

Bayesian Analysis In Mplus: A Brief Introduction

Bengt Muthén
Incomplete Draft, Version 1

May 2, 2010

Abstract

This paper uses a series of examples to give an introduction to how Bayesian analysis is carried out in Mplus. The examples are a mediation model with estimation of an indirect effect, a structural equation model, a two-level regression model with estimation of a random intercept variance, a multiple-indicator binary growth model with a large number of latent variables, a two-part growth model, and a mixture model. It is shown how the use of Mplus graphics provides information on estimates, convergence, and model fit. Comparisons are made with frequentist estimation using maximum likelihood and weighted least squares. Data and Mplus scripts are available on the Mplus website.

1 Introduction

Frequentist (e.g., maximum likelihood) and Bayesian analysis differ by the former viewing parameters as constants and the latter as variables. Maximum likelihood (ML) finds estimates by maximizing a likelihood computed for the data. Bayes combines prior distributions for parameters with the data likelihood to form posterior distributions for the parameter estimates. The priors can be diffuse (non-informative) or informative where the information may come from previous studies. The posterior provides an estimate in the form of a mean, median, or mode of the posterior distribution.

There are many books on Bayesian analysis and most are quite technical. Gelman et al. (2004) provides a good general statistical description, whereas Lynch (2010) gives a somewhat more introductory account. Lee (2007) gives a discussion from a structural equation modeling perspective. Schafer (1997) gives a statistical discussion from a missing data and multiple imputation perspective, whereas Enders (2010) gives an applied discussion of these same topics. Statistical overview articles include Gelfand et al. (1990) and Casella and George (1992). Overview articles of an applied nature and with a latent variable focus include Scheines et al. (1999), Rupp et al. (2004), and Yuan and MacKinnon (2009).

Bayesian analysis is firmly established in mainstream statistics. Its popularity is growing and currently appears to be featured at least half as often as frequentist analysis. Part of the reason for the increased use of Bayesian analysis is the success of new computational algorithms referred to as Markov chain Monte Carlo (MCMC) methods. Outside of statistics, however, application of Bayesian analysis lags behind. One possible reason is that Bayesian analysis is perceived as difficult to do, requiring complex statistical specifications such as those used in the flexible, but technically-oriented general Bayes program WinBUGS. These observations were the background for developing Bayesian analysis in Mplus (Muthén & Muthén, 1998-2010). In Mplus, simple analysis specifications

with convenient defaults allow easy access to a rich set of analysis possibilities. Diffuse priors are used as the default with the possibility of specifying informative priors. A range of graphics options are available to easily provide information on estimates, convergence, and model fit.

Three key points motivate taking an interest in Bayesian analysis:

1. More can be learned about parameter estimates and model fit
2. Analyses can be made less computationally demanding
3. New types of models can be analyzed

Point 1 is illustrated by parameter estimates that do not have a normal distribution. ML gives a parameter estimate and its standard error and assumes that the distribution of the parameter estimate is normal based on asymptotic (large-sample) theory. In contrast, Bayes does not rely on large-sample theory and provides the whole distribution not assuming that it is normal. The ML confidence interval $Estimate \pm 1.96 \times SE$ assumes a symmetric distribution, whereas the Bayesian credibility interval based on the percentiles of the posterior allows for a strongly skewed distribution. Bayesian exploration of model fit can be done in a flexible way using Posterior predictive checking (PPC; see, e.g., Gelman et al., 2004, Chapter 6; Lee, 2007, Chapter 5; Scheines et al., 1999). Any suitable test statistics for the observed data can be compared to statistics based on simulated data obtained via draws of parameter values from the posterior distribution, avoiding statistical assumptions about the distribution of the test statistics. Examples of non-normal posteriors are presented in Section 2 for single-level models as well as in Section 4 for multilevel models. Examples of PPC are given in Section 3.

Point 2 may be of interest for an analyst who is hesitant to move from ML estimation to Bayesian estimation. Many models are computationally cumbersome or impossible using ML, such as with categorical outcomes and many latent variables resulting in many dimensions of numerical integration. Such an analyst may view the Bayesian analysis

simply as a computational tool for getting estimates that are analogous to what would have been obtained by ML had it been feasible. This is obtained with diffuse priors, in which case ML and Bayesian results are expected to be close in large samples (Browne & Draper, 2006; p. 505). Examples of this are presented in Section 5.

