Proceedings of the 90th Annual International Meeting of the Psychometric Society, Minneapolis, United States (2025).

Some Standard Errors of the Graded Response Model

Seock-Ho Kim *

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

Abstract

A procedure to calculate the asymptotic standard errors of the estimated parameters of the graded response model in the logistic form is described. The asymptotic standard errors computed from the sampling variances of maximum likelihood estimates are tabulated for some common situations. It is shown that maximum likelihood estimation for the graded response model requires a large sample to achieve accuracy to one decimal place of the standard errors of the estimated parameters. The required sample size is related to the number of categories employed in the graded response item.

Keywords: asymptotic standard errors, graded response model, maximum likelihood estimation

1. Introduction

In typical applications of item response theory (IRT), such as computerized adaptive testing, metric linking or equating, and test scoring, item parameter estimates are treated as known and true values. The uncertainty of item parameter estimates, however, should be assessed to make correct inference for examinees, for example, in order not to produce overly optimistic ability estimates. In adaptive testing where sequential item selection is based on item information, imprecision leads to inappropriate branching of items and premature termination of testing. Under maximum likelihood estimation, the uncertainty in estimates is expressed in terms of asymptotic standard errors.

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. In this study, Thissen and Wainer's approach is extended to the graded response model (Samejima, 1969, 1972), one of the IRT models for polytomous items.

Estimation of item parameters of the graded response model has been implemented in various IRT computer programs, including PARSCALE (Muraki & Bock, 1993, 2002; see also du Toit, 2003), MULTILOG (Thissen, 1991; Thissen, Chen, & Bock, 2002), and flexMIRT (Houts & Cai, 2013). This paper considers marginal maximum likelihood estimation of item parameters.

2. Estimation and Standard Errors

2.1 Maximum Likelihood Estimation

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

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

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)

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

The log likelihood l is

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

The maximum likelihood estimates of item parameters in ξ are located where the partial derivatives

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

are zero.

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

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

The inverse of the negative value of the matrix of second derivatives of the log likelihood approximates the asymptotic variance and covariance matrix of the maximum likelihood estimates (Kendall & Stuart, 1979, p. 59).

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 Derivatives 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., $b_0 \to -\infty$) and $P_K^* = 0$ (i.e., $b_K \to \infty$) 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

$$\frac{\partial P_k}{\partial b_k} = \frac{\partial P_{k-1}^*}{\partial b_k} - \frac{\partial P_k^*}{\partial b_k} = P_k^* Q_k^* (-a).$$

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.

For the graded response model with any number of categories, specified values of the parameters, and a given sample size, the derivatives may be substituted in Equation 2 to give elements of the information matrix. The information matrix may be inverted to give the asymptotic variance-covariance matrix of the estimated parameters. The square roots of the diagonal elements of the inverse are the asymptotic standard errors of the estimated parameters. In the following sections, we will present implementation of the procedure to obtain the standard errors of the graded response model. We will then tabulate and discuss those standard errors of the graded response model as functions of the estimated parameters.

3. Method

The magnitudes of the approximate standard errors can be tabulated in a convenient form as in Thissen and Wainer (1982). A set of R code was developed to obtain the approximate standard errors for the graded response model with two to nine categories (GR2–GR9). An example of the R function for the graded response model with two categories (i.e., `gr2()`) is presented in Appendix 1. The function was written for a single item and additional lines were coded in R for GR2.

Gaussian quadrature formulas were used in the ability integration. The 20 quadrature points and quadrature weights with 14 decimal places were based on Stroud and Secrest (1966, p. 220, Table 5 of the Gauss-Hermite formulas). Stroud and Secrest (1966) contained numbers with 30 decimal places because the Fortran program they used had capability of handling numbers in triple precision and resulting numbers are correct up to 38 significant decimals. The quadrature points are multiplied by $\sqrt{2}$, and the quadrature weights are divided by $\sqrt{\pi}$ with π defined upto 14 decimal places in order to obtain the standard normal density (Bock & Lieberman, 1970, p. 189).

