Parameters are variables that have a prior distribution. Estimates have a possibly non-normal posterior distribution. Does not depend on large-sample theory
  - Diffuse (non-informative) priors vs informative priors



19

Bayes Theorem

- Probabilities of events A and B:

  $$P(A,B) = P(A|B)P(B) = P(B|A)P(A)$$

  Bayes theorem: $$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

- Applied to modeling:

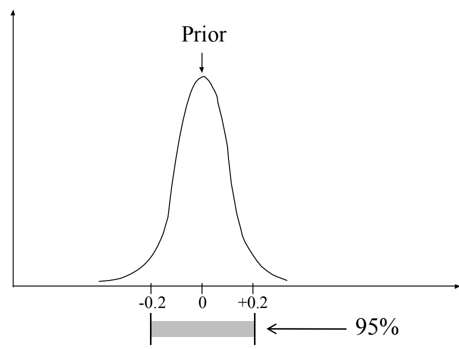
  $$[parameters|data] = \frac{[data|parameters][parameters]}{[data]} = \frac{likelihood \times prior}{[data]}$$

- Posterior $\propto likelihood \times prior$

20

Where Do Parameter Priors Come From?

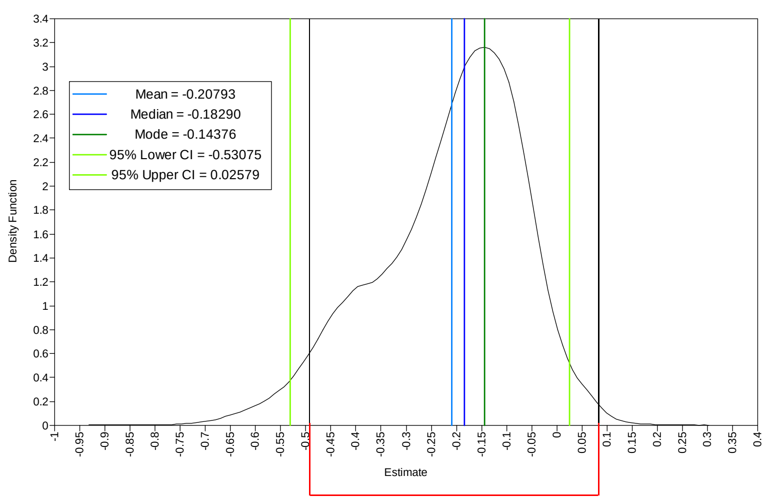
- Previous studies
- Hypotheses based on substantive theory
  - Example: Zero cross-loadings in CFA



$\lambda \sim N(0,0.01)$

21

Non-Normal Posterior Parameter Distribution



95% symmetric CI around the mean

22

Gibbs Sampler For A Bivariate Normal With Unknown Means And Uniform Prior

Gelman et al. (2004), p. 288): "Consider a single observation (y_1, y_2) from a bivariate normally distributed population with unknown mean $\theta = (\theta_1, \theta_2)$, and known covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. With a uniform prior distribution on θ, the posterior distribution is

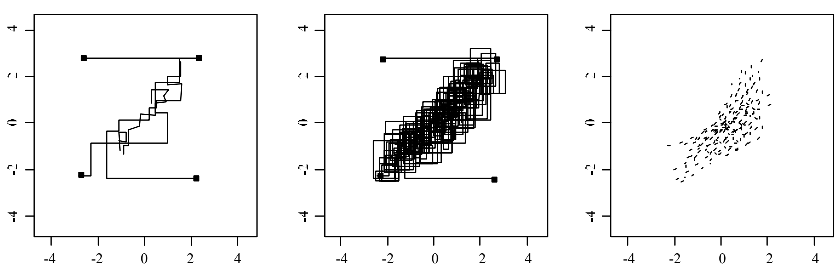
$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \bigg| y \sim N\left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

Consider the conditional posterior distributions

$$\theta_1 | \theta_2, y \sim N(y_1 + \rho(\theta_2 - y_2), 1 - \rho^2)$$
$$\theta_2 | \theta_1, y \sim N(y_2 + \rho(\theta_1 - y_1), 1 - \rho^2)$$

23

Gibbs Sampler For A Bivariate Normal With Unknown Means And Uniform Prior



Sampling from the conditional posterior distributions (y_1=0, y_2=0)

$$\theta_1 | \theta_2, y \sim N(y_1 + \rho(\theta_2 - y_2), 1 - \rho^2)$$
$$\theta_2 | \theta_1, y \sim N(y_2 + \rho(\theta_1 - y_1), 1 - \rho^2)$$

Source: Gelman et al. (2004)

24

Bayesian Estimation Using the Markov Chain Monte Carlo (MCMC) Algorithm

- θ_i : vector of parameters, latent variables, and missing observations at iteration *i*
- θ_i is divided into S sets:
  $\theta_i = (\theta_{1i}, \ldots, \theta_{Si})$
- Update θ using Gibbs sampling over *i* = 1, 2, …, n iterations:
  $\theta_{1i} \mid \theta_{2i-1}, \ldots, \theta_{Si-1}$, data, priors
  $\theta_{2i} \mid \theta_{1i}, \theta_{3i-1}, \ldots, \theta_{Si-1}$, data, priors
  …
  $\theta_{Si} \mid \theta_{1i}, \ldots, \theta_{S-1i-1}$, data, priors

Asparouhov & Muthén (2010a). Bayesian analysis using Mplus. Technical implementation.Technical appendix.

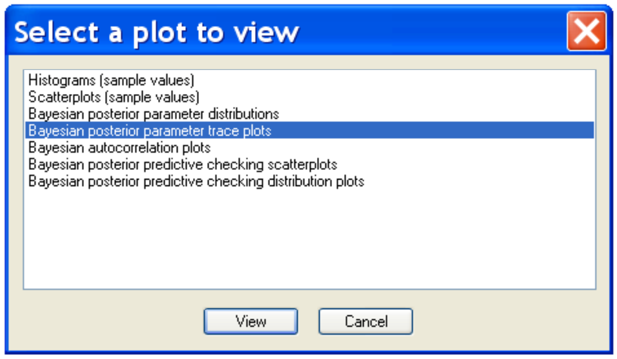
25

MCMC Iteration Issues

- Trace plot: Graph of the value of a parameter at different iterations
- Burnin phase: Discarding early iterations. Mplus discards first half
- Posterior distribution: Mplus uses the last half as a sample representing the posterior distribution
- Autocorrelation plot: Correlation between consecutive iterations for a parameter. Low correlation desired
- Mixing: The MCMC chain should visit the full range of parameter values, i.e. sample from all areas of the posterior density
- Convergence: Stationary process.

26

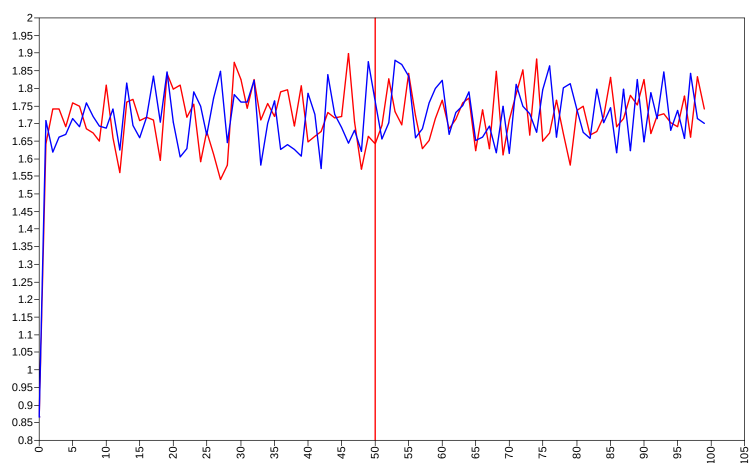
Bayes Graphs




```
PLOT:      TYPE = PLOT2;
```

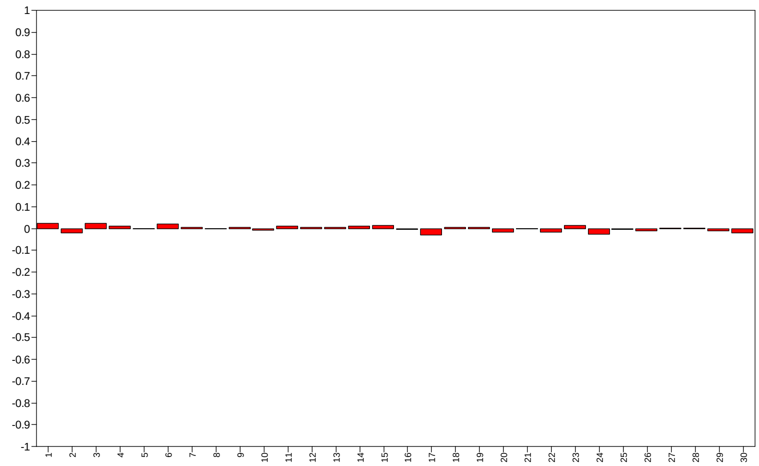
27

Trace Plot 1 (Two Chains Shown)



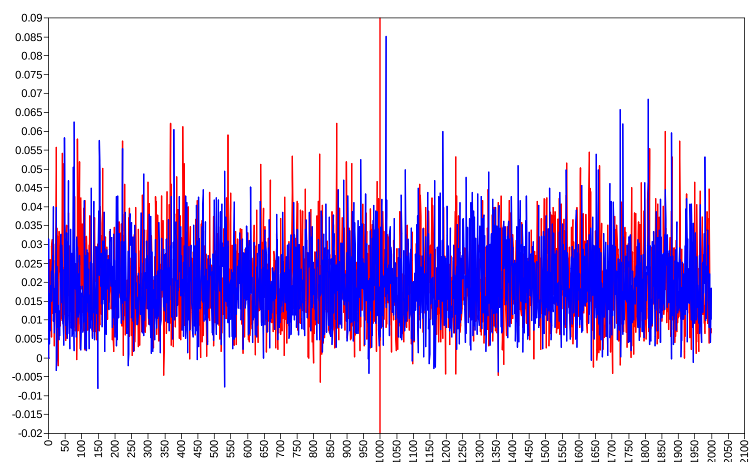
28

Auto Correlation Plot 1



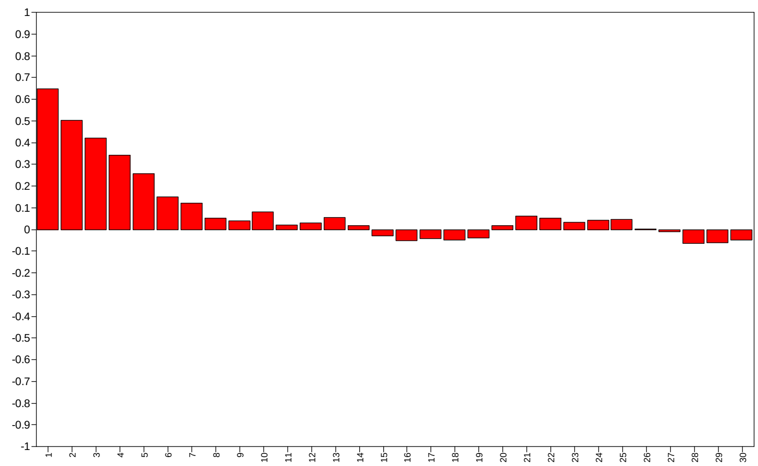
29

Trace Plot 2



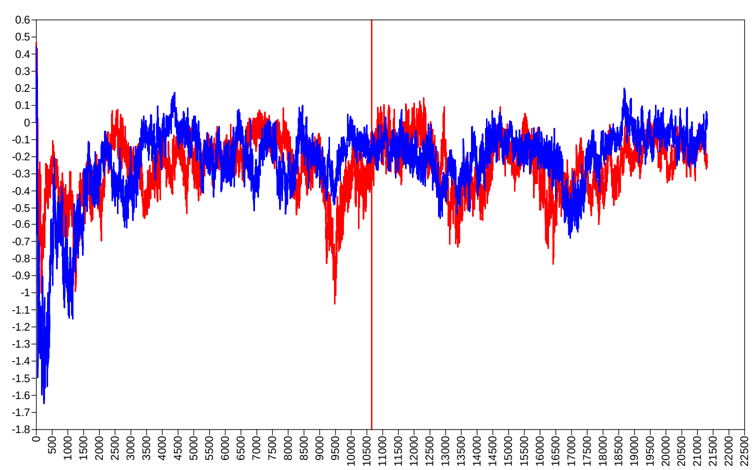
30

Auto Correlation Plot 2



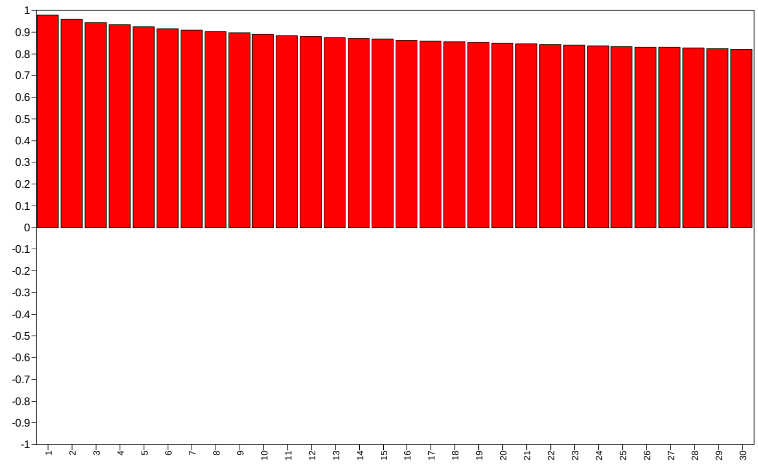
31

Trace Plot 3



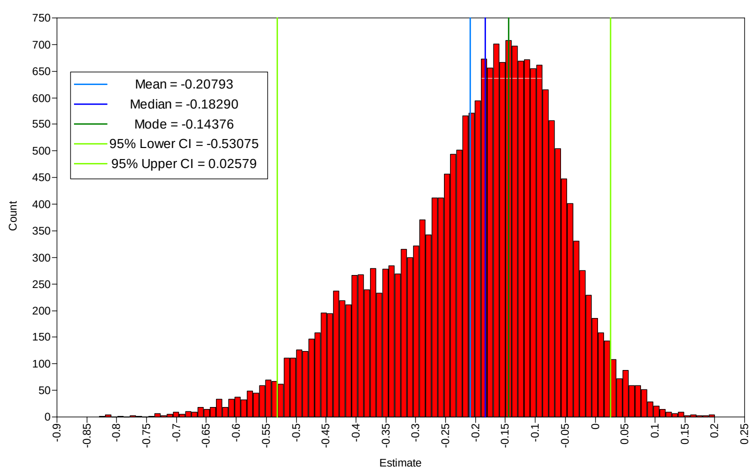
32

Auto Correlation Plot 3



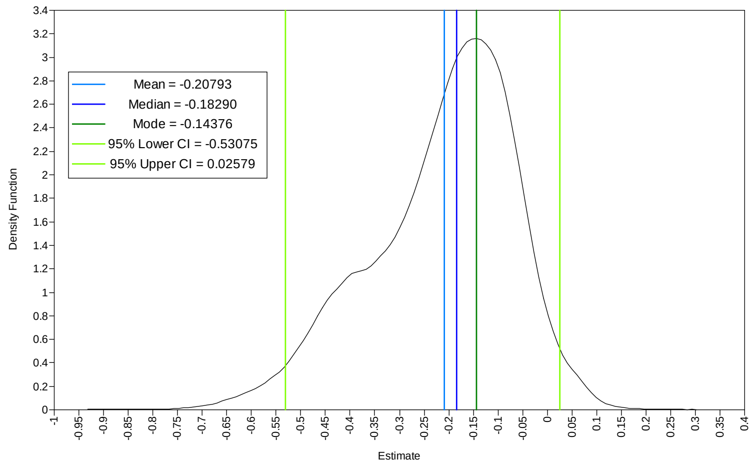
33

Posterior Using Histogram Option



34

Posterior Using Kernel Option



35

Convergence: Potential Scale Reduction Factor (PSR; TECH8)

- Several MCMC iterations carried out in parallel, independent chains. PSR considers n iterations in m chains, where θ_{ij} is the value of θ in iteration i of chain j:

$$\bar{\theta}_{.j} = \frac{1}{n}\sum_{i=1}^{n}\theta_{ij} \qquad \bar{\theta}_{..} = \frac{1}{m}\sum_{j=1}^{m}\bar{\theta}_{ij}$$

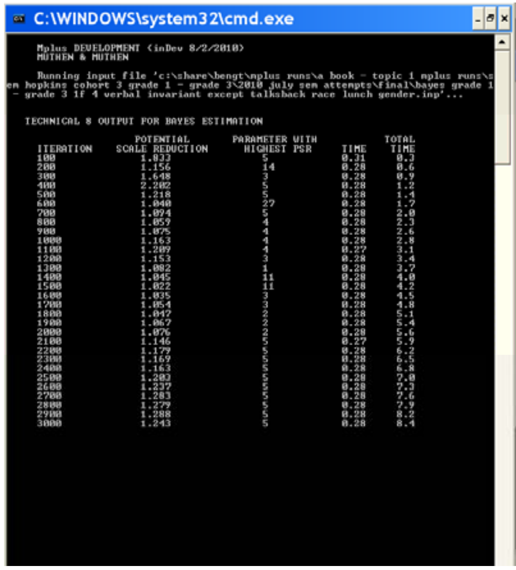
$$B = \frac{1}{m-1}\sum_{j=1}^{m}(\bar{\theta}_{.j} - \bar{\theta}_{..})^2 \qquad W = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{i=1}^{n}(\theta_{ij} - \bar{\theta}_{.j})^2$$

$$PSR = \sqrt{\frac{W+B}{W}}$$

- Convergence if PSR is not much larger than 1, e.g. less than 1.05 or 1.1.

