Proceedings of the 89th Annual International Meeting of the Psychometric Society, Prague, Czech Republic (2024).

Some Posterior Standard Deviations of the Graded Response Model

Seock-Ho Kim*

Department of Educational Psychology, University of Georgia, Athens, GA 30602, USA
*Corresponding author. Email: shkim@uga.edu

Abstract

The procedures required to obtain the approximate posterior standard deviations of the parameters of the graded response model for polytomously scored items from Bayes modal estimation are described and used to generate values for some common situations. The results were compared with those obtained from maximum likelihood estimation. It is shown that the use of priors may reduce the instability of Bayes modal estimates of the item parameters assuming the choice of priors is reasonable. The sample size required for acceptable accuracy for purposes of practical applications of item response theory may be inferred from the tables or the computer code in R. It is suggested that careful selection and specifications of priors be exercised to obtain the required precision when the graded response model is used in these situations.

Keywords: Bayes modal estimation, graded response model, item response theory, posterior standard deviations

1. Introduction

Thissen and Wainer (1982) presented the magnitudes of the asymptotic standard errors of estimated item parameters in the three commonly used IRT models for binary items. The standard errors were obtained using maximum likelihood estimation and presented in a convenient form assuming that one examinee, randomly sampled from a population of a Gaussian distribution with zero mean and unit variance, responded to the item. Based on the required precision and the values of item parameters, the readers can find a minimum sample size to achieve the accurate estimation of item parameters. Kim (2023) extended Thissen and Wainer's approach to the graded response model (Samejima, 1969, 1972), one of the IRT models for polytomous items.

The effect of the priors of item parameters on the approximate posterior standard deviations from Bayes modal estimation for the three IRT models for binary items is presented in Kim (2007). Such effect is not well known, however, for the IRT models for polytomous items. A number of different forms of priors have been proposed in the estimation of item parameters of the graded response model (see Kim, 2024, and references there in) and implemented in various IRT computer programs, for example, PARSCALE (Muraki & Bock, 1993, 2002; see also du Toit, 2003), MULTILOG (Thissen, 1991; Thissen, Chen, & Bock, 2002), and flexMIRT (Houts & Cai, 2013). This paper used the priors in the computer program PARSCALE.

2. Estimation and Posterior Standard Deviations

2.1 Bayes Modal Estimation

If the item response model is a function of ability θ_i for an examinee i and a set of item parameters ξ, the probability of a response k to the item can be written as

$$P_{ik} = P(r_i = k|\theta_i, \xi),$$

where r_i is the examinee's response. The likelihood function of the observed responses for N independent examinees is (with an indicator variable $u_{ik} = 1$ for $r_i = k$ and $u_{ik} = 0$ otherwise)

$$L = \prod_{i=1}^{N} \prod_{k=1}^{K} P_{ik}^{u_{ik}}.$$

Let $l = \sum_{i=1}^{N} \sum_{k=1}^{K} u_{ik} \log P_{ik}$ be the log likelihood and $\pi = \log p(\xi)$ be the log prior of ξ. From Bayes theorem, the log posterior is proportional to the log likelihood and the log prior,

$$\text{log posterior} \propto l + \pi = \sum_{i=1}^{N} \sum_{k=1}^{K} u_{ik} \log P_{ik} + \log p(\xi).$$

The Bayes modal estimates of item parameters in ξ are located where the partial derivatives

$$\frac{\partial(l+\pi)}{\partial \xi} = \sum_{i=1}^{N} \sum_{k=1}^{K} \frac{u_{ik}}{P_{ik}} \frac{\partial P_{ik}}{\partial \xi} + \frac{\partial \pi}{\partial \xi}$$

are zero.

If the log posterior has a well-defined maximum, the degree of concavity would determine the sampling variances of the modal estimates. The second partial derivatives of the log posterior can be obtained for any parameter ξ_s and ξ_t. The second partial derivatives of the log posterior have the form

$$\frac{\partial^2 l}{\partial \xi_s \partial \xi_t} + \frac{\partial^2 \pi}{\partial \xi_s \partial \xi_t} = \sum_{i=1}^{N} \sum_{k=1}^{K} \left\{ \frac{u_{ik}}{P_{ik}} \frac{\partial^2 P_{ik}}{\partial \xi_s \partial \xi_t} + \frac{\partial P_{ik}}{\partial \xi_s} \left(\frac{-u_{ik}}{P_{ik}^2} \right) \frac{\partial P_{ik}}{\partial \xi_t} \right\} + \frac{\partial^2 \pi}{\partial \xi_s \partial \xi_t}. \tag{1}$$

The inverse of the negative value of the matrix of second derivatives of the log posterior approximates the asymptotic variance and covariance matrix of the Bayes modal estimates (Berger, 1985; Mislevy, 1986; Tsutakawa & Johnson, 1990).