For the R functions, no protections are imposed to prevent some potential computational errors in programming (e.g., overflow or underflow). In order to obtain results with higher precision (e.g., use of the 30 decimal places for the Gauss-Hermite formulas), the procedure should be implemented with other lower-level computer languages equipped with better handling of numerical precision.

To assess the sample size effect on the asymptotic standard errors from the maximum likelihood method, five sample sizes (N = 100, 400, 900, 2000, 10000) are used to obtain the results presented for GR2. Other interesting results from GR2–GR9 with different sample sizes are also reported.

4. Results

The results obtained through the use of Equation 2 (i.e., standard errors) are shown for various situations in Tables 1 to 5 (n.b., Tables 4 and 5 are not presented here but available from the author or from `https://github.com/seockhokim/imps25.git`): Table 1 as a summary, Table 2 for the standard errors for GR2, Table 3 for the standard errors for GR3 with a = 1.5, Table 4 for the standard errors for GR4 with a = 1.5, and Table 5 for the standard errors for GR5 with a = 1.5. Other possible tables for GR6 to GR9 are too voluminous to present. Note that results from Tables 2 and 3 are summarized with Figures 2 and 3.

Shown in Table 1 are the sample sizes required to achieve accuracy to one decimal place for location and slope for GR2–GR9 when slope is 1.5 and the locations range –2 to +2 with increment of 0.5. The standard error of the order of 0.05 was considered for accuracy.

Table 1. Sample Size Required to Achieve Accuracy to One Decimal Place 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

Model	Parameter	Calculation	Sample Size	Example Set (a, b's)
GR2	Location	$(4.33/0.05)^2$	7500	(1.5, -2)
	Slope	$(4.72/0.05)^2$	8912	(1,5, -2)
GR3	Location	$(3.17/0.05)^2$	4020	(1.5, -2, 2)
	Slope	$(3.41/0.05)^2$	4652	(1.5, -2, -1.5)
GR4	Location	$(2.92/0.05)^2$	3411	(1.5, -2, -1.5, 2)
	Slope	$(2.67/0.05)^2$	2852	(1.5, -2, -1.5, -1)
GR5	Location	$(2.75/0.05)^2$	3025	(1.5, -2, -1.5, -1, 2)
	Slope	$(2.21/0.05)^2$	1954	(1.5, -2, -1.5, -1, -0.5)
GR6	Location	$(2.64/0.05)^2$	2788	(1.5, -2, -1.5, -1, -0.5, 2)
	Slope	$(1.91/0.05)^2$	1460	(1.5, -2, -1.5, -1, -0.5, 0)
GR7	Location	$(2.56/0.05)^2$	2622	(1.5, -2, -1.5, -1, -0.5, 0, 2)
	Slope	$(1.71/0.05)^2$	1170	(1.5, -2, -1.5, -1, -0.5, 0, 0.5)
GR8	Location	$(2.50/0.05)^2$	2500	(1.5, -2, -1.5, -1, -0.5, 0, 1, 2)
	Slope	$(1.57/0.05)^2$	986	(1.5, -2, -1.5, -1, -0.5, 0, 0.5, 1)
GR9	Location	$(2.49/0.05)^2$	2481	(1.5, -2, -1.5, -1, -0.5, 0, 0.5, 1, 2)
	Slope	$(1.49/0.05)^2$	889	(1.5, -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5)

The approximate sample sizes in Table 1 to achieve accuracy to one decimal place for parameters in the graded response model from maximum likelihood estimation are based on the results from the R code for GR2–GR9. The sample sizes required to achieve accuracy to one decimal place for GR2 from maximum likelihood estimation when slope is 1.5 for the location range –2 to +2 are 7500 for location and 8912 for slope. Note that for the example set of a = 1.5 and b = –2, the sample size of 7500 for location was computed from $(4.33/0.05)^2$ for which 4.33 was the value obtained from the R function for the asymptotic standard error of one examinee. Similarly, the sample size of 8912 for slope was based on $(4.72/0.05)^2$ for which 4.72 was the value obtain from the R function. The

sample sizes required to achieve accuracy to one decimal place for GR3 from maximum likelihood estimation when slope is 1.5 for 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. Maximum likelihood estimation seems to require very large sample sizes for the accuracy to one decimal place of parameter estimates, assessed by the size of standard errors.