36

TECH8 For Bayes



```
C:\WINDOWS\system32\cmd.exe

    Mplus DEVELOPMENT (inDev 8/2/2010)
    MUTHEN & MUTHEN

    Running input file 'c:\share\bengt\mplus runs\a book - topic 1 mplus runs\s
em hopkins cohort 3 grade 1 - grade 3\2010 july sem attempts\final\bayes grade 1
 - grade 3 1f 4 verbal invariant except talksback race lunch gender.inp'...

  TECHNICAL 8 OUTPUT FOR BAYES ESTIMATION

                 POTENTIAL        PARAMETER WITH              TOTAL
    ITERATION    SCALE REDUCTION    HIGHEST PSR     TIME      TIME
    100              1.833             5            0.31      0.3
    200              1.156             14           0.28      0.6
    300              1.648             3            0.28      0.9
    400              2.202             5            0.28      1.2
    500              1.218             5            0.28      1.4
    600              1.040             27           0.28      1.7
    700              1.094             5            0.28      2.0
    800              1.059             4            0.28      2.3
    900              1.075             4            0.28      2.6
    1000             1.163             4            0.28      2.8
    1100             1.209             4            0.27      3.1
    1200             1.153             3            0.28      3.4
    1300             1.082             1            0.28      3.7
    1400             1.045             11           0.28      4.0
    1500             1.022             11           0.28      4.2
    1600             1.035             3            0.28      4.5
    1700             1.054             3            0.28      4.8
    1800             1.047             2            0.28      5.1
    1900             1.067             2            0.28      5.4
    2000             1.076             2            0.28      5.6
    2100             1.146             5            0.27      5.9
    2200             1.179             5            0.28      6.2
    2300             1.169             5            0.28      6.5
    2400             1.163             5            0.28      6.8
    2500             1.203             5            0.28      7.0
    2600             1.237             5            0.28      7.3
    2700             1.283             5            0.28      7.6
    2800             1.279             5            0.28      7.9
    2900             1.288             5            0.28      8.2
    3000             1.243             5            0.28      8.4
```


Options Related To Bayes Estimation And Multiple Imputation

See User's Guide pages 559-563:

- POINT (mean, median, mode; default is median)
- CHAINS (default is 2)
- BSEED
- STVALUES (= ML, PERTURBED, UNPERTURBED)
- MEDIATOR (observed, latent; default is latent)
- ALGORITHM (GIBBS, MH; default is GIBBS)
- BCONVERGENCE (related to PSR)
- BITERATIONS (to go beyond 50K iterations)
- FBITERATIONS (fixed number of iterations)
- THIN (every kth iteration recorded; default is 1)
- DISTRIBUTION (how many iterations used for MODE)

Mplus Default Priors

- Intercepts, regression slopes, loadings: N(0, infinity), unless these parameters are in a probit regression in which case N(0, 5) is used
- Variances: IG(0, -1)
- Covariance matrices: IW(0, -p-1), unless the elements include parameters from a probit regression in which case IW(I, p+1) is used
- Imputation with an unrestricted model: IW(I, p+1)
- Thresholds: N(0, infinity)
- Class proportions: Dirichlet prior D(10, 10, ..., 10)

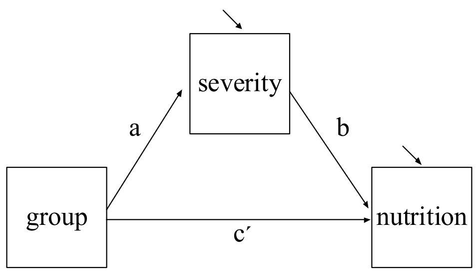
39

Path Analysis With Indirect Effects

40

Mediation Modeling With Diffuse Priors: ATLAS Example

Source: MacKinnon et al. (2004), MBR. n = 861



- Intervention aimed at increasing perceived severity of using steroids among athletes. Perceived severity of using steroids is in turn hypothesized to increase good nutrition behaviors
- Indirect effect: a x b

41

Input For Bayesian Analysis Of ATLAS Example

```
TITLE:          ATLAS
DATA:           FILE = mbr2004atlast.txt;
VARIABLE:       NAMES = obs group severity nutrit;
                USEV = group – nutrit;
ANALYSIS:       ESTIMATOR = BAYES;
                PROCESS = 2;
MODEL:          severity ON group (a);
                nutrit ON severity (b)
                group;
MODEL CONSTRAINT:
                NEW (indirect);
                indirect = a*b;
OUTPUT:         TECH1 TECH8 STANDARDIZED;
PLOT:           TYPE = PLOT2;
```

42

Output For Bayesian Analysis Of ATLAS Example

Parameter	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%
severity ON					
group	0.282	0.106	0.010	0.095	0.486
nutrit ON					
severity	0.067	0.031	0.000	0.015	0.125
group	-0.011	0.089	0.440	-0.180	0.155
Intercepts					
severity	5.641	0.072	0.000	5.513	5.779
nutrit	3.698	0.191	0.000	3.309	4.018

43

Output For Bayesian Analysis Of ATLAS Example (Continued)

Parameter	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%
Residual variances					
severity	1.722	0.072	0.000	1.614	1.868
group	1.331	0.070	0.000	1.198	1.468
New/Additional parameters					
indirect	0.016	0.013	0.010	0.002	0.052

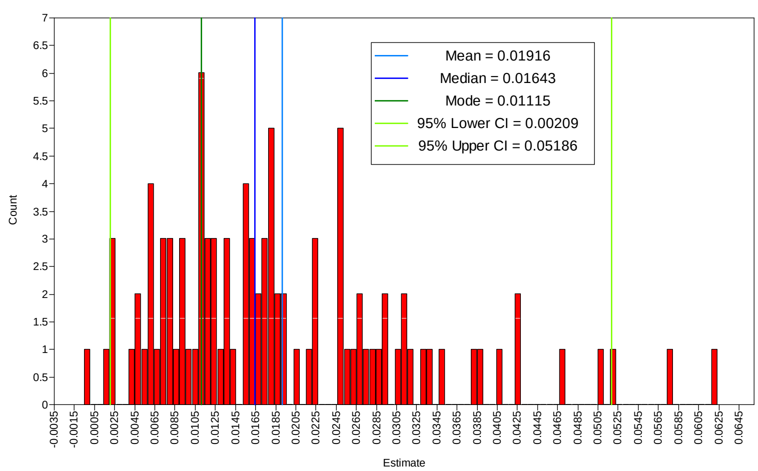
44

Output For Bayesian Analysis Of ATLAS Example (Continued)

```
TECHNICAL 8 OUTPUT FOR BAYES ESTIMATION

CHAIN        BSEED
1            0
2            285380

                  POTENTIAL        PARAMETER WITH
ITERATION      SCALE REDUCTION      HIGHEST PSR
100                 1.037                2
```

45

Posterior Distribution For Indirect Effect: 100 Iterations

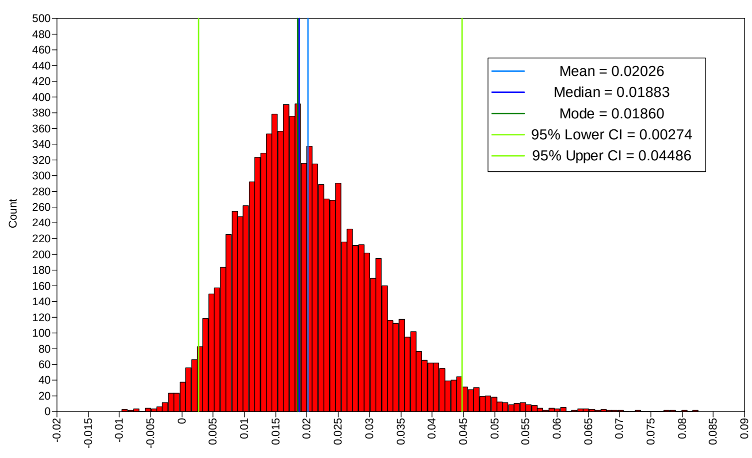


46

Mplus Input For Bayesian Analysis Of ATLAS Example (Continued)

```
ANALYSIS:    ESTIMATOR = BAYES;
             PROCESS = 2;
             FBITER = 10000;
```


Bayesian Posterior Distribution For The Indirect Effect: 10000 Iterations



Bayesian Posterior Distribution For The Indirect Effect

- Bayesian analysis: There is a mediated effect of the intervention
  - The 95% Bayesian credibility interval does not include zero
- ML analysis: There is not a mediated effect of the intervention
  - ML-estimated indirect effect is not significantly different from zero and the symmetric confidence interval includes zero
  - Bootstrap SEs and CIs can be used with ML

49

Mediation Modeling With Informative Priors: Firefighter Example. *N* = 354

- Source: Yuan & MacKinnon (2009). Bayesian Mediation Analysis. Psychological Methods, 14, 301-322.
  - Nice description of Bayesian analysis
- Informative priors based on previous studies: a~N (0.35, 0.04), b~N (0.1, 0.01)
- 95% credibility interval for indirect effect shrunken by 16%
- WinBUGS code in Yuan & MacKinnon (2009).
  Mplus code on next slides using MODEL PRIORS. Same results

50

Mplus Input For Bayesian Analysis With Priors: Firefighters

```
TITLE:      Yuan and MacKinnon firefighters mediation using
            Bayesian analysis
            Elliot DL., Goldberg L., Kuehl KS, et al. The PHLAME
            Study: process and outcomes of 2 models of behavior
            change. J Occup Environ Med. 2007; 49(2): 204-213.
DATA:       FILE = fire.dat;
VARIABLE:   NAMES = y m x;
MODEL:      m ON x (a);
            y ON m (b)
            x;
ANALYSIS:   ESTIMATOR = BAYES;
            PROCESS = 2;
            FBITER = 10000;
```

51

Mplus Input For Bayesian Analysis With Priors: Firefighters (Continued)

```
MODEL PRIORS:
            a ~ N(0.35, 0.04);
            b ~ N(0.1, 0.01);
MODEL CONSTRAINT:
            NEW (indirect);
            indirect = a*b;
OUTPUT:     TECH1 TECH8;
PLOT:       TYPE = PLOT2;
```

52

Model Fit

53

Posterior Predictive Checking

Gelman et al. (1996), Scheines et al. (1999)

- A posterior predictive p-value (PPP) of model fit can be obtained via a fit statistic f based on the usual chi-square test of H_0 against H_1. Low PPP indicates poor fit
- Let $f(Y, X, \theta_i)$ be computed for the data Y, X using the parameter values at iteration i
- At iteration i, generate a new data set Y_i^* (synthetic/replicated data) of the same size as the original data using the parameter values at iteration i and compute $f(Y_i^*, X, \theta_i)$ for these replicated data

54

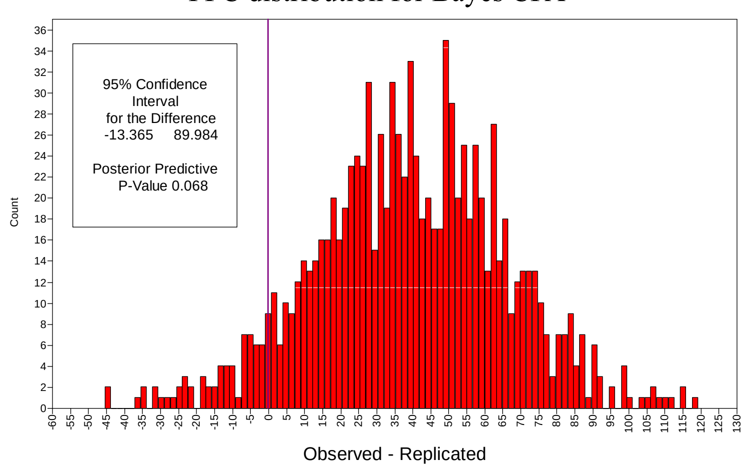
Posterior Predictive Checking (Continued)

- PPP is approximated by the proportion of iterations where $f(Y, X, \theta_i) < f(Y_i^*, X, \theta_i)$
- Mplus computes PPP using every 10th iteration among the iterations used to describe the posterior distribution of parameters
- A 95% confidence interval is produced for the difference in chi-square for the real and replicated data; negative lower limit is good

55

Posterior Predictive Checking For A Poorly Fitting Model

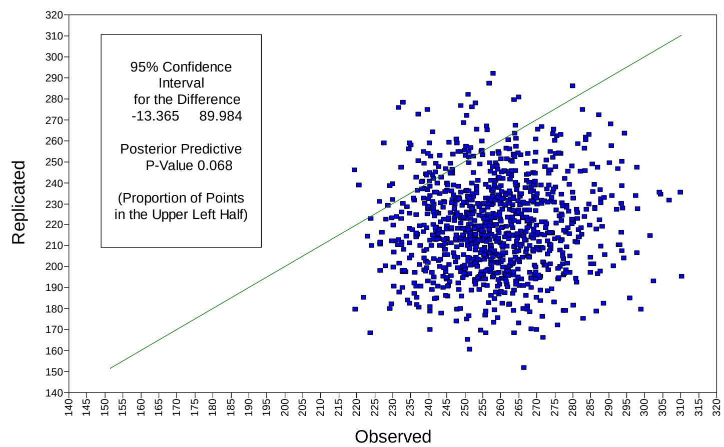
PPC distribution for Bayes CFA



56

Posterior Predictive Checking For A Poorly Fitting Model

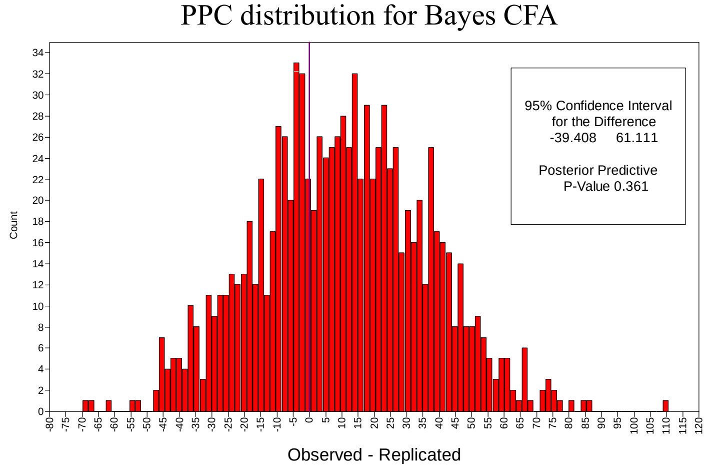
PPC scatterplot for Bayes CFA



57

Posterior Predictive Checking For A Well Fitting Model

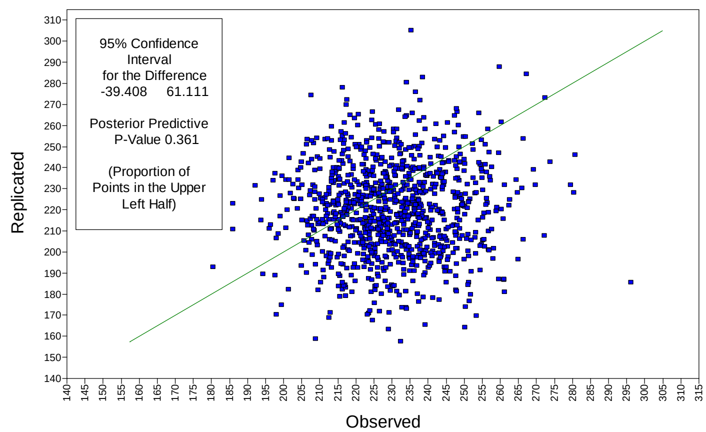
PPC distribution for Bayes CFA



58

Posterior Predictive Checking For A Well Fitting Model

PPC scatterplot for Bayes CFA



59

Deviance Information Criterion (DIC)

- DIC is a Bayesian generalization of the ML AIC and BIC (low value is good)
- DIC uses a number of parameters the effective number of parameters referred to as p_D

60

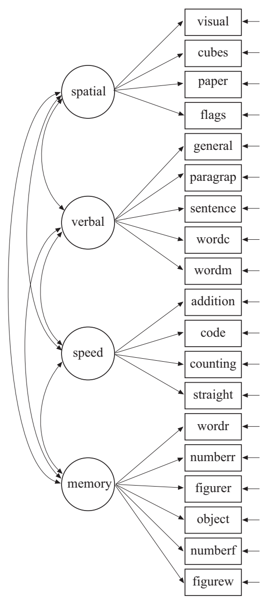
Factor Analysis

A Factor Analysis Example: Holzinger-Swineford Data

- 19 tests hypothesized to measure four mental abilities: Spatial, verbal, speed, and memory
- n=145 7th and 8th grade students from Grant-White elementary school
- n=156 7th and 8th grade students from the Pasteur elementary school

Source: Muthén & Asparouhov (2010). Bayesian SEM: A more flexible representation of substantive theory.

