

# An Implementation of the Gibbs Sampling Method under the Rasch Model

Sedat ŞEN\* Tuğba KARADAVUT\*\* Hyo Jin EOM\*\*\*  
Allan S. COHEN\*\*\*\* Seock-Ho KIM\*\*\*\*\*

## Abstract

A brief explication of the implementation of the Gibbs sampling method via rejection sampling to obtain Bayesian estimates of difficulty and ability parameters under the Rasch model is presented. The Gibbs sampling method via rejection sampling was used in conjunction with the computer program OpenBUGS. Examples that compared the estimation method with another Gibbs sampling method via data augmentation as well as conditional, marginal, and joint maximum likelihood estimation methods are presented using empirical data sets. The effects of prior specifications on the difficulty and ability estimates are illustrated with the empirical data sets. A discussion is presented for related issues of Bayesian estimation in item response theory.

*Key Words:* Bayesian estimation, data augmentation, Gibbs sampling, rejection sampling, Rasch model.

## INTRODUCTION

For the one-parameter logistic Rasch model (Rasch, 1980) many estimation methods can be used to obtain item difficulty and person's ability parameter estimates (Fischer & Molenaar, 1995; Hoijtink & Boomsma, 1995; Molenaar, 1995). Difficulty and ability parameters can be estimated jointly by maximizing the joint likelihood function (i.e., JML; Wright & Stone, 1979). Conditional maximum likelihood (CML; Andersen, 1980) seems to be the standard estimation method under the one-parameter logistic model for estimation of difficulty parameters (e.g., Molenaar, 1995). Also, marginal maximum likelihood (MML) estimation using the expectation and maximization algorithm can be used to obtain difficulty parameter estimates (du Toit, 2003; Thissen, 1982). In addition, joint Bayesian estimation and marginal Bayesian estimation can be employed to obtain parameter estimates under the one-parameter logistic model (e.g., Birnbaum, 1969; Mislevy, 1986; Swaminathan & Gifford, 1982; see also Tsutakawa, & Lin, 1986).

Point estimates of the Rasch model difficulty and ability parameters are obtained in these earlier maximum likelihood estimation and Bayesian estimation methods by maximizing some forms of the likelihood function or of the posterior distribution. Instead of obtaining point estimates, procedures to approximate the posterior distribution under the Bayesian framework have been proposed relatively recently. One such method, Gibbs sampling approaches the estimation of item and ability parameters using the joint posterior distribution rather than the marginal distribution (e.g., Albert, 1992; Johnson & Albert, 1999; Kim, 2001; Patz & Junker, 1999). It can be noted that there are several different versions and implementations of Gibbs sampling that can be used to estimate item and ability parameters. Even so, all Bayesian estimation methods should yield comparable item and ability

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

\*\* Asst. Prof. Dr., Kilis 7 Aralık University, Faculty of Education, Educational Sciences, Kilis-Turkey, e-mail: tugba-mat@hotmail.com, ORCID ID: orcid.org/0000-0002-8738-7177

\*\*\* Research Prof. Dr. Korea University, Office of Research Management, Seoul, South Korea, e-mail: heom@korea.ac.kr, ORCID ID: orcid.org/0000-0003-0611-1994

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

parameter estimates, especially when comparable priors are used or when ignorance or locally-uniform priors are used. This paper was designed to investigate this issue using the one-parameter logistic Rasch model. Specifically, difficulty and ability parameter estimates from a Gibbs sampling method that used the rejection sampling (GS1) is examined and compared with another Gibbs sampling method that used data augmentation (GS2) as well as CML, MML, and JML. Because there exists Swaminathan and Gifford's (1982) seminal paper for Bayesian estimation under the Rasch model, GS1 is explained below with their framework instead of employing new notations. The main issue that differentiates GS1 in the current paper and the implementation used in Swaminathan and Gifford (1982) lies in the notion of the posterior maximization and approximation.

It should be noted that in item response theory Gibbs sampling and the more general Markov chain Monte Carlo methods are originally proposed to estimate parameters in rather complicated item response models for that the usual estimation methods may not be readily available. Although Gibbs sampling and the Markov chain Monte Carlo methods have been successfully applied to the modeling of complex response data in some studies (e.g., Bolt, Cohen, & Wollack, 2001, 2002; Cohen & Bolt, 2005; Karabatsos & Batchelder, 2003; Sen, Cohen, & Kim, 2018) and some specialized computer programs (e.g., Baker, 1998; Johnson & Albert, 1999; Wang, Bradlow, & Wainer, 2005) as well as a general computer program (Spiegelhalter, Thomas, Best, & Gilks, 1997a) have been available, only limited studies are available that investigated the characteristics of parameter estimates from Gibbs sampling or the Markov chain Monte Carlo methods for the traditional item response theory models including the Rasch model. Wollack, Bolt, Cohen, and Lee (2002), for example, investigated the recovery characteristics of Gibbs sampling for the nominal response model, and Baker (1998) investigated the recovery characteristics for the two-parameter logistic model. Kim (2001) reported results from a comparison study for the one-parameter logistic model in which a Gibbs sampling method was contrasted with other maximum likelihood estimation methods. Öztürk and Karabatsos (2017) discussed Gibbs sampling methods for estimating difficulty and ability parameters along with item response outlier detection parameters under the Rasch model. Levy (2009) presented an excellent review of the Markov chain Monte Carlo methods and Gibbs sampling for estimating item response theory models and the discussion of prior specifications for the Bayesian estimation. Interested readers should consult with Levy (2009) and references therein for the various computational methods under the Bayesian framework. Recently, Sheng (2010, 2017) investigated the use or specification of priors on the Markov chain Monte Carlo estimates under the three-parameter normal ogive model. Natesan, Nandakumar, Minka, and Rubright (2016) investigated the effects of priors on the Markov chain Monte Carlo and variational Bayes estimates for the one-, two-, and three-parameter logistic models.

Note that, despite the importance of the specification of priors in Bayesian estimation and the Gibbs sampling method, there is not much transparency regarding the selection and use of priors in the literature. This paper also illustrates the role of priors in the context of hierarchical Bayesian framework of Swaminathan and Gifford (1982) under the Rasch model.

In the subsequent sections, various implementations of the estimation methods for the Rasch model are briefly presented for the maximum likelihood methods and the Bayesian methods with a detailed explication of prior specifications. Results from a comparison study for the various estimation methods for the Rasch model are reported using empirical data from a published article. In order to assess the effects of prior specifications on the parameter estimates in GS1, results from a comparison study for employing various prior specifications are reported. Discussion for the general issues related Bayesian estimation in item response theory is followed.

## ***Implementations of Estimation Methods***

### ***Methods of Maximum Likelihood***

This paper employed proprietary computer programs for the maximum likelihood estimation of the difficulty and ability parameters. Specifically, WINMIRA (van Davier, 2001) was used for CML, IRTPRO (Cai, Thissen, & du Toit, 2010) was used for MML, and Winsteps (Linacre, 2003) was used

for JML. Technical treatments of these estimation methods can be found in several original articles contained as references in the computer program manuals. Baker and Kim (2004) also contains some accounts of the implementations of the respective methods.

A main reference for CML is Andersen (1980) (see also Andersen, 1970, 1972; Baker & Harwell, 1994). Earlier FORTRAN code of CML can be found in Fischer (1968) and Fischer and Allerup (1968). Thissen (1982) presented detailed accounts for theoretical background and the implementation of MML of difficulty parameters under the Rasch model. The explication of the two versions of Thissen's (1982) MML can be found in Baker and Kim (2004, pp. 397–411) with BASIC and Java code. Wright and his colleagues published many papers that presented implementations of JML (e.g., Wright & Panchapakesan, 1969). FORTRAN code for the earlier predecessors of Winsteps can be found in Wright and Mead (1978) and Wright, Mead, and Bell (1980) (cf. Wright, Linacre, & Schultz, 1989). Although not treated in this manuscript, it should be noted that there are other recent implementations of these earlier methods in R (Venables, Smith, & The R Development Core Team, 2009). Examples of R packages for item response theory modeling include ltm (Rizopoulos, 2006), eRm (Mair & Hatzinger, 2007), and mirt (Chalmers, 2012).

### *Bayesian Methods*

Swaminathan and Gifford (1982) presented Bayesian<sup>1</sup> estimation for the Rasch model. There are other papers that presented Bayesian estimation methods for more general item response theory models (e.g., Leonard & Novick, 1985; Mislevy, 1986; Swaminathan & Gifford, 1985, 1986; Swaminathan, Hambleton, Sireci, Xing, & Rizavi, 2003; Tsutakawa & Lin, 1986). As indicated earlier, nearly all Bayesian methods in item response theory that were implemented on the computer programs were used to obtain parameter estimates by maximizing some form of the posterior distribution.

Only recently, for example, Fox (2010), Stone and Zhu (2015), Levy and Mislevy (2016), and Luo and Jiao (2017) presented Bayesian estimation of item and ability parameters based on the techniques for the approximation of the posterior distribution, although Albert (1992) presented such a method some time ago. Kim and Bolt (2007) presented excellent instructional material for the Markov chain Monte Carlo methods to estimate parameters in item response theory models.

This paper is based on Swaminathan and Gifford's framework and presents its implementation on OpenBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2014). It deals with two different Bayesian estimation cases; (1) ability parameter estimation with known difficulty parameters and (2) difficulty and ability parameter estimation. The first case may provide a good foundational information for the second case. These two cases are presented below without employing detailed equations because nearly all of them can be found in Swaminathan and Gifford (1982).