Point 3 is exemplified by models with a very large number of parameters or where ML does not provide a natural approach. Examples of the former include image analysis (see, e.g., Green, 1996)) and examples of the latter include random change-point analysis (see, e.g., Dominicus et al., 2008).

This paper gives a brief introduction to Bayesian analysis as implemented in Mplus. For a technical discussion of this implementation, see Asparouhov and Muthén (2010). Section 2 provides two mediation modeling examples which illustrate a non-normal posterior, how to use priors, and how to do a basic Bayes analysis in Mplus. Section 3 uses an SEM example to illustrate PPC. Section 4 uses two-level regression to illustrate how to test significance of a skewed random effect variance estimate and intraclass correlation. Section 5 uses multiple-indicator growth for binary items to illustrate high-dimensional analysis. Section 6 uses a two-part growth model to illustrate Bayes speed advantage over ML with many dimensions. Section 7 uses a mixture model to illustrate how to handle label switching. Section 8 discusses alternative approaches to missing data modeling. Data and Mplus scripts are available on the Mplus web site under Mplus Examples, Applications using Mplus.

2 Two mediation modeling examples

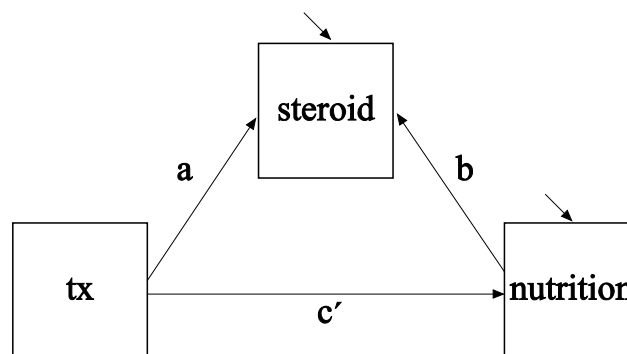
Two mediation modeling examples are considered. The first example uses the ATLAS data of MacKinnon et al. (2004) and illustrates how different conclusions about the intervention effect are arrived at using ML versus Bayes. The second example uses the firefighter data

of Yuan and MacKinnon (2009) to illustrate the use of priors based on information from previous studies to shorten the credibility interval (the Bayesian counterpart to confidence intervals) for the intervention effect.

2.1 The ATLAS example - different conclusions using Bayes vs ML

The mediational model in Figure 1 was considered in MacKinnon et al. (2004). The intervention program ATLAS (Adolescent Training and Learning to Avoid Steroids) was administered to high school football players to prevent use of anabolic steroids. MacKinnon et al. (2004) used a sample of $n = 861$ with complete data from 15 treatment schools and 16 control schools (the multilevel nature of the data was ignored and is ignored here as well; multilevel Bayesian mediational modeling is, however, available in Mplus). One part of the intervention aimed at increasing perceived severity of using steroids. This in turn was hypothesized to increase good nutrition behaviors. In Figure 1 these three variables are denoted tx , $steroid$, and $nutrition$, respectively.

Figure 1: Mediation model for the ATLAS example



A key parameter is the indirect effect of intervention on the nutrition outcome, $a \times b$.

Table 1: Input for Bayes analysis of the Atlas example

TITLE:	ATLAS, Step 1
DATA:	FILE = mbr2004atlas.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;

The ML point estimate (SE) for this is 0.020 (0.011) with an asymptotically-normal z test value of 1.913. Because this z value is not greater than 1.96, the indirect effect of the intervention is not deemed significant at the 5% level. Correspondingly, the ML 95% confidence interval obtained by the CINTERVAL option of the OUTPUT command is 0 – 0.041, that is, not excluding zero.

The Mplus input for the corresponding Bayesian analysis is shown in Table 1. The only change is to replace ESTIMATOR = ML with ESTIMATOR = BAYES. When two processors are used faster computations are obtained with the default of two MCMC chains to be discussed below. Note that the indirect effect is defined as the NEW parameter "indirect" in MODEL CONSTRAINT. Bayesian graphics are obtained with the option TYPE = PLOT2.