–The "BSEM" paper



CFA Factor Loading Pattern

	Spatial	Verbal	Speed	Memory
visual	x	0	0	0
cubes	x	0	0	0
paper	x	0	0	0
flags	x	0	0	0
general	0	x	0	0
paragrap	0	x	0	0
sentence	0	x	0	0
wordc	0	x	0	0
wordm	0	x	0	0
addition	0	0	x	0
code	0	0	x	0
counting	0	0	x	0
straight	0	0	x	0
wordr	0	0	0	x
numberr	0	0	0	x
figurer	0	0	0	x
object	0	0	0	x
numberf	0	0	0	x
figurew	0	0	0	x

63

ML CFA Testing Results For Holzinger-Swineford Data For Grant-White (n =145) And Pasteur (n=156)

Model	χ^2	df	P-value	RMSEA	CFI
Grant-White					
CFA	216	146	0.000	0.057	0.930
EFA	110	101	0.248	0.025	0.991
Pasteur					
CFA	261	146	0.000	0.071	0.882
EFA	128	101	0.036	0.041	0.972

64

	Grant-White Factor loading pattern for EFA				Pasteur Factor loading pattern for EFA			
	Spatial	Verbal	Speed	Memory	Spatial	Verbal	Speed	Memory
visual	**0.628***	0.065	0.091	0.085	**0.580***	0.307*	-0.001	0.053
cubes	**0.485***	0.050	0.007	-0.003	**0.521***	0.027	-0.078	-0.059
paper	**0.406***	0.107	0.084	0.083	**0.484***	0.101	-0.016	-0.229*
flags	**0.579***	0.160	0.013	0.026	**0.687***	-0.051	0.067	0.101
general	0.042	**0.752***	0.126	-0.051	-0.043	**0.838***	0.042	-0.118
paragrap	0.021	**0.804***	-0.056	0.098	0.026	**0.800***	-0.006	0.069
sentence	-0.039	**0.844***	0.085	-0.057	-0.045	**0.911***	-0.054	-0.029
wordc	0.094	**0.556***	0.197*	0.019	0.098	**0.695***	0.008	0.083
wordm	0.004	**0.852***	-0.074	0.069	0.143*	**0.793***	0.029	-0.023
addition	-0.302*	0.029	**0.824***	0.078	-0.247*	0.067	**0.664***	0.026
code	0.012	0.050	**0.479***	0.279*	0.004	0.262*	**0.552***	0.082
counting	0.045	-0.159	**0.826***	-0.014	0.073	-0.034	**0.656***	-0.166
straight	0.346*	0.043	**0.570***	-0.055	0.266*	-0.034	**0.526***	-0.056
wordr	-0.024	0.117	-0.020	**0.523***	-0.005	0.020	-0.039	**0.726***
numberr	0.069	0.021	-0.026	**0.515***	-0.026	-0.057	-0.057	**0.604***
figurer	0.354*	-0.033	-0.077	**0.515***	0.329*	0.042	0.168	**0.403***
object	-0.195	0.045	0.154	**0.685***	-0.123	-0.005	0.333*	**0.469***
numberf	0.225	-0.127	0.246*	**0.450***	-0.014	0.092	0.092	**0.427***
figurew	0.069	0.099	0.058	**0.365***	0.139	0.013	0.237*	**0.291***

65

Holzinger-Swineford ML EFA Using 19 Variables And Geomin Rotation: Four-Factor Solution (Continued)

Factor Correlations

	Grant-White				Pasteur			
	Spatial	Verbal	Speed	Memory	Spatial	Verbal	Speed	Memory
Spatial	1.000				1.000			
Verbal	0.378*	1.000			0.186*	1.000		
Speed	0.372*	0.386*	1.000		0.214	0.326*	1.000	
memory	0.307*	0.380*	0.375*	1.000	0.190*	0.100	0.242*	1.000

66

Bayesian CFA Using MCMC For Holzinger-Swineford

- CFA: Cross-loadings fixed at zero - the model is rejected
- A more realistic hypothesis: Small cross-loadings allowed
- Cross-loadings are not all identified in terms of ML
- Different alternative: Bayesian CFA with informative priors for cross-loadings: $\lambda \sim N(0, 0.01)$.

  This means that 95% of the prior is in the range -0.2 to 0.2

67

Input Bayes CFA 19 Items 4 Factors Crossloading Priors

```
VARIABLE:    NAMES = id female grade agey agem school
             ! school = 0/1 with Pasteur = 1
             visual cubes paper flags general paragrap sentence
             wordc wordm addition code counting straight wordr
             numberr figurer object numberf figurew deduct numeric
             problemr series arithmet;
             USEV = visual-figurew;
             USEOBS = school eq 0;
DEFINE:      STANDARDIZE visual-figurew;
ANALYSIS:    ESTIMATOR = BAYES;
             PROC = 2;
             FBITER = 10000;
```

68

Input Bayes CFA 19 Items 4 Factors Crossloading Priors (Continued)

```
MODEL:       spatial BY visual* cubes paper flags;
             verbal BY general* paragrap sentence wordc wordm;
             speed BY addition* code counting straight;
             memory BY wordr* numberr figurer object numberf
             figurew;

             spatial-memory@1;

             ! cross-loadings:

             spatial  BY general-figurew*0 (a1-a15);
             verbal BY visual-flags*0 (b1-b4);
             verbal BY addition-figurew*0 (b5-b14);
             speed BY visual-wordm*0 (c1-c9);
             speed BY wordr-figurew*0 (c10-c15);
             memory BY visual-straight*0 (d1-d13);
```

69

Input Bayes CFA 19 Items 4 Factors Crossloading Priors (Continued)

```
MODEL PRIORS:
             a1-d13 ~ N(0,.01);
OUTPUT:      TECH1 TECH8 STDY;
PLOT:        TYPE = PLOT2;
```

70

ML analysis

Model	χ^2	Df	P-value	RMSEA	CFI
Grant-White					
CFA	216	146	0.000	0.057	0.930
EFA	110	101	0.248	0.025	0.991
Pasteur					
CFA	261	146	0.000	0.071	0.882
EFA	128	101	0.036	0.041	0.972

Bayesian analysis

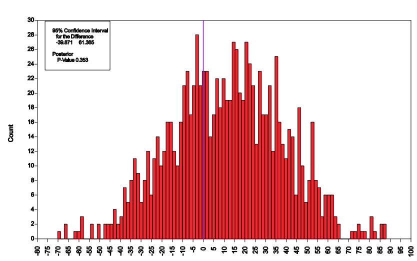
Model	Sample LRT	2.5% PP limit	97.5% PP limit	PP p-value
Grant-White				
CFA	219	12	112	0.006
CFA w/ cross-loadings	142	-39	61	0.361
Pasteur				
CFA	264	56	156	0.000
CFA w/ cross-loadings	156	-28	76	0.162

71

Bayesian Posterior Predictive Checking For The CFA Model For Grant-White

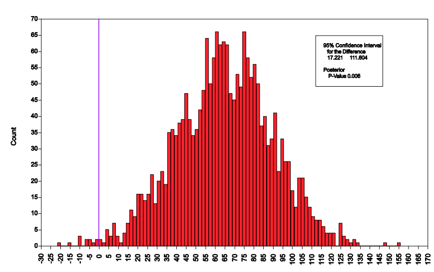
CFA with small cross-loadings not rejected by Bayes PPC:

p = 0.361



Conventional CFA model rejected by Bayes PPC:

p = 0.006:



72

	Grant-White Cross-Loadings Using Informative Priors				Pasteur Cross-Loadings Using Informative Priors			
	Spatial	Verbal	Speed	Memory	Spatial	Verbal	Speed	Memory
visual	**0.640***	0.012	0.050	0.047	**0.633***	0.145	0.027	0.039
cubes	**0.521***	-0.008	-0.010	-0.012	**0.504***	-0.027	-0.041	-0.030
paper	**0.456***	0.040	0.041	0.047	**0.515***	0.018	-0.024	-0.118
flags	**0.672***	0.046	-0.020	0.005	**0.677***	-0.095	0.026	0.093
general	0.037	**0.788***	0.049	-0.040	-0.056	**0.856***	0.027	-0.084
paragrap	-0.001	**0.837***	-0.053	0.030	0.015	**0.801***	-0.011	0.050
sentence	-0.045	**0.885***	0.021	-0.055	-0.063	**0.925***	-0.032	-0.036
wordc	0.053	**0.612***	0.096	0.029	0.055	**0.694***	0.013	0.063
wordm	-0.012	**0.886***	-0.086	0.020	0.092	**0.803***	0.001	0.012
addition	-0.172*	0.030	**0.795***	0.004	-0.147	-0.004	**0.655***	0.010
code	-0.002	0.054	**0.560***	0.130	-0.004	0.111	**0.655***	0.049
counting	0.013	-0.092	**0.828***	-0.049	0.025	-0.058	**0.616***	-0.057
straight	0.189*	0.043	**0.633***	-0.035	0.132	-0.067	**0.558***	0.001
wordr	-0.040	0.044	-0.031	**0.556***	-0.058	0.006	-0.090	**0.731***
numberr	0.003	-0.004	-0.038	**0.552***	0.006	-0.098	-0.106	**0.634***
figurer	0.132	-0.024	-0.049	**0.573***	0.156*	0.027	0.064	**0.517***
object	-0.139	0.014	0.029	**0.724***	-0.097	0.007	0.122	**0.545***
numberf	0.099	-0.071	0.095	**0.564***	-0.029	0.041	0.003	**0.474***
figurew	0.012	0.045	0.007	**0.445***	0.049	0.018	0.085	**0.397***

Bayes For Holzinger-Swineford: Four-factor Solution Using Informative Priors For Cross-loadings (Continued)

	Grant-White				Pasteur			
	Spatial	Verbal	Speed	Memory	Spatial	Verbal	Speed	Memory
Spatial	1.000				1.000			
Verbal	0.535*	1.000			0.348*	1.000		
Speed	0.471*	0.443*	1.000		0.307	0.457*	1.000	
Memory	0.526*	0.515*	0.557*	1.000	0.324*	0.179	0.405*	1.000

Effects Of Using Different Variances For The Informative Priors Of The Cross-Loadings For The Holzinger-Swineford Data: Grant-White

Prior variance	95% cross-loading limit	PPP	Cross-loading (Posterior SD)	Factor corr. range
0.01	0.20	0.361	0.189 (.078)	0.443 – 0.557
0.02	0.28	0.441	0.248 (.096)	0.439 – 0.542
0.03	0.34	0.457	0.275 (.109)	0.423 – 0.530
0.04	0.39	0.455	0.292 (.120)	0.413 – 0.521
0.05	0.44	0.453	0.303 (.130)	0.404 – 0.513
0.06	0.48	0.447	0.309 (.139)	0.400 – 0.510
0.07	0.52	0.439	0.315 (.148)	0.395 – 0.508
0.08	0.55	0.439	0.319 (.156)	0.387 – 0.508
0.09	0.59	0.435	0.323 (.163)	0.378 – 0.506
1.00	0.62	0.427	0.327 (.171)	0.369 – 0.504

75

Effects Of Using Different Variances For The Informative Priors Of The Cross-Loadings For The Holzinger-Swineford Data: Pasteur

Prior Variance	95% cross-loading limit	PPP	Cross-loading (Posterior SD)	Factor corr. range
0.01	0.20	0.162	0.132 (.076)	0.179 – 0.457
0.02	0.28	0.205	0.201 (.088)	0.184 – 0.441
0.03	0.34	0.219	0.223 (.098)	0.188 – 0.431
0.04	0.39	0.218	0.237 (.106)	0.189 – 0.424
0.05	0.44	0.205	0.247 (.115)	0.175 – 0.408
0.06	0.48	0.196	0.255 (.122)	0.175 – 0.402
0.07	0.52	0.195	0.261 (.128)	0.176 – 0.397
0.08	0.55	0.192	None	0.176 – 0.394
0.09	0.59	0.187	None	0.177 – 0.391
1.00	0.62	0.185	None	0.177 – 0.388

76

Summary Of Analyses Of Holzinger-Swineford Data

- Conventional, frequentist, CFA model rejected
- Bayesian CFA with informative cross-loadings not rejected
- The Bayesian approach uses an intermediate hypothesis:
  - Less strict than conventional CFA
  - Stricter than EFA, where the hypothesis only concerns the number of factors
  - Cross-loadings shrunken towards zero; acceptable degree of shrinkage monitored by PPP
- Bayes modification indices obtained by estimated cross-loadings
- Factor correlations: EFA < BSEM < CFA

77

Comparing BSEM And Target Rotation

- Target rotation: EFA rotation chosen to match zero target loadings using least-squares fitting
  - Similarities: Replaces mechanical rotation with judgement/hypotheses
  - Differences: For Target, specifying more than the necessary EFA restrictions does not affect fit and user-defined closeness to zero is replaced with least-squares fitting
- Results for Holzinger-Swineford data:
  - Results similar to EFA with 10 significant cross-loadings for Grant-White and 15 for Pasteur

78

Comparing BSEM And ESEM

- ESEM: Structural equation modeling with EFA measurement model
  - Similarities: Both ESEM and BSEM can be used for measurement models in SEM
  - Differences:
    - ESEM is EFA-oriented while BSEM is CFA-oriented
    - ESEM uses a mechanical rotation and the rotation is not based on information from other parts of the model
    - BSEM is applicable not only to measurement models

79

Bayes Factor Analysis Extensions Using Informative Priors

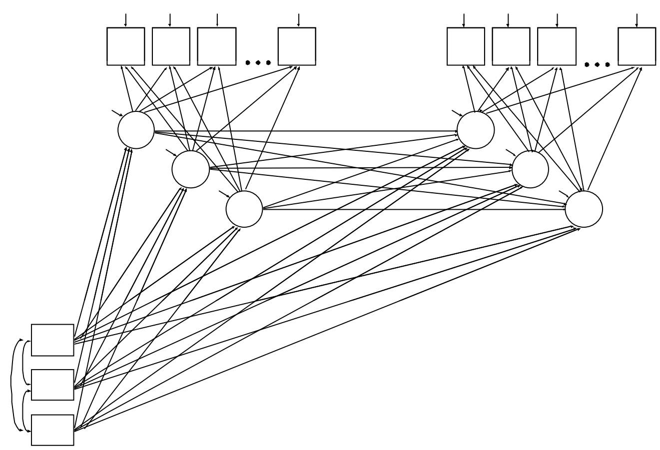
- BSEM paper:
  - Allowing small residual correlations
  - Allowing small deviations from measurement invariance using the MIMIC approach
- Allowing small deviations from measurement invariance using the multiple-group approach
- Other hypotheses than for measurement models
- Other models with non-identified parameters

80

Bayesian Analysis When ML Is Slow Or Intractable Due To Many Dimensions Of Numerical Integration

81

Structural Equation Model With Categorical Factor Indicators Of Three Factors At Two Timepoints



82

Johns Hopkins Aggression Study, 13 Items, Cohort 3 (N=678)

Factor loadings

	Grade 1			Grade 3		
Variables	Verbal	Person	Property	Verbal	Person	Property
stubborn	**0.892***	0.000	-0.286*	**0.913***	0.000	-0.220*
breaks rules	0.413*	0.301*	0.001	0.500*	0.286*	0.001
harms others & property	-0.007	0.508*	0.376*	-0.009	0.526*	0.372*
breaks things	0.016	0.006	**0.808***	0.017	0.057	**0.669***
yells at others	**0.803***	0.019	-0.053	**0.821***	0.015	-0.040
takes others' property	0.227	0.026	**0.589***	0.277	0.025	0.541*
fights	0.016	**0.886***	0.074	0.020	**0.838***	0.067
harms property	0.155	0.004	**0.819***	0.186	0.004	**0.742***
lies	**0.793***	-0.254	0.328*	**0.759***	-0.191	0.236*
talks back to adults	**1.112***	-0.355	0.009	**0.949***	-0.238	0.006
teases classmates	0.408*	0.372*	0.006	0.454*	0.326*	0.005
fights with classmates	0.153*	**0.792***	0.002	0.183*	**0.744***	0.002
loses temper	**0.863***	0.043	-0.149	**0.826***	0.032	-0.107