### *Ability Estimation with Known Difficulty Parameters*

In Bayesian ability estimation with known difficulty parameters, the posterior distribution can be defined as

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}, \quad (1)$$

where  $p(x|\theta) \equiv l(\theta)$  is the likelihood function of the ability parameter  $\theta$  with item response data  $x$ ,  $p(\theta)$  is the prior distribution, and  $p(x) = \int p(x|\theta)p(\theta)d\theta$ . Following Lindley and Smith (1972) and

---

<sup>1</sup>It is not known to us that what will be the Reverend Thomas Bayes's (1701–1761) answer to the question of “Are you a Bayesian?” He was the first by the eponymy to solve the inverse problem of passage from the sample to population using ideas that are very popular today (Dodge, 2003, p. 29; Trader, 1997; cf. Stigler, 1980). Bayes's (1763) original paper was reprinted (see Bayes, 1958) with a biographical note by Barnard (1958). It should be noted that there is a list of eight errata for the original paper (Bayes, 1763) on the supposedly page 543 of the *Philosophical Transactions*, Vol. 53. Barnard's (1958) note didn't indicate that there is the errata page, and the reprint on *Biometrika*, Vol. 45 with modern notation did not include two of the errata.

Novick, Lewis, and Jackson (1973), Swaminathan and Gifford (1982) used a hierarchical prior,  $p(\theta) = \prod_i p(\theta_i | \mu, \phi) p(\mu, \phi)$ , where  $i$  designates each person,  $p(\mu, \phi) = p(\phi)$  for which  $p(\mu)$  has an improper uniform distribution and  $p(\phi)$  has the inverse chi-square distribution with parameters  $\nu$  and  $\lambda$  (i.e.,  $\phi \sim \chi^{-2}(\nu, \lambda)$ ; Novick & Jackson, 1974, pp. 190–194). The nuisance parameters  $\mu$  and  $\phi$  are integrated out of the posterior distribution and then the resulting proportional posterior distribution is maximized with the Newton-Raphson scheme to obtain point estimates of the ability parameters. With a fixed  $\mu$  value, the kernel of the resulting ability distribution is that of the multivariate  $t$  distribution (Anderson, 1984, pp. 272–273), and all ability parameters are estimated simultaneously in the Newton-Raphson scheme. The specification of the hyperparameters  $\nu$  and  $\lambda$  is a key issue in such hierarchical Bayesian estimation.

In conjunction with the Markov chain Monte Carlo method for approximating the entire posterior distribution and in the context of the computer program OpenBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2014) used in this study, it is better to use a proper yet noninformative uniform or normal hyperprior distribution for  $\mu$  in addition to employing an independent hyperprior distribution for  $\phi$ . The specification of the hyperparameters for the hyperprior distributions seems to be a very important issue. A noninformative, diffuse hyperprior distribution can be used for  $\mu$  by specifying appropriate hyperparameters, and an informative hyperprior distribution can be used for  $\phi$  by specifying appropriate hyperparameters.

One problem frequently encountered when specifying the distributional characteristics is that there are too many different definitions of the specific distributions in Bayesian literature (cf. Segal's law; Block, 1977, p. 79). Because this paper is based on Swaminathan and Gifford's notation but uses OpenBUGS to obtain posterior distributional statistics in GS1, it is imperative to connect seemingly the same yet different notations from different sources. An illustration below is for the inverse chi-square distribution and the gamma distribution in essence.

Swaminathan and Gifford (1982, p. 178) used the scaled inverse chi-square distribution for  $\phi$ :

$$p(\phi | \nu, \lambda) \propto \frac{1}{\phi} \frac{1}{2^{\nu+1}} \exp\left[-\frac{\lambda}{2\phi}\right], \quad 0 < \phi < \infty, \quad \lambda > 0, \quad \nu > 0 \quad (2)$$

(see Novick & Jackson, 1974, pp. 190–194; Isaacs, Christ, Novick, & Jackson, 1974, 175–196). Hence  $\phi \sim \chi^{-2}(\nu, \lambda)$  and  $\phi^{-1} \sim \chi^2(\nu, \lambda^{-1}) = \chi^2(\nu, \omega)$ , where  $W = \phi^{-1}$  variable has a scaled chi-square density,

$$p(W | \nu, \omega) \propto \frac{W^{(\nu/2)-1}}{\omega^{\nu/2}} \exp\left[-\frac{W}{2\omega}\right], \quad W > 0, \quad \nu > 0, \quad \omega > 0 \quad (3)$$

(see Novick & Jackson, 1974, pp. 186–190). It is not good that functions are shown with proportionality because the exact density of the distribution is not explicit.

In terms of the exact density of the scaled inverse chi-square without employing proportionality (see e.g., Gelman, Carlin, Stern, & Rubin, 1995, pp. 474–475 with their  $\theta = \phi$  and  $\nu s^2 = \lambda$  of Novick & Jackson, 1974, p. 191),

$$p(\phi | \nu, \lambda) = \frac{(\lambda/2)^{\nu/2}}{\Gamma(\nu/2)} \frac{1}{\phi} \frac{1}{2^{\nu+1}} \exp\left[-\frac{\lambda}{2\phi}\right], \quad (4)$$

where  $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$  is a gamma function (Davis, 1964, p. 255). Note that this distribution is Berger's (1985, p. 561) inverse gamma density,  $IG(\alpha, \beta)$ , where  $\alpha = \nu/2$  and  $\beta = 2/\lambda$  (n.b., this  $\beta$  is not the difficulty parameter).

In prior specification, a different but better form of the distribution can be used. If  $v\lambda_l/\phi \sim \chi^2(v)$  (Lindley, 1965, p. 26; Leonard & Hsu, 1999, p. 214; subscript  $l$  designates  $\lambda$  from Lindley and Leonard & Hsu), then

$$p(\phi|v, \lambda_l) = \frac{(v\lambda_l/2)^{v/2}}{\Gamma(v/2)} \frac{1}{\phi} \exp\left[-\frac{v\lambda_l}{2\phi}\right] \phi^{-v+1} \quad (5)$$

where  $v$  is the prior sample size and  $\lambda_l^{-1}$  is the prior mean of  $\phi^{-1}$  with the prior mean of  $\phi$  to be  $v\lambda_l/(v-2)$  for  $v > 2$ . In terms of Berger's  $IG(\alpha, \beta)$ , the corresponding parameters should be  $\alpha = v/2$  and  $\beta = 2/(v\lambda_l)$ . In terms of Swaminathan and Gifford's (1982, p. 178)  $\chi^{-2}(v, \lambda)$ ,  $v = v$  and  $\lambda = v\lambda_l$  of Lindley (1965, p. 26), yielding the prior sample size is  $v$ , the prior mean of  $\phi^{-1}$  is  $v/\lambda$ , and the prior mean of  $\phi$  is  $\lambda/(v-2)$  for  $v > 2$ .

These distributions may not be directly used in available computer software. In OpenBUGS, WinBUGS, as well as BUGS (e.g., Lunn, Jackson, Best, Thomas, & Spiegelhalter, 2013, pp. 345–346),  $\phi \sim \text{dgamma}(a, b)$  denotes the density is

$$p(\phi|a, b) = b^a \phi^{a-1} e^{-b\phi} / \Gamma(a) \quad \text{for } \phi > 0, a, b > 0 \quad (6)$$

with mean  $a/b$  and variance  $a/b^2$ . In Berger's (1985, p. 560) gamma density,  $G(\alpha, \beta)$ , the parameters are  $\alpha = a$  and  $\beta = 1/b$  with mean  $\alpha\beta$  and variance  $\alpha\beta^2$ . Note that  $\phi \sim IG(v/2, 2/\lambda)$  means

$\phi^{-1} \sim G(v/2, 2/\lambda) = \text{dgamma}(v/2, 2/\lambda)$  in OpenBUGS with  $v = 2a$  to be the prior sample size,  $v/\lambda = a/b$  to be the prior mean of  $\phi^{-1}$ , and  $\lambda/(v-2) = b/(a-1)$  to be the prior mean of  $\phi$  for  $v = 2a > 2$ .

### Estimation of Both Difficulty and Ability Parameters

The posterior distribution in this case can be defined as

$$p(\theta, \beta|x) = \frac{p(x|\theta, \beta)p(\theta, \beta)}{p(x)} \quad (7)$$

where  $p(x|\theta, \beta) = l(\theta, \beta)$  is the likelihood function of the ability parameter  $\theta$  and the difficulty parameter  $\beta$  with item response data  $x$ ,  $p(\theta, \beta)$  is the prior distribution, and  $p(x) = \int p(x|\theta, \beta)p(\theta, \beta)d(\theta, \beta)$ . Again, following Lindley and Smith (1972) and Novick, Lewis, and Jackson (1973), Swaminathan and Gifford (1982) used independent hierarchical priors,  $p(\theta, \beta) = p(\theta)p(\beta) = \prod_i p(\theta_i|\mu_\theta, \phi_\theta)p(\mu_\theta, \phi_\theta) \times \prod_j p(\beta_j|\mu_\beta, \phi_\beta)p(\mu_\beta, \phi_\beta)$ , where  $i$  designates each person and  $j$  designates each item,  $p(\mu_\theta, \phi_\theta) = p(\phi_\theta)$  and  $p(\mu_\beta, \phi_\beta) = p(\phi_\beta)$  for which  $p(\mu_\theta)$  and  $p(\mu_\beta)$  have improper uniform distributions and  $p(\phi_\theta)$  and  $p(\phi_\beta)$  have the inverse chi-square distributions with parameters  $v_\theta, \lambda_\theta, v_\beta, \lambda_\beta$ , respectively (i.e.,  $\phi_\theta \sim \chi^{-2}(v_\theta, \lambda_\theta)$  and  $\phi_\beta \sim \chi^{-2}(v_\beta, \lambda_\beta)$ ). Again, the nuisance parameters  $\mu_\theta, \phi_\theta, \mu_\beta, \phi_\beta$  are integrated out of the posterior distribution and then the resulting proportional posterior distribution is maximized with the Newton-Raphson scheme to obtain point estimates of the ability and item parameters. An iterative Birnbaum paradigm is used to obtain a set of ability estimates and then a set of difficulty parameter estimates until the overall convergence criterion can be met (Swaminathan & Gifford, 1982, p. 184).