The analysis results are shown in Table 2. The first column gives the point estimate, which by default is the median of the posterior distribution. The mean or mode can be obtained using the POINT option of the ANALYSIS command. The second column gives

the standard deviation of the posterior distribution. A normally distributed z ratio is not used in Bayesian analysis. The third column gives a one-tailed p-value based on the posterior distribution. For a positive estimate, the p-value is the proportion of the posterior distribution that is below zero. For a negative estimate, the p-value is the proportion of the posterior distribution that is above zero. The fourth and fifth columns give the 2.5 and 97.5 percentiles in the posterior distribution, resulting in a 95% Bayesian credibility interval.

Using the default posterior median point estimate, the indirect effect estimate is 0.016, that is, slightly lower than the ML value with a slightly higher posterior distribution standard deviation of 0.013. Unlike the ML confidence interval, the Bayesian 95% credibility interval of 0.002 – 0.052 does not include zero, implying a positive intervention effect. The reason for this ML-Bayes discrepancy is found when studying the posterior distribution of the indirect effect.

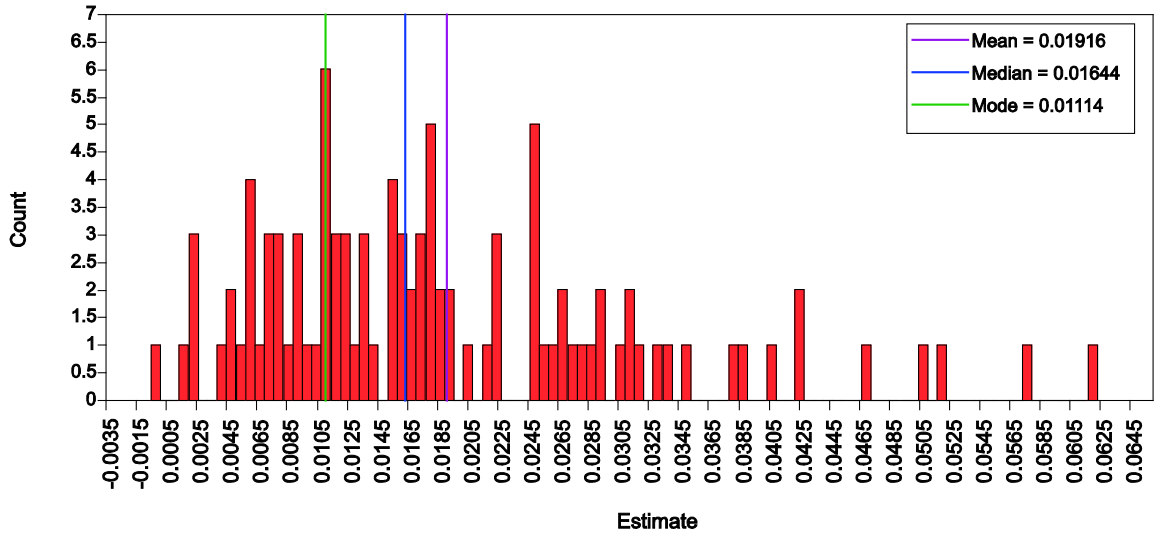
View Graphs is used to open the Bayes graphs. The first menu item is Bayesian posterior parameter distributions. Staying in histogram mode and selecting Parameter 8, indirect shows a skewed distribution for the indirect effect with a median of 0.01644; see Figure 2. A smoother picture of the posterior distribution is obtained by requesting the Kernel density (Botev et al., 2010) option instead of histogram; see Figure 3. The mean, median, and mode are marked as vertical lines in the distribution and are different due to the skewness of the distribution. It is clear that the ML normality assumption is not suitable for the indirect effect parameter (for technical arguments, see also MacKinnon, 2008). Therefore, the symmetric confidence interval that ML uses is not appropriate. Instead, the Bayesian credibility interval uses the 2.5 and 97.5 percentiles of the posterior distribution, allowing for skewness.