For the likelihood part, the density of θ with the subscript i omitted is taken to be $\phi(\theta)$, and we obtain the expectation of the second partial derivatives by substituting P_k for u_k with the subscript i omitted. We simplify and integrate, thus obtaining

$$-E\left(\frac{\partial^2 l}{\partial \xi_s \partial \xi_t}\right) = N \int_{-\infty}^{\infty} \sum_{k=1}^{K} \left\{ \left(\frac{1}{P_k}\right) \left(\frac{\partial P_k}{\partial \xi_s} \frac{\partial P_k}{\partial \xi_t}\right) \right\} \phi(\theta) d\theta. \tag{2}$$

The partial derivatives of P_k with respect to its parameters and the specification of $\phi(\theta)$ are needed for the evaluation of Equation 2. The density of θ, $\phi(\theta)$, is assumed to be Gaussian with mean zero and variance one.

2.2 Derivatibes of Item Response Functions

The graded response model in the logistic form for $k = 1, \ldots K$ is

$$P_k = P_{k-1}^* - P_k^*,$$

where

$$P^*_{k-1} = \frac{1}{1 + \exp\{-a(\theta - b_{k-1})\}}$$

and

$$P^*_k = \frac{1}{1 + \exp\{-a(\theta - b_k)\}}$$

in which a is the slope parameter of the boundary response function, and b_{k-1} and b_k are the location parameters. Note that $P^*_0 = 1$ (i.e., $b_0 \equiv -\infty$) and $P^*_K = 0$ (i.e., $b_K \equiv \infty$) by definition.

The partial derivatives of the graded response model for the respective parameters are

$$\frac{\partial P_k}{\partial a} = \frac{\partial P^*_{k-1}}{\partial a} - \frac{\partial P^*_k}{\partial a} = P^*_{k-1} Q^*_{k-1}(\theta - b_{k-1}) - P^*_k Q^*_k (\theta - b_k)$$

and

$$\frac{\partial P_k}{\partial b_k} = \frac{\partial P^*_{k-1}}{\partial b_k} - \frac{\partial P^*_k}{\partial b_k} = P^*_k Q^*_k(-a),$$

where $Q^*_{k-1} = 1 - P^*_{k-1}$ and $Q^*_k = 1 - P^*_k$.

With $w_k = P^*_k Q^*_k$ we can obtain:

$$-E\left(\frac{\partial^2 l}{\partial a^2}\right) = N \int_{-\infty}^{\infty} \sum_{k=1}^{K} \left[\left(\frac{1}{P_k}\right) \{w_{k-1}(\theta - b_{k-1}) - w_k(\theta - b_k)\}^2\right] \phi(\theta) d\theta$$

$$-E\left(\frac{\partial^2 l}{\partial b_k^2}\right) = N \int_{-\infty}^{\infty} (-a w_k)^2 \left(\frac{1}{P_k} + \frac{1}{P_{k+1}}\right) \phi(\theta) d\theta$$

$$-E\left(\frac{\partial^2 l}{\partial b_{k-1} \partial b_k}\right) = -N \int_{-\infty}^{\infty} a^2 w_{k-1} w_k \left(\frac{1}{P_k}\right) \phi(\theta) d\theta$$

$$-E\left(\frac{\partial^2 l}{\partial b_k \partial b_{k+1}}\right) = -N \int_{-\infty}^{\infty} a^2 w_k w_{k+1} \left(\frac{1}{P_{k+1}}\right) \phi(\theta) d\theta$$

$$-E\left(\frac{\partial^2 l}{\partial a \partial b_k}\right) = -N \int_{-\infty}^{\infty} a(\theta - b_k) w_k \left(\frac{(w_{k-1} - w_k)}{P_k} - \frac{(w_k - w_{k+1})}{P_{k+1}}\right) \phi(\theta) d\theta$$

The similar derivatives are given by Baker (1992, pp. 233–236) and elsewhere.

2.3 Derivatives of Priors

For the graded response model, a lognormal distribution (Hogg & Craig, 1978, p. 180) with $\mu_{\log a}$ and $\sigma^2_{\log a}$ for a,

$$p(a) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma_{\log a}} \exp\left[-\dfrac{1}{2}\left(\dfrac{\log a - \mu_{\log a}}{\sigma_{\log a}}\right)^2\right] \dfrac{1}{a} & \text{if } 0 < a < \infty \\ 0 & \text{otherwise,} \end{cases}$$

can be employed in PARSCALE. The second derivative of the log prior for a is

$$\frac{\partial^2 \log p(a)}{\partial a^2} = -\frac{1 - (\log a - \mu_{\log a} + \sigma^2_{\log a})}{\sigma^2_{\log a} a^2}.$$

A normal distribution (Hogg & Craig, 1978, p. 111) with μ_{b_k} and $\sigma^2_{b_k}$ can be taken as the prior for b_k in PARSCALE,

$$p(b_k) = \frac{1}{\sqrt{2\pi}\sigma_{b_k}} \exp\left[-\frac{1}{2}\left(\frac{b_k - \mu_{b_k}}{\sigma_{b_k}}\right)^2\right] \quad \text{for } -\infty < b_k < \infty.$$

The second partial derivative of the log prior for b_k is

$$\frac{\partial^2 \log p(b_k)}{\partial b_k^2} = -\frac{1}{\sigma^2_{b_k}}.$$

It is assumed later that σ^2_b are the same for all k.