83

Johns Hopkins Aggression Study, 13 Items, Cohort 3 (N=678)

Factor correlations with variances on the diagonal

	Grade 1			Grade 3		
Variables	Verbal	Person	Property	Verbal	Person	Property
Grade 1						
Verbal	1.000					
Person	0.856	1.000				
	(0.037)					
Property	0.727	0.660	1.000			
	(0.082)	(0.079)				
Grade 3						
Verbal	0.338	0.253	0.148	1.512		
	(0.068)	(0.069)	(0.078)	(0.246)		
Person	0.262	0.224	0.109	0.819	0.936	
	(0.077)	(0.073)	(0.076)	(0.068)	(0.184)	
Property	0.119	0.117	0.075	0.670	0.675	0.855
	(0.078)	(0.072)	(0.072)	(0.105)	(0.089)	(0.216)

84

Computational Issues

- Maximum-likelihood estimation with categorical indicators requires numerical integration with six dimensions which is not practical (problems of computing time, memory, numerical precision). Computational time grows exponentially with the number of continuous latent variables (factors, random effects).
- Bayes is feasible. Computational time grows linearly with the number of continuous latent variables.
  - Hopkins aggression study: Convergence after 7 minutes using two processors and the default of two MCMC chains, converging in 30K iterations

85

Statistical Issues

- Measurement invariance
  - (1) None
  - (2) Configural
  - (3) Factor loadings (factors on the same scale)
  - (4) Factor loadings and intercepts (growth studies possible)

- (3) chosen here

86

Input Excerpts Structural Equation Model With Three Factors At Two Timepoints

```
              USEVARIABLES = y1-y13 y301-y313 black lunch312 male;
              CATEGORICAL = y1-y313;
DEFINE:       CUT y1-y313 (1.5);
ANALYSIS:     ESTIMATOR = BAYES;
              PROCESSORS = 2;
MODEL:        f11 BY y1@1
              y2*.5 (1)
              y3@0
              y4@0
              y5*1 (2)
              y6*0 (3)
              y7*0 (4)
              y8*0 (5)
              y9*1 (6)
              y10*1 (7)
              y11*.5 (8)
              y12*0 (9)
              y13*1 (10);
```

87

Input Excerpts Structural Equation Model With Three Factors At Two Timepoints (Continued)

```
              f12 by y1@0
              y2*.5 (11)
              y3*.5 (12)
              y4*0 (13)
              y5*0 (14)
              y6*0 (15)
              y7@1
              y8@0
              y9*0 (16)
              y10*0 (17)
              y11*0 (18)
              y12*1 (19)
              y13*0 (20);
```

88

Input Excerpts Structural Equation Model With Three Factors At Two Timepoints (Continued)

```
f13 by y1*0 (31)
y2*0 (32)
y3*.5 (33)
y4*1 (34)
y5*0 (35)
y6*.5 (36)
y7*0 (37)
y8@1
y9*.5 (38)
y10@0
y11*8 (39)
y12@0
y13*0 (40);
```

89

Input Excerpts Structural Equation Model With Three Factors At Two Timepoints (Continued)

```
f21 by y301@1
y302*.5 (1)
y303@0
y304@0
y305*1 (2)
y306*0 (3)
y307*0 (4)
y308*0 (5)
y309*1 (6)
y310*1 (7)
y311*.5 (8)
y312*0 (9)
y313*1 (10);
```

90

Input Excerpts Structural Equation Model With Three Factors At Two Timepoints (Continued)

```
f22 by y301@0
y302*.5 (11)
y303*.5 (12)
y304*0 (13)
y305*0 (14)
y306*0 (15)
y307@1
y308@0
y309*0 (16)
y310*0 (17)
y311*0 (18)
y312*1 (19)
y313*0 (20);
```

91

Input Excerpts Structural Equation Model With Three Factors At Two Timepoints (Continued)

```
            f23 by y301*0 (31)
            y302*0 (32)
            y303*.5 (33)
            y304*1 (34)
            y305*0 (35)
            y306*.5 (36)
            y307*0 (37)
            y308@1
            y309*.5 (38)
            y310@0
            y311*8 (39)
            y312@0
            y313*0 (40);

            f11-f23 ON black lunch312 male;
OUTPUT:     TECH1 STDY TECH8;
PLOT:       TYPE = PLOT3;
```

92

Output Excerpts Structural Equation Model With Three Factors At Two Timepoints

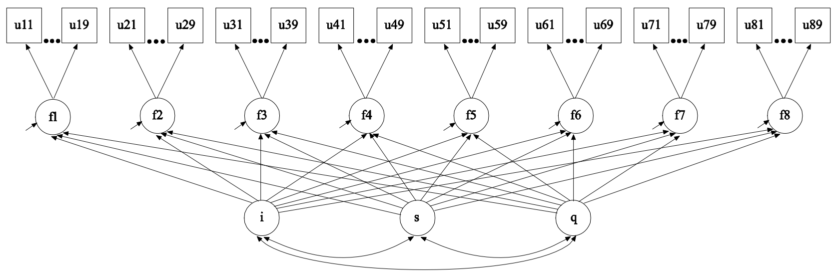
```
Test of model fit


Number of Free Parameters                                              95

Bayesian Posterior Predictive Checking using Chi-Square

     95% Confidence Interval for the Difference Between the
     Observed and the Replicated Chi-Square Values

                                                      -67.231   108.112

     Posterior Predictive P-Value                                   0.324
```

93

Multiple Indicator Growth Modeling With Categorical Variables

94

Growth Model With 9 Categorical Indicators Of A Factor Measured At 8 Time Points



95

Hopkins Aggression Study, Cohort 1, Grade 1 – 7, N = 1174

- Nine binary items
- Measurement invariant loadings and thresholds across time points

96

Input For Bayes Multiple-Indicator Growth Modeling

```
TITLE:        Hopkins Cohort 1 All time points with Classroom
              Information
DATA:         FILE = Cohort1 classroom ALL.DAT;
VARIABLE:     NAMES = PRCID stub1F bkRule1F harmO1F bkThin1F
              yell1F takeP1F ght1F lies1F tease1F stub1S bkRule1S
              harmO1S bkThin1S yell1S takeP1S ght1S lies1S tease1S
              stub2S bkRule2S harmO2S bkThin2S yell2S takeP2S
              ght2S lies2S tease2S stub3S bkRule3S harmO3S bk-
              Thin3S yell3S takeP3S ght3S lies3S tease3S stub4S
              bkRule4S harmO4S bkThin4S yell4S takeP4S ght4S
              lies4S tease4S stub5S bkRule5S harmO5S bkThin5S
              yell5S takeP5S ght5S lies5S tease5S stub6S bkRule6S
              harmO6S bkThin6S yell6S takeP6S ght6S lies6S tease6S
              stub7S bkRule7S harmO7S bkThin7S yell7S takeP7S
              ght7S lies7S tease7S gender race des011 sch011
              sec011 juv99 violchld antisocr conductr athort1F
              harmP1S athort1S harmP2S athort2S harmP3S athort3S
```

97

Input For Bayes Multiple-Indicator Growth Modeling (Continued)

```
              harmP4S athort4S harmP5S athort5S harmP6S harmP7S
              athort7S stub2F bkRule2F harmO2F bkThin2F yell2F
              takeP2F ght2F harmP2F lies2F athort2F tease2F
              classrm;
              USEVAR = stub1f-tease7s male;
              CATEGORICAL = categorical = stub1f-tease7s;
              MISSING = ALL (999);
DEFINE:       cut stub1f-tease7s(1.5);
              MALE = 2 - gender;
ANALYSIS:     PROCESS = 2;
              ESTIMATOR = BAYES;
              FBITER = 20000;
MODEL:        f1 BY stub1f
              bkrule1f-tease1f (1-8);
              f2 BY stub1s
              bkrule1s-tease1s (1-8);
              f3 BY stub2s
              bkrule2s-tease2s (1-8);
```

98

Input For Bayes Multiple-Indicator Growth Modeling (Continued)

```
f4 BY stub3s
bkrule3s-tease3s (1-8);
f5 BY stub4s
bkrule4s-tease4s (1-8);
f6 BY stub5s
bkrule5s-tease5s (1-8);
f7 BY stub6s
bkrule6s-tease6s (1-8);
f8 BY stub7s
bkrule7s-tease7s (1-8);
[stub1f$1 stub1s$1 stub2s$1 stub3s$1 stub4s$1] (11);
[stub5s$1 stub6s$1 stub7s$1] (11);
[bkrule1f$1 bkrule1s$1 bkrule2s$1 bkrule3s$1] (12);
[bkrule4s$1 bkrule5s$1 bkrule6s$1 bkrule7s$1] (12);
[harmo1f$1 harmo1s$1 harmo2s$1 harmo3s$1] (13);
[harmo4s$1 harmo5s$1 harmo6s$1 harmo7s$1] (13);
[bkthin1f$1 bkthin1s$1 bkthin2s$1 bkthin3s$1] (14);
[bkthin4s$1 bkthin5s$1 bkthin6s$1 bkthin7s$1] (14);
```

99

Input For Bayes Multiple-Indicator Growth Modeling (Continued)

```
          [yell1f$1 yell1s$1 yell2s$1 yell3s$1 yell4s$1
          yell5s$1] (15);
          [yell6s$1 yell7s$1] (15);
          [takeP1f$1 takeP1s$1 takeP2s$1 takeP3s$1] (16);
          [takeP4s$1 takeP5s$1 takeP6s$1 takeP7s$1] (16);
          [ght1f$1 ght1s$1 ght2s$1 ght3s$1 ght4s$11] (17);
          [ght5s$1 ght6s$1 ght7s$1] (17);
          [lies1f$1 lies1s$1 lies2s$1 lies3s$1 lies4s$1
          lies5s$1] (18);
          [lies6s$1 lies7s$1] (18);
          [tease1f$1 tease1s$1 tease2s$1 tease3s$1 tease4s$1]
          (19);
          [tease5s$1 tease6s$1 tease7s$1] (19);
          [f1-f8@0];
          i s q | f1@0 f2@.5 f3@1.5 f4@2.5 f5@3.5 f6@4.5 f7@5.5
          f8@6.5;
          i-q ON male;
OUTPUT:   TECH1 TECH4 TECH8 TECH10 STANDARDIZED SVALUES;
PLOT:     TYPE = PLOT2;
```

100

Output For Bayes Multiple-Indicator Growth Modeling

```
Tests of model fit

Number of Free Parameters                                   36

Bayesian Posterior Predictive Checking using Chi-Square

     95% Confidence Interval for the Difference Between
     the Observed and the Replicated Chi-Square Values

                                                 129.488   547.650

     Posterior Predictive P-Value                            0.002
```

101

Mixture Modeling

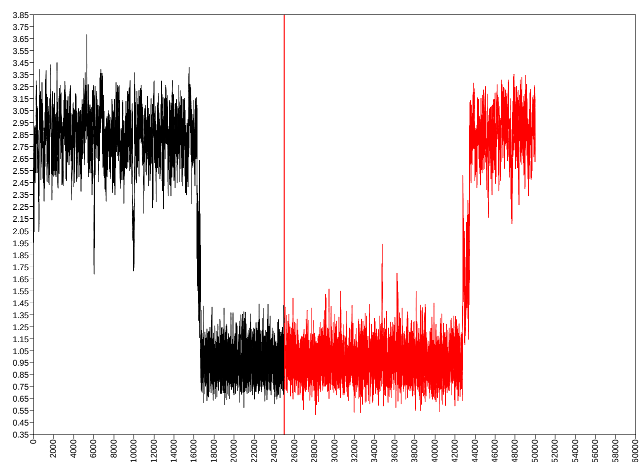
102

Growth Mixture Modeling

```
TITLE:      this is an example of a GMM for a continuous outcome
DATA:       FILE = ex8.1.dat;
            NOBS = 600;
VARIABLE:   NAMES = y1-y4 x;
            CLASSES = c(2);
ANALYSIS:   TYPE= MIXTURE;
            ESTIMATOR = BAYES;
            CHAIN = 1;
            FBITER = 50000;
MODEL:      %OVERALL%
            i s | y1@0 y2@1 y3@2 y4@3;
            i s ON x;
OUTPUT:     TECH1 TECH8;
PLOT:       TYPE = PLOT3;
```

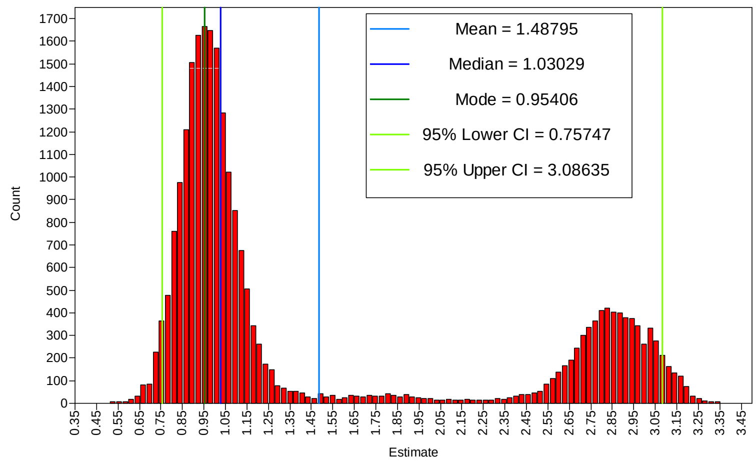
103

Trace Plot For Intercept Mean In Class 1 Mixture Label Switching



104

Posterior Distribution For Intercept Mean In Class 1 Mixture Label Switching



105

Growth Mixture Modeling With Parameter Constraint

```
TITLE:      this is an example of a GMM for a continuous outcome
DATA:       FILE = ex8.1.dat;
            NOBS = 600;
VARIABLE:   NAMES = y1-y4 x;
            CLASSES = c(2);
ANALYSIS:   TYPE = MIXTURE;
            ESTIMATOR = BAYES;
            FBITER = 50000;
            CHAIN = 1;
MODEL:      %OVERALL%
            i s | y1@0 y2@1 y3@2 y4@3;
            i s ON x;
            %C#1%
            [i*1] (m1);
            %C#2%
            [i*0] (m2);
```

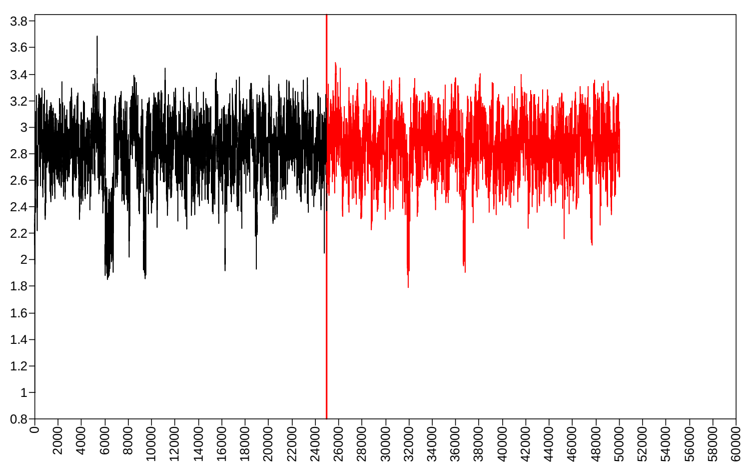
106

Growth Mixture Modeling With Parameter Constraint (Continued)

```
OUTPUT:      TECH1 TECH8;
PLOT:        TYPE = PLOT3;
MODEL CONSTRAINT:
             m1 > m2;
```