It is important to carefully consider convergence in Bayesian analysis. The default convergence criterion is that a Proportional Scale Reduction (PSR) factor is close enough

Table 2: Output excerpts for Bayes analysis of the 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	4.108
Residual Variances					
severity	1.722	0.072	0.000	1.614	1.868
nutrit	1.331	0.070	0.000	1.198	1.468
New/Additional Parameters					
indirect	0.016	0.013	0.010	0.002	0.052

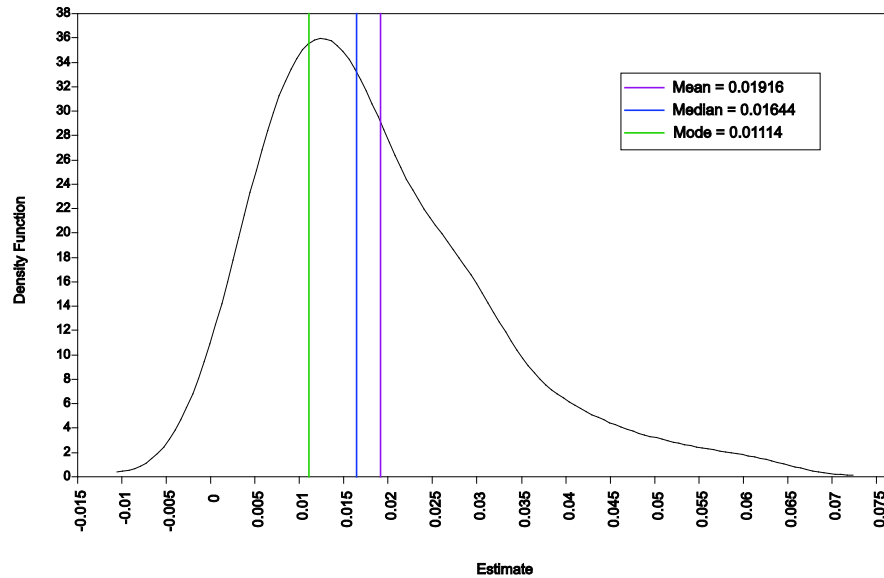
Figure 2: Indirect effect posterior distribution for the ATLAS example



to 1 for each parameter. For a technical definition, see Gelman and Rubin (1992), Gelman et al. (2004), and Asparouhov and Muthén (2010). Briefly stated, Bayesian analysis uses Markov chain Monte Carlo (MCMC) algorithms to iteratively obtain an approximation to the posterior distributions of the parameters from which the estimates are obtained as means, medians, or modes. Such iterations are referred to as a chain. In Mplus several such chains are carried out in parallel when using multiple processors. The PSR approach to determining convergence compares the parameter variation within each chain to that across chains to make sure that the different chains do not converge to different values. The PSR criterion essentially requires the between-chain variation to be small relative to the total of between- and within-chain variation. Mplus uses the default of two chains which usually gives good PSR information that compares well with using more chains. The first half of the iterations are considered as a "burn-in" phase and are not used to represent the posterior distribution.

Figure 4 shows a trace plot of the two chains for the residual variance of the *severity*

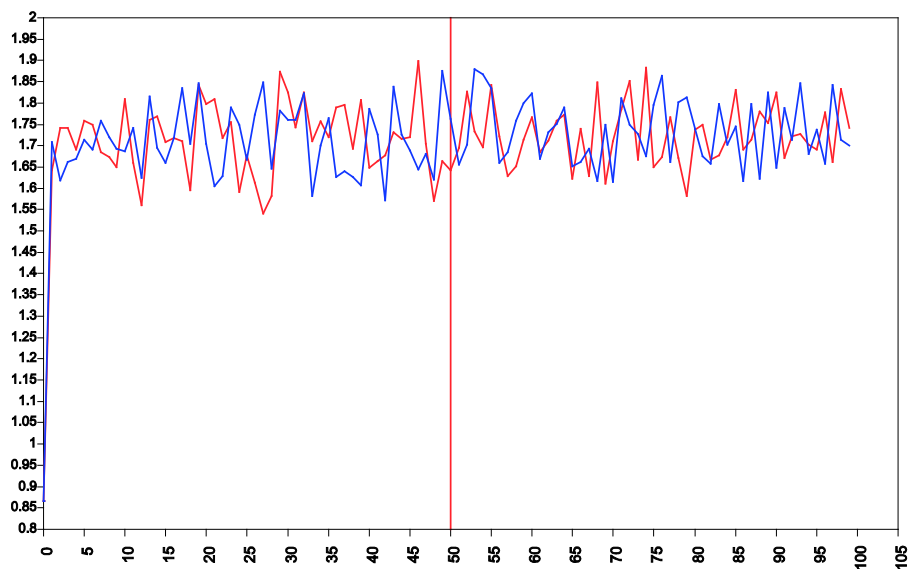
Figure 3: Kernel density estimate of indirect effect posterior distribution for the ATLAS example