3. Method

Since priors do not depend on the N examinees, the magnitudes of the approximate posterior standard deviations cannot be tabulated in a convenient form as in Thissen and Wainer (1982) and in Kim (2023). The approximate posterior standard deviations depend on both N and priors. The contribution of the priors to the magnitude of the approximate posterior standard deviations will be affected by sample size rather unsystematically because of the inversion of the matrix of the second derivatives. A set of R code was developed to obtain the approximate posterior standard deviations for the graded response model with two to nine categories.

To compare the asymptotic standard errors from the maximum likelihood method and the approximate posterior standard deviations, two sample sizes (N = 100 and N = 1000) are used to obtain the main results presented in the next section. Other interesting results from different sample sizes are also reported. For the graded response model with k categories, the priors were a lognormal prior on a with $\mu_{\log a} = 0$ and $\sigma^2_{\log a} = 0.5^2$ and normal priors on b_k with $\mu_{b_k} = 0$ and $\sigma^2_{b_k} = 2^2$.

4. Results

The results obtained through the use of Equation 1 (i.e., posterior standard deviations) and Equation 2 (i.e., standard errors) could be shown for various situations. Table 1 is presented here as a summary (n.b., Tables 2–9 are not presented here but available from the author): For example, Tables 2–3 for the posterior standard deviations for the graded response model with two categories (GR2) with 100 and 1000 examinees, Tables 4–5 for the standard errors for GR2 with 100 and 1000 examinees, Tables 6–7 for the posterior standard deviations for GR3 with 100 and 1000 examinees, and Tables 8–9 for the standard errors for GR3 with 100 and 1000 examinees. Other possible tables for GR4 to GR9 are too voluminous to present. Note that results from Tables 2–9 are summarized with Figures 1–5.

Shown in Table 1 are the sample sizes required to achieve one decimal place of accuracy for location and slope for the graded response model with two to nine categories (GR2–GR9) when slope is 1.5 and the location range −2 to +2 with increment of 0.5. The posterior standard deviation of the order of 0.05 was considered for accuracy. The sample sizes from maximum likelihood estimation were also reported in Table 1 based on Kim (2023). The standard error of the order of 0.05 was considered for accuracy.

The approximate sample sizes to achieve one decimal place of accuracy for parameters in the graded response model from Bayes modal estimation are based on the the results from the R code for GR2–GR9. The sample sizes required to achieve one decimal place of accuracy for GR2 from Bayes modal estimation when slope is 1.5 and the location range −2 to +2 are 7493 for location and 8897 for slope. The sample sizes required to achieve one decimal place of accuracy for GR2 from maximum

Table 1. *Sample Size Required to Achieve One Decimal Place of Accuracy for Location and Slope for the Graded Response Model with Two to Nine Categories (GR2–GR9) When Slope Is 1.5 and the Locations Range –2 to +2 with Increment of* 0.5 *from Bayes Modal Estimation with Normal Priors of* $N(0, 2^2)$ *on Location and Lognormal Prior of* $LN(0, 0.5^2)$ *on Slope and from Maximum Likelihood Estimation (MLE)*

Model	Parameter	Sample Size	MLE Sample Size	Example Set (*a*, *b*'s)
GR2	Location	7493	7500	(1.5, -2.0)
	Slope	8897	8912	(1,5, -2.0)
GR3	Location	4026	4020	(1.5, -2.0, 2.0)
	Slope	4640	4652	(1.5, -2.0, -1.5)
GR4	Location	3414	3411	(1.5, -2.0, -1.5, 2.0)
	Slope	2844	2852	(1.5, -2.0, -1.5, -1.0)
GR5	Location	3024	3025	(1.5, -2.0, -1.5, -1.0, 2.0)
	Slope	1945	1954	(1.5, -2.0, -1.5, -1.0, -0.5)
GR6	Location	2777	2788	(1.5, -2.0, -1.5, -1.0, -0.5, 2.0)
	Slope	1450	1460	(1.5, -2.0, -1.5, -1.0, -0.5, 0.0)
GR7	Location	2623	2622	(1.5, -2.0, -1.5, -1.0, -0.5, 0.0, 2.0)
	Slope	1163	1170	(1.5, -2.0, -1.5, -1.0, -0.5, 0.0, 0.5)
GR8	Location	2494	2500	(1.5, -2.0, -1.5, -1.0, -0.5, 0.0, 1.0, 2.0)
	Slope	991	986	(1.5, -2.0, -1.5, -1.0, -0.5, 0.0, 0.5, 1.0)
GR9	Location	2483	2481	(1.5, -2.0, -1.5, -1.0, -0.5, 0.0, 0.5, 1.0, 2.0)
	Slope	888	889	(1.5, -2.0, -1.5, -1.0, -0.5, 0.0, 0.5, 1.0, 1.5)

Note. Values of the MLE sample size are from Kim (2023).

likelihood estimation when slope is 1.5 and the location range –2 to +2 are 7500 for location and 8912 for slope.

The sample sizes required to achieve one decimal place of accuracy for GR3 from Bayes modal estimation when slope is 1.5 and the location range –2 to +2 are 4026 for location and 4640 for slope. The sample sizes required to achieve one decimal place of accuracy for GR3 from maximum likelihood estimation when slope is 1.5 and the location range –2 to +2 are 4020 for location and 4652 for slope.

