

A Short Note on Obtaining Item Parameter Estimates of IRT Models with Bayesian Estimation in Mplus

Sedat ŞEN * Allan COHEN ** Seock-ho KIM ***

Abstract

Parameter estimation of Item Response Theory (IRT) models can be applied using both Bayesian and non-Bayesian methods. Although maximum likelihood estimation (MLE), a non-Bayesian method, has predominated since the 1970s, there is an increasing use of Bayesian methods, due to their capability for estimating complex models and for their implementation in commercially available software. In view of the recent increase in the popularity of these methods, a comparison between model parameter estimates from the two types of methods would be useful for practitioners. In this study, we compare MLE and Bayesian estimation, two popular methods for obtaining parameter estimates for dichotomous IRT models, using the MLE and Bayes estimator options as implemented in the Mplus software package. Results indicated Bayesian and MLE estimates differed only slightly, clearly demonstrating the consistency between estimates from the two methods. Further, Bayes estimator option in Mplus can be a viable and relatively easy to use tool for calibrations of IRT models.

Key Words: Item response theory, dichotomous models, Bayesian estimation, Mplus.

INTRODUCTION

Item response theory (IRT) models have been used for testing over the last half-century (van der Linden & Hambleton, 2013). Parameter estimation is considered one of the important processes of IRT modeling. Estimates of IRT model parameters have typically been done using methods such as maximum likelihood estimation (MLE; Bock & Aitkin, 1981) and Markov chain Monte Carlo estimation (MCMC; Patz & Junker, 1999a). MLE methods are based on a frequentist approach, and MCMC is a Bayesian method. MLE-based estimation methods have been widely used in IRT modeling since the development of software such as BILOG (Zimowski, Muraki, Mislevy, & Bock, 2003), MULTILOG (Thissen, 1991) and PARSCALE (Muraki & Bock, 1996). Implementations of MCMC methods for estimation of IRT models began to be reported in the early 1990s (e.g., Albert, 1992; Albert & Chib, 1993). MLE generally requires large samples to produce reliable results (e.g., Asparouhov & Muthén, 2010a; Meuleman & Billiet, 2009), a condition not necessarily required by Bayesian methods.

In this regard, there are often advantages with Bayesian methods that can overcome some of the problems associated with MLE methods. Lee and Song (2004) note that Bayesian methods can provide asymptotically distribution-free estimates, as well as more accurate results with smaller samples with non-normal ability distribution (see also Gao, & Chen, 2005). Fox (2010) suggests that, unlike MLE, Bayesian methods enable the use of additional information for estimating model parameters in addition to providing smaller standard errors than those of marginal maximum likelihood estimate, when reasonable prior information is available. Thus, MCMC implementations, such as Metropolis-Hastings and Gibbs sampling, became increasingly popular for IRT modeling after 1990s, particularly for

* Assoc. Prof. Dr., Harran University, Faculty of Education, Educational Sciences, Şanlıurfa-Turkey, sedatsen@harran.edu.tr, ORCID ID: orcid.org/0000-0001-6962-4960

** Prof. Dr., University of Georgia, College of Education, Educational Psychology Department, Athens-Georgia-USA, acohen@uga.edu, ORCID ID: orcid.org/0000-0002-8776-9378

*** Prof. Dr., University of Georgia, College of Education, Educational Psychology Department, Athens-Georgia-USA, shkim@uga.edu, ORCID ID: orcid.org/0000-0002-2353-7826

To cite this article:

Şen, S., Cohen, A., & Kim, S.(2020). *A Short Note on Obtaining Item Parameter Estimates of IRT Models with Bayesian Estimation in Mplus*, 11(3), 266-282. doi: 10.21031/epod.693719.

Received: 24.02.2020

Accepted: 15.08.2020

complex models (Rupp, Dey & Zumbo, 2004). Albert (1992), and Albert and Chib (1993), for example, demonstrated the application of the Gibbs sampler for two-parameter normal-ogive models. More generally, Patz and Junker (1999a, 1999b) provided details of Metropolis-Hastings procedures within Gibbs for logistic IRT models. Baker (1998), Ghosh, Ghosh, Chen, and Agresti (2000), and Sheng (2010) provide applications of Gibbs sampling. In addition to these applications, Bayesian estimation has been reported for other IRT models, including the 4PL model (Culpepper, 2016), the multi-dimensional graded response IRT model (Kuo & Sheng, 2015), and the mixture IRT model (Bolt, Cohen, & Wollack, 2001).

MCMC estimation of IRT requires respondents' data and prior distributions for the model parameters. Gibbs sampling is one of the MCMC algorithms that can be used to estimate the parameters of IRT models. This method summarizes the joint posterior distribution of (θ, ξ) by simulation (Albert, 1992). After collecting the examinee responses in the data matrix \mathbf{Y} , suppose we have a vector (ω) including item (ξ) and person (θ) parameters. Implementation of Gibbs sampling starts with initial guesses on this vector. Let $\omega^{(0)} = (\omega_1^{(0)}, \dots, \omega_k^{(0)})$ denote initial values of ω . As explained by Wollack, Bolt, Cohen, and Lee (2002, p.340), "a single sample from the joint posterior distribution, $\pi(\omega|\mathbf{Y})$, is approximated, by sampling each parameter from its conditional posterior distribution." A cycle is completed after sampling the first set of parameters. The parameters of the previous cycle serve as conditional values for the next cycle. We need to iterate the cycles t times until convergence is achieved. Mean, mode, or median of each parameter's marginal posterior distribution can be used as the final estimates (Baker, 1998). Readers are referred to Albert (1992) and Baker (1998) for more details on Gibbs Sampling.

One characteristic of the implementation of MCMC algorithms is that they can be quite technically complex. Fortunately, implementations of Bayesian estimation algorithms are often available from authors of published articles (Curtis, 2010). Early applications of Bayesian IRT estimation were mainly implemented in WINBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2003). Thus, most of the Bayesian estimation programs have been written in the BUGS language (Curtis, 2010). Over the years, other software packages have also been developed for Bayesian estimation of IRT including JAGS (Plummer, 2003), STAN (Stan Development Team, 2016), and OPENBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2010), which is the more recent version of WinBUGS. These packages are designed primarily for Bayesian estimation. Bayesian algorithms have become available in several general purpose software packages including SAS (e.g., Proc MCMC; SAS Institute, 2017), S-PLUS, R packages including DPpackage (Jara, Hanson, Quintana, Mueller, & Rosner 2012), mlirt (Fox, 2007), MCMCpack (Martin, Quinn, & Park, 2011), pscl (Jackman et al., 2017) and MATLAB (The MathWorks Inc., 2010).

Mplus (Muthén, & Muthén, 1998-2019) recently implemented a Bayesian MCMC algorithm option for latent variable models (Asparouhov & Muthén, 2010). Mplus is a general latent variable modeling program that implements the estimation of several statistical model families, including structural equation models, latent class analysis, factor analysis, mixture models with both single- and multi-level data structures. Mplus also estimates a range of IRT models (Embretson, & Reise, 2000) including the one-, two- and three-parameter logistic models (1PL, 2PL, and 3PL; Birnbaum, 1968), the four-parameter logistic model (4PL; Barton & Lord, 1981), the partial credit model (PCM; Masters & Wright, 1997), and the generalized partial credit model (GPCM; Muraki, 1997). Multi-level and mixture extensions of these models are also possible within Mplus. A relatively small number of studies have reported results use of Mplus for Bayesian estimation of IRT models (e.g., Luo & Dimitrov, 2019). Luo and Dimitrov (2019) have shown how to obtain estimates for MCMC/EAP of the ability parameters when using Bayes estimator in Mplus. Their study has focused on ability parameters. However, studies investigating the Bayesian estimation of item parameters of IRT models using Mplus do not yet appear to have been reported. As the popularity of the Mplus software package increases, the user may want to learn how to create syntaxes to be able to estimate item parameters of IRT models using Mplus software package. A didactic paper on this issue would be very helpful for practitioners. To this end, the purpose of this study is to introduce Bayes estimator of Mplus and to demonstrate its application in obtaining item parameter estimates of dichotomous IRT models.

METHOD

Empirical Example

Estimation of dichotomous IRT models on a sample data set using the Bayesian estimation algorithm implemented in Mplus will be demonstrated in this part. In addition, Bayesian estimates were compared with ML based estimates as implemented in Mplus.