Before discussing Tables 2–5 in detail a few general remarks are in order. First, Tables 2–5 omit the factor $N^{-1/2}$. In order to obtain the asymptotic standard error for any particular test administration one must divide the number in the appropriate table by $N^{1/2}$. Second, the slope parameter used here is for a logistic function. If the user is interested in transforming this to be comparable to a normal ogive, divide the expressed values of the slope by 1.702. Third, Tables 2–5 are presented for reference purposes since they may span most situations found in practice. The parameter values are chosen to represent the range encountered in testing situations. The R functions can be used to obtain the standard errors for other parameter values, not shown in these tables.

For GR2 in Table 2 the values for the slope parameter range from 0.25 which would be unacceptably low for most applications, to 3 which is very high with increment of 0.25. The location range of –3 to +3 with increment of 0.5 spans the reasonable range for a fixed test. One can safely interpolate between these values for intermediate ones.

Table 2. Minimal Asymptotic Standard Errors ($\times N^{-1/2}$) for Locations (Upper Cell) and Slopes (Lower Cell) for the Graded Response Model with Two Categories

	Location												
	-3.00	-2.50	-2.00	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50	2.00	2.50	3.00
Slope													
0.25	26.85	22.43	18.34	14.62	11.40	9.00	8.06	9.00	11.40	14.62	18.34	22.43	26.85
	2.18	2.14	2.11	2.08	2.06	2.05	2.05	2.05	2.06	2.08	2.11	2.14	2.18
0.50	15.45	12.35	9.75	7.57	5.81	4.58	4.12	4.58	5.81	7.57	9.75	12.35	15.45
	2.68	2.52	2.39	2.30	2.23	2.19	2.18	2.19	2.23	2.30	2.39	2.52	2.68
0.75	12.20	9.26	7.00	5.27	3.98	3.14	2.83	3.14	3.98	5.27	7.00	9.26	12.20
	3.45	3.09	2.82	2.62	2.48	2.40	2.38	2.40	2.48	2.62	2.82	3.09	3.45
1.00	10.83	7.81	5.67	4.14	3.08	2.43	2.20	2.43	3.08	4.14	5.67	7.81	10.83
	4.47	3.82	3.35	3.02	2.80	2.67	2.63	2.67	2.80	3.02	3.35	3.82	4.47
1.25	10.15	6.98	4.87	3.47	2.55	2.01	1.83	2.01	2.55	3.47	4.87	6.98	10.15
	5.74	4.71	3.99	3.49	3.17	2.99	2.94	2.99	3.17	3.49	3.99	4.71	5.74
1.50	9.77	6.42	4.33	3.02	2.20	1.74	1.59	1.74	2.20	3.02	4.33	6.42	9.77
	7.30	5.76	4.72	4.03	3.60	3.35	3.28	3.35	3.60	4.03	4.72	5.76	7.30
1.75	9.51	6.02	3.94	2.69	1.95	1.54	1.41	1.54	1.95	2.69	3.94	6.02	9.51
	9.14	6.97	5.55	4.63	4.06	3.75	3.65	3.75	4.06	4.63	5.55	6.97	9.14
2.00	9.32	5.69	3.63	2.44	1.76	1.40	1.28	1.40	1.76	2.44	3.63	5.69	9.32
	11.31	8.35	6.47	5.29	4.57	4.19	4.06	4.19	4.57	5.29	6.47	8.35	11.31
2.25	9.15	5.42	3.39	2.25	1.61	1.29	1.19	1.29	1.61	2.25	3.39	5.42	9.15
	13.81	9.89	7.50	6.01	5.12	4.66	4.50	4.66	5.12	6.01	7.50	9.89	13.81
2.50	8.99	5.18	3.18	2.09	1.50	1.20	1.11	1.20	1.50	2.09	3.18	5.18	8.99
	16.67	11.61	8.62	6.79	5.71	5.16	4.96	5.16	5.71	6.79	8.62	11.61	16.67
2.75	8.84	4.97	3.01	1.97	1.40	1.12	1.04	1.12	1.40	1.97	3.01	4.97	8.84
	19.90	13.48	9.84	7.61	6.34	5.70	5.44	5.70	6.34	7.61	9.84	13.48	19.90
3.00	8.69	4.77	2.86	1.86	1.32	1.06	0.99	1.06	1.32	1.86	2.86	4.77	8.69
	23.52	15.51	11.18	8.47	7.00	6.27	5.94	6.27	7.00	8.47	11.18	15.51	23.52