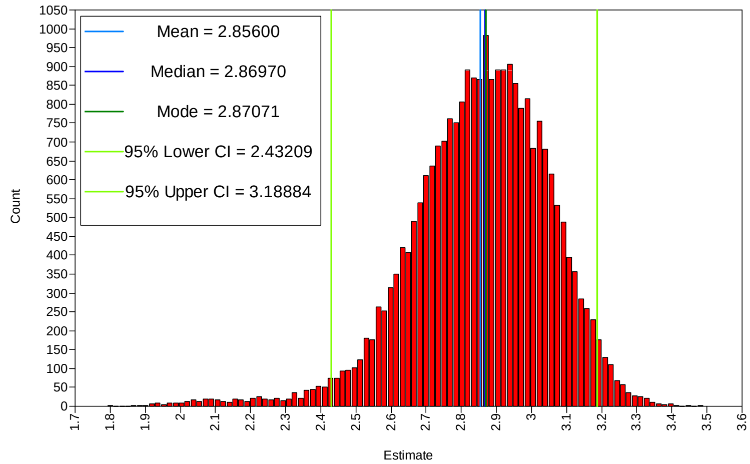
107

Trace Plot For Intercept Mean In Class 1



108

Posterior Distribution For Intercept Mean In Class 1



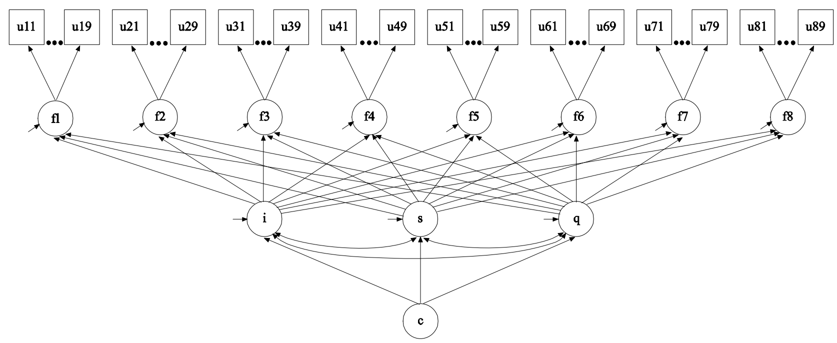
109

Bayes Mixture Starting Values Via ML

```
ANALYSIS:   TYPE = MIXTURE;
            STARTS = 100 20; ! do ML from multiple starts
            STVALUES = ML; ! start Bayes from the best ML solution
            PROCESSORS = 4 (STARTS);
            ESTIMATOR = BAYES;
            FBITER = 10000;
```

110

Growth Mixture Model With Multiple Categorical Indicators



111

Multilevel Regression With A Continuous Dependent variable

112

Multilevel Regression With A Random Intercept

Consider a two-level regression model for individuals $i = 1, 2 \ldots n_j$ in clusters $j = 1, 2 \ldots, J$,

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + r_{ij}, \quad (1)$$

$$\beta_{0j} = \gamma_{00} + u_{0j}, \quad (2a)$$

$$\beta_{1j} = \gamma_{10} \quad (2b)$$

113

Multilevel Regression With A Continuous Dependent Variable And A Small Number Of Clusters

- 10 schools, each with 50 students
- Intra-class correlation 0.10

114

Input For Random Intercept Regression

```
TITLE:
DATA:          FILE = c10n50icc1.dat;
VARIABLE:      NAMES = y x clus;
               WITHIN = x;
               CLUSTER = clus;
ANALYSIS:      TYPE = TWOLEVEL;
               ESTIMATOR = BAYES;
               PROCESS = 2;
               FBITER = 10000;
MODEL:         %WITHIN%
               y ON x;
               y (w);
               %BETWEEN%
               y (b);
```

115

Input For Random Intercept Regression (Continued)

```
MODEL PRIORS:
               b ~ IG (.001, .001);
MODEL CONSTRAINT:
               NEW(icc);
               icc = b/(b+w);
OUTPUT:        TECH1 TECH8;
PLOT:          TYPE = PLOT2;
```

116

Output For Random Intercept Regression

Parameter	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%
Within Level					
y ON					
x	0.909	0.069	0.000	0.777	1.042
Residual variances					
y	2.105	0.135	0.000	1.866	2.394

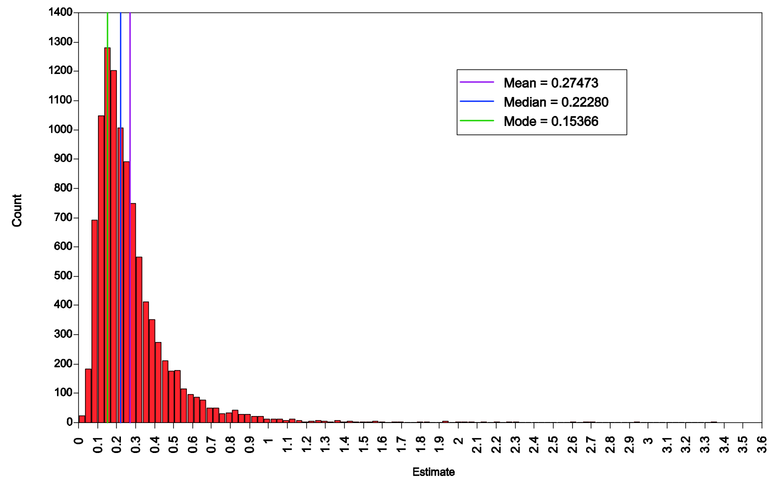
117

Output For Random Intercept Regression (Continued)

Parameter	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%
Between Level					
Means					
y	0.145	0.178	0.191	-0.209	0.493
Variances					
y	0.223	0.205	0.000	0.073	0.805
New/Additional Parameters					
ICC	0.096	0.063	0.00	0.033	0.276

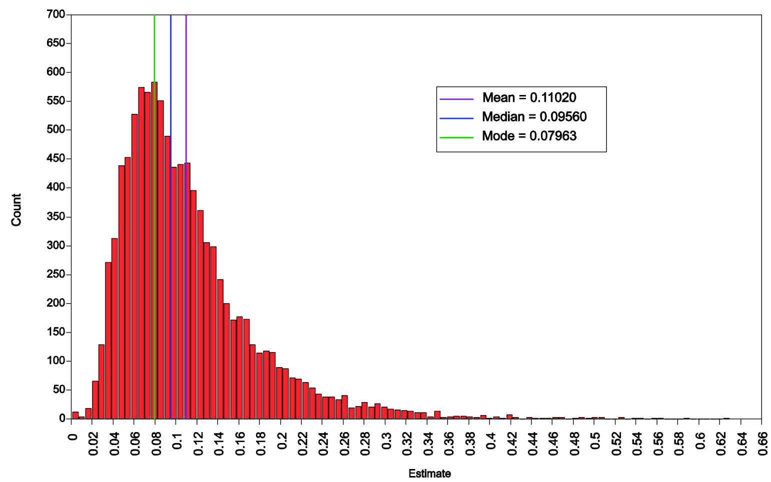
118

Posterior Distribution Of Between-Level Intercept Variance



119

Posterior Distribution Of Intra-Class Correlation



120

Output For
ML Twolevel Regression

Parameter	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
y ON				
x	0.909	0.069	13.256	0.000
Residual variances				
y	2.089	0.133	15.653	0.000

121

Output For
ML Twolevel Regression (Continued)

Parameter	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Between Level				
Means				
y	0.143	0.152	0.942	0.346
Variances				
y	0.190	0.104	1.828	0.067
New/Additional Parameters				
ICC	0.083	0.042	1.975	0.048

122

Simulation Studies: Multilevel Regression With A Small Number Of Clusters

Browne-Draper Simulation Study

Browne, W.J. & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Bayesian Analysis, 3, 473-514.

- Priors studied
  a) Variance ~ U(0, 1/ε) Gelman-Rubin 1992
  b) Variance ~ Inverse-Gamma(ε, ε) Spiegelhalter 1997
- Monte Carlo study for a random intercept model with number of clusters (total sample):
  6 (108), 12 (216), 24 (432), and 48 (864)
  with intraclass correlations varying from 0.012 to 0.5.

Browne-Draper Simulation Study (Continued)

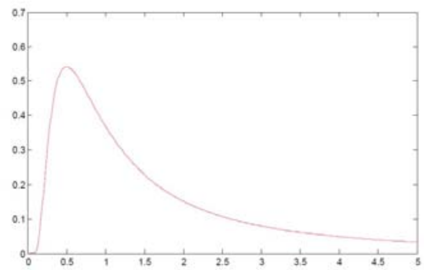
- Browne-Draper (2006) findings:
  - generally good results (bias, coverage) for prior a) using the mode and prior b) using the median
  - poorer results for smaller number of clusters and smaller intraclass correlations

Mplus Monte Carlo Study

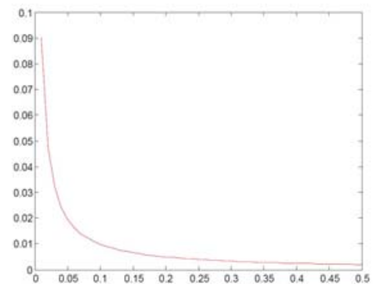
- 10 schools, each with 50 students
- Intra-class correlation 0.10

Source: Muthén (2010). Bayesian analysis in Mplus: A brief introduction

Different Inverse-Gamma Priors For Variance Parameters



IG(1,1) density



IG(0.001,0.001) density

The mean for IG(α, β) is $\beta/(\alpha$-1)
The mode is $\beta/(\alpha+1)$
α: shape parameter
β: scale parameter

127

Output Excerpts For ML In A Monte Carlo Study Of Twolevel Regression

Parameter	Population	Estimates Average	Std. Dev.	S.E. Average	M.S.E.	95% Cover	% Sig Coeff
Within Level							
y ON							
x	1.000	0.9957	0.0630	0.0639	0.0040	0.948	1.000
Residual variances							
y	2.000	2.0052	0.1291	0.1281	0.0167	0.946	1.000

128

Output Excerpts For ML In A Monte Carlo Study Of Twolevel Regression (Continued)

Parameter	Population	Estimates Average	Std. Dev.	S.E. Average	M.S.E.	95% Cover	% Sig Coeff
Between Level							
Means							
y	0.000	-0.0035	0.1624	0.1485	0.0263	0.892	0.108
Variances							
y	0.222	0.1932	0.1155	0.1045	0.0141	0.808	0.180
New/Additional Parameters							
ICC	0.100	0.0860	0.0458	0.0422	0.0023	0.812	0.498

129

Input For Bayes Monte Carlo Twolevel Regression Using An IG (0.001, 0.001) Prior

```
TITLE:          Bayes IG (eps,eps)
MONTECARLO:     NAMES = y x;
                NOBS = 500;
                NREP = 500;
                NCSIZES = 1;
                CSIZES = 10 (50); ! 10 clusters of size 50
                WITHIN = x;
MODEL POPULATION:
                %WITHIN%
                x*1;
                y ON x*1;
                y*2;
                %BETWEEN%
                y*.222; !icc = .222/2.222 = 0.1
ANALYSIS:       TYPE = TWOLEVEL;
                ESTIMATOR = BAYES;
                PROCESS = 2;
                FBITER = 10000;
```

130

Input For Bayes Monte Carlo Twolevel Regression Using An IG (0.001, 0.001) Prior (Continued)

```
MODEL:        %WITHIN%
              y ON x*1;
              y*2 (w);
              %BETWEEN%
              y*.22(b); !icc = .222/2.222 =0.1
MODEL PRIORS:
              b ~ IG (.001, .001);
MODEL CONSTRAINT:
              NEW(icc*.1);
              icc = b/(w+b);
OUTPUT:       TECH9;
```

131

Output For Bayes Monte Carlo Twolevel Regression Using An IG (0.001, 0.001) Prior

Parameter	Population	Estimates Average	Std. Dev.	S.E. Average	M.S.E.	95% Cover	% Sig Coeff
Between Level							
Means							
y	0.000	0.0128	0.1646	0.1746	0.0272	0.928	0.072
Variances							
y	0.222	0.2322	0.1373	0.2036	0.0189	0.942	1.000
New/Additional Parameters							
ICC	0.100	0.1011	0.0532	0.0618	0.0028	0.934	1.000

132

Output For Bayes Monte Carlo Twolevel Regression Using A U(0, 1000) Prior (POINT=MODE)

Parameter	Population	Estimates Average	Std. Dev.	S.E. Average	M.S.E.	95% Cover	% Sig Coeff
Between Level							
Means							
y	0.000	0.0190	0.1645	0.2106	0.0277	0.966	0.034
Variances							
y	0.222	0.2304	0.1246	0.3255	0.0156	0.930	1.000
New/Additional Parameters							
ICC	0.100	0.1012	0.0484	0.0841	0.0023	0.928	1.000

133

Output For Bayes Monte Carlo Twolevel Regression Using An IG (-1, 0) Prior

Parameter	Population	Estimates Average	Std. Dev.	S.E. Average	M.S.E.	95% Cover	% Sig Coeff
Between Level							
Means							
y	0.000	0.0123	0.1645	0.2122	0.0272	0.966	0.034
Variances							
y	0.222	0.3348	0.1784	0.3767	0.0445	0.928	1.000
New/Additional Parameters							
ICC	0.100	0.1391	0.0626	0.0876	0.0054	0.928	1.000

134

Conclusions

- ML does poorly
- Bayes
  - Good coverage in all cases
  - Low bias except for IG(-1,0) prior
  - 10 clusters of size 50 and icc = 0.1 can be handled well in Bayes

135

Multilevel Regression With A Categorical Dependent Variable And Several Random Slopes

136

Computational Advantage Of Bayes Over ML In Random Effects Models

- With categorical outcomes and continuous latent variables (random effects, factors), ML requires numerical integration
- Each random effect accounts for one dimension of integration
- ML is computationally impractical with several dimensions (>4?), so only a few random slopes can be included in addition to the random intercept
- Monte Carlo integration is possible but can lead to non-convergence and poor precision for the log likelihood
- Estimation time for ML grows exponentially as a function of the number of random effects, but for Bayes it grows linearly

137

Monte Carlo Study For Twolevel Probit Regression With Many Random Slopes

Source: Asparouhov, T. & Muthén, B. (2010). Bayesian analysis of latent variable models using Mplus. Technical Report. Version 4. Section 5.

Choice of Inverse-Wishart priors for the covariance matrix of the q random effects:

- IW(0,-q-1) . Constant density, improper uniform prior
- IW(I, q+1). Default, proper prior. Marginal prior for the correlations is uniform on [-1, 1]. Marginal for variances is IG(1, 0.5) with mode 0.25. Weakly informative prior.
- IW(2I, q+1). Marginal variance prior IG(1, 1) with mode 0.5
- Data are generated with true random effects variances of 0.5 and 200 clusters of size 20. The number of random effects vary from 1 (only the intercept is random) to 6.

138

Bias (Percent Coverage) For Random Intercept Variance (= 0.5) In Two-Level Probit Regression With *q* Random Effects

- 200 clusters of size 20

Prior	q =1	q = 2	q = 3	q = 4	q = 5	q = 6
IW(0, -q -1)	0.03 (90)	0.04 (92)	0.04 (96)	0.08 (81)	0.10 (79)	0.19 (60)
IW(*I*, q +1)	0.03 (89)	0.02 (93)	-0.01 (97)	-0.01 (95)	-0.04 (97)	-0.05 (92)
IW(2*I*, q +1)	0.03 (90)	0.03 (93)	0.01 (96)	0.02 (97)	-0.01 (97)	-0.01 (96)

- q = 1: Random intercept only
- Prior 1 is the default in Mplus 6.0. Prior 2 is the default in Mplus 6.1 and later
- Source: Asparouhov & Muthén, B. (2010). Bayesian analysis of latent variable models using Mplus.

139

Conclusions

- The dependence on the prior increases with increasing number of random effects q
- The first prior IW(0, -q-1) is the worst
- The third prior IW(2I, q+1) has its prior variance mode equal to the true value 0.5, so approximately the ML estimate. Using an ML-based prior, Bayes can be viewed as a computing device to get estimates akin to ML

140

Multilevel Regression With A Categorical Dependent Variable And Small Random Slope Variance

Monte Carlo Simulation With A Small Random Slope Variance

- Random intercept and two random slopes with the second random slope variance of zero
- 200 clusters of size 20
- Bayes: 5 sec/rep
- ML: 19 minutes/rep due to slow convergence with random slope variance zero
- Bayes variance-covariance prior: IW(I, 4)

Source: Asparouhov & Muthén, B. (2010). Bayesian analysis of latent variable models using Mplus.

Bias (Percent Coverage) For Small Random Slope Variance Estimation

Parameter	ML	Bayes
α_1	0.01 (90)	0.00 (91)
α_2	0.01 (95)	0.00 (91)
α_3	0.00 (96)	0.01 (97)
β_1	0.01 (96)	0.00 (94)
β_2	0.00 (98)	0.01 (93)
β_3	0.00 (95)	0.03 (91)
ψ_{11}	0.01 (96)	0.03 (94)
ψ_{22}	0.01 (93)	0.03 (94)
ψ_{33}	0.01 (99)	0.06 (0)
ψ_{12}	0.00 (97)	0.00 (94)
ψ_{13}	0.00 (97)	0.00 (98)
ψ_{23}	0.00 (97)	0.01 (97)

143

Conclusions

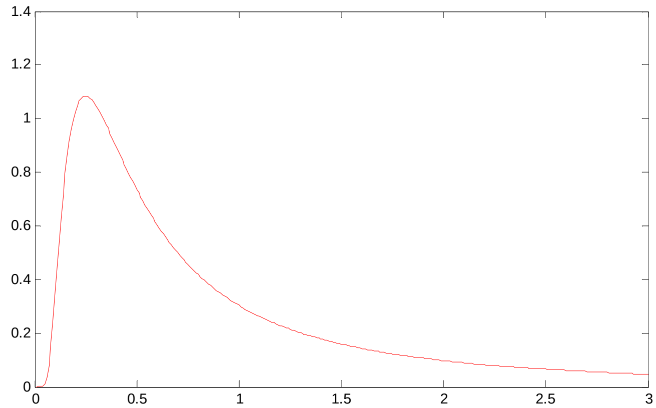
- Bayes avoids the slow convergence of ML
- Bayes has small bias
- Bayes has acceptable coverage

- But, what about $\psi_{33} = 0.06$?