Data

Section 6 of the Law School Admission Test data set (LSAT6; Bock & Lieberman, 1970) was analyzed in this study. The LSAT6 data set consists of the item responses of 1,000 examinees to five multiple choice five-choice items on Figure Classification. Multiple choice items are coded as 0 and 1 for wrong and right answers, respectively. LSAT6 data set was preferred in this study as it has been studied in many IRT studies and it is available with most of the IRT software packages. This data set is known to be useful example of testing out IRT procedures and showing the use of Bayesian estimation.

IRT Models

Only Rasch, 1PL, and 2PL dichotomous models were considered in the present study as the Bayes estimator is not currently available in Mplus for other IRT models (e.g., 3PL, 4PL, and polytomous models). Under the 2PL IRT model, the probability of a correct response to an item can be given as follows:

$$P(X_{is} = 1 | \theta_s, \beta_i, \alpha_i) = \frac{\exp[\alpha_i(\theta_s - \beta_i)]}{1 + \exp[\alpha_i(\theta_s - \beta_i)]} \quad (1)$$

where θ_s is the person ability parameter for examinee s , and β_i and α_i are the difficulty and discrimination parameters, respectively, for item i . The 1PL model can be written from Equation 1 by setting the item discrimination parameters to a constant value (e.g., the average item discrimination). Similarly, the Rasch model equation may be written by setting the item discrimination parameters to 1. Given that conditional independence assumption, the conditional likelihood of a response pattern would be given as

$$L(u_{s1}, u_{s2}, \dots, u_{sI}) = \prod_{i=1}^I P_i(\theta_s)^{u_{si}} Q_i(\theta_s)^{1-u_{si}}, \quad (2)$$

where $Q_i(\theta_s) = 1 - P_i(\theta_s)$.

Estimation

Parameter estimates of the model in Equation 1 can be obtained with either MLE or Bayesian estimation. The MLE algorithm focuses on finding the ability-level estimates that maximize the log-likelihood function with iterative procedures like Newton-Raphson (Embretson & Reise, 2000). Bayesian estimation integrates the prior distribution into the likelihood function. In Mplus, Bayesian estimation is on the probit scale, and MLE is on the logistic scale. When using the probit model, the posterior distribution of θ_s can be defined as

$$p(\theta_s | Y) \propto L \times p(\theta_s) \quad (3)$$

where $p(\theta_s)$ is the prior. As can be seen in Equation 3 (right side), prior distributions required to define the posterior distributions. These are used with the specified model in determining the posterior. The central tendency statistics of the posterior distribution can be reported as the final estimates.

Bayesian Estimation with Mplus

Bayesian estimation in Mplus is not currently available for count, continuous-time survival, censored or nominal data, nor is Bayesian estimation currently available for the 3PL, graded response model, partial credit model, or the generalized partial credit model.

The ESTIMATOR option should be defined as “ESTIMATOR=BAYES;” to obtain Bayes estimates in Mplus. Other Bayesian related options in the ANALYSIS command include ALGORITHM, BCONVERGENCE, BITERATIONS, BSEED, CHAINS, FBITERATIONS, POINT, PREDICTOR, PRIOR, STVALUES, and THIN. Descriptions of these options can be found in Mplus User’s Guide (Muthén & Muthén, 1998-2019). The non-informative priors are the default for Bayes estimator in Mplus. However, users can specify informative priors such as inverse Gamma, Inverse Wishart, and Dirichlet (Muthén & Asparouhov, 2013).

Analyses

Three dichotomous IRT models (i.e., Rasch, 1PL, and 2PL models) were fit to the LSAT6 data set using MLE and Bayesian options in Mplus. For comparison, the MLE estimator was used as implemented in the Mplus software. Bayesian estimation with Mplus was applied with non-informative default priors $N(0, 5)$ (Muthén & Muthén, 1998-2019, p. 775) for item difficulty and discrimination parameters. A detailed list of the alternative prior distributions can be found in Mplus User’s Guide (Muthén & Muthén, 1998-2019, p. 775). The Mplus syntaxes for Bayesian estimation of the three models are presented in Figures 1 to 3, respectively. All model parameters were estimated from the posterior distribution. LSAT6 data have been previously analyzed with MCMC by Kim (2003). In that study, 5,000 iterations were used as burn-in based on Gelman and Rubin diagnostic information. In this study, the same number of iterations was used for burn-in stage. Thus, a total of 10,000 MCMC iterations were run; the first 5,000 iterations were used as burn-in. Posterior means of the sampled values for each parameter were taken as parameter estimates.

In the Mplus syntax for the Bayesian estimation of the Rasch model (see Figure 1), the TITLE command was used to describe the problem. The FILE option was added for the data set (FILE = LSAT6.txt;). As the LSAT6 data set consists of five dichotomous items, the “NAMES = item1-item5;” option was used to label these five items under the VARIABLE command. The CATEGORICAL option was used to specify these items as categorical (CATEGORICAL = item1-item5;). USEVAR option was used to specify to use all of the items in the analysis (USEVAR = item1-item5;). The ANALYSIS command, ESTIMATOR = BAYES; option was specified to obtain Bayesian estimates. The Bayes estimator in Mplus employs Gibbs sampling (line 10) with two processors (line 11) to run two parallel chains (line 12). FBITERATIONS option was used to set the number of iterations to be 10,000 (line 13). That is, the posterior is based on the last half (i.e., 5,000). STVAL=ML; option was used to specify starting value information as ML-based values. In the syntax, we specified the mean of the posterior distribution to be reported in the output (POINT=MEAN;). THIN indicates the number of iterations to be used for estimating the posterior. In this example, thinning was set to be 5 (see line 16), which means every 5th iteration is used for estimating the posterior. The MODEL command (see line 17) indicates a general factor f1 (see line 18) and use the command “by” to indicate the relationship between items and factor f1. As can be seen in Figure 1, f1 by item1@0.587544 item2-item5@0.587544; (line 18) was used to link five items (items1-item5) with factor f1. The @0.587544 part was used to fix item discrimination parameters at 1, which enable us to obtain a model based on Rasch framework. f1@1 option was used fix factor variance to one. In addition, the mean of f1 was fixed at zero by writing [f1@0;]. Specifications between lines 17 and 20 were used to obtain Rasch model. The PLOT command was also used to create plots after estimation. As mentioned in the Mplus User’s Guide, TYPE = PLOT3; can be used to request graphical displays of several results. STANDARDIZED option was used to obtain standardized solution, and TECH8 was added to request diagnostic information regarding model convergence in the OUTPUT section (line 24). TECH8 also shows the total number of iterations, including the discards. Mplus reports the Potential Scale Reduction (PSR) computed based on Gelman and Rubin’s convergence diagnostic (Gelman & Rubin, 1992) when included TECH8 under the OUTPUT command.

In addition, a posterior predictive checking (PPC) statistics and its p-value (PPP) can be obtained for model fit assessment in Mplus. This statistics is based on the usual χ^2 test, which is computed using the replicated and the observed data in MCMC iteration t . The χ^2 fit function difference between these two values is calculated using every 5th iteration. For an excellent-fitting model, 95% CI for the difference in χ^2 value should include zero (Wang & Wang, 2020). Poor fit is observed with low PPP values (e.g. <0.05).

1	TITLE: Bayesian Estimation of Rasch Model
2	DATA:
3	FILE = LSAT6.txt;
4	VARIABLE:
5	NAMES = item1-item5;
6	CATEGORICAL = item1-item5;
7	USEVAR = item1-item5;
8	ANALYSIS:
9	ESTIMATOR = BAYES;
10	ALGORITHM=GIBBS;
11	PROCESS=2;
12	CHAINS=2;
13	FBITERATIONS=10000;
14	STVAL=ML;
15	POINT=MEAN;
16	THIN=5;
17	MODEL:
18	f1 by item1@0.587544 item2-item5@0.587544; !1/1.702=0.587544
19	f1@1;
20	[f1@0];
21	PLOT:
22	TYPE = PLOT3;
23	OUTPUT:
24	STANDARDIZED TECH8;

Figure 1. Mplus Syntax for Bayesian Estimation of Rasch Model.