The specification of the hyperparameters (i.e.,  $v_\theta, \lambda_\theta, v_\beta, \lambda_\beta$ ) is a key issue in hierarchical Bayesian estimation. In conjunction with the Markov chain Monte Carlo method for approximating the entire



posterior distribution and in the context of the computer program OpenBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2014) used in this study, it is better to use a proper yet noninformative uniform or normal hyperprior distribution for  $\mu_\theta$  or  $\mu_\beta$  in addition to employ an independent hyperprior for  $\phi_\theta$  or  $\phi_\beta$ . A noninformative, diffuse hyperprior distribution can be used for each  $\mu$  by specifying appropriate hyperparameters, and an informative hyperprior distribution can be used for each  $\phi$  by specifying appropriate hyperparameters.

## METHOD

Without loss of generality, we present below a comparison study for estimation of both difficulty and ability parameters under Rasch model. Ability estimation can also be done by modifying the programs in a trivial manner and hence not presented.

To compare GS1, GS2, CML, MML, and JML, illustrations using (1) the Law School Admission Test-Section 6 (LSAT6; Bock & Aitkin, 1981; Bock & Lieberman, 1970) data and (2) the Law School Admission Test-Section 7 (LSAT7) are presented below. It should be noted that the LSAT6 and LSAT7 data have been analyzed in many published articles and books (e.g., Andersen, 1980; McDonald, 1999). Use of these data instead of employing simulation data, hence, may provide a familiar baseline to make comparisons of different estimation methods.

GS1 estimates were obtained using OpenBUGS. GS2 estimates were obtained using MATLAB (The MathWorks, 1996) employing the code from Johnson and Albert (1999). Instead of OpenBUGS, WinBUGS or BUGS (e.g., Spiegelhalter et al., 1997a) can also be used. Difficulty parameter estimates are reported first and ability parameter estimates are subsequently reported for LSAT6 and LSAT7, respectively. It is not necessary to show the listings of the input lines of CML, MML, and JML. Also for GS2, the MATLAB function presented in Johnson and Albert (1999, p. 248) was used without any modification. However, it is necessary to present the input lines for OpenBUGS. The portions of the input lines are contained in Appendix. Note that in Appendix the inverse of the hyperparameter variance was specified with  $d\gamma$  ( $a=2.5$ ,  $b=5$ ) for both ability and difficulty prior distributions. This prior specification is equivalent to Swaminathan and Gifford's (1982)  $v=5$  and  $\lambda=10$ . Also note that the centered value of the log odds of the classical item facilities denoted as  $p_j$  (i.e., values of  $\log[(1-p_j)/p_j]$  centered at 0) were used for the initial values for difficulty parameters. Similar initial values were specified for the ability parameters.

Based on the suggestions from Kim and Bolt (2007) and Kim (2001), burn-in was set to 1000 and the next 10,000 iterations were used for GS1 to construct the posterior distributions that showed convergence of the simulated draws (see Gilks, Richardson, & Spiegelhalter, 1996). The convergence of the chains was visually monitored by checking history and autocorrelation plots. It should be noted that there are many different ways to summarize the sampled values in GS1 or GS2. Instead of using the actual posterior credibility interval, the posterior means and the posterior standard deviations are used in this study. The marginal posterior densities of the samples values for respective parameters all followed unimodal and likely normal distributions in GS1. GS2 also yielded similar results for the sampled values.

## RESULTS

### *Comparison of Estimation Methods*

#### *LSAT6 Estimation Results*

For the LSAT6 data that contained responses of 1000 subjects to five items, all five methods yielded practically the same results for the difficulty estimates. Table 1 presents difficulty parameter estimates based on the usual Rasch model scaling (i.e., the mean of difficulties is zero) that is the default setting

for nearly all Rasch model calibration computer programs. Note that some differences still exist among the difficulty parameter estimates and the accompanied standard errors or posterior standard deviations. Although results from this simple data set may not be sufficient for fully evaluating different estimation methods, these may provide good enough information about the agreement in estimation results.

Table 1. LSAT6 Difficulty Estimates

Item	GS1	GS2 <sup>a</sup>	CML	MML <sup>a</sup>	JML
	$b_j$ (p.s.d.)	$b_j$ (p.s.d.)	$b_j$ (s.e.)	$b_j$ (s.e.)	$b_j$ (s.e.)
1	-1.26 (0.11)	-1.38 (0.10)	-1.26 (0.13)	-1.26 (0.13)	-1.24 (0.11)
2	0.48 (0.07)	0.52 (0.07)	0.47 (0.08)	0.48 (0.08)	0.45 (0.07)
3	1.25 (0.07)	1.43 (0.07)	1.24 (0.08)	1.24 (0.07)	1.30 (0.07)
4	0.17 (0.07)	0.16 (0.08)	0.17 (0.09)	0.17 (0.09)	0.13 (0.07)
5	-0.63 (0.09)	-0.72 (0.09)	-0.62 (0.11)	-0.63 (0.11)	-0.64 (0.08)

Note. p.s.d. = posterior standard deviation; s.e. = standard error

<sup>a</sup>Estimates were transformed onto the zero centered logistic metric.

LSAT6 ability estimates and either the accompanied standard errors or the posterior standard deviations are reported in Table 2 for each number-correct raw score from 0 to 5. In GS1 and GS2 there were different posterior means for examinees with the same response pattern or the same raw score. In reporting of the ability estimates, the first examinees who got the respective raw scores were used to obtain the estimates (i.e., examinees 1, 4, 12, 28, 62, and 703). Although the estimates who got the same raw score were trivially different in the consideration of the magnitude of the posterior standard deviation, obtaining such odd results were not seen in other maximum likelihood based estimation procedures.

The most pronounced pattern in Table 2 is that estimates from GS1 and MML/EAP (i.e., expected a posteriori) were very similar. Other estimation methods look somewhat different due to the extremely small test size. Except for the scores 0 and 5, however, ability estimates from CML/ML and JML were very similar. Because in the Rasch model with conditional maximum likelihood estimation the weighted likelihood estimation (WLE; Warm, 1989) is popular, the results for such a case were reported in the CML/WLE column.

Table 2. LSAT6 Ability Estimates

Score	GS1	GS2 <sup>a</sup>	CML/ML	CML/WLE	MML/EAP <sup>a</sup>	JML <sup>b</sup>
	$\theta_i$ (p.s.d.)	$\theta_i$ (p.s.d.)	$\theta_i$ (s.e.)	$\theta_i$ (s.e.)	$\theta_i$ (p.s.d.)	$\theta_i$ (s.e.)
0	-0.09 (0.64)	-1.61 (0.98)		-2.79 (1.72)	0.03 (1.05)	-3.22 (1.93)
1	0.31 (0.64)	-0.74 (0.91)	-1.60 (1.18)	-1.34 (1.11)	0.40 (1.05)	-1.72 (1.21)
2	0.71 (0.64)	0.02 (0.87)	-0.47 (0.99)	-0.41 (0.99)	0.76 (1.07)	-0.52 (1.03)
3	1.12 (0.66)	0.79 (0.85)	0.48 (0.99)	0.42 (0.98)	1.14 (1.11)	0.51 (1.21)
4	1.56 (0.67)	1.48 (0.91)	1.60 (1.18)	1.34 (1.11)	1.54 (1.11)	1.72 (1.21)
5	2.02 (0.70)	3.32 (1.24)		2.78 (1.71)	1.95 (1.13)	3.28 (1.93)

Note. p.s.d. = posterior standard deviation; s.e. = standard error. GS1 and GS2 estimates were from examinees 1, 4, 12, 28, 62, and 703.

<sup>a</sup>Estimates were transformed onto the zero centered logistic metric of item difficulty.

<sup>b</sup>Ad hoc estimates were inserted to scores 0 and 5, respectively.

### LSAT7 Estimation Results

For the LSAT7 data, all five methods yielded practically the same results for the difficulty estimates as did for the LSAT6 data. Table 3 presents difficulty parameter estimates based on the usual Rasch model scaling. Note that some differences still exist among the difficulty parameter estimates and the accompanied standard errors or posterior standard deviations.