The required sample sizes are gradually reduced when the number of categories of the item increases. Both Bayes modal estimation and maximum likelihood estimation are expected to use very large but nearly the same sample sizes for the one decimal place of accuracy of parameter estimates, respectively, assessed by the size of posterior standard deviations and the size of standard errors.

Let us compare the posterior standard deviations and the standard errors of location for GR2 for two sample sizes (N = 100 and N = 1000) in order to get a better feel for the sample size issue and the estimation method issue. Shown in Figure 1 are the Bayes modal posterior standard deviations and the maximum likelihood standard errors of location (i.e., left figure) and slope (i.e., right figure) for GR2 when slope is 1.5 which is a representative value for most serious testing applications. In the left figure for a sample of 100 for the location range –3 to +3, the posterior standard deviations of location are smaller, sometimes trivially so in the middle of the range, than the standard errors of location. For a sample size of 100 for the location range –1 to +1 both posterior standard deviations of location and standard errors of location are below of the order of 0.2. For a sample of 1000 for the location range –3 to +3 the posterior standard deviations of location are the same as the standard errors of location. For a sample of 1000 for the location range –2.5 to +2.5 both posterior standard deviations of location and standard errors of location are below of the order of 0.2. As inferred from Table 1, for the sample size of 1000 both posterior standard deviations of location and standard errors of location are much larger than of the order of 0.05 for the entire location range.

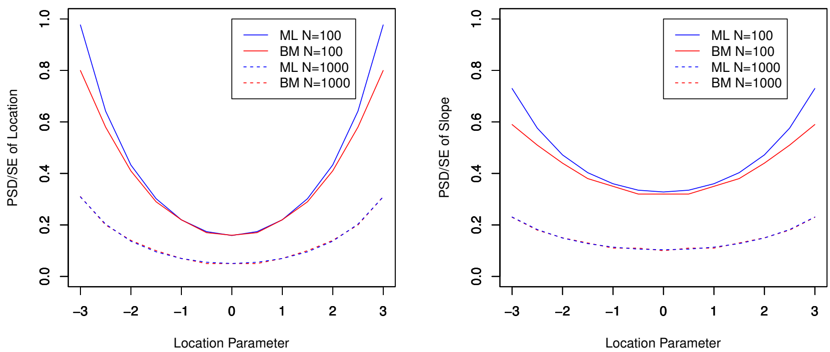

Figure 1. Bayes Modal Posterior Standard Deviations (PSDs) in Red and Maximum Likelihood Standard Errors (SEs) in Blue of Location (Left Figure) and Slope (Right Figure) for the Graded Response Model with Two Categories When Slope Is 1.5 with Sample Sizes of 100 and 1000

In the right figure for a sample of 100 for the location range –3 to +3, the posterior standard deviations of slope are smaller than the standard errors of location. For a sample of 100 for the location range –1 to +1 both posterior standard deviations of slope and standard errors of slope are below of the order of 0.4. For a sample of 1000 for the location range –3 to +3 the posterior standard deviations of slope are the same as the standard errors of slope. For a sample of 1000 for the location range –2.5 to +2.5 both posterior standard deviations of slope and standard errors of slope are below of the order of 0.2. For a sample size of 1000 none of the posterior standard deviations of slope and the standard errors of slope are below of the order of 0.1. As also inferred from Table 1, for the sample size of 1000 both posterior standard deviations of slope and standard errors of slope are much larger than of the order of 0.05 for the entire location range.

It is disquieting to see in Figure 1 that adequate estimates of location or slope cannot be obtained with the sample of 1000 regardless of whether Bayes modal estimation or maximum likelihood estimation is used. The sample size to achieve the one decimal accuracy of location or slope if the location range is extended to –3 and +3 would be more than 10000 which is as large a sample as plausible with only truly large scale assessment cases.

Let us examine three-dimensional plots in order to better understand the relationship that the slope and the location parameter for GR2 has with posterior standard deviation. Shown in Figure 2 are the posterior standard deviations taken from Table 2 of location (i.e., top left figure) and slope (i.e., top right figure), respectively, for GR2 with a sample of 100.

The perspective plot in the top left figure, conditioned on each slope value, shows that the posterior standard deviation of location reaches its minimum at a point where location is 0 which is the center of the ability distribution. All conditional functions of the posterior standard deviation of location are concave down with location 0 to have their minimum values. The plot also shows that the posterior standard deviation of location gradually decreases as the value of slope increases, except for the two extreme location regions.

The perspective plot in the top right figure, conditioned on each location value, shows that the posterior standard deviation of slope sharply increases as the value of slope increases. For relatively larger values of slope, the posterior standard deviation of slope reaches its minimum at a point where location is 0. The magnitude of the posterior standard deviations of slope gets larger when its value is large and location is far away from the center of the ability distribution. Nevertheless all posterior standard deviations of slope reach their minimum when location is 0.