For GR3 to GR5 in Tables 3 to 5 the value for slope is 1.5, which is a representative value for most serious testing applications. For the graded response item with K categories, there are $(K-1)$ ordered location parameters and one slope parameter. The values for the first location parameter (location

1) in Tables 2–5 range –3 to $(4 - K/2)$ where K is the number of categories because of increment of 0.5. The range $(-4 + K/2)$ to 3 is the range of the last, which is the $(K - 1)$th location parameter. The values of the kth location parameter in general for which $k = 1, \ldots, (K - 1)$ is $(-3.5 + k/2)$ to $(3.5 - (K - k)/2)$ in these tables. Rather than try to discuss all aspects of these findings separately we will make some general remarks, and then move on to the discussion of certain details of interest.

Table 3. Minimal Asymptotic Standard Errors ($\times N^{-1/2}$) for Location 1s (Upper Cell), Location 2s (Middle Cell), and Slopes (Lower Cell) for the Graded Response Model with Three Categories When Slope Is 1.5

Slope = 1.5												
	Location 2											
	-2.50	-2.00	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50	2.00	2.50	3.00
Location 1												
-3.00	6.41	5.16	4.72	4.60	4.62	4.72	4.89	5.13	5.47	5.91	6.47	7.14
	5.90	3.90	2.76	2.09	1.72	1.58	1.67	2.01	2.63	3.62	5.09	7.14
	5.02	3.82	3.13	2.73	2.54	2.50	2.59	2.79	3.12	3.58	4.16	4.84
-2.50		4.40	3.70	3.49	3.45	3.51	3.62	3.79	4.02	4.32	4.68	5.09
		4.03	2.78	2.09	1.72	1.58	1.66	1.95	2.51	3.39	4.68	6.47
		4.06	3.16	2.66	2.41	2.32	2.37	2.54	2.82	3.19	3.66	4.16
-2.00			3.13	2.76	2.67	2.68	2.74	2.85	2.99	3.17	3.39	3.62
			2.86	2.11	1.73	1.59	1.64	1.90	2.39	3.17	4.32	5.91
			3.41	2.72	2.36	2.20	2.19	2.31	2.52	2.83	3.19	3.58
-1.50				2.34	2.16	2.13	2.15	2.21	2.29	2.39	2.51	2.63
				2.15	1.74	1.59	1.62	1.85	2.29	2.99	4.02	5.47
				2.98	2.44	2.17	2.09	2.13	2.29	2.52	2.82	3.12
-1.00					1.86	1.80	1.79	1.81	1.85	1.90	1.95	2.01
					1.75	1.59	1.62	1.81	2.21	2.85	3.79	5.13
					2.72	2.28	2.08	2.05	2.13	2.31	2.54	2.79
-0.50						1.62	1.61	1.62	1.62	1.64	1.66	1.67
						1.59	1.61	1.79	2.15	2.74	3.62	4.89
						2.60	2.23	2.08	2.09	2.19	2.37	2.59
0.00							1.59	1.59	1.59	1.59	1.58	1.58
							1.62	1.80	2.13	2.68	3.51	4.72
							2.60	2.28	2.17	2.20	2.32	2.50
0.50								1.75	1.74	1.73	1.72	1.72
								1.86	2.16	2.67	3.45	4.62
								2.72	2.44	2.36	2.41	2.54
1.00									2.15	2.11	2.09	2.09
									2.34	2.76	3.49	4.60
									2.98	2.72	2.66	2.73
1.50										2.86	2.78	2.76
										3.13	3.70	4.72
										3.41	3.16	3.13
2.00											4.03	3.90
											4.40	5.16
											4.06	3.82
2.50												5.90
												6.41
												5.02