Table 3. LSAT7 Difficulty Estimates

Item	GS1	GS2 <sup>a</sup>	CML	MML <sup>a</sup>	JML
	$b_j$ (p.s.d.)	$b_j$ (p.s.d.)	$b_j$ (s.e.)	$b_j$ (s.e.)	$b_j$ (s.e.)
1	-0.54 (0.08)	-0.59 (0.14)	-0.54 (0.10)	-0.54 (0.13)	-0.55 (0.08)
2	0.54 (0.07)	0.59 (0.12)	0.54 (0.08)	0.54 (0.09)	0.53 (0.07)
3	-0.13 (0.07)	-0.17 (0.14)	-0.13 (0.09)	-0.13 (0.11)	-0.15 (0.07)
4	0.81 (0.07)	0.90 (0.11)	0.81 (0.08)	0.80 (0.09)	0.83 (0.07)
5	-0.67 (0.08)	-0.73 (0.15)	-0.67 (0.10)	-0.66 (0.14)	-0.67 (0.08)

Note. p.s.d. = posterior standard deviation; s.e. = standard error

<sup>a</sup>Estimates were transformed onto the zero centered logistic metric.

Table 4 shows the ability estimates and either the accompanied standard errors or the posterior standard deviations for each number-correct raw score from 0 to 5 for LSAT7. As was the case for LSAT6, in GS1 and GS2 there were different posterior means for examinees with the same response pattern or the same raw score. In reporting of the ability estimates, the first examinees who got the respective raw scores were used to obtain the estimates (i.e., examinees 1, 13, 33, 65, 145, and 693).

Note that ability estimates from GS1 and MML/EAP were very similar in Table 4. Other estimation methods yielded somewhat different ability estimates partly due to the extremely small test size. Except for the scores 0 and 5, however, ability estimates from CML/ML and JML were very similar.

Table 4. LSAT7 Ability Estimates

Score	GS1	GS2 <sup>a</sup>	CML/ML	CML/WLE	MML/EAP <sup>a</sup>	JML <sup>b</sup>
	$\theta_i$ (p.s.d.)	$\theta_i$ (p.s.d.)	$\theta_i$ (s.e.)	$\theta_i$ (s.e.)	$\theta_i$ (p.s.d.)	$\theta_i$ (s.e.)
0	-0.63 (0.73)	-1.72 (1.00)		-2.57 (1.66)	-0.59 (0.70)	-2.96 (1.90)
1	-0.12 (0.71)	-0.81 (0.91)	-1.49 (1.14)	-1.21 (1.07)	-0.10 (0.69)	-1.54 (1.16)
2	0.38 (0.72)	0.11 (0.90)	-0.44 (0.95)	-0.38 (0.94)	0.39 (0.70)	-0.47 (0.97)
3	0.91 (0.73)	0.78 (0.91)	0.44 (0.95)	0.37 (0.95)	0.89 (0.72)	0.45 (0.97)
4	1.47 (0.77)	1.54 (0.94)	1.49 (1.15)	1.21 (1.07)	1.44 (0.75)	1.54 (1.16)
5	2.11 (0.83)	2.86 (1.16)		2.59 (1.67)	2.05 (0.80)	2.98 (1.91)

Note. p.s.d. = posterior standard deviation; s.e. = standard error. GS1 and GS2 estimates were from examinees 1, 13, 33, 65, 145, and 693.

<sup>a</sup>Estimates were transformed onto the zero centered logistic metric of item difficulty.

<sup>b</sup>Ad hoc estimates were inserted to scores 0 and 5, respectively.

### Comparison of Prior Specifications

To assess the effects of prior specifications on the difficulty and ability parameter estimates, the same LSAT6 and LSAT7 data were analyzed with OpenBUGS. Four prior specifications with four different sets of hyperparameters were used for both ability and difficulty prior distributions; (1)  $d\gamma(a=2.5, b=5)$ , (2)  $d\gamma(a=4, b=5)$ , (3)  $d\gamma(a=7.5, b=5)$ , and (4)  $d\gamma(a=12.5, b=5)$ . Because the first specification was the same as in the earlier calibration condition, only three additional OpenBUGS runs were performed for LSAT6 and LSAT7, respectively. Except for the prior specification, all other settings to obtain the estimates remained the same for the OpenBUGS runs.



Note that these prior specifications of  $a=2.5, 4, 7.5, 12.5$  with  $b=5$  are fully equivalent to Swaminathan and Gifford's (1982)  $v=5, 8, 15, 25$  with  $\lambda=10$  used in their study.

*LSAT6 Prior Specification Results*

For the LSAT6 data, all four prior specifications yielded practically the same results for the difficulty estimates, but a bit different results for the ability estimates. Table 5 presents difficulty parameter estimates based on the usual Rasch model scaling. Note that only trivial differences exist among the difficulty parameter estimates and the posterior standard deviations, that occur in the second decimal places. Because each difficulty parameter was estimated with the sample size of 1000, shrinkage toward the mean of the difficulty estimates might exist with the increasing hyperparameter  $a$  values but barely noticeable. In Figure 1(a) LSAT6 difficulty estimates are plotted with the four different values of the hyperparameter  $a=2.5, 4, 7.5, 12.5$  (because the hyperparameter  $b=5$  for all cases only the four hyperparameters of  $a$  were used). The numbers in the plot designate the item numbers.

Table 5. LSAT6 Difficulty Estimates from Prior Specifications

Item	GS1 Hyperparameters							
	$a=2.5, b=5$		$a=4, b=5$		$a=7.5, b=5$		$a=12.5, b=5$	
	$b_j$	(p.s.d.)	$b_j$	(p.s.d.)	$b_j$	(p.s.d.)	$b_j$	(p.s.d.)
1	-1.26	(0.11)	-1.25	(0.10)	-1.24	(0.10)	-1.22	(0.10)
2	0.48	(0.07)	0.48	(0.07)	0.47	(0.07)	0.46	(0.07)
3	1.25	(0.07)	1.24	(0.07)	1.23	(0.07)	1.21	(0.07)
4	0.17	(0.07)	0.17	(0.07)	0.16	(0.07)	0.16	(0.07)
5	-0.63	(0.09)	-0.63	(0.08)	-0.62	(0.08)	-0.61	(0.08)

Note. p.s.d. = posterior standard deviation

LSAT6 ability estimates from the four prior specifications and the posterior standard deviations are reported in Table 6 for each number-correct raw score from 0 to 5. In GS1 there were different posterior means for examinees with the same response pattern or the same raw score. In reporting of the ability estimates, the first examinees who got the respective raw scores were used to obtain the estimates (i.e., examinees 1, 4, 12, 28, 62, and 703).

Considering the magnitude of the posterior standard deviations, it can be noted in Table 6 that practically trivial differences exist among the ability estimates and the posterior standard deviations. Nevertheless, because each ability parameter was estimated with the truly small number of items, shrinkage toward the mean of ability estimates with the increasing hyperparameter  $a$  values was quite noticeable. In Figure 1(b) LSAT6 ability estimates are plotted with the four different values of the hyperparameter  $a=2.5, 4, 7.5, 12.5$  (because the hyperparameter  $b=5$  for all cases only the four hyperparameters of  $a$  were used). The numbers in the plot designate the raw scores from 0 to 5.

Table 6. LSAT6 Ability Estimates from Four Prior Specifications

Score	GS1 Hyperparameters							
	$a=2.5, b=5$		$a=4, b=5$		$a=7.5, b=5$		$a=12.5, b=5$	
	$\theta_i$	(p.s.d.)	$\theta_i$	(p.s.d.)	$\theta_i$	(p.s.d.)	$\theta_i$	(p.s.d.)
0	-0.09	(0.64)	-0.07	(0.64)	0.01	(0.62)	0.12	(0.60)
1	0.31	(0.64)	0.33	(0.63)	0.39	(0.62)	0.45	(0.59)
2	0.71	(0.64)	0.72	(0.64)	0.77	(0.62)	0.79	(0.59)
3	1.12	(0.66)	1.13	(0.64)	1.13	(0.62)	1.15	(0.60)
4	1.56	(0.67)	1.56	(0.65)	1.54	(0.64)	1.50	(0.61)
5	2.02	(0.70)	2.02	(0.68)	1.97	(0.66)	1.89	(0.63)