The Mplus syntax created for Bayesian estimation of 1PL model is presented in Figure 2. Since most parts of this syntax are similar to the syntax of the previous model, only the different parts are explained here. The main difference between the syntaxes in Figure 1 and Figure 2 is that the MODEL section was modified to be able to estimate 1PL model. The f1 by item1-item5* (loading); option (line 18) was used to fix all discrimination parameters to be equal, and the [item1\$1-item5\$1*] option (line 19) was used to freely estimate the item difficulty parameters.

```
1 TITLE: Bayesian Estimation of 1PL Model
2 DATA:
3 FILE = LSAT6.txt;
4 VARIABLE:
5 NAMES = item1-item5;
6 CATEGORICAL = item1-item5;
7 USEVAR = item1-item5;
8 ANALYSIS:
9 ESTIMATOR = BAYES;
10 ALGORITHM=GIBBS;
11 PROCESS=2;
12 CHAINS=2;
13 FBITERATIONS=10000;
14 STVAL=ML;
15 POINT=MEAN;
16 THIN=5;
17 MODEL:
18 f1 by item1-item5* (loading);
19 [item1$1-item5$1*]
20 [f1@0]; f1@1;
21 PLOT:
22 TYPE = PLOT3;
23 OUTPUT:
24 STANDARDIZED TECH8;
```

Figure 2. Mplus Syntax for Bayesian Estimation of 1PL IRT Model.

The Mplus syntax created for Bayesian estimation of 2PL model is presented in Figure 3. Since most parts of this syntax are similar to the syntax of the previous model, only the different parts are explained here. The most noticeable difference between the syntaxes in Figure 1 and Figure 3 is that the MODEL section was modified to be able to estimate 2PL model. Lines 18 and 19, `f1 by item1-item5;` and `[item1$1-item5$1*]` options were used to freely estimate the item discrimination and difficulty parameters, respectively.

```
1 TITLE: Bayesian Estimation of 2PL Model
2 DATA:
3 FILE = LSAT6.txt;
4 VARIABLE:
5 NAMES = item1-item5;
6 CATEGORICAL = item1-item5;
7 USEVAR = item1-item5;
8 ANALYSIS:
9 ESTIMATOR = BAYES;
10 ALGORITHM=GIBBS;
11 PROCESS=2;
12 CHAINS=2;
13 FBITERATIONS=10000;
14 STVAL=ML;
15 POINT=MEAN;
16 THIN=5;
17 MODEL:
18 f1 by item1-item5;
19 [item1$1-item5$1*]
20 [f1@0]; f1@1;
21 PLOT:
22 TYPE = PLOT3;
23 OUTPUT:
24 STANDARDIZED TECH8;
```

Figure 3. Mplus Syntax for Bayesian Estimation of 2PL IRT Model.

ML estimates of item parameters of three IRT models with LSAT6 data set were also obtained with Mplus software package for comparison. In order to transform the Mplus derived parameter estimates (i.e., location and thresholds) to typical IRT estimates (i.e., discrimination and difficulty), we followed the transformation formula provided by Asparouhov and Muthén (2016, p.6). Given that μ represents factor mean and ψ denotes the factor variance, item discrimination parameter (α_i) and the item difficulty parameter (β_i) can be calculated as below using the location (λ_i) and threshold (τ_i) parameters from Mplus output:

$$\alpha_i = \lambda_i \sqrt{\psi} \quad (4)$$

$$\beta_i = \frac{\tau_i - \lambda_i \mu}{\lambda_i \sqrt{\psi}} \quad (5)$$

Forero and Maydeu-Olivares (2009) use the normal and logistic versions of the IRT models to place parameter estimates and standard errors in the same metric (within .01 units). The constant D (i.e., 1.702) proposed by Haley (1952) was used in this study as well to convert normal ogive parameters to the logistic parameters.

RESULTS

Tables 1-3 contain approximate posterior means using the MCMC algorithm for the five item parameters for the three IRT models fit the LSAT6 data set. Bayesian estimation and the MLE estimates of item parameters for Rasch, 1PL, and 2PL models were compared. In order to rescale parameters obtained from the Bayesian estimation, all of the parameters were multiplied by 1.7, since the Bayesian estimation in Mplus uses the probit link function.

Traditional fit indices cannot be used when IRT models are estimated with Bayesian estimation. In this case, the convergence of the Markov chain should be checked using Bayesian diagnostic statistics. Mplus reports the potential scale reduction (PSR) for convergence assessment (Asparouhov & Muthén, 2010b). PSR values between 1 and 1.1 indicate good convergence (Wang & Wang, 2020). In this study, PSR values were found to be 1.024, 1.018, and 1.042 for Rasch, 1PL, and 2PL models, respectively. Under the MCMC estimation, the PPC statistics can be used to assess model fit. Mplus provides a χ^2 fit function difference between observed values and replicated estimates, a p-value (PPP), scatter plot and histogram. χ^2 fit function differences were found to be [-14.585, 22.300], [-18.232, 16.381], and [-17.844, 17.061] for Rasch, 1PL, and 2PL models, respectively. Associated p-values were 0.346, 0.545, and 0.505, respectively. All of the 95% CI's include zero, and associated p-values are high indicating that three IRT models fit the data very well. In addition, plots based posterior predictive checking (see Figures A3 and A6) also indicate that the fit of the model is reasonable. Table 1 lists item parameter estimates of the Rasch model from Mplus for the MLE and Bayes estimators. Mplus-ML columns present MLE estimates, and the last column displays Bayesian estimates. The item parameter estimates from Bayesian estimation in Mplus are not on the same scale as those from MLE. Therefore, to convert the item difficulty and item discrimination parameters to the logit scale, we applied Equations 4 and 5 to the Bayesian estimates. The results presented in Table 1 in the right column (headed Mplus-Bayes) are on the logit scale. As shown in Table 1, the transformed Bayesian estimates differed in the first and second decimal places compared to the MLE estimates. The Pearson correlation value between Bayes estimates and ML estimates was .999.

Table 1. Item Parameter Estimates of Rasch Model

Item parameter	Mplus-ML	Mplus-Bayes
β_1	-2.871	-2.791
β_2	-1.062	-1.076
β_3	-0.257	-0.260
β_4	-1.387	-1.396
β_5	-2.218	-2.194
α	1.000	1.000

Note. β =item difficulty; α =item discrimination.

Table 2 presents the item parameter estimations for 1PL IRT model from the MLE and Bayes estimators in Mplus. As shown in Table 2, the Bayesian estimates from Mplus yielded similar estimates to those obtained with the MLE estimator. As shown in Table 2, the transformed Bayesian estimates differed in the first and second decimal places compared to the MLE estimates. The Pearson correlation value between Bayes estimates and ML estimates was .999.

Table 2. Item Parameter Estimates of 1PL IRT Model

Item parameter	Mplus-ML	Mplus-Bayes
β_1	-3.606	-3.571
β_2	-1.319	-1.372
β_3	-0.317	-0.333
β_4	-1.726	-1.783
β_5	-2.773	-2.804
α	0.757	0.745

Note. β =item difficulty; α =item discrimination.

Table 3 presents the item parameter estimates of the 2PL IRT model in Mplus with the two estimators. As can be seen, the Bayes estimator in Mplus also yielded similar estimates to those obtained with MLE (see Table 3). For item discrimination parameters, the Pearson correlation value between Bayes estimates and ML estimates was .87. For item difficulty parameters, the Pearson correlation value between Bayes estimates and ML estimates was .999. As shown in Table 3, the transformed Bayesian estimates differed in the first and second decimal places compared to the MLE estimates.

Several plots are possible in the “View plots” submenu under the “Plots” menu of Mplus screen. A screen opens and shows several options that can be used to request graphical displays of several results. As an example, Bayesian related plots for item 1 estimated with the 1PL IRT model are presented in the Appendix (see Figures A1-A7).

Table 3. Item Parameter Estimates of the 2PL IRT Model

Item parameter	Mplus-ML	Mplus-Bayes
α_1	0.825	0.727
α_2	0.722	0.749
α_3	0.891	0.934
α_4	0.688	0.689
α_5	0.659	0.616
β_1	-3.360	-3.672
β_2	-1.371	-1.373
β_3	-0.280	-0.279
β_4	-1.868	-1.916
β_5	-3.117	-3.323

Note. β =item difficulty; α =item discrimination.

DISCUSSION and CONCLUSION

Dichotomous IRT models are typically estimated with both Bayesian and MLE estimation algorithms. Bayesian estimation of IRT models is sometimes preferable to MLE as MLE needs numerical integration, which can be slow or prohibitive depending on the numbers of dimensions of integration as a function of the numbers of latent variables. Tutorials do not yet appear to be presented in the psychometric literature for estimating dichotomous IRT models in Mplus using Bayesian estimation. In this paper, we provide a simplified step-by-step method for the estimation of dichotomous IRT models with Bayesian estimation.