Before presenting results from GR3 to GR5, let us compare the standard errors of location for GR2 for five sample sizes (N = 100, 400, 900, 2500, 10000) in order to get a better feel for the sample size issue. Shown in left of Figure 1 are the standard errors of location for GR2 when slope is 1.5

which is a representative value for most serious testing applications. For a sample of 2500 for the location range –1 to +1 the standard errors of location are below of the order of 0.05. For a sample of 10000 for the location range –2 to +2 the standard errors of location are below of the order of 0.05. Even with a sample of 10000, the standard errors of location are larger than of the order of 0.05 for the location ranges beyond –2.5 or +2.5. It is disquieting to see that adequate estimates of location cannot be obtained with the sample of 10000 which is as large a sample as plausible with only truly large scale assessment cases.

A similar plot is constructed on the right side of Figure 1 with the standard errors of slope for GR2 when slope is 1.5. For a sample of 2500 for the location range –3 to +3 none of the standard errors of slope are below of the order of 0.05. For a sample of 10000 for the location range –2 to +2 the standard errors of slope are below of the order of 0.05. Even the sample of 10000 the standard errors of slope are larger than of the order of 0.05 for the location ranges beyond –2.5 or +2.5. A sample needed to achieve accuracy to one decimal place of slope for GR2 when slope is 1.5 may be unattainable in usual practice.

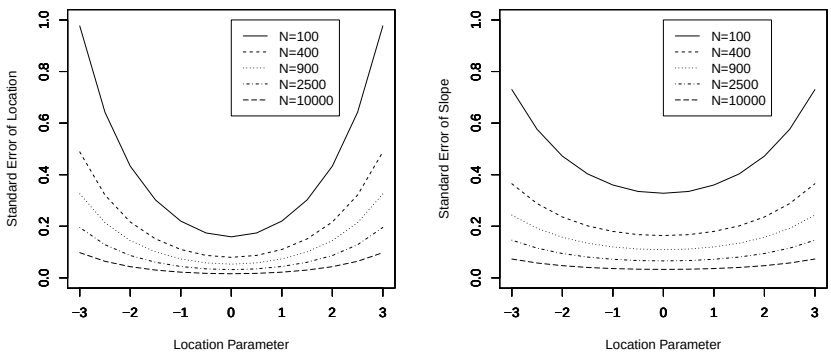

Figure 1. Comparing Standard Errors of Location (Left Figure) and Slope (Right Figure) for the Graded Response Model with Two Categories When Slope Is 1.5

Let us examine three-dimensional plots in order to better understand the relationship that the slope and the location parameter for GR2 has with standard error. Shown in Figure 2 are the standard errors taken from Table 2 of location and slope, respectively, for GR2.

The perspective plot in left side of Figure 2, 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 errors 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 right side of Figure 2, 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 is large when its value is large and location is far away from the center of the ability distribution. Nevertheless all standard errors of slope reach their minimum when location is 0.

Understanding the magnitudes of standard errors of location and slope as in Figure 2 for 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 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 to GR9 (n.b., trivially smaller values of standard errors of location may be observed). 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 to GR9 (i.e., the upper bounds).