Note. p.s.d. = posterior standard deviation

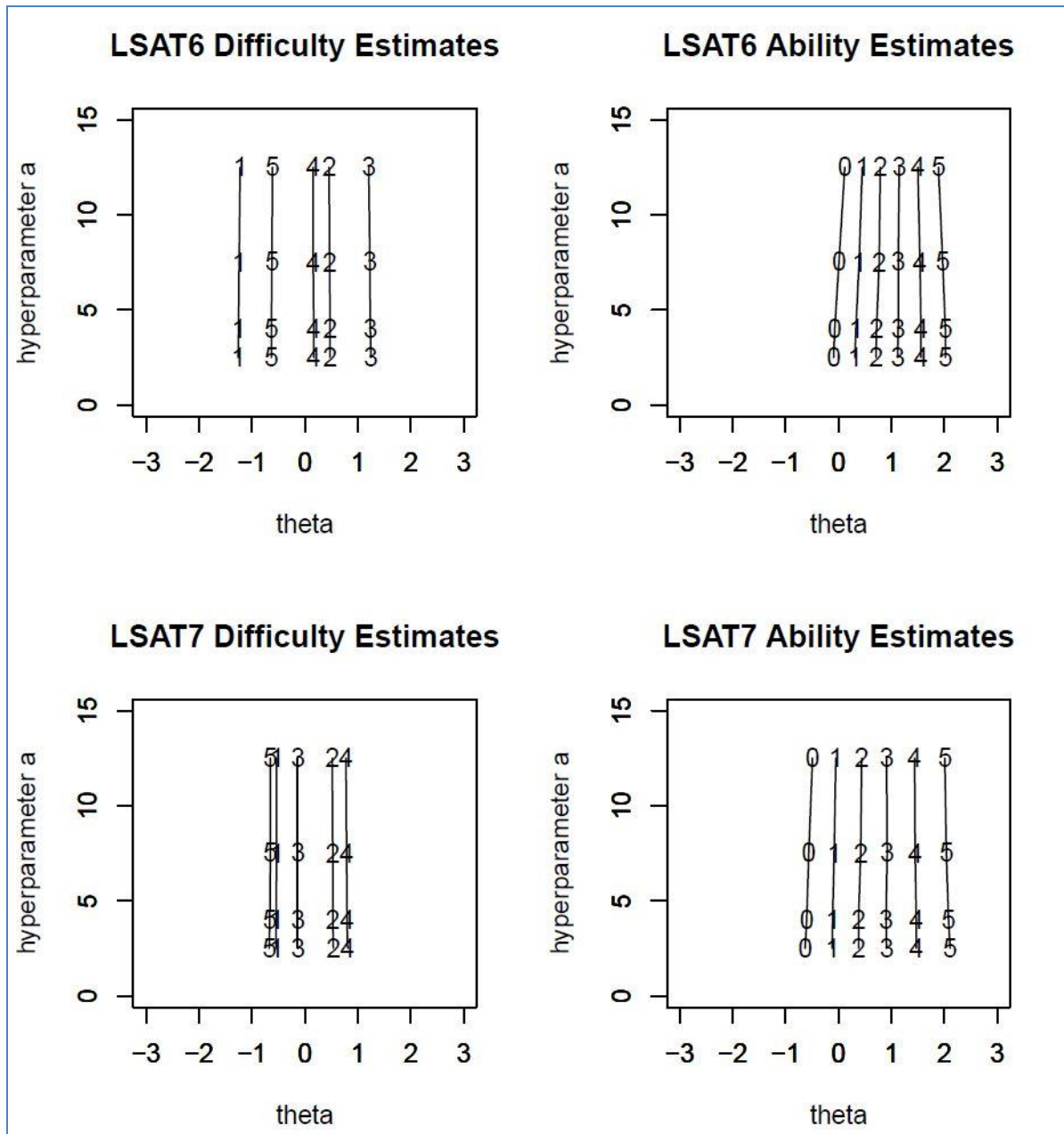


Figure 1. Plots of (a) LSAT6 difficulty estimates, (b) LSAT6 ability estimates, (c) LSAT7 difficulty estimates, and (d) LSAT7 ability estimates for the hyperparameter values of  $a=2.5, 4, 7.5, 12.5$  with  $b=5$ .

#### LSAT7 Prior Specification Results

For the LSAT7 data, all four prior specifications yielded practically the same results for the difficulty estimates, but a bit different results for the ability estimates. Table 7 presents difficulty parameter estimates based on the usual Rasch model scaling. Note that only trivial differences exist among the difficulty parameter estimates and the posterior standard deviations, that occur in the second decimal places. Because each difficulty parameter was estimated with the sample size of 1000, shrinkage toward the mean of difficulty estimates might exist but not really noticeable. In Figure 1(c) LSAT7

difficulty estimates are plotted with the four different values of the hyperparameter  $a=2.5, 4, 7.5, 12.5$ . The numbers in the plot designate the item numbers.

Table 7. LSAT7 Item Difficulty Estimates from Four Prior Specifications

Item	GS1 Hyperparameters							
	$a=2.5, b=5$		$a=4, b=5$		$a=7.5, b=5$		$a=12.5, b=5$	
	$b_j$ (p.s.d.)		$b_j$ (p.s.d.)		$b_j$ (p.s.d.)		$b_j$ (p.s.d.)	
1	-0.54	(0.08)	-0.54	(0.08)	-0.53	(0.08)	-0.53	(0.08)
2	0.54	(0.07)	0.53	(0.07)	0.53	(0.07)	0.52	(0.07)
3	-0.13	(0.07)	-0.13	(0.07)	-0.13	(0.07)	-0.13	(0.07)
4	0.81	(0.07)	0.80	(0.07)	0.79	(0.07)	0.78	(0.07)
5	-0.67	(0.08)	-0.66	(0.08)	-0.65	(0.08)	-0.65	(0.08)

Note. p.s.d. = posterior standard deviation

LSAT7 ability estimates from the four prior specifications and the posterior standard deviations are reported in Table 8 for each number-correct raw score from 0 to 5. In GS1 there were different posterior means for examinees with the same response pattern or the same raw score. In reporting of the ability estimates, the first examinees who got the respective raw scores were used to obtain the estimates (i.e., examinees 1, 13, 33, 65, 145, and 693).

It can be noted that practically trivial differences exist among the ability estimates and the posterior standard deviations, considering the magnitude of the posterior standard deviations. Nevertheless, each ability parameter was estimated with the truly small number of items, shrinkage toward the mean of ability estimates with the increasing hyperparameter  $a$  values was quite noticeable. In Figure 1(d) LSAT7 ability estimates are plotted with the four different values of the hyperparameter  $a=2.5, 4, 7.5, 12.5$ . The numbers in the plot designate the raw scores from 0 to 5.

Table 8. LSAT7 Ability Estimates from Four Prior Specifications

Score	GS1 Hyperparameters							
	$a=2.5, b=5$		$a=4, b=5$		$a=7.5, b=5$		$a=12.5, b=5$	
	$\theta_i$	(p.s.d.)	$\theta_i$	(p.s.d.)	$\theta_i$	(p.s.d.)	$\theta_i$	(p.s.d.)
0	-0.63	(0.73)	-0.60	(0.73)	-0.56	(0.71)	-0.49	(0.69)
1	-0.12	(0.71)	-0.11	(0.72)	-0.08	(0.69)	-0.05	(0.69)
2	0.38	(0.72)	0.38	(0.72)	0.42	(0.69)	0.44	(0.70)
3	0.91	(0.73)	0.90	(0.73)	0.92	(0.72)	0.91	(0.71)
4	1.47	(0.77)	1.47	(0.77)	1.45	(0.75)	1.44	(0.73)
5	2.11	(0.83)	2.08	(0.83)	2.04	(0.80)	2.01	(0.78)

Note. p.s.d. = posterior standard deviation

## DISCUSSION and CONCLUSION

The main difference between the two Gibbs sampling methods, GS1 and GS2, lies in both the specifications of prior distributions and the underlying sampling procedures. The prior distributions used in GS1 had the hierarchical form following Swaminathan and Gifford (1982). For example, the hyperparameter mean of the normal prior distribution for ability had a noninformative uniform distribution and the inverse of the hyperparameter variance of the normal prior had a gamma distribution. In GS1 with  $\text{gamma}(a=2.5, b=5)$  the prior sample size of the gamma distribution was specified as  $2(2.5)=5$  and the prior expected value was  $2.5/5=0.5$  (i.e., the expected value of the hyperparameter variance to be  $5/1.5=3.33$ ). Note that this prior specification is equivalent to Swaminathan and Gifford's (1982)  $v=5$  and  $\lambda=10$ , one of the prior specifications in their paper. They

used three other prior specifications that were converted to the equivalent specifications in the second study. The use of  $\text{gamma}(2.5, 5)$  seems reasonable among the choices. Swaminathan and Gifford (1982) concluded similarly. Note that there are also other ways of specifying priors for the Rasch model (see Kim, 2001; Levy & Mislevy, 2016; Spiegelhalter et al., 1997b; Stone & Zhu, 2015) instead of using priors in the hierarchical form. In Johnson and Albert's (1999) `item_r1` function for GS2 the hyperparameters of the theta prior was set to have a standard normal distribution while prior standard deviation of the item difficulty parameters was set to unity. See Johnson and Albert (1999, pp. 202–204) for the detailed Gibbs sampling for GS2. Hence GS1 and GS2 differ not only the mathematical forms of the model but also the priors employed.

Because the full conditional distributions for the Rasch model are log-concave (Ghosh, Ghosh, Chen, & Agresti, 1999), the sampling in GS1 used the derivative-free adaptive rejection sampling algorithm (Gilks, 1996; Gilks & Wild, 1992). Due to the use of hierarchical prior distributions, more general sampling procedures can be employed for various parameters in GS1 (see Lunn et al., 2013, pp. 68–70) that include slice sampling (Neal, 2003) and Metropolis-within-Gibbs (Metropolis et al., 1953; Hastings, 1970). In GS2, direct Gibbs sampling method was used with data augmentation because the actual item response theory model was that of the normal ogive instead of the logistic ogive (Albert, 1992; Baker, 1998). The resulting parameter estimates in GS2 were initially expressed on the normal ogive metric but placed onto the logistic metric.

When difficulty and ability are estimated together in GS1 or GS2, the ability estimate for specific case is not unique. The same response pattern may yield different ability estimates and that is not acceptable in practice. In addition, because of employing the exchangeability concept, all ability estimates are estimated simultaneously and there exists some dependency in the resulting estimates. Although estimates are not independent in general, it seems troublesome that estimating ability even with known item parameters may yield different estimates for a specific response pattern. Hence, Gibbs sampling methods or some other estimation methods based on Markov chain Monte Carlo may not be seen as viable methods for the usual item and ability parameter estimation for the usual item response theory models for dichotomous items that include the Rasch model.