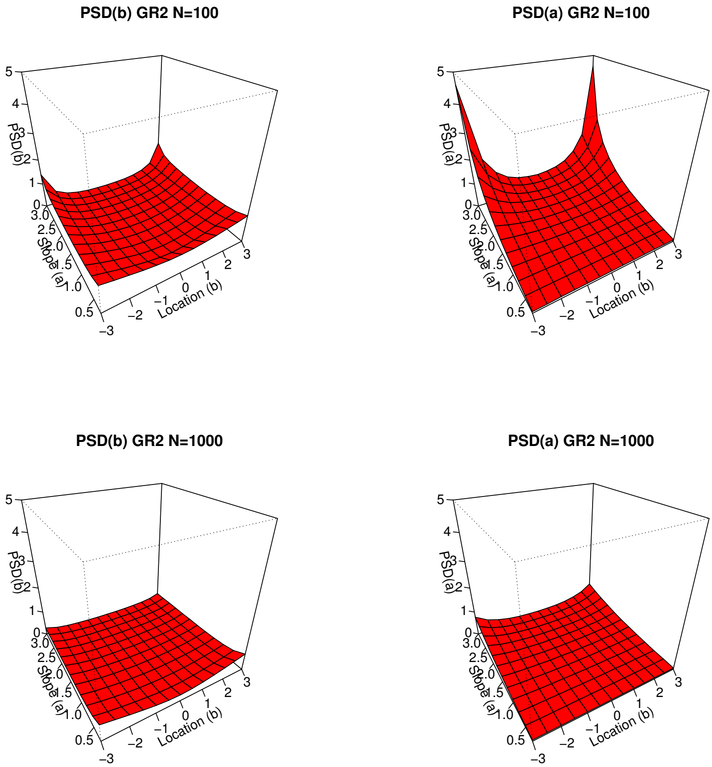

Figure 2. The Posterior Standard Deviation of Location (Left Side) and Slope (Right Side) Shown as a Bivariate Function of Slope and Location for the Graded Response Model with Two Categories with Sample Size 100 (Top Figures) and 1000 (Bottom Figures)

Shown also in Figure 2 are the posterior standard deviations taken from Table 3 of location (i.e., bottom left figure) and slope (i.e., bottom right figure), respectively, for GR2 with a sample of 1000. The general sizes of the posterior standard deviations from the two bottom figure are much smaller than those from the two top figures.

The perspective plot in the bottom left figure, conditioned on each slope value, shows that the posterior standard deviation of location reaches its minimum at a point where location is 0 which is the center of the ability distribution. All conditional functions of the posterior standard deviation of location are concave down with location 0 to have their minimum values. The plot also shows that the posterior standard deviation of location gradually decreases as the value of slope increases.

The perspective plot in the bottom right figure, conditioned on each location value, shows that the posterior standard deviation of slope gradually increases as the value of slope increases. For relatively larger values of slope, the posterior standard deviation of slope reaches its minimum at a point where location is 0. The magnitude of the posterior standard deviations of slope gets larger when its value is large and location is far away from the center of the ability distribution. All posterior standard deviations of slope reach their minimum when location is 0, but such a pattern is not really noticeable for $a = 0.5$.

In order to compare the posterior standard deviation from Bayes modal estimation and the standard errors from maximum likelihood estimation, let us examine three-dimensional plots in

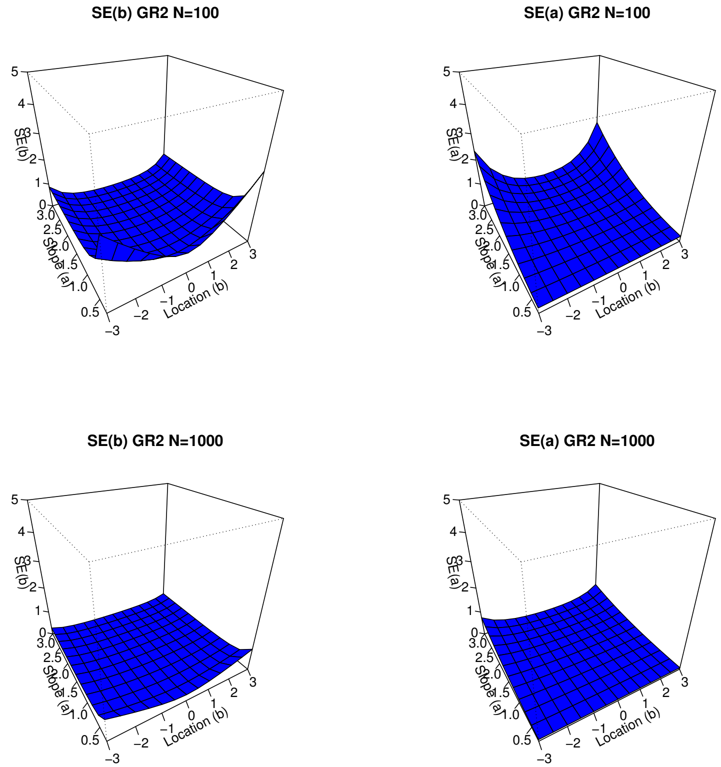

Figure 3. The Standard Error of Location (Left Side) and Slope (Right Side) Shown as a Bivariate Function of Slope and Location for the Graded Response Model with Two Categories with Sample Size 100 (Top Figures) and 1000 (Bottom Figures)

order to better understand the relationship that the slope and the location parameter for GR2 has with standard error. Shown in Figures 3 are the standard errors taken from Table 4 of location (i.e., top left figure) and slope (top right figure) for GR2 with a sample of 100.