variable. It is seen how a starting value is improved in a few iterations. A total of 100 iterations are used at which point the convergence criterion is fulfilled. The TECH8 output shown on the screen and printed at the end of the output shows that for this example the largest PSR value at 100 iterations is 1.037. In the trace plot the burn-in phase is denoted by a vertical line at 50 iterations. The last 50 iterations appear to show a stable process with no upward or downward trend with the two chains overlapping in their variation. The posterior distribution shown in Figure 2 is determined by the remaining 50 values for each of the two chains, resulting in a distribution based on 100 points.

A good approach to gain evidence of convergence is to run longer chains and check that the parameter values have not changed in important ways and that the PSR still remains close to 1. The Mplus option FBITERATIONS in the ANALYSIS command can be used to request a fixed number of Bayes iterations. Requesting 10,000 iterations gives the progression of PSR values in TECH8 seen in Table 3. Only the first 1000 iterations

Figure 4: Trace plot of severity residual variance for the ATLAS example



are shown.

When FBITERATIONS is used the PSR convergence criterion is not applied by Mplus. Although Mplus prints THE MODEL ESTIMATION TERMINATED NORMALLY convergence has to be verified by the user when using FBITERATIONS. The results are shown in Table 4. The estimates are rather close to those shown earlier for 100 iterations in Table 2.

Figure 5 shows the posterior distribution for the indirect effect. It is now smoother than in Figure 2 due to using longer chains, but still shows the skewness.

A further check of the posterior distributions is obtained by an autocorrelation plot. Figure 6 shows this for the indirect effect in the 10,000 iteration analysis. Autocorrelations show the degree of correlatedness of parameter values across iterations for different lags (intervals in the chain). A small value is desirable to obtain approximately independent draws from the posterior. A value of 0.1 or lower has been suggested. If the autocorrelation is high for small lags but decreases with increasing lags, using only every kth iteration can

Table 3: TECH8 iterations with PSR for the first 1000 iterations of the 10,000 iterations analysis of the ATLAS example

Iteration	Potential Scale Reduction	Parameter With Highest PSR
100	1.037	2
200	1.014	4
300	1.002	2
400	1.003	3
500	1.002	7
600	1.002	6
700	1.000	6
800	1.003	1
900	1.002	1
1000	1.002	1

Figure 5: Indirect effect posterior distribution from 10,000 iterations for the ATLAS example