In this study, the Rasch model was employed without addressing the problem of model selection, choice of link function, or model fit. Kim and Bolt (2007) contains an excellent introductory review of these issues. Interested readers should refer to Kim and Bolt (2007) and other general references including Lunn et al. (2013).

Note that although Gibbs sampling methods and some computer programs which implemented such procedures have been available sometime, the accuracy of the methods has not been thoroughly studied. Obviously these techniques have been applied to some complicated modeling situations where the traditional maximum likelihood based methods are too difficult to implement, and hence have not been thoroughly tested and compared. Because maximum likelihood based methods have not been implemented at all in such applications, still we need to investigate the relevant estimation procedures. In addition, because there are many different ways of implementing Gibbs sampling methods in item response theory and many different prior distributions can be employed with many different specifications in Bayesian estimation, the illustrative implementation of the Gibbs sampling method and comparing results with other existing Bayesian and likelihood based methods should provide measurement specialists and test developers as well as the users of the computer programs with guidelines for using the Gibbs sampling method under the Rasch item response theory model.

In this study, explications of nearly all estimation methods for the Rasch model were presented together with the two methods based on Gibbs sampling. The specification of priors for ability and difficulty parameters in Bayesian estimation and the Gibbs sampling method was fully explained with detailed mathematical statistical formulas, basically following the framework of Swaminathan and Gifford (1982). Illustrations about the effects of prior specifications on the estimates were presented with empirical data. It should be noted that additional, full scale simulation studies as well as more cumulative experience with regard to prior specifications for Bayesian estimation are definitely needed.

## REFERENCES

- Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics, 17*, 251–269.
- Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society, Series B, 32*, 283–301.
- Andersen, E. B. (1972). The numerical solution to a set of conditional estimation equations. *Journal of the Royal Statistical Society, Series B, 34*, 42–54.
- Andersen, E. B. (1980). *Discrete statistical models with social science applications*. Amsterdam,: North-Holland.
- Anderson, T. W. (1984). *An introduction to multivariate statistical analysis* (2nd ed.). New York, NY: Wiley.
- Baker, F. B. (1998). An investigation of the item parameter recovery characteristics of a Gibbs sampling approach. *Applied Psychological Measurement, 22*, 153–169.
- Baker, F. B., & Harwell, M. R. (1994). *Estimation of item parameters in the Rasch model via conditional maximum likelihood: A didactic*. Unpublished manuscript, Department of Educational Psychology, University of Wisconsin, Madison, WI.
- Baker, F. B., & Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques* (2nd ed.). New York, NY: Dekker.
- Barnard, G. A. (1958). Thomas Bayes - a biographical note. *Biometrika, 45*, 293–295.
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions (of the Royal Society of London), 53*, 370-418; Errata, c. 543.
- Bayes, T. (1958). An essay towards solving a problem in the doctrine of changes. *Biometrika, 45*, 296–315.
- Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (2nd ed.). New York, NY: Springer.
- Birnbaum, A. (1969). Statistical theory for logistic mental test models with a prior distribution of ability. *Journal of Mathematical Psychology, 6*, 258–276.
- Block, A. (1977). *Murphy's law and other reasons why things go wrong!* Los Angeles, CA: Price/Stern/Sloan.
- 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*, 381–409.
- Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2002). Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement, 39*, 331–348.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Applications of an EM algorithm. *Psychometrika, 46*, 443–459.
- Bock, R. D., & Lieberman, M. (1970). Fitting a response model for  $n$  dichotomously scored items. *Psychometrika, 35*, 179–197.
- Cai, L., Thissen, D., & du Toit, S. (2010). *IRTPRO: Item response theory for patient-reported outcomes* [Computer software]. Skokie, IL: Scientific Software International.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29.
- Cohen, A. S., & Bolt, D. M. (2005). A mixture model analysis of differential item functioning. *Journal of Educational Measurement, 42*, 133–148.
- Davis, P. J. (1964). Gamma function and related functions. In M. Abramowitz & I. A. Stegun (Eds.), *Handbook of mathematical functions with formulas, graphs, and mathematical tables* (pp. 253–293). Washington, DC: National Bureau of Standards.
- Dodge, Y. (Ed.). (2003). *The Oxford dictionary of statistical terms*. Oxford, Great Britain: Oxford University Press.
- du Toit, M. (Ed.). (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Chicago, IL: Scientific Software International.
- Fischer, G. H. (1968). *Einführung in die theorie psychologischer tests: Grundlagen und anwendungen* [Introduction to the theory of psychological tests: Foundations and applications]. Bern, Switzerland: Huber.
- Fischer, G. H., & Allerup, P. (1968). Rechentechnische fragen zu raschs eindimensionalem modell [Computational questions on Rasch's unidimensional model]. In G. H. Fischer (Hrsg. [Ed.]), *Psychologische Testtheorie [Psychological test theory]* (pp. 269–280). Bern, Switzerland: Huber.
- Fischer, G. H., & Molenaar, I. W. (Eds.). (1995). *Rasch models: Foundations, recent developments, and applications*. New York, NY: Springer-Verlag.
- Fox, J.-P. (2010). *Bayesian item response modeling: Theory and applications*. New York, NY: Springer.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London, Great Britain: Chapman & Hall.



- Ghosh, M., Ghosh, A., Chen, M.-H., & Agresti, A. (1999). *Bayesian estimation for item response model* (Tech. Rep.). Gainesville, FL: University of Florida, Department of Statistics.
- Gilks, W. R. (1996). Full conditional distribution. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice* (pp. 75–88). London, England: Chapman and Hall.
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov chain Monte Carlo in practice*. London, England: Chapman and Hall.
- Gilks, W. R., & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41, 337–348.
- Hasting, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97–109.
- Hojtink, H., & Boomsma, A. (1995). On person parameter estimation in the dichotomous Rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 53–68). New York, NY: Springer-Verlag.
- Isaacs, G. I., Christ, D. E., Novick, M. R., & Jackson, P. H. (1974). *Tables for Bayesian statisticians*. The Iowa Testing Program, The University of Iowa, Iowa City, IA.
- Johnson, V. E., & Albert, J. H. (1999). *Ordinal data modeling*. New York, NY: Springer.
- Karabatsos, G., & Batchelder, W. H. (2003). Markov chain Monte Carlo estimation for test theory without an answer key. *Psychometrika*, 68, 373–389.
- Kim, J.-S., & Bolt, D. M. (2007). Estimating item response theory models using Markov chain Monte Carlo methods. *Educational Measurement: Issues and Practice*, 26(4), 38–51.
- Kim, S.-H. (2001). An evaluation of a Markov chain Monte Carlo method for the Rasch model. *Applied Psychological Measurement*, 25, 163–176.
- Leonard, T., & Hsu, J. S. J. (1999). *Bayesian methods: An analysis for statisticians and interdisciplinary researchers*. New York, NY: Cambridge University Press.
- Leonard, T., & Novick, M. R. (1985). *Bayesian inference and diagnostics for the three parameter logistic model* (ONR Technical Report No. 85-5). Iowa City, IA: The University of Iowa, Cada Research Group. (ERIC Document Reproduction Service No. ED261068)
- Levy, R. (2009). The rise of Markov chain Monte Carlo estimation for psychometric modeling. *Journal of Probability and Statistics*, 2009, ID 537139. doi:10.155/2007/537139
- Levy, R., & Mislavy, R. J. (2016). *Bayesian psychometric modeling*. Boca Raton, FL: CRC Press.
- Linacre, J. M. (2003). WINSTEPS Rasch measurement computer program [Computer software]. Chicago, IL: Winsteps.com.
- Lindley, D. V. (1965). *Introduction to probability and statistics from a Bayesian viewpoint: Part 2, Inference*. London, Great Britain: Cambridge University Press.
- Lindley, D. V., & Smith, A. F. (1972). Bayesian estimates for the linear model. *Journal of the Royal Statistical Society, Series B*, 34, 1–41.
- Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2013). *The BUGS book: A practical introduction the Bayesian analysis*. Boca Raton, FL: CRC Press.
- Luo, Y., & Jiao, H. (2017). Using the Stan program for Bayesian item response theory. *Educational and Psychological Measurement*, Advance Online Publication. <https://doi.org/10.1177/0013164417693666>
- Mair, P., & Hatzinger, R. (2007). Extended Rasch modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software*, 20(9), 1–20.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21, 1087–1092.
- Mislavy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika*, 51, 177–195.
- Molenaar, I. W. (1995). Estimation of item parameters. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 39–51). New York, NY: Springer-Verlag.
- 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. doi:10.3389/fpsyg.2016.01422
- Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, 31, 705–741.
- Novick, M. R., & Jackson, P. H. (1974). *Statistical methods for educational and psychological research*. New York, NY: McGraw-Hill.
- Novick, M. R., Lewis, C., & Jackson, P. H. (1973). The estimation of proportions in  $n$  groups. *Psychometrika*, 38, 19–46.
- Öztürk, N., & Karabatsos, G. (2017). A Bayesian robust IRT outlier detection model. *Applied Psychological Measurement*, 41, 195–208.
- Patz, R. J., & Junker, B. W. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, 24, 146–178.