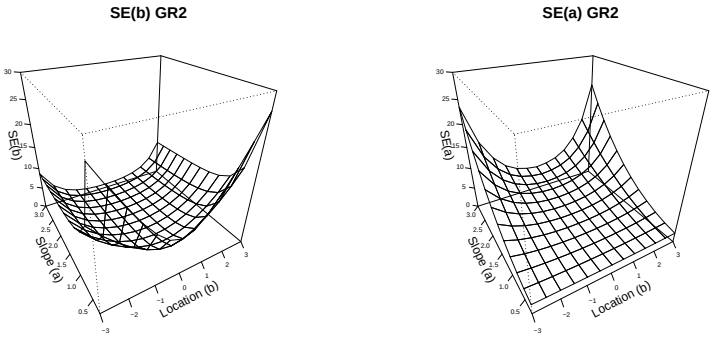

Figure 2. The Standard Error of Location (Left Side) and Slope (Right Side) Shown as a Bivariate Function of Slope and Location for the Graded Response Model with Two Categories

Shown in Figure 3 are the standard errors taken from Table 3 of location 1, location 2, and slope, respectively, for GR3 when slope is 1.5. Slope can be specified by any value from 0.25 to 3. But proportionally similar plots can be obtained to Figure 3.

The perspective plot in left side of Figure 3 with the slope value of 1.5, shows that the standard error of location 1 reaches its minimum at a point where location 1 is 0 which is the center of the ability distribution. For all other cases where location 1 is not 0, the standard error of location 1 gets smaller at a point where location 2 is 0. Because the location parameters are ordered (i.e., $b_1 < b_2$) the plot looks half of the concave down paraboloidal reflector.

The perspective plot in middle of Figure 3 with the slope value of 1.5, shows that the standard error of location 2 reaches its minimum at a point where location 2 is 0 which is the center of the ability distribution. For all other cases where location 2 is not 0, the standard error of location 2 gets smaller at a point where location 1 is 0. Because the location parameters are ordered (i.e., $b_1 < b_2$) the plot looks half of the concave down paraboloidal reflector.

The perspective plot in right side of Figure 3 with the slope value of 1.5, shows that the standard error of slope reaches its minimum at a point where both location 1 and location 2 are close to 0, but both locations are neither too close nor too separated (e.g., location 1 is –1 and location 2 is +1; see also Table 3). Because the location parameters are ordered (i.e., $b_1 < b_2$) the plot resembles portions of a concave-down elliptic paraboloid.

Similar plots to that in Figure 3 for the standard errors may be constructed for GR4 and beyond. For example, Table 4 (n.b., not presented here) for GR4 contains 11 subtables, each with a respective value of $b_1 = -3(0.5)2$ where the parentheses designate increment. Each subtable is similar to Table 3 with row entries of location 2 and column entries of location 3. Each cell for the combination of location 2 and location 3 contains four values of standard errors for b_1, b_2, b_3, and a. Again, for brevity as in GR3, the value of slope is fixed as 1.5 in Table 4.

For each subtable of Table 4 perspective plots can be constructed. There will be four perspective plots of the standard errors for each subtable. Assuming that $a = 1.5$ and $b_1 = -3$, Figure 4 for GR4 was constructed with four perspective plots. 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