The perspective plot in the top left figure, conditioned on each slope value, shows that the standard error of location reaches its minimum at a point where location is 0 which is the center of the ability distribution. All conditional functions of the standard error of location are concave down with location 0 to have their minimum values. The plot also shows that the standard error of location gradually decreases as the value of slope increases.

The perspective plot in the top right figure, conditioned on each location value, shows that the standard error of slope gradually increases as the value of slope increases. For relatively larger values of slope, the standard error of slope reaches its minimum at a point where location is 0. The magnitude of the standard errors of slope gets larger when its value is large and location is far away from the center of the ability distribution. All standard errors of slope reach their minimum when location is 0, but such a pattern is not really noticeable for a = 0.5.

Shown also in Figure 3 are the standard errors taken from Table 5 of location (i.e., bottom left figure) and slope (i.e., bottom right figure), respectively, for GR2 with a sample of 1000. The general sizes of the standard errors from the two bottom figures are much smaller than those from the two top figures.

The perspective plot in the bottom left figure, conditioned on each slope value, shows that the standard error of location reaches its minimum at a point where location is 0 which is the center of the ability distribution. All conditional functions of the standard error of location are concave down with location 0 to have their minimum values. The plot also shows that the standard error of location gradually decreases as the value of slope increases.

The perspective plot in the bottom right figure, conditioned on each location value, shows that the standard error of slope gradually increases as the value of slope increases. For relatively larger values of slope, the standard error of slope reaches its minimum at a point where location is 0. The magnitude of the standard errors of slope gets larger when its value is large and location is far away from the center of the ability distribution. All standard errors of slope reach their minimum when location is 0, but such a pattern is not really noticeable for $a = 0.5$.

Understanding the magnitudes of posterior standard deviations of location and slope as in Figure 2 and the magnitudes of standard errors of location and slope as in Figure 3 in GR2 is important because similar patterns are observed for other graded response model with more response categories, although the magnitudes get relatively smaller as the number of categories increases. The posterior standard deviation of location and the standard error of location for GR2 when slope is 1.5 and location is 0 seems to provide the minimum value (i.e., the lower bound) for GR3–GR9 (n.b., trivially smaller values of posterior standard deviations of location and standard errors of location, respectively, might be observed). The posterior standard deviations of slope and the standard errors of slope for GR2 for all corresponding points on the location range −3 to +3 seem to provide the maximum values for GR3–GR9 (i.e., the upper bounds).

Comparing the two top figures of Figures 2–3 reveals that the posterior standard deviations from large values of slope are relatively larger than the standard errors from the corresponding values of slope for a sample of 100. For a sample of 100 the posterior standard deviations are relatively, trivially smaller than the standard errors if reasonable priors are employed (i.e., priors are matched with the respective parameters). Figures 2–3 show, however, that practically the same precision measures are obtained for a sample of 1000.

Shown in Figure 4 are the posterior standard deviations taken from Table 6 of location 1 (i.e., top left figure), location 2 (i.e., top middle figure), and slope (i.e., top right figure), respectively, for GR3 when slope is 1.5 with a sample of 100. Slope can be specified by any value from 0.25 to 3 with a sample of 100. But proportionally similar plots can be obtained to the three top figures.

Shown also in Figure 4 are the posterior standard deviations taken from Table 7 of location 1 (i.e., bottom left figure), location 2 (i.e., bottom middle figure), and slope (i.e., bottom right figure) for GR3 when slope is 1.5 with a sample of 1000. Again, slope can be specified by any value from 0.25 to 3 with a sample of 1000. But proportionally similar plots can be obtained to the three bottom figures. The general sizes of the posterior standard deviations from the three bottom figures are much smaller than those from the three top figures.

In order to compare the posterior standard deviation from Bayes modal estimation and the standard errors from maximum likelihood estimation, let us examine three-dimensional plots in order to better understand the relationship that the slope and the location parameter for GR3 has with standard error.

Shown in Figure 5 are the standard error taken from Table 8 of location 1 (i.e., bottom left figure), location 2 (i.e., bottom middle figure), and slope (i.e., bottom right figure), respectively, for GR3 when slope is 1.5 with a sample of 100. Slope can be specified by any value from 0.25 to 3 with a sample of 100. But proportionally similar plots can be obtained to the top three figures.