- Rasch, G. (1980). *Probabilistic model for some intelligence and attainment tests* (With a foreword and afterword by B. D. Wright). Chicago, IL: The University of Chicago Press.
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analysis. *Journal of Statistical Software*, 17(5), 1–25.
- Sen, S., Cohen, A. S., & Kim, S. H. (2018). Model selection for multilevel mixture Rasch models. *Applied Psychological Measurement*, 1-18. doi: 10.1177/0146621618779990
- Sheng, Y. (2010). A sensitivity analysis of Gibbs sampling for 3PNO IRT models: Effects of prior specifications on parameter estimates. *Behaviormetrika*, 37, 87–110.
- Sheng, Y. (2017). Investigating a weakly informative prior for item scale hyperparameters in hierarchical 3PNO IRT models. *Frontiers in Psychology*, 8:123. doi:10.3389/fpsyg.2017.00123
- Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W. R. (1997a). *BUGS: Bayesian inference using Gibbs sampling* (Version 0.6) [Computer software]. Cambridge, UK: University of Cambridge, Institute of Public Health, Medical Research Council Biostatistics Unit.
- Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W. R. (1997b). *BUGS 0.5 examples* (Vol. 1, Version i). Cambridge, UK: University of Cambridge, Institute of Public Health, Medical Research Council Biostatistics Unit.
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2014). *OpenBUGS user manual*. Cambridge, UK: MRC Biostatistics Unit, Institute of Public Health.
- Stigler, S. M. (1980). Stigler's law of eponymy. *Transactions of the New York Academy of Sciences, Series 2*, 39, 147–157.
- Stone, C. A., & Zhu, X. (2015). *Bayesian analysis of item response theory models using SAS*. Cary, NC: SAS Institute.
- Swaminathan, H., & Gifford, J. A. (1982). Bayesian estimation in the Rasch model. *Journal of Educational Statistics*, 7, 175–191.
- Swaminathan, H., & Gifford, J. A. (1985). Bayesian estimation in the two-parameter logistic model. *Psychometrika*, 50, 349–364.
- Swaminathan, H., & Gifford, J. A. (1986). Bayesian estimation in the three-parameter logistic model. *Psychometrika*, 51, 589–601.
- Swaminathan, H., Hambleton, R. K., Sireci, S. G., Xing, D., & Rizavi, S. M. (2003). Small sample estimation in dichotomous item response models: Effect of priors based on judgmental information on the accuracy of item parameter estimates. *Applied Psychological Measurement*, 27, 27–51.
- The MathWorks. (1996). *MATLAB* (Version 5) [Computer program]. Natick, MA: Author.
- Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, 47, 175–186.
- Trader, R. L. (1997). Bayes, Thomas. in N. L. Johnson & S. Kotz (Eds.), *Leading personalities in statistical sciences: From the seventeenth century to the present* (pp. 11–14). New York, NY: John Wiley & Sons.
- Tsutakawa, R. K., & Lin, H. Y. (1986). Bayesian estimation of item response curves. *Psychometrika*, 51, 251–267.
- Venables, W. N., Smith, D. M., & The R Development Core Team. (2009). *An introduction to R* (2nd ed.). La Vergne, TN: Network Theory.
- von Davier, M. (2001). *WINMIRA 2001* [Computer program]. St. Paul, MN: Assessment Systems Corporation.
- Wang, X., Bradlow, E. T., & Wainer, H. (2005). *User's guide for SCORIGHT (Version 3.0): A computer program for scoring tests built of testlets including a module for covariate analysis* (ETS Research Rep. RR-04-49). Princeton, NJ: Educational Testing Service.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response models. *Psychometrika*, 54, 427–450.
- 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, 339–352.
- Wright, B. D., & Linacre, J. M., & Schultz, M. (1989). *A user's guide to BIGSCALE: Rasch-model rating scale analysis computer program*. Chicago, IL: MESA Press.
- Wright, B. D., & Mead, R. J. (1978). *BICAL: Calibrating items and scales with the Rasch model* (Research Memorandum No. 23A). Chicago, IL: University of Chicago, Department of Education, Statistical Laboratory.
- Wright, B. D., Mead, R. J., & Bell, S. R. (1980). *BICAL: Calibrating items with the Rasch model* (Research Memorandum No. 23C). Chicago, IL: University of Chicago, Department of Education, Statistical Laboratory.
- Wright, B. D., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement*, 29, 23–48.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: MESA Press.

## Rasch Modelinde Gibbs Örnekleme Yönteminin Uygulanması

### Giriş

Tek parametrelili lojistik Rasch modelinde (Rasch, 1980), madde güçlüğü ve kişi yetenek parametre kestirimlerini elde etmek için birçok kestirim metodu kullanılabilir (Fischer ve Molenaar, 1995; Molenaar, 1995; Hoijtink ve Boomsma, 1995). Madde güçlük ve kişi yetenek parametreleri, ortak olabilirlik fonksiyonunu maksimize ederek ortak olarak kestirilebilir (yani, JML; Wright ve Stone, 1979). Koşullu maksimum olabilirlik (CML; Andersen, 1980), madde güçlük parametrelerinin tahmini için tek parametrelili lojistik modelin altında standart kestirim metodu olarak görünmektedir (ör. Molenaar, 1995). Ayrıca, beklenti ve maksimizasyon algoritmasını kullanarak marjinal maksimum olabilirlik (MML) kestirimi, madde güçlük parametre kestirimlerini elde etmek için kullanılabilir (du Toit, 2003; Thissen, 1982). Ek olarak, tek parametrelili lojistik model altında parametre kestirimlerini elde etmek için ortak Bayes kestirimi ve marjinal Bayes kestirimi kullanılabilir (ör. Birnbaum, 1969; Mislevy, 1986; Swaminathan & Gifford, 1982; ayrıca bkz. Tsutakawa, & Lin, 1986).

Rasch modeli madde güçlük ve kişi yetenek parametrelerinin nokta tahminleri, bu olasılık fonksiyonlarını veya sonsal (posterior) dağılımın bazı formlarını maksimize ederek, maksimum olasılık kestirimi ve Bayes kestirimi yöntemlerinden elde edilir. Nokta tahminleri elde etmek yerine, Bayesci çerçevedeki sonsal dağılımı tahmin etmeye yönelik prosedürler nispeten yakın zamanda önerilmiştir. Böyle bir yöntem olan Gibbs örnekleme, marjinal dağılımdan ziyade ortak sonsal dağılımı kullanarak madde ve yetenek parametrelerini kestiren bir yaklaşımdır (ör. Albert, 1992; Johnson & Albert, 1999; Kim, 2001; Patz & Junker, 1999). Madde ve yetenek parametrelerini kestirmek için kullanılacak Gibbs örneklemesinin birkaç farklı versiyonu ve uygulamasının olduğu unutulmamalıdır. Yine de, tüm Bayesci kestirim metodları, özellikle karşılaştırılabilir önseller kullanıldığında veya yerel olarak tekdüze önseller kullanıldığında karşılaştırılabilir madde ve yetenek parametre kestirimleri vermelidir. Bu çalışma, tek parametrelili lojistik Rasch modelini kullanarak bu sorunu araştırmak için tasarlanmıştır. Özellikle, reddetme örnekleme (GS1) kullanılan bir Gibbs örnekleme yönteminin madde güçlük ve kişi yetenek parametre kestirimleri incelenmiş ve veri artırma (GS2) yönteminin yanı sıra CML, MML ve JML kullanılan başka bir Gibbs örnekleme yöntemi ile karşılaştırılmıştır. Bu çalışmada GS1 için yeni notasyonlar kullanmak yerine Swaminathan ve Gifford'un (1982) Rasch modelinde Bayes kestirimi ile ilgili önermiş olduğu notasyon takip edilmiştir. GS1'i mevcut çalışmada farklılaştıran temel konu ve Swaminathan ve Gifford (1982)'da kullanılan uygulama, sonsal maksimizasyon ve yakınsama kavramında yatmaktadır. Bayes kestiriminde ve Gibbs örnekleme yönteminde önsellerin belirlenmesinin önemine rağmen, literatürde önsel seçimi ve kullanımı konusunda fazla bir şeffaflık olmadığı gözlenmiştir. Bu çalışma aynı zamanda, Rasch modelinde Swaminathan ve Gifford'un (1982) hiyerarşik Bayes çerçevesi bağlamında önsel seçiminin rolünü de göstermektedir.

### Yöntem

Bu çalışmada Rasch modeli altında hem madde güçlük hem de kişi yetenek parametrelerinin kestirimi için bir karşılaştırma yapılmıştır. GS1, GS2, CML, MML ve JML'yi karşılaştırmak için, (1) Hukuk Fakültesi Kabul Testi 6. Bölüm (LSAT6; Bock & Aitkin, 1981; Bock & Lieberman, 1970) ve (2) Hukuk Fakültesi Kabul Testi 7. Bölüm (LSAT7) verileri kullanılmıştır. LSAT6 (1000 kişi ve 5 madde) ve LSAT7 verileri yayınlanmış birçok makale ve kitapta daha önce analiz edilmiştir (ör., Andersen, 1980; McDonald, 1999). Simülasyon verileri yerine bu verilerin kullanılması, farklı kestirim yöntemlerinin karşılaştırılmasını yapmak için okuyuculara bir temel sağlamaktadır.