Specifically, three dichotomous IRT models were analyzed using the five-item LSAT6 data set. Parameter estimates of these five items were compared for MLE and Bayesian estimations. Results suggested that there were some differences between item parameter estimates from MLE and Bayesian

estimation. This is consistent with the results of previous research that showed comparable estimation results for MLE and MCMC (e.g., Kieftenbeld & Natesan, 2012; Luo, 2018; Paek, Cui, Öztürk Gübeş, & Yang, 2018; Wollack, Bolt, Cohen, & Lee, 2002). For instance, Luo (2018) found very close estimates between MCMC and robust ML (MLR) estimation of 2PL testlet model in Mplus. Kieftenbeld and Natesan (2012) have found little difference between the estimates obtained from MML and MCMC estimation of the graded response model. Similarly, Wollack et al. (2002) have also demonstrated that item parameter estimates from MMLE and MCMC methods were very similar under the nominal response model.

Bayesian estimation with non-informative priors should give asymptotically similar estimates as MLE. As shown in this study, non-informative priors are the default for the Bayes estimator in Mplus, although several informative priors can be defined. The prior distribution is one of the most important aspects of Bayesian estimation. This is because prior distributions can substantially affect the posterior distribution especially for small sample sizes (Natesan, Nandakumar, Minka, & Rubright, 2016). Only non-informative priors were used in this study. However, it is possible to add informative or slightly informative priors to the estimation in Mplus software. Further studies may reveal the differential effect of using different priors in Mplus. The addition of the Bayesian estimation algorithm in Mplus makes Mplus a viable and useful software package for the estimation of dichotomous IRT models.

REFERENCES

- Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics, 17*(3), 251–269.
- Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association, 88*(422), 669–679.
- Asparouhov, T., & Muthén, B. (2010a). *Bayesian analysis of latent variable models using Mplus*. (Technical Report). Version 4. Retrieved from <http://www.statmodel.com/download/BayesAdvantages18.pdf>
- Asparouhov, T., & Muthén, B. (2010b). *Bayesian analysis using Mplus, version 4*. Technical report. Los Angeles: Muthén and Muthén. www.statmodel.com.
- Asparouhov, T., & Muthén, B. (2016). *IRT in Mplus* (Technical report). Los Angeles, CA: Muthén & Muthén.
- Baker, F. B. (1998). An investigation of the item parameter recovery characteristics of a Gibbs sampling procedure. *Applied Psychological Measurement, 22*(2), 153–169.
- Barton, M.A., & Lord, F. M. (1981). *An upper asymptote for the three-parameter logistic item-response model* (ETS RR- 81-20). Princeton, NJ: Educational Testing Service.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord and M.R. Novick, *Statistical theories of mental test scores* (pp. 395-479). Reading, Massachusetts: Addison-Wesley.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*(4), 443–459.
- Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika, 35*(2), 179–197.
- Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2001). A mixture item response model for multiple-choice data. *Journal of Educational and Behavioral Statistics, 26*(4), 381–409.
- Culpepper, S. A. (2016). Revisiting the 4-parameter item response model: Bayesian estimation and application. *Psychometrika, 81*(4), 1142–1163.
- Curtis, S. M. (2010). BUGS code for item response theory. *Journal of Statistical Software, 36*(1), 1–34.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Hillsdale, NJ: Erlbaum.
- Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods, 14*(3), 275–299.
- Fox, J. P. (2007). Multi-level IRT modeling in practice with the Package mlirt. *Journal of Statistical Software, 20*(5), 1–16.
- Fox, J. P. (2010). *Bayesian item response modeling: Theory and applications*. New York, NY: Springer.
- Gao, F. & Chen, L. (2005). Bayesian or non-Bayesian: A comparison study of item parameter estimation in the three-parameter logistic model. *Applied Measurement in Education, 18*(4), 351–380.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457–472.
- Ghosh, M., Ghosh, A., Chen, M. H., & Agresti, A. (2000). Non-informative priors for one-parameter item response models. *Journal of Statistical Planning and Inference, 88*(1), 99–115.

- Haley, D. C. (1952). *Estimation of the dosage mortality relationship when the dose is subject to error*. (Tech. Rep. No.15, Office of Naval Research Contract No. 25140, NR-342-022). Stanford, CA: Stanford University, Applied Mathematics and Statistics Laboratory.
- Jackman, S., Tahk, A., Zeileis, A., Maimone, C., Fearon, J., Meers, Z., ... & Imports, M. A. S. S. (2017). Package 'pscl'. See <http://github.com/atahk/pscl>.
- Jara, A., Hanson, T., Quintana, F. A., Mueller, P., & Rosner, G. (2012). DPpackage: Bayesian nonparametric modeling in R. R package version, 1-1.
- Kieftenbeld, V., & Natesan, P. (2012). Recovery of graded response model parameters: A comparison of marginal maximum likelihood and Markov chain Monte Carlo estimation. *Applied Psychological Measurement*, 36(5), 399–419.
- Kim, S. H. (2003). An investigation of Bayes estimation procedures for the two-parameter logistic model. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, & J. J. Meulman (Eds.), *New developments in psychometrics* (pp. 389-396). Tokyo: Springer.
- Kuo, T. C., & Sheng, Y. (2015). Bayesian estimation of a multi-unidimensional graded response IRT model. *Behaviormetrika*, 42(2), 79–94.
- Lee, S. Y., & Song, X. Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research*, 39(4), 653–686.
- Luo, Y. (2018). A short note on estimating the testlet model with different estimators in Mplus. *Educational and Psychological Measurement*, 78(3), 517–529.
- Luo, Y., & Dimitrov, D. M. (2019). A short note on obtaining point estimates of the IRT ability parameter with MCMC estimation in Mplus: How many plausible values are needed?. *Educational and Psychological Measurement*, 79(2), 272–287.
- Martin, A. D., Quinn, K. M., & Park, J. H. (2011). Mcmcpack: Markov chain Monte Carlo in R. *Journal of Statistical Software*, 42(9), 1–21.
- Masters, G. N., & Wright, B. D. (1997). The partial credit model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 101–121). New York, NY: Springer.
- Meuleman, B., & Billiet, J. (2009). A Monte Carlo sample size study: How many countries are needed for accurate multi-level SEM?. *Survey Research Methods*, 3(1), 45–58.
- Muraki, E. (1997). A generalized partial credit model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 153–164). New York, NY: Springer.
- Muraki, E., & Bock, R. D. (1996). PARSCALE: IRT based test scoring and item analysis for graded open-ended exercises and performance tasks (Version 3) [Computer software]. Chicago, IL: Scientific Software.
- Muthén, B., & Asparouhov, T. (2013). Item response modeling in Mplus: A multi-dimensional, multi-level, and multi-timepoint example. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of item response theory: Models, statistical tools, and applications* (Vol 1, pp. 527-539). Boca Raton, FL: Chapman & Hall/CRC Press.
- Muthén, L. K., & Muthén, B. O. (1998-2019). Mplus (version 8.3)[computer software]. Los Angeles, CA: Muthén & Muthén.
- Natesan, P., Nandakumar, R., Minka, T., & Rubright, J. D. (2016). Bayesian prior choice in IRT estimation using MCMC and variational Bayes. *Frontiers in Psychology*, 7, 1422.
- Paek, I., Cui, M., Öztürk Gübeş, N., & Yang, Y. (2018). Estimation of an IRT model by Mplus for dichotomously scored responses under different estimation methods. *Educational and Psychological Measurement*, 78(4), 569–588.
- Patz, R. J., & Junker, B. W. (1999a). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, 24(2), 146–178.
- Patz, R. J., & Junker, B. W. (1999b). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics*, 24(4), 342–366.
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In Proceedings of the 3rd international workshop on distributed statistical computing (Vol. 124, p. 125). Technische Universität Wien, Vienna, Austria.
- 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(3), 424–451.
- SAS Institute. (2017). *Base SAS 9.4 procedures guide: Statistical procedures*. SAS Institute.
- Sheng, Y. (2010). A sensitivity analysis of Gibbs sampling for 3PNO IRT models: Effects of prior specifications on parameter estimates. *Behaviormetrika*, 37(2), 87–110.
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). WinBUGS version 1.4 [Computer program]. Cambridge, UK: MRC Biostatistics Unit, Institute of Public Health.
- Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2010). *OpenBUGS Version 3.1.1 user manual*. Retrieved from <http://www.openbugs.net/Manuals/Manual.html>