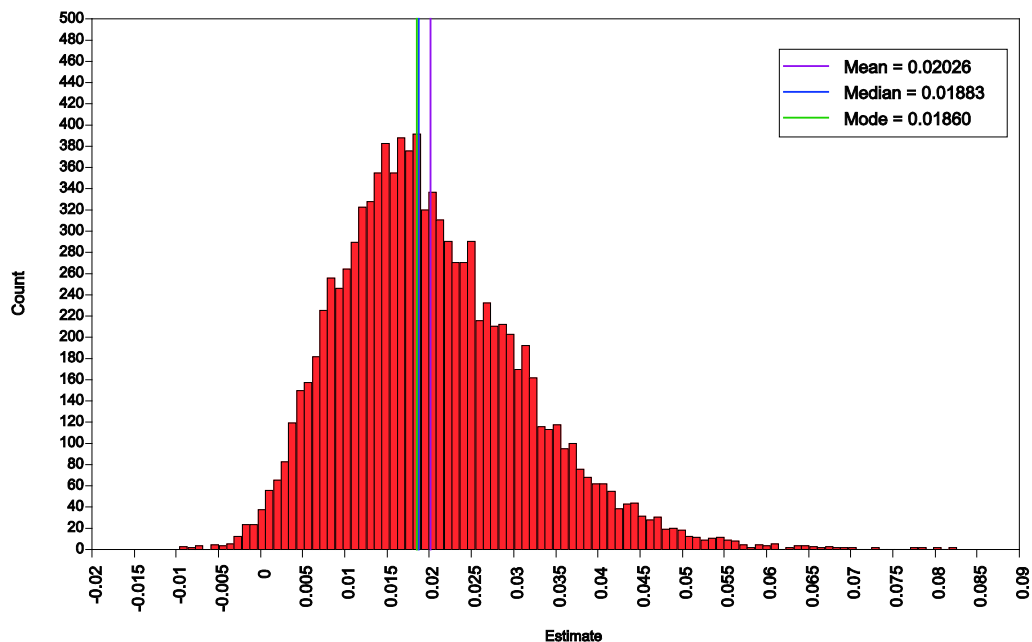
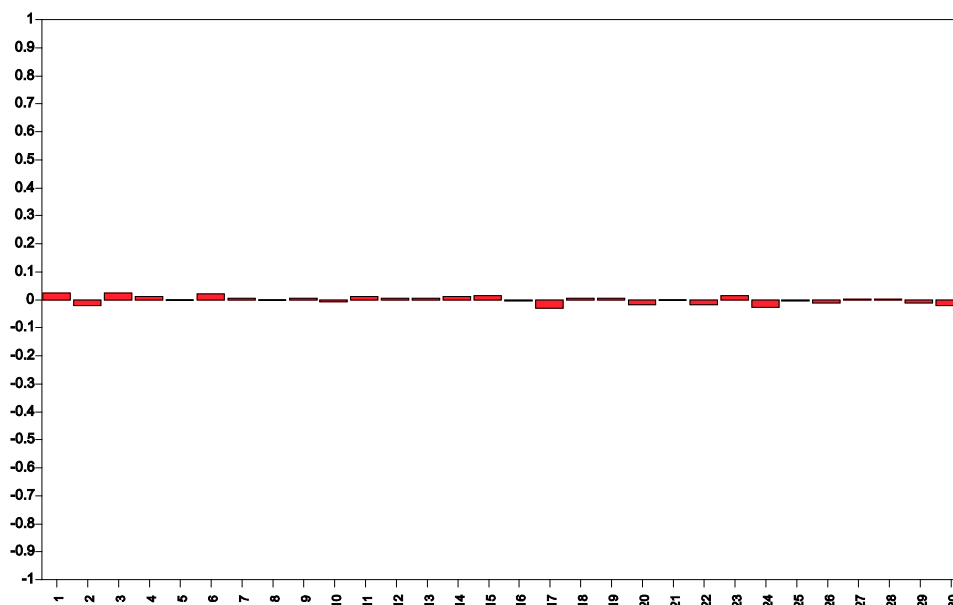


Table 4: Output excerpts from Bayesian analysis using 10,000 iterations for the ATLAS example

Parameter	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I.	
				Lower 2.5%	Upper 2.5%
severity ON					
group	0.272	0.089	0.001	0.098	0.448
nutrit ON					
severity	0.074	0.030	0.008	0.014	0.133
group	-0.018	0.080	0.408	-0.177	0.140
Intercepts					
severity	5.648	0.062	0.000	5.525	5.768
nutrit	3.663	0.177	0.000	3.313	4.014
Residual Variances					
severity	1.719	0.083	0.000	1.566	1.895
nutrit	1.333	0.065	0.000	1.215	1.467
New/Additional Parameters					
indirect	0.019	0.011	0.009	0.003	0.045

Figure 6: Autocorrelation plot for indirect effect from 10,000 iterations for the ATLAS example



be accomplished by thinning using the THIN option of the ANALYSIS command. Thinning is also useful when a large number of iterations is needed for convergence, but a smaller number is desired for displaying the posterior distributions to reduce computer storage (this influences the size of the Mplus .gph file).