144

Small Random Slope Variance: Prior Versus Posterior

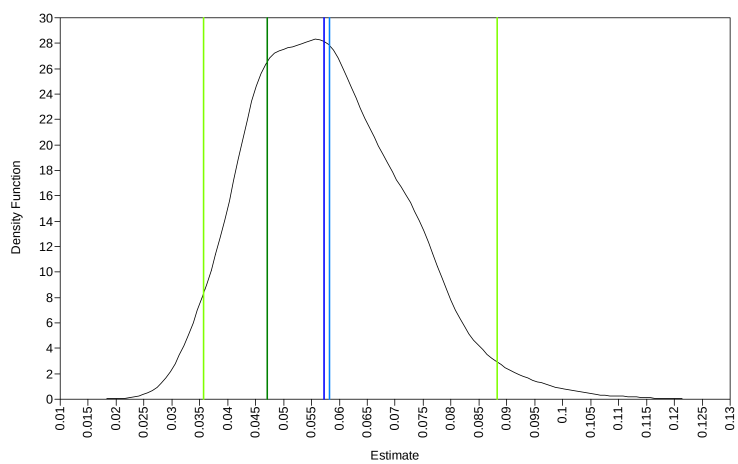
Prior distribution using IW(I, 4):



145

Small Random Slope Variance: Prior Versus Posterior (Continued)

Posterior distribution:



146

Small Random Slope Variance: Prior Versus Posterior (Continued)

- Posterior suggests that prior is not appropriate
- Change the prior to a weakly informative prior with mode at the mode of the previous posterior:

```
MODEL PRIORS:
      v1~IW(1,4);
      v2~IW(1,4);
      v3~IW(0.2,4);
      r1~IW(0,4);
      r2~IW(0,4);
      r3~IW(0,4);
```

- New estimate for the random slope variance = 0.02

147

Testing For Zero Variance Components

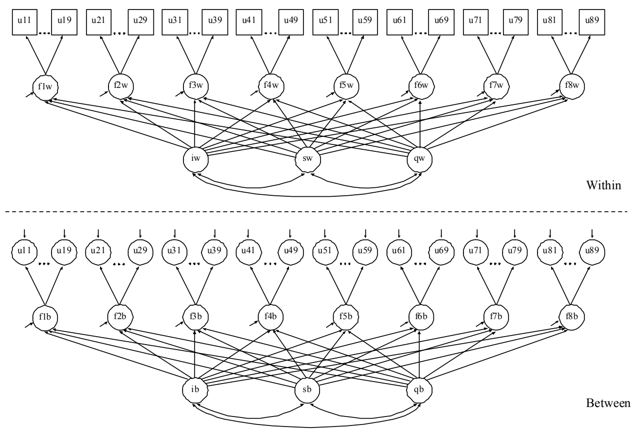
- ML standard errors can not be used because the testing is for a value at the border of the admissible parameter space
- ML LRT complicated
- Bayes inherently will always have significant variances due to prior
- Bayes can be used iteratively based on adjusting the prior

148

Twolevel Growth: 3/4 -Level Analysis

Multiple-indicator growth modeling (T occasions, p items/occ.):

- Number of dimensions: 2 x *T*, or *T* + *p* x *T* (2-level growth with between-level residuals)



149

Meta-Analysis

150

Hierarchical Structure Of Meta-Analysis

- Meta-analysis pools information from several studies designed to address the same scientific question
- Data frequently are in the form of summary statistics from each study, such as effect measures, means, (log) odds ratios, relative risks, z-transformed correlations, and the associated sampling variances
- A normal model for the summary statistic y_j in study j assumes
  $y_j \sim N(\theta_j, \sigma^2_j)$,
  where σ^2_j is assumed known, estimated from data
- A random-effects model specifies
  $\theta_j \sim N(\mu, \tau^2)$
- A Bayesian model adds priors such as
  $\mu \sim N(0, 1000)$, $\tau^2 \sim U(0,1000)$

151

Beta-Blocker Trials (Yusuf et al., 1985; Gelman et al., 2004)

- Mortality after myocardial infarction in 22 clinical trials
- Patients randomized to receive or not receive beta-blockers (relaxing heart muscles)
- The outcome is defined as the log odds

$$y_j = log\left(\frac{y_{1_j}}{n_{1_j} - y_{1_j}}\right) - log\left(\frac{y_{0_j}}{n_{0_j} - y_{0_j}}\right)$$

- With approximate sampling variance

$$\sigma_j^2 = \frac{1}{y_{1_j}} + \frac{1}{n_{1_j} - y_{1_j}} + \frac{1}{y_{0_j}} + \frac{1}{n_{0_j} - y_{0_j}}$$

152

Beta-Blocker Data From 22 Trials

Study	Raw Data (deaths/total)				Log-Odds	sd
j	Control		Treated		y_j	σ_j
1	3/39	(8%)	3/38	(8%)	0.028	0.850
2	14/116	(12%)	7/114	(6%)	-0.741	0.483
3	11/93	(12%)	5/69	(7%)	-0.541	0.565
4	127/1520	(8%)	102/1533	(7%)	-0.246	0.138
5	27/365	(7%)	28/355	(8%)	0.069	0.281
6	6/52	(12%)	4/59	(7%)	-0.584	0.676
7	152/939	(16%)	98/945	(10%)	-0.512	0.139
8	48/471	(10%)	60/632	(9%)	-0.079	0.204
9	37/282	(13%)	25/278	(9%)	-0.424	0.274
10	188/1921	(10%)	138/1916	(7%)	-0.335	0.117
11	52/283	(9%)	64/873	(7%)	-0.213	0.195

153

Beta-Blocker Data From 22 Trials (Continued)

Study	Raw Data (deaths/total)				Log-Odds	sd
j	Control		Treated		y_j	σ_j
12	47/266	(18%)	45/263	(17%)	-0.039	0.229
13	16/293	(5%)	9/291	(3%)	-0.593	0.425
14	45/883	(5%)	57/858	(7%)	0.282	0.205
15	31/147	(21%)	25/154	(16%)	-0.321	0.298
16	38/213	(18%)	33/207	(16%)	-0.135	0.261
17	12/122	(10%)	28/251	(11%)	0.141	0.364
18	6/154	(4%)	8/151	(5%)	0.322	0.553
19	3/134	(2%)	6/174	(3%)	0.444	0.717
20	40/218	(18%)	32/209	(15%)	-0.218	0.260
21	43/364	(12%)	27/391	(7%)	-0.591	0.257
22	39/674	(6%)	22/680	(3%)	-0.608	0.272

154

Random Slope Approach In Mplus

- Heterogeneous variances can be handled by random slopes as in Mplus Web Note # 3. See also User's Guide ex 3.9 (random coefficient regression)
- A similar approach is used in

  Cheung (2008). A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling. Psychological Methods, 13, 182-202

  $y_j = \theta_j + \varepsilon_j$; $\varepsilon_j \sim N(0, \sigma^2_j)$

  Dividing by σ_j,

  $y^*_j = 0 + \theta_j * x_j + e_j$; $e_j \sim N(0, 1)$

  where $y^*_j = y_j / \sigma_j$, $x_j = 1/ \sigma_j$ and θ_j is a random slope ~ $N(\mu, \tau^2)$

155

Input For ML Single Level

```
TITLE:     Yusuf (1985) data
           ML using a single-level random slope approach
DATA:      FILE = yusuf.txt;
VARIABLE:  NAMES = id y sd;
           USEVARIABLES = y x;
DEFINE:    ! transform to unit error variance:
           y = y/sd;
           x = 1/sd;
ANALYSIS:  TYPE = RANDOM;
MODEL:     [y@0.0];              ! Intercept is fixed at 0
           y@1.0;                ! Error variance is fixed at 1
           theta | y ON x;
           theta;                ! var(theta): tau^2
           [theta];              ! mean(theta): mu
```

156

Input For ML Twolevel

```
TITLE:      Yusuf (1985) beta blocker data
            ML twolevel approach
DATA:       FILE = yusuf.txt;
VARIABLE:   NAMES = id y sd;
            USEVARIABLES = y x;
            WITHIN = ALL;
            CLUSTER = id;
DEFINE:     y = y/sd;
            x = 1/sd;
ANALYSIS:   TYPE = TWOLEVEL RANDOM;
MODEL:      %WITHIN%
            [y@0.0];              ! Intercept is fixed at 0
            y@1.0;                ! Error variance is fixed at 1
            theta | y ON x;
            %BETWEEN%
            theta;                ! var(theta): tau^2
            [theta];              ! mean(theta): mu
```

157

Output For ML Twolevel

Parameter	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
Intercepts				
y	0.000	0.000	999.000	999.000
Residual variances				
y	1.000	0.000	999.000	999.000
Between Level				
Means				
theta	-0.249	0.064	-3.876	0.000
Variances				
theta	0.010	0.020	0.519	0.604

158

Input For Bayes Twolevel

```
TITLE:      Yusuf (1985) beta blocker trials
            Bayes twolevel approach
DATA:       FILE = yusuf.txt;
VARIABLE:   NAMES = id y sd;
            USEV = y x;
            WITHIN = ALL;
            CLUSTER = id;
DEFINE:     y = y/sd;
            x = 1/sd;
ANALYSIS:   TYPE = TWOLEVEL RANDOM;
            ESTIMATOR = BAYES;
            PROCESSOR = 2;
            FBITER = 10000;   ! To make sure the posterior is well
                              ! captured
            THIN = 30;      ! Slower, but avoids high autocorrelation
            POINT = MODE;   ! Report the mode instead of the median
                            ! with small number of clusters and
                            ! uniform prior U(0,1/eps) according to
                            ! Browne-Draper (2006)
```

159

Input For Bayes Twolevel (Continued)

```
MODEL:          %WITHIN%
                [y@0.0];            ! Intercept is fixed at 0
                y@1.0;              ! Error variance is fixed at 1
                theta | y ON x;
                %BETWEEN%
                theta (v);          ! var(theta): tau^2
                [theta];            ! mean(theta): mu
MODEL PRIORS:   v~U(0,1000);
OUTPUT:         TECH8;
PLOT:           TYPE = PLOT2;
```

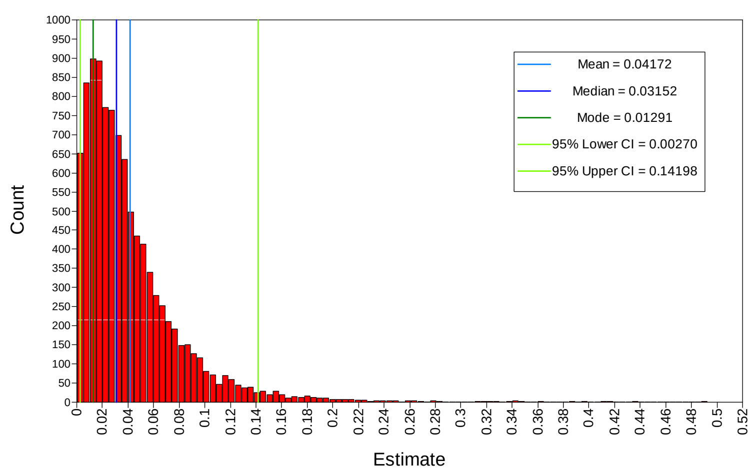
160

Output For Bayes Twolevel

Parameter	Estimate	Posterior S.D.	One-Tailed P-Value	95% Lower 2.5%	95% Upper 2.5%
Within Level					
Intercepts					
y	0.000	0.000	1.000	0.000	0.000
Residual variances					
y	1.000	0.000	0.000	1.000	1.000
Between Level					
Means					
theta	-0.255	0.074	0.002	-0.386	-0.092
Variances					
theta	0.013	0.038	0.000	0.003	0.142

161

Posterior Distribution For The Between-Level Variance



162

Estimated Beta-Blocker Model

Overall effect: Estimated median μ = -0.255

The Bayesian posterior gives the 95% credibility interval for μ as [-0.386, -0.092], or (exponentiating) in odds ratio scale [0.68, 0.91], favoring beta-blockers

In contrast, complete pooling of all 22 studies, that is, assuming all studies are identical so that there is no effect variation across studies, i.e. $\tau^2 = 0$, gives the narrower odds ratio interval [0.70, 0.85].

Variance: Estimated mode for $\tau^2 = 0.013$

τ^2 may be very close to zero but may also plausibly be as high as 0.14 (sd = 0.38).

163

Estimating The Study-Specific θ_j Effects: Plausible Values

- The study-specific θ_j values are latent variable values that can be estimated in line with factor scores
- In Bayesian analysis θ_j values can be drawn by
  - 1. drawing μ and τ^2 from the posterior distribution
  - 2. drawing θ_j from $N(\mu, \tau^2)$
- Add to the previous Mplus input:

```
DATA IMPUTATION:

    NDATASETS = 100;              ! 100 draws
    PLAUSIBLE = yusufPlaus.dat;   ! this contains summaries for each
                                  ! theta_j over the 100 draws, e.g.
                                  ! the median
    SAVE = yusufimp*.dat;         ! This contains all 100 values of each
                                  ! theta_j
```

- The plausible values for θ_j are shrunken toward the overall mean, with stronger shrinkage for smaller studies.

164

Beta-Blocker Data From 22 Trials

Study	Raw Data (deaths/total)				Log-Odds	sd	Posterior for θ_j		
j	Control		Treated		y_j	σ_j	2.5%	Median	97.5%
1	3/39	(8%)	3/38	(8%)	0.028	0.850	-0.604	-0.200	0.126
2	14/116	(12%)	7/114	(6%)	-0.741	0.483	-0.724	-0.288	0.090
3	11/93	(12%)	5/69	(7%)	-0.541	0.565	-0.716	-0.254	0.017
4	127/1520	(8%)	102/1533	(7%)	-0.246	0.138	-0.449	-0.241	-0.043
5	27/365	(7%)	28/355	(8%)	0.069	0.281	-0.530	-0.164	0.131
6	6/52	(12%)	4/59	(7%)	-0.584	0.676	-0.732	-0.280	0.225
7	152/939	(16%)	98/945	(10%)	-0.512	0.139	-0.609	-0.398	-0.183
8	48/471	(10%)	60/632	(9%)	-0.079	0.204	-0.433	-0.191	0.083
9	37/282	(13%)	25/278	(9%)	-0.424	0.274	-0.690	-0.286	-0.043
10	188/1921	(10%)	138/1916	(7%)	-0.335	0.117	-0.492	-0.306	-0.136
11	52/283	(9%)	64/873	(7%)	-0.213	0.195	-0.581	-0.245	-0.014

165

Beta-Blocker Data From 22 Trials (Continued)

Study	Raw Data (deaths/total)				Log-Odds	sd	Posterior for θ_j		
j	Control		Treated		y_j	σ_j	2.5%	Median	97.5%
12	47/266	(18%)	45/263	(17%)	-0.039	0.229	-0.502	-0.191	0.186
13	16/293	(5%)	9/291	(3%)	-0.593	0.425	-0.647	-0.254	-0.038
14	45/883	(5%)	57/858	(7%)	0.282	0.205	-0.261	-0.015	0.332
15	31/147	(21%)	25/154	(16%)	-0.321	0.298	-0.670	-0.299	-0.005
16	38/213	(18%)	33/207	(16%)	-0.135	0.261	-0.534	-0.220	0.044
17	12/122	(10%)	28/251	(11%)	0.141	0.364	-0.617	-0.178	0.220
18	6/154	(4%)	8/151	(5%)	0.322	0.553	-0.565	-0.199	0.159
19	3/134	(2%)	6/174	(3%)	0.444	0.717	-0.593	-0.208	0.257
20	40/218	(18%)	32/209	(15%)	-0.218	0.260	-0.485	-0.207	0.088
21	43/364	(12%)	27/391	(7%)	-0.591	0.257	-0.759	-0.336	-0.040
22	39/674	(6%)	22/680	(3%)	-0.608	0.272	-0.734	-0.362	-0.031

166

Meta Analysis With A Covariate: Hox Data

```
TITLE:      Hox data
            Bayes twolevel approach with a covariate
DATA:       FILE = hoxID.txt;
VARIABLE:   NAMES = id d varofd dum weeks;
            USEVARIABLES = d weeks intercpt;
            WITHIN = ALL;
            CLUSTER = id;
DEFINE:     w2 = SQRT(varofd**(-1));
            d = w2*d;
            one = 1;
            intercpt = w2*one;
            weeks = w2*weeks;
ANALYSIS:   TYPE = TWOLEVEL RANDOM;  ! Use random slope analysis
            ESTIMATOR = BAYES;
            PROCESSOR = 2;
            FBITER = 10000;   ! To make sure the posterior is well
                              ! captured
```

167

Meta Analysis With A Covariate: Hox Data

```
         THIN = 30;       ! Slower, but avoids high autocorrelation
         POINT = MODE;    ! Report the mode instead of the median
                          ! with uniform prior U(0,1/eps) according
                          ! to Browne-Draper (2006)
MODEL:   %WITHIN%
         [d@0.0];             ! Intercept is fixed at 0
         d@1.0;               ! Error variance is fixed at 1
         u | d ON intercpt;
         d ON weeks;
         %BETWEEN%
         u (v);               ! var(u): tau^2
         [u];                 ! mean(u): intercept
MODEL
PRIORS:  v~U(0,1000);         ! An alternative is the Spiegelhalter
                              ! prior Inverse-Gamma (eps,eps)
                              ! V~IG(.001,.001); reporting the
                              ! median
```

168

Meta Analysis With A Covariate: Hox Data

		Posterior	One-Tailed	95%	
Parameter	Estimate	S.D.	P-Value	Lower 2.5%	Upper 2.5%
Within Level					
d ON					
weeks	0.144	0.039	0.001	0.064	0.218
Intercepts					
d	0.000	0.000	1.000	0.000	0.000
Residual variances					
d	1.000	0.000	0.000	1.000	1.000
Between Level					
Means					
u	-0.232	0.234	0.163	-0.682	0.256
Variances					
u	0.030	0.064	0.000	0.006	0.245

169

Multiple Imputation

170

Analysis With Missing Data

Used when individuals are not observed on all outcomes in the analysis to make the best use of all available data and to avoid biases in parameter estimates, standard errors, and tests of model fit.