- Stan Development Team. (2016). *Stan modeling language users guide and reference manual* (Version 2.12.0). Retrieved from <http://mc-stan.org/documentation/>
- The MathWorks Inc. (2010). MATLAB. [Computer software]. Natick, MA: The MathWorks Inc.
- Thissen, D. (1991). MULTILOG user's guide: Multiple, categorical item analysis and test scoring using item response theory (Version 6.0) [Software manual]. Chicago, IL: Scientific Software.
- van der Linden, W. J., & Hambleton, R. K. (Eds.). (2013). *Handbook of modern item response theory*. Springer Science & Business Media.
- Wang, J., & Wang, X. (2020). *Structural equation modeling: Applications using Mplus*. Hoboken, NJ: John Wiley & Sons.
- Wollack, J. A., Bolt, D. M., Cohen, A. S., & Lee, Y. S. (2002). Recovery of item parameters in the nominal response model: A comparison of marginal maximum likelihood estimation and Markov chain Monte Carlo estimation. *Applied Psychological Measurement*, 26(3), 339–352.
- Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). BILOG-MG [computer software]. Lincolnwood, IL: Scientific Software.

Mplus'ta Bayes Kestirimi ile IRT Modellerinin Madde Parametre Kestirimlerinin Elde Edilmesine İlişkin Kısa Bir Not

Giriş

Madde tepki kuramı (MTK) modelleri 1970'lerden beri çeşitli test uygulamalarında kullanılmaktadır (van der Linden ve Hambleton, 2013). MTK modellemesinde en önemli adımlardan birini parametre tahmini oluşturmaktadır. MTK modeline ait madde ve yetenek parametrelerinin tahmini tipik olarak maksimum olabilirlik tahmini (MLE; Bock & Aitkin, 1981) ve Markov zinciri Monte Carlo tahmini (MCMC; Patz ve Junker, 1999a) gibi yöntemler kullanılarak yapılmaktadır. MLE tabanlı tahmin yöntemleri, BILOG (Zimowski, Muraki, Mislevy ve Bock, 2003), MULTILOG (Thissen, 1991) ve PARSCALE (Muraki ve Bock, 1996) gibi yazılımların geliştirilmesinden bu yana MTK modellemesinde yaygın olarak kullanılmaktadır. MTK modellerinin tahmini için MCMC yöntemlerinin uygulamaları 1990'ların başında gösterilmeye başlanmıştır (ör. Albert, 1992; Albert ve Chib, 1993). Kararlı sonuçlar üretmek adına MLE yöntemi, genellikle Bayesci yöntemlerin gerektirdiği bir koşul olmayan, büyük örneklemelere ihtiyaç duyar (örneğin, Asparouhov ve Muthén, 2010; Meuleman ve Billiet, 2009). Bu bağlamda, Bayesci kestirim yöntemleri MLE yöntemleriyle ilişkili bazı sorunların üstesinden gelebilecek avantajlara sahiptir. Lee ve Song (2004) Bayesci kestirim yöntemlerinin asimptotik olarak dağılımsız serbest tahminler sağlayabileceğini ve normal olmayan yetenek dağılımına sahip daha küçük örneklemelerle daha doğru sonuçlar verebileceğini belirtmektedir. Fox (2010), MLE'den farklı olarak Bayesci kestirim yöntemlerinin, makul önsel (prior) bilgiler mevcut olduğunda marjinal maksimum olabilirlik tahmininden daha küçük standart hatalar sunmasının yanı sıra model parametrelerini tahmin etmek için ek bilgilerin kullanılmasını sağladığını ileri sürmektedir. Böylece, Metropolis-Hastings ve Gibbs örnekleme gibi MCMC uygulamaları, 1990'lardan sonra, özellikle nispeten karmaşık modeller için veya verilerin az olduğu ve asimptotik teoreminin tutulmasının pek mümkün olmadığı durumlarda MTK modellemesi için giderek daha popüler hale gelmiştir (Rupp, Dey ve Zumbo, 2004). Örneğin Albert (1992) ve Albert ve Chib (1993), iki parametrelilik normal-ogive modelleri için Gibbs örnekleme yönteminin uygulanabileceğini göstermişlerdir. Daha genel olarak, Patz ve Junker (1999a, 1999b) lojistik MTK modelleri için Gibbs içinde Metropolis-Hastings prosedürlerinin ayrıntılarını göstermiştir. Baker (1998), Ghosh, Ghosh, Chen ve Agresti (2000) ve Sheng (2010) çalışmaları Gibbs örnekleme uygulamaları sunmaktadır. Bu uygulamalara ek olarak, 4 parametrelilik MTK modeli (Culpepper, 2016), çok boyutlu derecelendirilmiş yanıt MTK modeli (Kuo ve Sheng, 2015) ve karma MTK modeli (Bolt, Cohen, ve Wollack, 2001) dahil olmak üzere diğer MTK modelleri için Bayesci kestirimi kullanılmıştır. MTK'nin MCMC tahmini, yanıtlar için bir olasılık modelinden, model parametreleri için önsel dağılımlardan ve muhtemelen hiperparametreler için önsel dağılımlardan oluşur. MCMC algoritmalarının uygulanmasının bir özelliği, teknik olarak oldukça karmaşık olmalarıdır. Neyse ki, Bayesci kestirim algoritmalarının uygulamaları genellikle yayınlanmış

makalelerin yazarlarından elde edilebilmektedir (Curtis, 2010). Bayesci MTK tahmininin ilk uygulamaları esas olarak WINBUGS yazılımında uygulanmıştır (Spiegelhalter, Thomas, Best ve Lunn, 2003). Bu nedenle, Bayesci tahmin programlarının çoğu BUGS dilinde yazılmıştır (Curtis, 2010). Yıllar içinde, JAGS (Plummer, 2003), STAN (Stan Geliştirme Ekibi, 2016) ve OPENBUGS (WinBUGS'ın yeni sürümü; Spiegelhalter, Thomas, Best ve Lunn, 2010) dahil olmak üzere MTK'nin Bayesci tahmini için başka yazılım paketleri de geliştirilmiştir. Bu paketler öncelikle Bayesci kestirim için tasarlanmıştır. Bayesci algoritmaları, MATLAB (The MathWorks Inc., 2010), SAS (ör. Proc MCMC; SAS Enstitüsü, 2017), S-PLUS, MCMCpack (Martin, Quinn ve Park, 2011), pscl (Jackman ve diğerleri, 2017), DPpackage (Jara, Hanson, Quintana, Mueller ve Rosner 2012), mlIRT (Fox, 2007) gibi R paketleri de dahil olmak üzere çeşitli genel amaçlı yazılım paketlerinde kullanılabilir hale gelmiştir.

Mplus (Muthén ve Muthén, 1998-2019) kısa bir süre önce örtük değişken modelleri için bir Bayesci MCMC algoritma seçeneği uygulamaya başlamıştır (Asparouhov ve Muthén, 2010). Mplus, yapısal eşitlik modelleri, örtük sınıf analizi, faktör analizi, hem tek hem de çok seviyeli veri yapılarına sahip karma modelleri de dâhil olmak üzere birçok istatistiksel model ailesinin tahminini uygulayan genel bir örtük değişken modelleme programıdır. Mplus ayrıca 1 parametrelili lojistik model (1PL), 2 parametrelili lojistik model (2PL), 3 parametrelili lojistik model (3PL; Birnbaum, 1968), 4 parametrelili lojistik model (4PL; Barton ve Lord, 1981), kısmi kredi modeli (PCM; Masters ve Wright, 1997) ve genelleştirilmiş kısmi kredi modeli (GPCM; Muraki, 1997) dahil olmak üzere bir dizi MTK modelini (Embretson ve Reise, 2000) tahmin edebilmektedir. Bu modellerin çok düzeyli ve karma uzantıları Mplus içerisinde de mümkündür. Nispeten az sayıda çalışma, MTK model parametrelerinin Bayesci kestirimi için Mplus kullanımı sonuçlarını raporlamıştır (örn. Luo ve Dimitrov, 2019). Mplus kullanarak MTK modellerinin madde parametrelerinin Bayesci tahminini araştıran çalışmalar henüz rapor edilmemiştir. Bu çalışmanın amacı Mplus kullanarak Bayesci kestirimine bir giriş sağlamak ve bunun iki kategorili (dikotomus) MTK modellerinin tahmininde uygulanmasını göstermektir.