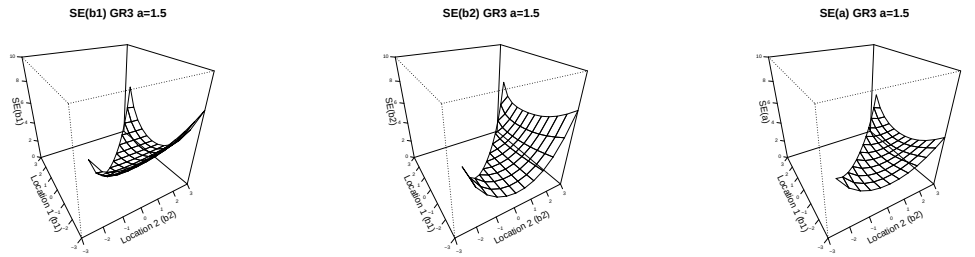

Figure 3. The Standard Error of Location 1 (Left Side), Location 2 (Middle), and Slope (Right Side) Shown as a Bivariate Function of Location 1 and Location 2 for the Graded Response Model with Three Categories When Slope is 1.5

location 3 from the range –2 to +3.

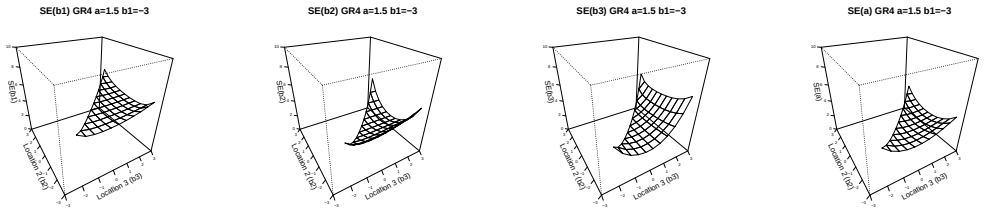

Figure 4. The Standard Error of Location 1 (Left Side), Location 2 (Middle Left), Location 3 (Middle Right), and Slope (Right Side) Shown as a Bivariate Function of Location 2 and Location 3 for the Graded Response Model with Four Categories When Slope is 1.5 and Location 1 is -3

Table 5 (n.b., not presented here) for GR5 contains 55 subtables, each with a respective value of b_1 = –3(0.5)1.5 and a respective value of b_2 = –2.5(0.5)2 where $b_1 < b_2$. Again, each subtable is similar to Table 3 with row entries of location 3 and column entries of location 4. Each cell for the combination of location 3 and location 4 contains five values of standard errors for b_1, b_2, b_3, b_4, and a. For brevity as in GR3 and GR4, the value of slope is fixed as 1.5 in Table 5.

For each subtable of Table 5 perspective plots can be constructed. There will be five perspective plots of the standard errors for each subtable. For example, assuming that a = 1.5, b_1 = –3, and b_2 = –2.5, Figure 5 (n.b., not actually created) for GR5 can be constructed with five perspective plots. The x-dimension of each perspective plot designates the values of location 3 from the range –2.0 to +2.5, and the y-dimension designates the values of location 4 from the range –1.5 to +3.

It can be found that the same general structure would continue to hold for the standard errors for GR6 to GR9. As locations became close to 0 which is the center of the ability distribution, the standard errors of location got smaller. As locations became far away from 0 the standard errors of location got larger. Similar patterns would be observed for the standard errors of slope. In addition, as slopes became more gradual in size, the standard errors of slope got larger. As the number of locations increased, the standard errors of slope got smaller.

One coherent finding from the examination of the tables and figures is that the use of maximum likelihood estimation for the parameters of the graded response model yields results too inexact to be of practical use for a sample size less than, say, 2500. Maximum likelihood estimation requires 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 especially for items with locations away

from the center of the ability distribution. When an item has adjacent categories in extreme ends (e.g., relatively not endorsed, selected, or assigned), there are few observations available to estimate the location parameters thus making the accompanied standard errors very large.

5. Discussion

The sample sizes shown in Table 1 are very large because these are based on the standard errors of the order of 0.05 for the one decimal place accuracy. For the sample sizes required to obtain the standard errors of the order of 0.1, the sample sizes shown in Table 1 can be divided by four. The inverse relation between the sample size and the standard error is clear. The larger the sample size, the smaller the standard error.