2.2 The firefighter example - using informative priors

Yuan and MacKinnon (2004) discusses the benefits of Bayesian analysis for mediational analysis. Here it is shown how their firefighter example is carried out in Mplus. The example uses x to represent exposure to the randomized experiment, m to represent change in knowledge of the benefits of eating fruit and vegetables, and y to represent reported eating of fruits and vegetables. The model diagram is analogous to Figure 1. The sample size is $n = 354$. The focus is on the indirect effect $a \times b$. This example illustrates the use of priors. As a first step the analysis is done using the default of diffuse priors. The Mplus

Table 5: Input for Bayesian analysis of Firefighter example

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;
MODEL CONSTRAINT:	NEW(indirect); indirect = a*b;
OUTPUT:	TECH1 TECH8;
PLOT:	TYPE = PLOT2;

input is shown in Table 5 and the output in Table 6. Yuan and MacKinnon (2009) give the corresponding WinBUGS code in their appendix. Table 6 shows a positive indirect effect with 95% credibility interval 0.011 – 0.117.

The firefighter example is used to illustrate the benefit of informative priors. Yuan and MacKinnon (2009, p. 311) put priors on the a and b slopes based on previous studies to show how this can reduce the width of the credibility interval for the indirect effect of the intervention. A normal prior is used for both slopes. For a the prior has mean 0.35 and variance 0.04 and for b the prior has mean 0.1 and variance 0.01. The prior variances are four times larger than what had been observed in previous studies in order to take into account possible differences in the current study. The Mplus input is shown in Table 7. The MODEL PRIORS command uses the labels a and b to apply the normal priors.

Table 6: Output excerpts for Bayesian analysis of Firefighter example

Parameter	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I.	
				Lower 2.5%	Upper 2.5%
m ON					
x	0.395	0.121	0.000	0.160	0.634
y ON					
m	0.142	0.052	0.003	0.040	0.243
x	0.108	0.117	0.176	-0.127	0.339
Intercepts					
y	0.418	0.057	0.000	0.308	0.530
m	0.000	0.059	0.499	-0.115	0.116
Residual Variances					
y	1.144	0.089	0.000	0.987	1.338
m	1.218	0.093	0.000	1.054	1.419
New/Additional Parameters					
indirect	0.053	0.028	0.004	0.011	0.117

Table 7: Input for Bayesian analysis with priors for Firefighter example

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;
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;

The output given in Table 8 shows a 95% credibility interval of 0.012–0.102. Compared to the Table 6 analysis with the default diffuse priors, this represents a 16% shortening of the credibility interval. The results agree with those of Yuan and MacKinnon (2009). With a smaller sample, the prior has a stronger effect. Figure 7 shows that the indirect effect has a skewed posterior distribution.

Figure 7: Indirect effect posterior from 10,000 iterations for the Firefighter example