Yöntem

Bu bölümde, Mplus'ta uygulanan Bayesci kestirim algoritmasını kullanarak ampirik bir veri seti üzerinde iki kategorili MTK modellerinin madde parametre tahminleri gösterilmektedir. Ayrıca, bu madde parametre tahminleri Mplus'ta uygulanan maksimum olabilirlik tahminininin elde edilen değerlerle karşılaştırılmaktadır. Bu çalışmada analiz edilen Hukuk Fakültesi Kabul Testi 6 (LSAT6; Bock ve Lieberman, 1970) veri seti 1000 adayın beş seçenekli beş maddeye verdiği cevaplardan oluşmaktadır. Çoktan seçmeli maddeler yanlış yanıt için 0 ve doğru yanıt için 1 olarak kodlanmıştır. Bu çalışmada, Bayes kestiricisi şu anda Mplus'ta 3PL, 4PL ve polytomous modeller gibi diğer MTK modelleri için mevcut olmadığından sadece Rasch, 1PL ve 2PL MTK modellerine odaklanılmıştır. Bu 3 modele ait yetenek ve madde parametresi tahminleri, MLE veya Bayes kestirimi ile elde edilebilir. MLE algoritması, Newton-Raphson (Embretson & Reise, 2000) gibi yinelemeli kestirimlerle olabilirlik fonksiyonunu en üst düzeye çıkararak yetenek ve madde düzeyi tahminlerini bulmaya odaklanır. Bayes kestiricisi ise, önsel dağılımı olabilirlik fonksiyonuna entegre eder. Mplus'ta Bayesci kestirimi probit ölçeğinde ve MLE kestirimi lojistik skalada gerçekleştirilmektedir. Bayesci kestirim yönteminde posterior dağılımın modu, medyanı veya ortalaması, merkezi eğilimin nihai tahmini olarak alınmaktadır. Mplus'ta Bayesci kestirimi şimdilik sayım, sürekli zamanlı yaşam kalım, sansürlü veya nominal veriler için mevcut değildir ayrıca 3PL, kademeli yanıt modeli, kısmi kredi modeli veya genelleştirilmiş kısmi kredi modeli için de Bayesci kestirimi mevcut değildir. Bu sebeple, bu çalışmada Rasch, 1PL ve 2PL modelleri Mplus'taki MLE ve Bayes seçenekleri kullanılarak LSAT6 veri setine uygulanmıştır. Karşılaştırma için Mplus yazılımında uygulandığı şekliyle MLE tahmincisi kullanılmıştır. Mplus ile Bayesci kestirimi bilgilendirici olmayan varsayılan önsellerle uygulanmıştır. Üç modelin Bayesci kestirimi için Mplus sözdizimleri sırasıyla Şekil 1 ile 3'te sunulmuştur. Tüm model parametreleri sonsal (posterior) dağılımdan hesaplanmıştır. LSAT6 verileri daha önce Kim (2003) tarafından MCMC ile analiz edilmiş ve Gelman ve Rubin tanılayıcı bilgilerine dayanarak burn-in (yanma) sayısı olarak 5000 iterasyon (yineleme) belirlenmiştir. Bu çalışmada da, yanma aşaması için aynı sayıda iterasyon kullanılmıştır. Böylece toplam 10000 MCMC yinelemesi gerçekleştirilmiş ve ilk 5000 yineleme burn-

in olarak kullanılmıştır. Her bir parametre için örneklenen değerlerin sonsal ortalamaları nihai tahminler olarak kullanılmıştır. LSAT6 veri setine sahip üç MTK modelinin madde parametrelerinin ML tahminleri de karşılaştırma için Mplus yazılım paketi ile elde edilmiştir. Mplus türevli parametre tahminlerini (yani konum [location] ve eşikler [thresholds]) tipik MTK tahminlerine (yani ayırt edicilik [discrimination] ve güçlük [difficulty] değerlerine) dönüştürmek için Asparouhov ve Muthén (2016, s.6) tarafından sunulan dönüşüm formülleri kullanılmıştır. Forero ve Maydeu-Olivares (2009), $D = 1.702$ ölçekleme sabitini kullanarak parametre tahminlerini ve standart hataları aynı metriğe (.01 birim içinde) yerleştirmek için MTK modellerinin normal ve lojistik varyantlarını kullanır (Haley, 1952). Bu çalışmada da D sabiti, kullanılan modelin normal ogive parametrelerini lojistik parametrelerin ölçeğine koymak için kullanılmıştır. Tablo 1-3, LSAT veri setine uyan Rasch, 1PL ve 2PL MTK modelleri için beş madde parametresi için MCMC algoritmasını kullanan yaklaşık sonsal (posterior) ortalamaları göstermektedir. Bulgular bölümünde, Rasch, 1PL ve 2PL modelleri için Bayesci kestirimi ile madde parametrelerinin MLE tahminleri karşılaştırılmaktadır. Bayesci kestiriminden elde edilen parametreleri yeniden ölçeklendirmek için Mplus'taki Bayesci kestirimi probit bağlantı fonksiyonunu kullandığından tüm parametreler 1.7 ile çarpılmıştır. Tablo 1-3'te MLE ve Bayesci tahminleri için Mplus'tan Rasch, 1PL ve 2PL modellerinin madde parametre tahminleri listelenmektedir. Mplus-ML sütunları MLE tahminlerini sunar ve son sütun Bayesci tahminlerini gösterir. Mplus'ta Bayesci kestiriminden alınan madde parametresi tahminleri, MLE ile aynı ölçekte değildir. Bu nedenle, madde güçlüğü ve madde ayırt edicilik parametrelerini logit ölçeğine dönüştürmek için Bayesci tahminlere dönüşüm formülleri uygulanmıştır. Tablo 1-3'te görülebileceği gibi, dönüştürülmüş Bayesci kestirimleri, birinci ve ikinci ondalık basamaklarda MLE tahminlerine göre farklılaşmıştır. Bayesci tahminleri ile ML tahminleri arasındaki Pearson korelasyon değeri .999 civarında bulunmuştur. MLE ve MCMC arasındaki en büyük farklılıklar 2PL MTK modeline ait parametrelerde bulunmuştur.

Sonuç ve Tartışma

İki kategorili MTK modelleri tipik olarak hem Bayesci hem de MLE tahmin algoritmaları ile tahmin edilir. MTK modellerinin Bayesci kestirimi bazen MLE'ye tercih edilir, çünkü MLE'nin örtük değişkenlerin sayısının bir fonksiyonu olan entegrasyon boyutlarının sayısına bağlı olarak yavaş veya engelleyici olabilen sayısal entegrasyona ihtiyacı vardır. İki kategorili MTK modellerinin Bayesci kestirimi kullanarak Mplus'ta nasıl tahmin edileceğini psikometri literatüründe gösteren didaktik bir çalışma yer almamaktadır. Bu çalışmada, Bayesci kestirimi ile iki kategorili MTK modellerinin tahmini için gerekli işlemler adımlar halinde sunulmaktadır. Spesifik olarak, beş maddelik LSAT6 veri seti kullanılarak iki kategorili üç MTK modeli analiz edilmiştir. Bu beş maddenin parametre tahminleri MLE ve Bayesci tahminleri için karşılaştırılmıştır. Sonuçlar, MLE ile Bayesci kestirimlerinden elde edilen madde parametre tahminleri arasında bazı farklılıklar olduğunu göstermiştir. Bu, MLE ve MCMC için karşılaştırılabilir tahmin sonuçları gösteren önceki araştırmaların sonuçlarıyla tutarlıdır (örneğin, Kieftenbeld & Natesan, 2012; Luo, 2018; Paek, Cui, Öztürk Gübeş ve Yang, 2018; Wollack, Bolt, Cohen ve Lee, 2002). Bilgilendirici olmayan (non-informative) önsellerle Bayesci kestirimi, MLE ile asimptotik olarak benzer tahminler vermelidir. Bu çalışmada gösterildiği gibi, Mplus'ta Bayesci kestirimi için varsayılan, bilgilendirici (informative) önseller kolayca belirtilmesine rağmen, bilgilendirici olmayan önseller kullanılmaktadır. Önsel dağılım, Bayesci tahminin en önemli kısımlarından biridir. Bunun nedeni, önsel dağılımların özellikle küçük örneklem büyüklükleri için sonsal dağılımı önemli ölçüde etkileyebilmesidir (Natesan, Nandakumar, Minka ve Rubright, 2016). Bu çalışmada sadece bilgilendirici olmayan önseller kullanılmıştır. Bununla birlikte, Mplus yazılımındaki tahminlere bilgilendirici veya biraz bilgilendirici önseller eklemek mümkündür. Mplus'a Bayesci kestirim algoritmasının eklenmesi, Mplus'ı iki kategorili MTK modellerinin kestirimi için uygun ve kullanışlı bir yazılım paketi haline getirmiştir.