A method of parameter estimation other than unrestricted maximum likelihood can be used if not enough observations are available but some previous knowledge about the behavior of parameters is available. Estimating the parameters from a Bayesian viewpoint places prior probability distributions (i.e., priors) on the slope and location parameters. As long as the prior mean may not be far away from the resulting estimate, the precision of the parameter estimate would be enhanced.

For GR2–GR9, as slopes gradually increase, the tandard errors become larger. The prior on the slope parameter in Bayesian estimation may prevent the outward drift of the slope estimate when its value gets excessively large under maximum likelihood estimation. The priors on the location parameters may also play important roles in Bayesian estimation.

There are other IRT models for polytomously scored items. Some examples are the partial credit model, the rating scale model, the generalized partial credit model, the continuation ratio model, and so on (see de Ayala, 2022, Chapters 7–9; Thissen et al., 2001). Because the parameters in the respective models are somewhat related, some asymptotic standard errors of the estimated parameters of these models can be obtained based on the procedure described in this paper. Judicious care should be exercised when small samples are used to estimate parameters of the graded response model.

Acknowledgement

The current study can be viewed as a derivative or sequel of Thissen and Wainer (1982). All creativity, ingenuity, and originality may be attributed to their seminal work.

Funding Statement This research was supported in part by funds from the Omer Clyde and Elizabeth Parr Aderhold Professorship in Research Methodology of the Mary Frances Early College of Education at the University of Georgia.

Competing Interests None.

References

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

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika, 35,* 179–197.

de Ayala, R. J. (2022). *The theory and practice of item response theory* (2nd ed.). New York, NY: The Guilford Press.

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

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

Kendall, M., & Stuart, A. (1979). *The advanced theory of statistics: Vol. 2. Inference and relationship* (4th ed.). New York, NY: Oxford University Press.

Muraki, E., & Bock, R. D. (1993). PARSCALE: IRT based test scoring and item analysis for graded open-ended exercises and performance tasks [Computer software]. 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 [Computer software]. 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.

Stroud, A. H., & Secrest, D. (1966). *Gaussian quadrature formulas.* Englewood Cliffs, NJ: Prentice-Hall.

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.

Thissen, D., Nelson, L., Rosa, K., & McLeod, L. D. (2001). Item response theory for items scored in more than two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 141–186). Mahwah, NJ: Lawrence Erlbaum Associates.

Appendix 1. R Function for GR2

```
gr2 <- function(a,b){
x <- c(
-5.3874808900112,
-4.6036824495507,
-3.9447640401156,
-3.3478545673832,
-2.7888060584281,
-2.2549740020893,
-1.7385377121166,
-1.2340762153953,
-0.7374737285454,
-0.2453407083009,
0.2453407083009,
0.7374737285454,
1.2340762153953,
1.7385377121166,
2.2549740020893,
2.7888060584281,
3.3478545673832,
3.9447640401156,
4.6036824495507,
5.3874808900112
)
w <- c(
2.229393645534/10^13,
4.399340992273/10^10,
1.086069370769/10^7,
7.802556478532/10^6,
2.283386360163/10^4,
3.243773342238/10^3,
2.481052088746/10^2,
1.090172060220/10,
2.866755053628/10,
4.622436696006/10,
4.622436696006/10,
2.866755053628/10,
1.090172060220/10,
2.481052088746/10^2,
3.243773342238/10^3,
```

```
2.283386360163/10^4,
7.802556478532/10^6,
1.086069370769/10^7,
4.399340992273/10^10,
2.229393645534/10^13
)
x <- sqrt(2)*x
w <- w/sqrt(3.14159265358979)
p <- 1/(1+exp(-a*(x-b)))
aa <- w*(x-b)^2*p*(1-p)
bb <- w*(-a)^2*p*(1-p)
ab <- w*(-a)*(x-b)*p*(1-p)
det <- sum(aa)*sum(bb)-sum(ab)^2
sea <- sqrt(sum(bb)/det)
seb <- sqrt(sum(aa)/det)
se <- c(seb,sea)
se
}
```