Types of Missingness

- MCAR -- missing completely at random
  - Variables missing by chance
  - Missing by randomized design
  - Multiple cohorts assuming a single population

171

Analysis With Missing Data (Continued)

- MAR -- missing at random
  - Missingness related to observed variables
  - Missing by selective design
- Non-Ignorable (NMAR)
  - Missingness related to values that would have been observed
  - Missingness related to latent variables

172

Correlates Of Missing Data

- MAR is more plausible when the model includes covariates influencing missing data
- Correlates of missing data may not have a "causal role" in the model, i.e. not influencing dependent variables, in which case including them as covariates can bias model estimates.

173

Correlates Of Missing Data (Continued)

- Two solutions:

  (1) Modeling (ML)
    - Including missing data correlates not as x variables but as "y variables," freely correlated with all other observed variables. AUX (M) does this automatically

  (2) Multiple imputation (Bayes; Schafer, 1997) with two different sets of observed variables
    - Imputation model
    - Analysis model

Overview in Enders (2010).

174

Bayesian Imputation Of Missing Data

- 3 Steps:
  - (1) Estimate the model using Bayes
  - (2) Draw a set of parameter values from the posterior distribution
  - (3) For each set of parameter values, generate missing data according to the model
- Choice of model in step (1):
  - H1: Unrestricted model
  - H0: Restricted model: Factor model, growth model, latent class model, twolevel model, etc

175

Three H1 Imputation Approaches In Mplus: DATA IMPUTATION: MODEL =

- COVARIANCE: Default in all cases
  - (1) Bayes estimation of an H1 model, (2) Do multiple draws of (2a) parameters, (2b) missing data generated from those parameters
- SEQUENTIAL: (Ragunathan et al., 2001; chained equations)
- REGRESSION
- Asparouhov, T. & Muthén, B. (2010). Multiple imputation with Mplus.

176

Plausible Values

- Multiple imputations for latent variable values (H0 imputation)
- Available for both continuous and categorical latent variables
- DATA IMPUTATION command saves two data sets:
  - SAVE = imp*.dat; saves all observations and latent variables for all imputations. Can be used to produce a distribution for each latent variable for each individual, not just a mean and SE
  - PLAUSIBLE = latent.dat; saves for each observation and latent variable a summary over imputed data sets
- Two uses:
  - Interest in individual scores
  - Interest in secondary analysis

177

Bayes Plausible Values Versus ML Factor Scores For Each Individual: Small Residual Variances With Small Sample

Source: Asparouhov, T. & Muthén, B. (2010). Plausible values for latent variables using Mplus. Technical Report.

Section 2 generates a data set for a 1-factor model with 3 indicators and n=45. Indicator reliability =0.5.

178

Bayes Plausible Values Versus ML Factor Scores For Each Individual: Small Residual Variances With Small Sample (Continued)

ML gives a small residual variance for the 3rd indicator:

Parameter	Pop. Value	ML	Bayes
μ_1	0	0.307	0.270
μ_2	0	0.008	-0.053
μ_3	0	0.296	0.212
λ_2	1	0.873	0.976
λ_3	1	1.665	1.543
θ_1	1	1.298	1.407
θ_2	1	1.587	1.603
θ_3	1	0.034	0.589
ψ	1	0.966	0.946

179

Bayes Plausible Values Versus ML Factor Scores For Each Individual: Small Residual Variances With Small Sample (Continued)

- ML factor scores of poor quality given high correlation with indicator having small residual variance
- Factor score (plausible value) MSE:

$$MSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{\eta}_i - \eta_i)^2}$$

MSE: ML = 0.636, Bayes = 0.563

- 95% coverage (over subjects): ML = 20%, Bayes = 89%

180

Bayes Plausible Values Versus ML Factor Scores For Each Individual: Small Residual Variances With Small Sample (Continued)

- Bayes correlation between factor score SE and absolute factor score value = 0.76, due to the tails of the factor score distribution having fewer observations. In contrast, the ML SE is constant by assuming parameter estimates have zero variation

181

Bayes Plausible Values Versus ML Factor Scores: Secondary Analysis

Asparouhov, T. & Muthén, B. (2010). Plausible values for latent variables using Mplus. Technical Report.

- Section 4.2 simulation of a data set with n=10,000 using a model with two factors and 3 indicators each
- Bayes uses 5 imputations and ML the regression method.

182

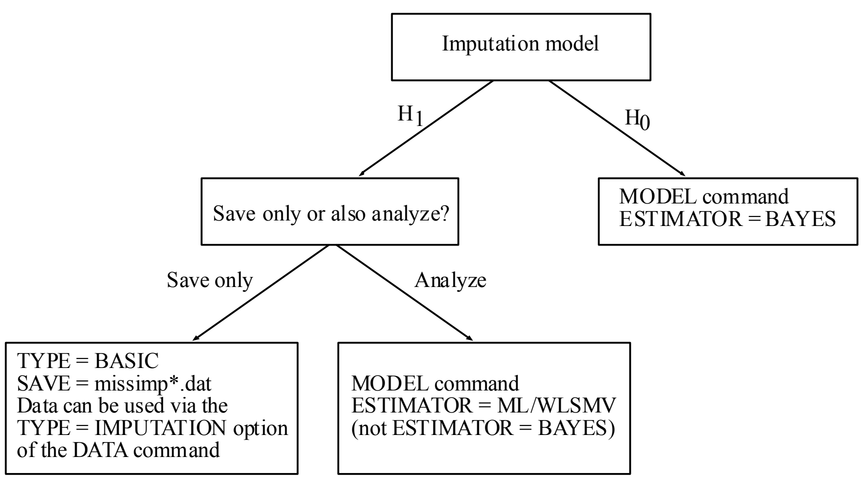
Bayes Plausible Values Versus ML Factor Scores: Secondary Analysis (Continued)

- Results are shown for factor means, variances, covariance, and correlation

Parameter	True Value	Factor Scores	Plausible Values
α_1	0	0.00	0.00
α_2	0	0.00	0.00
ψ_{11}	1	0.76	1.03
ψ_{22}	1	0.80	1.05
ψ_{12}	0.6	0.57	0.63
ρ	0.6	0.73	0.61

183

Multiple Imputation Choices



184

Multiple Imputation Using H1: Practical Advice

Asparouhov, T. & Muthén, B. (2010). Multiple imputation with Mplus. Technical Report. Version 2.

Section 4 gives 14 tips. The top four are:

- Analyze data first with TYPE=BASIC (no imputation). First, treat categorical as continuous, then categorical. If relevant, do TYPE=BASIC TWOLEVEL
- Check for perfectly correlated variables such as sums and individual components, or dummy variables for all categories
- Don't use all variables on the NAMES list for multiple imputation. Instead, choose a subset using USEVARIABLES.
- Don't use as imputation variables that you know have no predictive value for the missingness. For example, ID or SS numbers, or ZIP code treated as a continuous variable.

185

Multiple Imputation Using H1: Practical Advice (Continued)

Note that unlike other software, Mplus imputes missing data only after successfully estimating a general/unrestricted model using Bayes. This means that a convergence criterion needs to be satisfied, which may be hindered by slow mixing/non-convergence, e.g. due to model non-identification (see points 5 and onwards).

186

Input For H1 Multiple Imputation Without Further Analysis

```
TITLE:      this is an example of multiple imputation for a set of
            variables with missing values
DATA:       FILE  =  miss.dat;
VARIABLE:   ! the following are all the variables in the data
            ! set:
            NAMES = x1 x2 y1-y4 z1-z5 v1-v50;
            ! the following variables will be used to create the
            ! imputed data sets:
            USEVARIABLES = x1 x2 y1-y4 z1-z5;
            ! the following variables are saved with the imputed
            ! data sets, but not used to create the imputed data
            ! sets:
            AUXILIARY = v1 – v10;
            MISSING = ALL (999);
```

187

Input For H1 Multiple Imputation Without Further Analysis (Continued)

```
DATA IMPUTATION:
            ! the following are the variables for which missing
            ! data will be imputed:
            IMPUTE = y1-y4 x1 (c) x2;
            NDATASETS = 10;
            ! the following data sets will contain data for the
            ! variables x1 x2 y1-y4 z1-z5 v1-v10:
            SAVE = missimp*.dat;
ANALYSIS:   TYPE = BASIC;
OUTPUT:     TECH8;
```

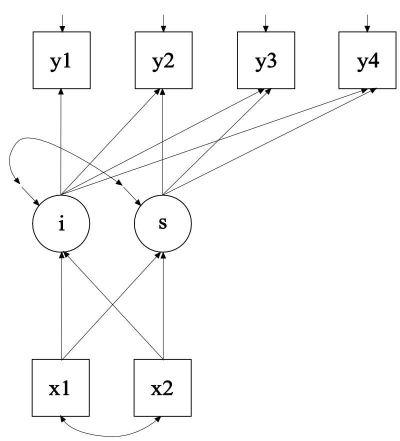
188

Input For Multiple Imputation For A Set Of Variables With Missing Values Followed By The Estimation Of A Growth Model (Ex 11.5)

```
TITLE:      this is an example of multiple imputation for a set of
            variables with missing values
DATA:       FILE = ex11.5.dat;
VARIABLE:   NAMES = x1 x2 y1-y4 z;
            MISSING = ALL (999);
DATA IMPUTATION:
            IMPUTE = y1-y4 x1 (c) x2;
            NDATASETS = 10;
            SAVE = ex11.5imp*.dat;
ANALYSIS:   TYPE = BASIC;
OUTPUT:     TECH8;
```

189

Linear Growth Model



190

Input For Multiple Imputation Followed By The Estimation Of A Growth Model (Ex 11.5), Continued

```
TITLE:          This is an example of growth modeling using
                multiple imuptation data
DATA:           FILE = ex11.5implist.dat;
                TYPE = IMPUTATION;
VARIABLES:      NAMES = x1 x2 y1-y4 z;
                USEVARIABLES = y1-y4 x1 x2;
ANALYSIS:       ESTIMATOR = ML;
MODEL:          i s | y1@0 y2@1 y3@2 y4@3;
                i s ON x1 x2;
OUTPUT:         TECH1 TECH4;
```

191

Input For H1 Multiple Imputation With Further H0 Analysis

```
DATA:          FILE = tlevimp.dat;
VARIABLE:      NAMES = y1-y10 c;
               BETWEEN = y1 y2;
               MISSING = ALL(999);
               CLUSTER = c;
DATA IMPUTATION:
               NDATASETS = 5;
               IMPUTE = y1 y2-y3 y4-y6 y10;
               SAVE = tlevimp*.dat;
ANALYSIS:      TYPE = TWOLEVEL;
               ESTIMATOR = ML;
```

192

Input For H1 Multiple Imputation With Further H0 Analysis (Continued)

```
MODEL:        %WITHIN%

              ETAW1 BY Y3-Y6*0.8;
              ETAW2 BY Y7-Y10*0.8;
              Y3-Y10*0.36;
              ETAW1-ETAW2 @1;
              ETAW1 WITH ETAW2 * .5;

              %BETWEEN%

              ETAB1 BY Y1*0.8 Y3-Y6*0.8;
              ETAB2 BY Y2*0.8 Y7-Y10*0.8;
              Y1-Y10*0.36;
              ETAB1-ETAB2@1 ;
              ETAB1 WITH ETAB2 *0.3;
```

193

Multiple Imputation With A Categorical Outcome In A Growth Model With MAR Missing Data: Using WLSMV On Imputed Data

194

Choice Of Estimators With Categorical Outcomes And Missing Data

- ML: Intractable with many continuous latent variables (factors, random effects)
- WLSMV: Fast also with many continuous latent variables, but does not handle missing data under MAR
- New alternative: Bayes multiple imputation + WLSMV

195

Monte Carlo Study Of Growth Modeling With Dichotomous Outcomes And Missing At Random (MAR)

- Monte Carlo simulation for 5 time points, binary outcome, linear growth model, n=1000, and MAR missingness as a function of the first outcome, varying the number of imputations as 5 versus 50. Unrestricted (H1) imputation using SEQUENTIAL (Ragunathan et al., 2001)

Source: Asparouhov, T. & Muthén, B. (2010). Multiple imputation with Mplus.

196

Bias (Coverage) For MAR Dichotomous Growth Model: WLSMV Versus Imputation+WLSMV

Estimator	WLSMV	WLSMV (5 Imput.)	WLSMV (50 Imput.)
μ_1	-0.03 (.92)	-0.01 (.96)	-0.01 (.93)
μ_2	-0.16 (.02)	0.00 (.92)	0.00 (.93)
ψ_{11}	-0.23 (.62)	0.06 (.94)	0.05 (.95)
ψ_{22}	0.09 (.96)	0.04 (.91)	0.04 (.91)
ψ_{12}	-0.08 (.68)	0.00 (.93)	0.00 (.94)

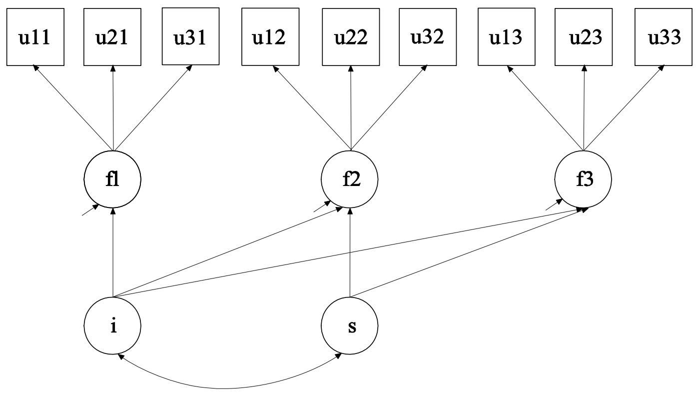
μ are means of the two growth factors. True values: 0, 0.20
ψ are variances and covariance of the two growth factors. True values: 0.50, 0.50, 0.30
Source: Asparouhov, T. & Muthén, B. (2010). Multiple imputation with Mplus.