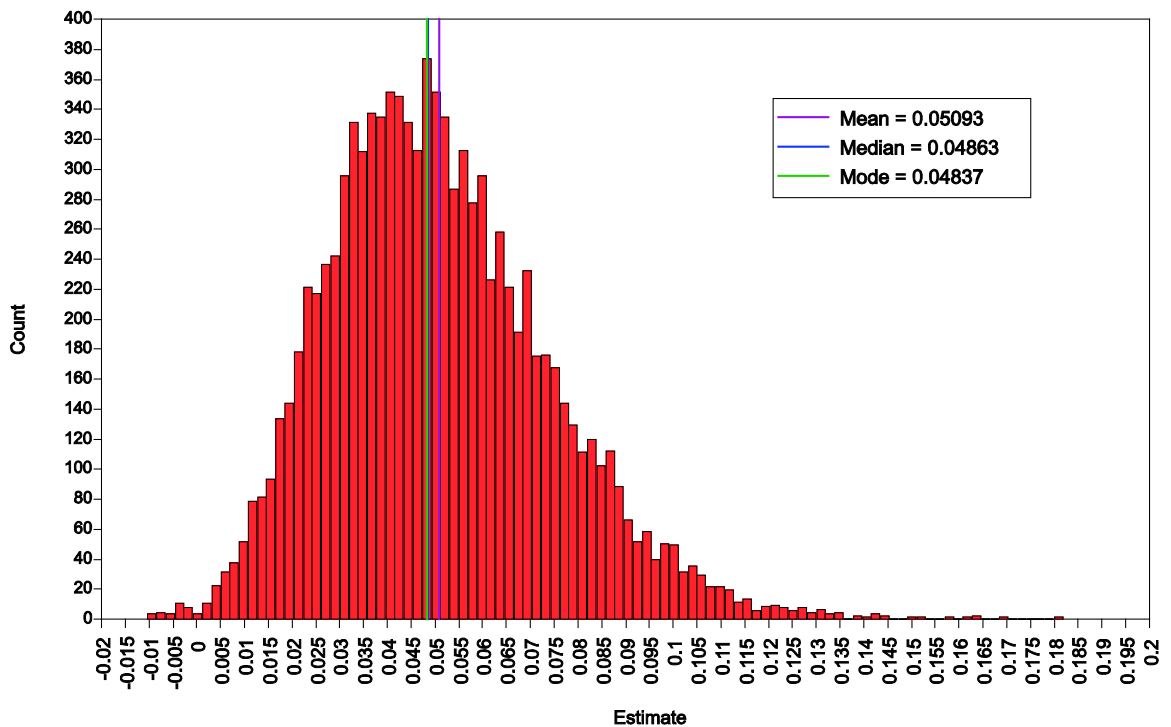


Table 8: Output excerpts for Bayesian analysis with priors for Firefighter example

Parameter	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I.	
				Lower 2.5%	Upper 2.5%
m ON					
x	0.383	0.104	0.000	0.182	0.588
y ON					
m	0.133	0.046	0.003	0.042	0.223
x	0.112	0.117	0.169	-0.124	0.341
Intercepts					
y	0.418	0.056	0.000	0.308	0.530
m	0.000	0.059	0.499	-0.115	0.116
Residual Variances					
y	1.143	0.089	0.000	0.986	1.338
m	1.218	0.093	0.000	1.053	1.418
New/Additional Parameters					
indirect	0.049	0.023	0.003	0.012	0.102

References

- [1] Asparouhov, T. & Muthén, B. (2010). Bayesian analysis using Mplus. Technical appendix. Los Angeles: Muthén & Muthén. www.statmodel.com
- [2] Botev, Z.I., Grotowski, J.F. & Kroese, D.P. (2010). Kernel density estimation via diffusion. Forthcoming in *Annals of Statistics*.
- [3] Browne, W.J. & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models.
- [4] Casella, G. & George, E.I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46, 167-174.
- [5] 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.
- [6] Enders, C.K. (2010). *Applied missing data analysis*. New York: Guilford Press.
- [7] 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.
- [8] Gelman, A. & Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-511.
- [9] Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (2004). *Bayesian data analysis*. Second edition. Boca Raton: Chapman & Hall.
- [10] 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.

- [11] Lee, S.Y. (2007). Structural equation modeling. A Bayesian approach. Chichester: John Wiley & Sons.
- [12] Little, R. J. & Rubin, D. B. (2002). *Statistical analysis with missing data*. Second edition. New York: John Wiley and Sons.
- [13] Lynch, S.M. (2010). Introduction to applied Bayesian statistics and estimation for social scientists. New York: Springer.
- [14] MacKinnon, D.P. (2008). Introduction to statistical mediation analysis. New York: Erlbaum.
- [15] MacKinnon, D.P., Lockwood, C.M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39, 99-128.
- [16] McLachlan, G. J. & Peel, D. (2000). *Finite mixture models*. New York: Wiley and Sons.
- [17] Muthén 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.
- [18] Muthén, B. & Muthén, L. (1998-20010). *Mplus User's Guide*. Sixth Edition. Los Angeles, CA: Muthén & Muthén.
- [19] 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.
- [20] 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.

- [21] Schafer, J.L. (1997). Analysis of incomplete multivariate data. London: Chapman & Hall.
- [22] Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, 64, 37-52.
- [23] Yuan, Y. & MacKinnon, D.P. (2009). Bayesian mediation analysis. *Psychological Methods*, 14, 301-322.