Bu çalışmada GS1 kestirimleri OpenBUGS programı kullanılarak elde edilmiştir. GS2 tahminleri, Johnson ve Albert (1999)'dan gelen kodu içeren MATLAB (MathWorks, 1996) kullanılarak elde edilmiştir. LSAT6 ve LSAT7 için önce madde güçlük parametre kestirimleri daha sonra da kişi

yetenek parametre tahminleri rapor edilmiştir. CML, MML ve JML sözdizimlerini göstermek gerekli değildir. Ayrıca GS2 için Johnson ve Albert (1999, s. 248)'de sunulan MATLAB fonksiyonu herhangi bir modifikasyon olmaksızın kullanılmıştır. Bununla birlikte, OpenBUGS sözdizimini sunmak gerekli görülmüştür. Sözdiziminin gerekli bölümleri ekte yer almaktadır. Ekte, hiperparametre varyansının tersi hem yetenek hem de madde güçlük parametreleri için  $d\gamma$  ( $a = 2.5, b = 5$ ) ile belirtilmiştir. Bu önsel belirleme Swaminathan ve Gifford'un (1982)  $v = 5$  ve  $\lambda = 10$  değerlerine eşdeğerdir. Ayrıca, güçlük parametrelerinin başlangıç değerleri için,  $p_j$  olarak gösterilen klasik madde güçlüğü'nün log oranlarının ortalanmış değerinin (yani, 0'da ortalanmış olan  $\log [(1-p_j) / p_j]$  değerleri) kullanıldığı dikkate alınmalıdır. Yetenek parametreleri için benzer başlangıç değerleri belirtilmiştir.

Kim ve Bolt (2007) ve Kim (2001)'in önerilerine dayanarak burn-in kısmındaki tekrar sayısı 1000'e ayarlanmış ve sonraki 10,000 tekrarı simüle edilmiş çekilişlerin yakınlaşmasını gösteren sonsal dağılımları oluşturmak için GS1 ve GS2'de kullanılmıştır (bkz. Gilks, Richardson & Spiegelhalter, 1996). Zincirlerin yakınsaklığı, geçmiş ve otokorelasyon çizimleri kontrol edilerek görsel olarak izlenmiştir. GS1 veya GS2'deki örneklenmiş değerleri özetlemenin birçok farklı yolu olduğuna dikkat edilmelidir. Gerçek sonsal güvenilirlik aralığını kullanmak yerine, bu çalışmada sonsal ortalamalar ve sonsal standart sapmalar kullanılmıştır. İlgili parametreler için örneklerin marjinal sonsal yoğunlukları, GS1'de tek modlu ve normal dağılım göstermiştir. GS2'de örneklenen değerler de benzer sonuçlar vermiştir.

### ***Sonuç ve Tartışma***

Bu çalışmada farklı kestirim metotları ve farklı önsel dağılımlar aynı veriler üzerinden karşılaştırılmıştır. LSAT6 verisi ile elde edilen madde güçlük parametresi tahminleri ve eşlik eden standart hatalar veya sonsal standart sapmalar arasında bazı farklılıklar gözlenmiştir. Bu bulgular arasında en belirgin olanı GS1 ve MML/EAP kestirimlerinin çok benzer çıkmasıdır. Diğer kestirim yöntemleri küçük test büyüklüğü nedeniyle biraz farklılık göstermiştir. LSAT7 verileri için, tüm metotlar, LSAT6 verileri için olduğu gibi, madde güçlük kestirimleri için pratik olarak aynı sonuçları vermiştir. Önsel belirlemelerin (prior specifications) madde güçlük ve yetenek parametre kestirimleri üzerindeki etkilerini değerlendirmek için, aynı LSAT6 ve LSAT7 verileri OpenBUGS ile analiz edilmiştir. LSAT6 ve LSAT7 verileri için, önsel belirlemelerin hepsi, madde güçlük tahminleri için pratik olarak aynı sonuçları vermiştir, fakat yetenek tahminleri için biraz farklı sonuçlar elde edilmiştir.

İki Gibbs örnekleme yöntemi, GS1 ve GS2, arasındaki ana fark, hem önsel dağılımların özelliklerinde hem de temel örnekleme prosedürlerinde yatmaktadır. GS1'de kullanılan önsel dağılımlar, Swaminathan ve Gifford (1982)'un önerisini takip eden hiyerarşik forma sahiptir. Örneğin, yetenek parametresinin normal olan önsel dağılımına ait hiperparametrenin ortalaması, bilgi-verici olmayan (non-informative) bir tekdüze dağılıma sahip iken önsel normal olanın hiperparametre varyansının tersi, bir gama dağılımına sahiptir. Gama ( $a = 2.5, b = 5$ ) dağılımlı GS1'de, gama dağılımının önsel örneklem büyüklüğü  $2*(2.5) = 5$  olarak belirlendi ve önsel beklenen değer  $2.5 / 5 = 0.5$  idi (yani, hipermetre varyansının beklenen değeri  $5 / 1.5 = 3.33$ ). Bu önsel belirlemenin, Swaminathan ve Gifford'un (1982)  $v = 5$  ve  $\lambda = 10$  değerlerine eşdeğer olduğunu unutmayın. Swaminathan ve Gifford ikinci bir çalışmada, eşdeğer belirlemelere dönüştürülmüş olan başka üç özellik daha kullanmıştır. Bu çalışmada Gamma (2.5, 5) kullanımı makul bir seçenek olarak görünmektedir. Swaminathan ve Gifford (1982) da benzer sonuçları raporlamıştır. Hiyerarşik formda önselleri kullanmanın yanında Rasch modeli için önselleri belirlemenin başka yolları da vardır (bkz. Kim, 2001; Levy & Mislevy, 2016; Spiegelhalter ve ark., 1996b; Stone & Zhu, 2015). Johnson ve Albert'in (1999) `item_r1` fonksiyonunda GS2 için, önsel teta hiperparametreleri standart bir normal dağılıma ayarlanmış, öte yandan standart sapma parametrelerinin birliği olarak ayarlanmıştır. GS2'ye ait ayrıntılı Gibbs örnekleme için Johnson ve Albert (1999, s. 202–204)'e bakılabilir. Dolayısıyla GS1 ve GS2 sadece modelin matematiksel formlarında değil, aynı zamanda kullanılan önsellerde de farklılık göstermektedir.

## Appendix: OpenBUGS Code

```
model {
# patterned data to individual responses
  for (i in 1:cof[1]) {
    for (j in 1:J) { x[i, j] <- pattern[1, j] }
  }
  for (g in 2:G) {
    for (i in cof[g-1]+1:cof[g]) {
      for (j in 1:J) { x[i, j] <- pattern[g, j] }
    }
  }
# Rasch model
  for (i in 1:I) {
    for (j in 1:J) {
      logit(p[i, j]) <- theta[i] - beta[j]
      x[i, j] ~ dbern(p[i, j])
    }
  }
# ability prior
  theta[i] ~ dnorm(mut, taut)
  t[i] <- theta[i] - mean(beta[])
}
# item prior
  for (j in 1:J) {
    beta[j] ~ dnorm(mub, taub)
    b[j] <- beta[j] - mean(beta[])
  }
# hyperpriors
  mut ~ dunif(-5, 5)
  taut ~ dgamma(2.5, 5)
  phit <- 1 / sqrt(taut)
  mub ~ dunif(-5, 5)
  taub ~ dgamma(2.5, 5)
}

# lsat6 patterned data with cumulative observed frequencies
list(I = 1000, G = 32, J = 5,
  cof = c(3, 9, 11, 22, 23, 24, 27, 31, 32, 40,
    40, 56, 56, 59, 61, 76, 86, 115, 129, 210,
    213, 241, 256, 336, 352, 408, 429, 602, 613, 674,
    702, 1000),
  pattern = structure(.Data = c(
    0, 0, 0, 0, 0,
    0, 0, 0, 0, 1,
    0, 0, 0, 1, 0,
    0, 0, 0, 1, 1,
    0, 0, 1, 0, 0,
    0, 0, 1, 0, 1,
    0, 0, 1, 1, 0,
    0, 0, 1, 1, 1,
    0, 1, 0, 0, 0,
    0, 1, 0, 0, 1,
    0, 1, 0, 1, 0,
    0, 1, 0, 1, 1,
    0, 1, 1, 0, 0,
    0, 1, 1, 0, 1,
    0, 1, 1, 1, 0,
    0, 1, 1, 1, 1,
    1, 0, 0, 0, 0,
    1, 0, 0, 0, 1,
    1, 0, 0, 1, 0,
```



```
1, 0, 0, 1, 1,
1, 0, 1, 0, 0,
1, 0, 1, 0, 1,
1, 0, 1, 1, 0,
1, 0, 1, 1, 1,
1, 1, 0, 0, 0,
1, 1, 0, 0, 1,
1, 1, 0, 1, 0,
1, 1, 0, 1, 1,
1, 1, 1, 0, 0,
1, 1, 1, 0, 1,
1, 1, 1, 1, 0,
1, 1, 1, 1, 1), .Dim = c(32, 5))
)

# initial values
list(
  beta = c(-1.163685322, 0.44376115, 1.121494003, 0.165095519, -0.566665352),
  mut = 0, taut = 1,
  mub = 0, taub = 1,
  theta = c(-2.1972246, -2.1972246, -2.1972246, -1.3862944, -1.3862944,
  .
  .
  .
  2.1972246) # 1000 initial theta values
)
```