Appendix. Mplus Plots for 1PL IRT Model with Bayes Estimator

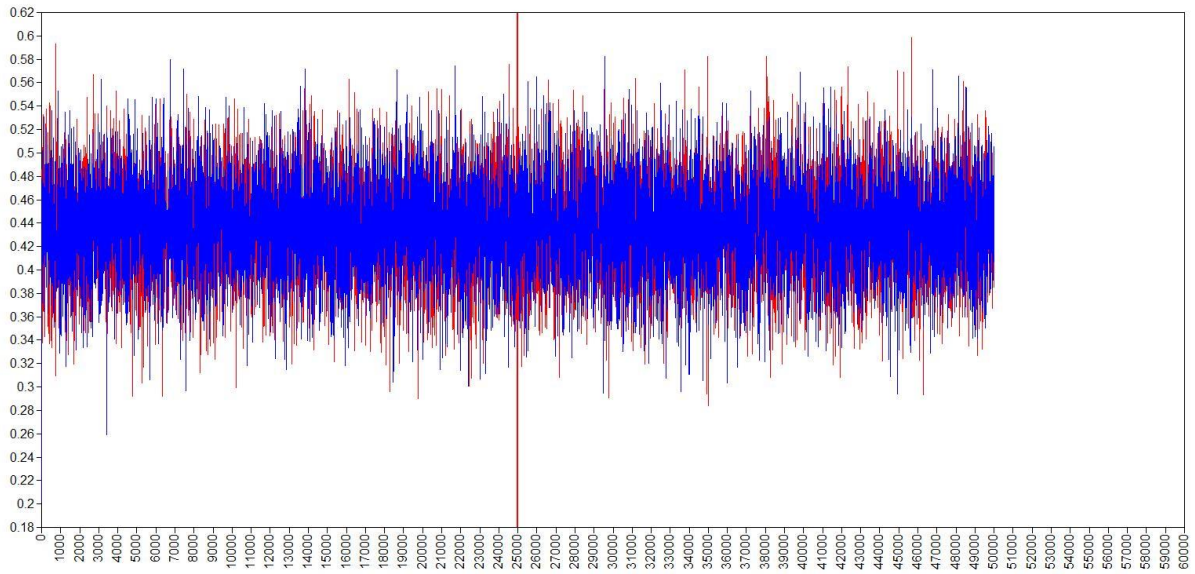


Figure A1. Bayesian posterior parameter trace plots.

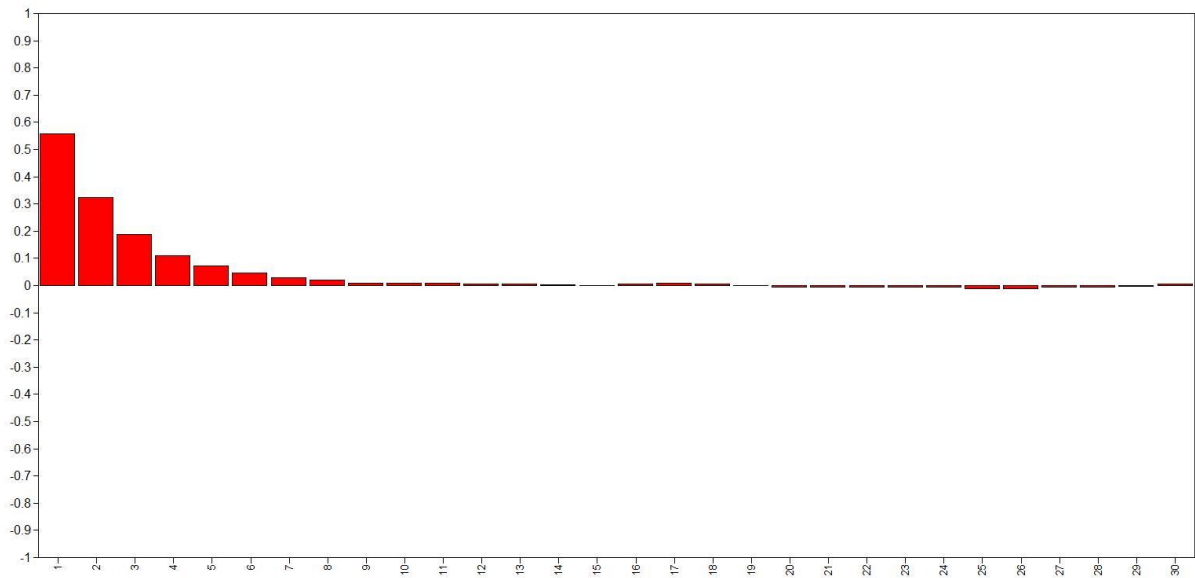


Figure A2. Bayesian autocorrelation plot.

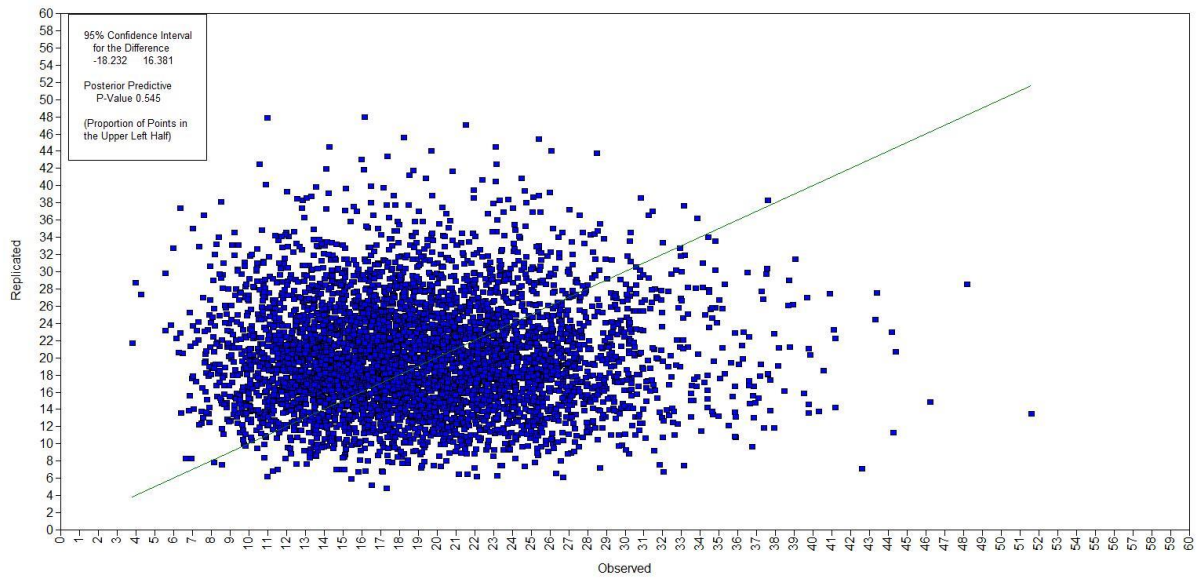


Figure A3. Bayesian posterior predictive checking scatterplots.