197

H0 Imputation

198

Multiple Indicator Linear Growth Model



199

Input For Multiple Imputation Of Plausible Values Using Bayesian Estimation Of A Growth Model (Ex 11.6). H0 Imputation

```
TITLE:      this is an example of multiple imputation of plausible
            values generated from a multiple indicator linear
            growth model for categorical outcomes using Bayesian
            estimation
DATA:       FILE = ex11.6.dat;
VARIABLE:   NAMES = u11 u21 u31 u12 u22 u32 u13 u23 u33;
            CATEGORICAL = u11-u33;
ANALYSIS:   ESTIMATOR = BAYES;
            PROCESSORS = 2;
```

200

Input For Multiple Imputation Of Plausible Values Using Bayesian Estimation Of A Growth Model (Continued)

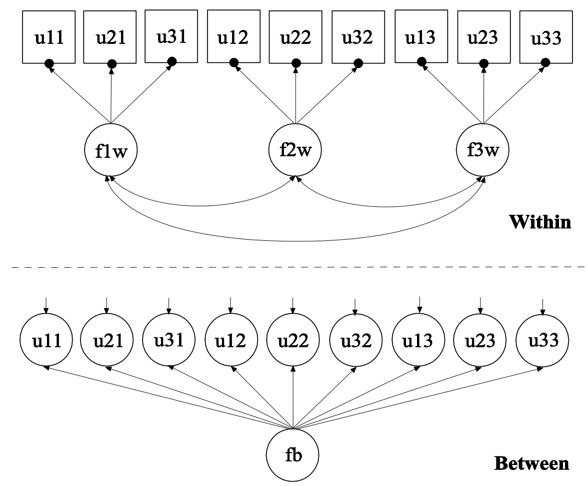
```
MODEL:      f1 BY u11
            u21-u31 (1-2);
            f2 BY u12
            u22-u32 (1-2);
            f3 BY u13
            u23-u33 (1-2);
            [u11$1 u12$1 u13$1] (3);
            [u21$1 u22$1 u23$1] (4);
            [u31$1 u32$1 u33$1] (5);
            i s | f1@0 f2@1 f3@2;
DATA IMPUTATION:
            NDATASETS = 20;
            PLAUSIBLE = ex11.6plaus.dat;
            SAVE = ex11.6imp*.dat;
OUTPUT:     TECH1 TECH8;
```

201

Twolevel Imputation

202

Twolevel Factor Model



203

Input For Multiple Imputation Using A Two-Level Factor Model With Categorical Outcomes Followed By The Estimation Of A Growth Model (Ex 11.7)

```
TITLE:      this is an example of multiple imputation using a two-
            level factor model with categorical outcomes
DATA:       FILE = ex11.7.dat;
VARIABLE:   NAMES = u11 u21 u31 u12 u22 u32 u13 u23 u33 clus;
            CATEGORICAL = u11-u33;
            CLUSTER = clus;
            MISSING = ALL (999);
ANALYSIS:   TYPE = TWOLEVEL;
            ESTIMATOR = BAYES;
            PROCESSORS = 2;
```

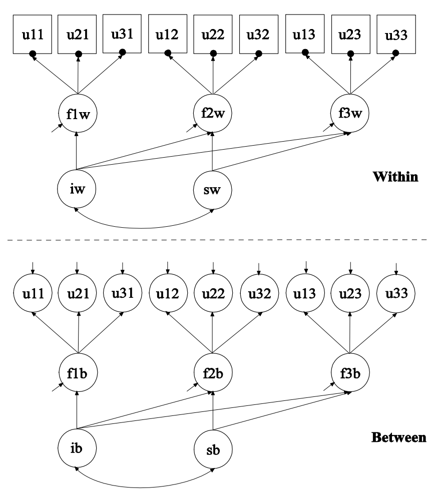
204

Multiple Imputation: Two-Level Factor Model With Categorical Outcomes Followed By The Estimation Of A Growth Model

```
MODEL:      %WITHIN%
            f1w BY  u11
                    u21 (1)
                    u31 (2);
            f2w BY u12
                    u22 (1)
                    u32 (2);
            f3w BY u13
                    u23 (1)
                    u33 (2);
            %BETWEEN%
            fb BY u11-u33*1;
            fb@1;
DATA IMPUTATION:
            IMPUTE = u11-u33(c);
            SAVE = ex11.7imp*.dat;
OUTPUT:     TECH1 TECH8;
```

205

Twolevel Growth Model



206

Input Multiple Imputation Using A Two-Level Factor Model With Categorical Outcomes Followed By The Estimation Of A Growth Model (Second Part)

```
TITLE:     this is an example of a two-level multiple indicator
           growth model with categorical outcomes using multiple
           imputation data
DATA:      FILE = ex11.7implist.dat;
           TYPE = IMPUTATION;
VARIABLE:  NAMES = u11 u21 u31 u12 u22 u32 u13 u23 u33 clus;
           CATEGORICAL = u11-u33;
           CLUSTER = clus;
ANALYSIS:  TYPE = TWOLEVEL;
           ESTIMATOR = WLSMV;
           PROCESSORS = 2;
```

207

Input Multiple Imputation Using A Two-Level Factor Model With Categorical Outcomes Followed By The Estimation Of A Growth Model (Second Part)

```
MODEL:     %WITHIN%
           f1w BY  u11
                   u21 (1)
                   u31 (2);
           f2w BY u12
                   u22 (1)
                   u32 (2);
           f3w BY u13
                   u23 (1)
                   u33 (2);
           iw sw | f1w@0 f2w@1 f3w@2;
           %BETWEEN%
           1b BY   u11
                   u21 (1)
                   u31 (2);
```

208

Input Multiple Imputation Using A Two-Level Factor Model With Categorical Outcomes Followed By The Estimation Of A Growth Model (Second Part)

```
          f2b BY u12
                 u22 (1)
                 u32 (2);
          f3b BY u13
                 u23 (1)
                 u33 (2);
          [u11$1 u12$1 u13$1] (3);
          [u21$1 u22$1 u23$1] (4);
          [u31$1 u32$1 u33$1] (5);
          u11-u33;
          ib sb | f1b@0 f2b@1 f3b@2;
          [f1b-f3b@0 ib@0 sb];
          f1b-f3b (6);
OUTPUT:   TECH1 TECH8;
SAVEDATA: SWMATRIX = ex11.7sw*.dat;
```

209

Technical Aspects Of Analyzing Multiple Imputation Data

210

Standard Errors For Parameters Estimated From Multiple Imputation Data

$$\bar{\theta} = \frac{1}{m}\sum_{j=1}^{m}\hat{\theta}_j$$

where $\hat{\theta}_j$ is an estimate for the j^{th} imputed data set

$$\bar{v} = \frac{1}{m}\sum_{j=1}^{m} v_j$$

where v_j is the corresponding estimated squared standard error

$$B = \frac{1}{m-1}\sum_{j=1}^{m}(\hat{\theta}_j - \bar{\theta})^2$$

$$T = \bar{v} + \left(1 + \frac{1}{m}\right)B$$

Rubin (1987), Schafer (1997)

211

Chi-Square Model Testing Using Multiple Imputation Data:

Source: Asparouhov-Muthen (2010): Chi-square statistics with multiple imputation. Technical appendix.

212

Formulas

T_m: LRT test statistic for the m-th imputed data set
Q_{0m}: estimates for the m-th imputed data set under the H_0 model
Q_{1m}: estimates for the m-th imputed data set under the H_1 model

$$\bar{T} = \frac{1}{M}\sum_{m=1}^{M} T_m$$

$$\overline{Q_0} = \frac{1}{M}\sum_{m=1}^{M} Q_{0m}$$

$$\overline{Q_1} = \frac{1}{M}\sum_{m=1}^{M} Q_{1m}$$

213

Formulas (Continued)

T'_m: LRT test statistic with estimates fixed at $\overline{Q}_0$ for H_0 and $\overline{Q}_1$ for H_1

$$\overline{T'} = \frac{1}{M}\sum_{m=1}^{M} T'_m$$

$$T_{imp} = \frac{\overline{T'}}{1 + r_3}$$

$$r_3 = \frac{M+1}{(M-1)(p_1 - p_0)}(\bar{T} - \overline{T'})$$

214

Monte Carlo Study Of Imputation Chi-Square Type I Error: Comparing the Naive And Correct Chi-2

25% missing	Average chi-square (rejection rate) with 8 d.f.	
N	Naïve $\overline{T}$	Correct T_{imp}
100	18.0 (.45)	9.2 (.12)
500	16.2 (.45)	7.8 (.08)
1000	15.7 (.46)	8.1 (.05)

40% missing	Average chi-square (rejection rate) with 8 d.f.	
N	Naïve $\overline{T}$	Correct T_{imp}
100	26.5 (.90)	18.8 (.15)
500	25.9 (.86)	8.7 (.09)
1000	25.5 (.78)	8.3 (.09)

215

Monte Carlo Study Of Imputation Chi-Square Power: Comparing Imputation And ML

Power study results for 25% missing data case. Percentage rejection rate.

N	100	150	200	250	300
T_{imp}	34	53	68	75	85
T_{FIML}	50	60	76	86	92

Power study results for 40% missing data case. Percentage rejection rate.

N	100	150	200	250	300
T_{imp}	30	32	44	51	69
T_{FIML}	40	52	55	75	84

216

References

Asparouhov T. & Muthen B. (2010). Multiple Imputation with Mplus. Technical Report. www.statmodel.com

Asparouhov & Muthén (2010). Plausible values for latent variable using Mplus. Technical Report.

Asparouhov, T. & Muthen, B. (2010a). Bayesian analysis using Mplus. Technical implementation. Technical appendix. Los Angeles: Muthen & Muthen. www.statmodel.com

Asparouhov, T. & Muthen, B. (2010b). Bayesian analysis of latent variable models using Mplus. Technical Report. Version 4. Los Angeles: Muthen & Muthen. www.statmodel.com

Asparouhov, T. & Muthen, B. (2009). Exploratory structural equation modeling. Structural Equation Modeling, 16, 397-438.

217

References (Continued)

Barnard, J., McCulloch, R. E., and Meng, X. L. (2000). Modeling co-variance matrices in terms of standard deviations and correlations, with application to shrinkage. Statistica Sinica 10, 1281-1311.

Botev, Z.I., Grotowski, J.F. & Kroese, D.P. (2010). Kernel density estimation via diffusion. Forthcoming in Annals of Statistics.

Browne, W.J. & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Bayesian Analysis, 3, 473-514.

Casella, G. & George, E.I. (1992). Explaining the Gibbs sampler. The American Statistician, 46, 167-174.

Cheung, M. (2008). A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling. Psychological Methods, 13, 182-202

218

References (Continued)

Dominicus, A., Ripatti, S., Pedersen. N.L., & Palmgren, J. (2008). A random change point model for assessing the variability in repeated measures of cognitive function. Statistics in Medicine, 27, 5786-5798.

Dunson D., Palomo J., and Bollen K. (2005) Bayesian structural equation modeling. SAMSI TR2005-5. http:nnwww.samsi.infonTRntr2005-05.pdf

Enders, C.K. (2010). Applied missing data analysis. New York: Guilford.

Fox, J.P., and Glas, C. A.W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. Psychometrika, 66, 271-288.

Gelfand, A.E., Hills, S.E., Racine-Poon, A., & Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. Journal of the American Statistical Association, 85, 972-985.

219

References (Continued)

Gelman A., Jakulin A., Pittau M. G., Su Y. S. (2008a). A weakly informative default prior distribution for logistic and other regression models. The Annals of Applied Statistics, 2, 1360-1383.

Gelman A., van Dyk, D. A., Huang Z., Boscardin J. (2008b) Using Redundant Parameterizations to Fit Hierarchical Models. Journal Of Computational And Graphical Statistics, 17, 95-122.

Gelman A. (2006) Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1, 515-533.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (2004). Bayesian data analysis. Second edition. Boca Raton: Chapman & Hall.

Gelman, A., Meng, X.L., Stern, H.S. & Rubin, D.B. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). Statistica Sinica, 6, 733-807.

220

References (Continued)

Gelman, A. & Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. Statistical Science, 7, 457-511.

Green, P. (1996). MCMC in image analysis. In Gilks, W.R., Richardson, S., & Spiegelhalter, D.J. (eds.), Markov chain Monte Carlo in Practice. London: Chapman & Hall.

Harman, H.H. (1976). Modern factor analysis. Third edition. Chicago: The University of Chicago Press.

Hjort, N. L., Dahl, F. A. and Steinbakk, G. H. (2006). Post-processing posterior predictive p-values. J. Amer. Statist. Assoc. 101, 1157-1174.

Holzinger, K.J. & Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution. Supplementary Educational Monographs. Chicago.: The University of Chicago Press.

221

References (Continued)

Lee S. Y., Song X. Y., Cai J. H. (2010) A Bayesian Approach for Non-linear Structural Equation Models With Dichotomous Variables Using Logit and Probit Links. Structural Equation Modeling, 17, 280-302.

Lee, S.Y. (2007). Structural equation modeling. A Bayesian approach. Chichester: John Wiley & Sons.

Little, R. J. & Rubin, D. B. (2002). Statistical analysis with missing data. Second edition. New York: John Wiley and Sons.

Lynch, S.M. (2010). Introduction to applied Bayesian statistics and estimation for social scientists. New York: Springer.

MacKinnon, D.P. (2008). Introduction to statistical mediation analysis. New York: Erlbaum.

222

References (Continued)

MacKinnon, D.P., Lockwood, C.M., & Williams, J. (2004). Condence limits for the indirect effect: Distribution of the product and resampling methods. Multivariate Behavioral Research, 39, 99-128.

McLachlan, G. J. & Peel, D. (2000). Finite mixture models. New York: Wiley and Sons.

Meng, X.L. & Rubin, D.B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. Biometrika, 79, 103-111.

Mislevy R., Johnson E., & Muraki E. (1992). Scaling Procedures in NAEP. Journal of Educational Statistics, Vol. 17, No. 2, Special Issue: National Assessment of Educational Progress, pp. 131-154.

Muthen, B. (2010) Bayesian Analysis In Mplus: A Brief Introduction. Mplus Technical Report. http://www.statmodel.com

Muthén, B. & Asparouhov, T. (2010). Bayesian SEM: A more flexible representation of substantive theory. Under review in Psychological Methods.

References (Continued)

Muthen, B., Asparouhov, T., Hunter, A. & Leuchter, A. (2010). Growth modeling with non-ignorable dropout: Alternative analyses of the STAR*D antidepressant trial. Psychological Methods, 16, 17-33.

Muthen B. & Asparouhov, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In Fitzmaurice, G., Davidian, M., Verbeke, G. & Molenberghs, G. (eds.), Longitudinal Data Analysis, pp. 143-165. Boca Raton: Chapman & Hall/CRC Press.

Muthen, B. & Muthen, L. (1998-2010). Mplus User's Guide. Sixth Edition. Los Angeles, CA: Muthen & Muthen.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation. Structural Equation Modeling, 14, 535-569.

References (Continued)

Olsen, M.K. & Schafer, J.L. (2001). A two-part random effects model for semicontinuous longitudinal data. Journal of the American Statistical Association, 96, 730-745.

Patz R. and Junker B. (1999) A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models Journal of Educational and Behavioral Statistics Summer, 24, 146-178.

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., and Solenberger, P. (2001) A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. Survey Methodology, Vol. 27, No 1. 85-95.

Rubin, D.B. (1987) Multiple Imputation for Nonresponse in Surveys. J. Wiley & Sons, New York.

225

References (Continued)

Rupp, A.A., Dey, D.K., & Zumbo, B.D. (2004). To Bayes or not to Bayes, from whether to when: Applications of Bayesian methodology to modeling. Structural Equation Modeling, 11, 424-451.

Savalei, V. (2010) Small Sample Statistics for Incomplete Nonnormal Data: Extensions of Complete Data Formulae and a Monte Carlo Comparison. Structural Equation Modeling, 17, 241-264.

Schafer, J.L. (1997). Analysis of incomplete multivariate data. London: Chapman & Hall.

Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. Psychometrika, 64, 37-52.

Segawa E., Emery S., and Curry S. (2008) Extended Generalized Linear Latent and Mixed Model, Journal of Educational and Behavioral Statistics, 33, 464-484.

226

References (Continued)

Shrout, P.E. & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. Psychological Methods, 7, 422-445.

Spiegelhalter, D.J., Thomas, A., Best, N. & Gilks, W.R.(1997). BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.60. Cambridge: Medical Research Council Biostatistics Unit.

Spiegelhalter, D.J., Best, N. G., Carlin, B.P., & van der Linde, A. (2002). Bayesian measures of model complexity and t (with discussion). Journal of the Royal Statistical Society, Series B (Statistical Methodology) 64, 583-639.

Song X.Y., Xia Y. M., Lee S. Y. (2009) Bayesian semiparametric analysis of structural equation models with mixed continuous and unordered categorical variables. Statistics in Medicine, 28, 2253-2276.

227

References (Continued)

van Dyk, D. A., and Meng, X. L. (2001), The Art of Data Augmentation. Journal of Computational and Graphical Statistics, 10, 1-111.

von Davier M., Gonzalez E. & Mislevy R. (2009) What are plausible values and why are they useful? IERI Monograph Series Issues and Methodologies in Large-Scale Assessments. IER Institute. Educational Testing Service. 12.

Yuan, Y. & MacKinnon, D.P. (2009). Bayesian mediation analysis. Psychological Methods, 14, 301-322.

Yusuf, Peto, Lewis, Collins & Sleight (1985). Beta blockade during and after myocardial infarction: An overview of the randomized trials. Progress in Cardiovascular Diseases, 27, 335-371.

228