Shown in also Figure 5 are the standard error taken from Table 9 of location 1 (i.e., top left figure), location 2 (i.e., top middle figure), and slope (i.e., top right figure) for GR3 when slope is 1.5 with a sample of 1000. Again, slope can be specified by any value from 0.25 to 3 with a sample of 1000. But proportionally similar plots can be obtained to the three bottom figures. The general sizes

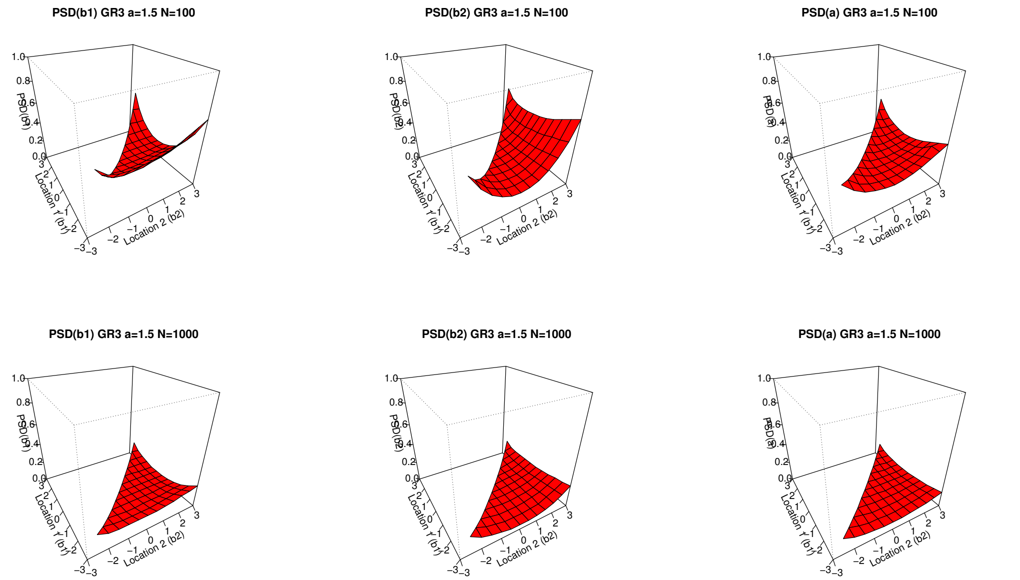

Figure 4. The Posterior Standard Deviation of Location 1 (Left Side), Location 2 (Middle), and Slope (Right Side) Shown as a Bivariate Function of Locaion 1 and Location 2 for the Graded Response Model with Three Categories When Slope is 1.5 with Sample Size 100 (Top Figures) and 1000 (Bottom Figures)

of the posterior standard deviations from the three top figures are much smaller than those from the the three top figures.

Figures 4–5 show that practically the same precision measures for location and slope are obtained for a sample of 100 as well as for a sample of 1000. Note that the value for the slope is 1.5 in Figures 4–5 as well as Tables 6–9.

Similar plots to that in Figure 4 for a sample of 100 and for a sample of 1000 for the posterior standard deviations as well as similar plots in that in Figure 5 for a sample of 100 and for a sample of 1000 for the standard errors may be constructed for GR4 and beyond.

For an example, a table can be obtained for GR4 for the posterior standard deviations of location 1, location 2, location 3, and slope with a sample of 100. Again, for brevity as in GR3, the value of slope can be fixed as 1.5. In addition, location 1 can be fixed as –3 in order to obtain such a table. Another table can be obtained for the posterior standard deviation of location 1, location 2, location 3, and slope with a sample of 1000. Again slope can be fixed as 1.5 and location 1 can also be fixed as –3 in order to obtain such a table. From the two tables, perspective plots and be constructed to make, say Figure 6 (n.b., not actually created). The x-dimension of each perspective plot designates the values of location 2 from the range –2.5 to +2.5, and the y-dimension designates the values of location 3 from the range –2 to +3. There would be four plots, that is, one for each parameter for GR4 for a sample of 100. A similar four perspective plots can be constructed for a sample of 1000.

Using the same logic, two tables can be construction for GR4 for the standard errors of location 1, location 2, location 3, and slope with a sample of 100 and with a sample of 1000, respectively. From the two tables, a total of eight perspective plots of standard errors can be constructed and combined to yield, say Figure 7 (n.b., not actually created).

We may find that the same general structure would continue to hold for the posterior standard deviations and the standard errors. As locations became close to 0 which is the center of the ability