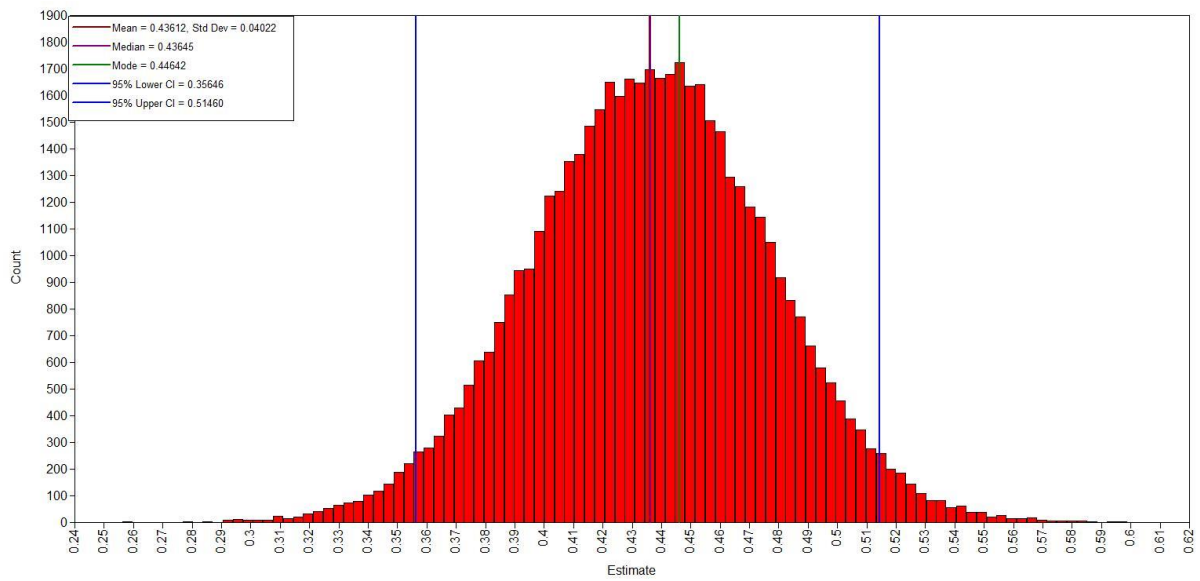


Figure A4. Bayesian posterior parameter distributions.

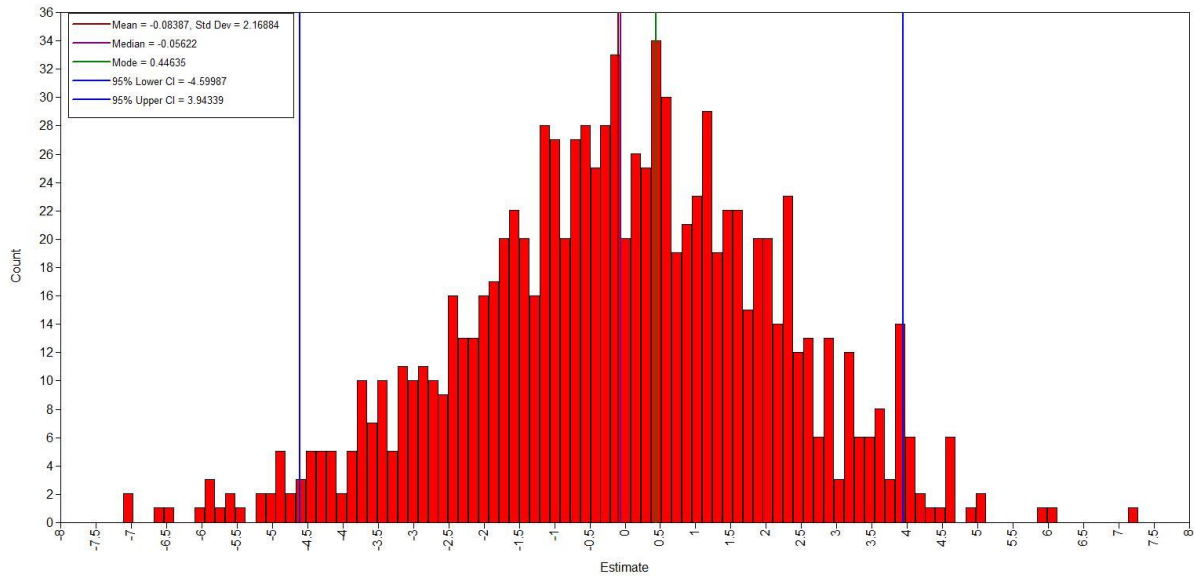


Figure A5. Bayesian prior parameter distributions.

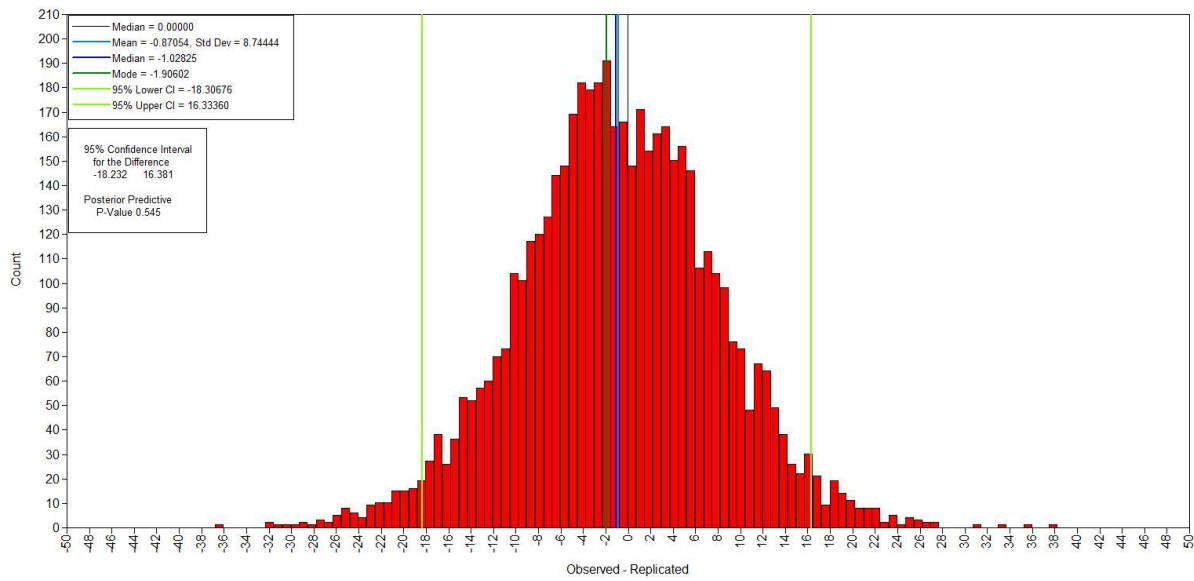


Figure A6. Bayesian posterior predictive checking distributions plots.

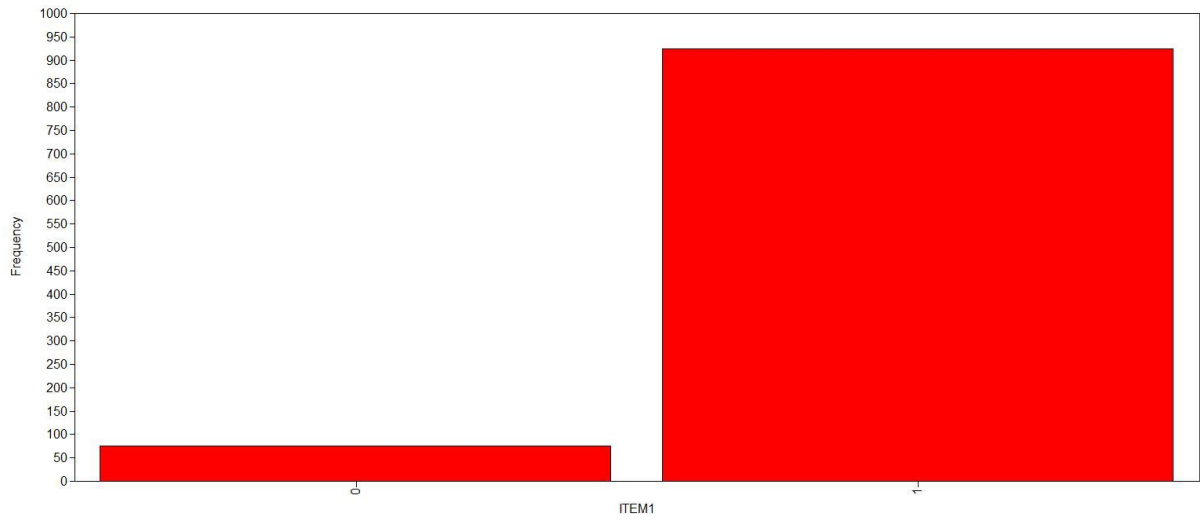


Figure A7. Histogram of sample values.