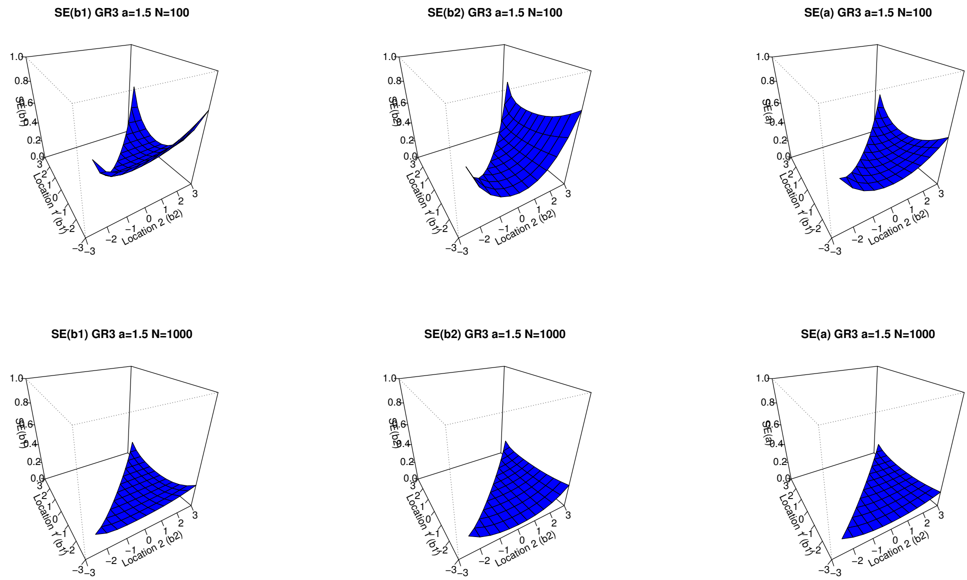

Figure 5. The Standard Error of Location 1 (Left Side), Location 2 (Middle), and Slope (Right Side) Shown as a Bivariate Function of Locaion 1 and Location 2 for the Graded Response Model with Three Categories When Slope is 1.5 with Sample Size 100 (Top Figures) and 1000 (Bottom Figures)

distribution, the posterior standard deviations of location and the standard errors of location get smaller. As locations became far away from 0 the posterior standard deviations of location and the standard errors of location get larger. Similar patterns would be observed for the posterior standard deviations of slope and the standard errors of slope. In addition, as slopes became more gradual in size, the posterior standard deviations of slope and the standard errors of slope get larger. As the number of locations increased the posterior standard deviations of slope and the standard errors of slope get smaller.

One coherent finding from the examination of the tables and figures is that the use of either Bayes modal estimation or maximum likelihood estimation for location and for slope of the graded response model yields results too inexact to be of practical use for a sample size less than, say, 2500. Both methods of estimation require samples of enormous size, especially for location beyond ± 2 the ability range so as to make it prohibitively expensive in terms of data collection. This problem arises for items with locations away from the center of the ability distribution. When an item has adjacent categories in extreme ends relatively not endorsed, selected, or assigned, there are few observations available to estimate the location parameters thus making the accompanied posterior standard deviations as well as the standard errors very large.

5. Discussion

Many different forms of priors (e.g., empirical, hierarchical, and multivariate normal) have been suggested in Bayesian inference for parameters in item response category functions of the graded response model (see Kim, 2024). Priors having different forms can certainly be used to estimate item parameters and to obtain approximate posterior standard deviations in Bayes modal estimation. In such a case we should obtain a different set of partial derivatives of the log priors.

Although the distributional forms of priors are the same as those used in this study, different prior parameters (i.e., hyperparameters) can be used in calculation of approximate posterior standard deviations. When the forms of priors are the same as ones presented earlier, the accompanying R code can be used to obtain approximate posterior standard deviations using various combinations of priors (e.g., lognormal for slope and uninformative uniform for location by commenting out the location portions of R code) and their parameter specifications. In this sense, attention should be given not only to the selection of the functional forms of priors but also to the values of prior parameters when a Bayesian approach is used.

There remain some issues for future research related to this paper and use of Bayesian inference in IRT. Some examples are prior forms and specifications, robustness, model selection, and posterior intervals. When using a reasonable sample size to calibrate items, however, it seems unlikely that totally different or incomparable results of item parameter estimates and variability measures will be obtained from different estimation methods.

References

Baker, F. B. (1992). *Item response theory: Parameter estimation techniques.* New York: Dekker.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (2nd ed.). New York: Springer-Verlag.

du Toit, M. (Ed.). (2003). *IRT from SSI.* Lincolnwood, IL: Scientific Software International.

Hogg, R. V., & Craig, A. T. (1978). *Introduction to mathematical statistics* (4th ed.). New York: Macmillan.

Houts, C. R., & Cai, L. (2013). *flexMIRT user's manual version 2.* Chapel Hill, NC: Vector Psychometric Group.

Kim, S.-H. (2007). Some posterior standard deviations in item response theory. *Educational and Psychological Measurement, 67*(2), 258-279.

Kim, S.-H. (2023). *Some standard errors of the graded response model* [Manuscript submitted for publication]. Department of Educational Psychology, University of Georgia.

Kim, S.-H. (2024). Priors in Bayesian estimation under the graded response model. In M. Wiberg, J.-S. Kim, H. Hwang, H. Wu, & T. Sweet (Eds.), *Quantitative psychology research* (pp. 13–23). Cham, Switzerland: Springer.

Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika, 51,* 177–195.

Muraki, E., & Bock, R. D. (1993). *PARSCALE: IRT based test scoring and item analysis for graded open-ended exercises and performance tasks.* Chicago, IL: Scientific Software International.

Muraki, E., & Bock, R. D. (2002). *PARSCALE: IRT based test scoring and item analysis for graded open-ended exercises and performance tasks.* Lincolnwood, IL: Scientific Software International.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement,* No. 17.

Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph Supplement,* No. 18.

Thissen, D. (1991). MULTILOG user's guide: Multiple, categorical item analysis and test scoring using item response theory [Computer software]. Chicago: Scientific Software.

Thissen, D., Chen, W.-H., & Bock, R. D. (2002). MULTILOG [Computer software]. Lincolnwood, IL: Scientific Software International.

Thissen, D., & Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika, 47,* 397–412.

Tsutakawa, R. K., & Johnson, J. C. (1990). The effect of uncertainly of item parameter estimation on ability estimates. *Psychometrika, 55,* 371–390.