Supplemental Materials for "Generalized Fiducial Inference for Logistic Graded Response Models"

Yang Liu[1] Jan Hannig[2]

[1]University of California, Merced

[2]The University of North Carolina at Chapel Hill

January 29, 2017

Appendix A: A general family of graded response models

The following item response function (IRF), a generalization of Equation 5 in ? (?) denoted by $f_j(\boldsymbol{\theta}_j, k|\mathbf{z}_i)$, describes respondent i's response to a K_j-category ordinal item j by an ordinal logistic regression on an r-dimensional latent variable $\mathbf{Z}_i$:

$$f_j(\boldsymbol{\theta}_j, k|\mathbf{z}_i) = P\{Y_{ij} = k|\mathbf{Z}_i = \mathbf{z}_i\}$$
$$= \begin{cases} 1 - \Psi(\alpha_{j1} + \boldsymbol{\beta}_j{}^\top \mathbf{z}_i), \ k = 0; \\ \Psi(\alpha_{j,K_j-1} + \boldsymbol{\beta}_j{}^\top \mathbf{z}_i), \ k = K_j - 1; \\ \Psi(\alpha_{jk} + \boldsymbol{\beta}_j{}^\top \mathbf{z}_i) - \Psi(\alpha_{j,k+1} + \boldsymbol{\beta}_j{}^\top \mathbf{z}_i), \ \text{otherwise.} \end{cases} \tag{S.1}$$

In Equation S.1, α_{jk}'s denote strictly ordered intercept parameters, i.e., $\alpha_{j,1} > \cdots > \alpha_{j,K_j-1}$, and $\boldsymbol{\beta}_j$ the slopes. All intercept parameters are assumed to be free, whereas some slopes must be fixed for model identification when $r > 1$. The r-dimensional latent variables are assumed to be standard normal, $\mathbf{Z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{r\times r})$; this includes unidimensional, bifactor, and exploratory graded response models (GRMs) as special cases. Inference for models with unknown covariance structure among latent dimensions (e.g., simple-structure models) is beyond the scope of the present work. Also note that when $K_j = 2$, the model reduces to the binary logistic model discussed in ? (?).

Arguments extending that used for the trichotomous unidimensional example can be employed to obtain the data generating equation (DGE) and set inverse of general graded items. The DGE describes how item response $Y_{ij} \in \{0, \dots, K_j - 1\}$ is generated from the multinomial model characterized by Equation S.1:

$$Y_{ij} = \sum_{k=1}^{K_j-1} \mathbb{I}\{A_{ij} \le \alpha_{jk} + \boldsymbol{\beta}_j{}^\top \mathbf{Z}_i\}. \tag{S.2}$$

Assume r_j slopes are free ($r_j \le r$) for item j. Denote by $\boldsymbol{\theta}_j$ all free item parameters that calibrates this item; the dimension of $\boldsymbol{\theta}_j$ is then $q_j = r_j + K_j - 1$. Let $\Theta_j \subset \{(\alpha_{j,1}, \ldots, \alpha_{j,K_j-1})^\top : \alpha_{j,1} > \cdots > \alpha_{j,K_j-1}\} \times \mathbb{R}^{r_j} \subset \mathbb{R}^{q_j}$ be the parameter space of $\boldsymbol{\theta}_j$. The set inverse function of Equation S.2 is the following subset of the q_j-dimensional parameter space:

$$\begin{aligned} Q_{ij}(y_{ij}, a_{ij}, \mathbf{z}_i) = \{\boldsymbol{\theta}_j \in \Theta_j : \; & a_{ij} > \alpha_{j1} + \boldsymbol{\beta}_j{}^\top \mathbf{z}_i, \quad \text{if } y_{ij} = 0; \\ & a_{ij} \leq \alpha_{j,K-1} + \boldsymbol{\beta}_j{}^\top \mathbf{z}_i, \quad \text{if } y_{ij} = K; \\ & \alpha_{j,k+1} + \boldsymbol{\beta}_j{}^\top \mathbf{z}_i < a_{ij} \leq \alpha_{jk} + \boldsymbol{\beta}_j{}^\top \mathbf{z}_i, \text{ otherwise.}\} \end{aligned} \tag{S.3}$$

Set inverses for individual response entries can be assembled into the overall set inverse for i.i.d. responses to a multiple-item test in the same manner as described in the previously discussed three-category example, i.e., Equations 7 and 8 in ? (?). The resulting GFQ is also defined by Equation 9.

Appendix B: Outline of proof of Theorem 1

The selected extremal point along a predetermined direction $\mathbf{d} = (\mathbf{d}_j)_{j=1}^m$ for each possibly unbounded polyhedron $Q_j(\mathbf{y}_{(j)}, \mathbf{a}_{(j)}, \mathbf{z})$ is either a vertex[1] or infinity. The proof consists of two major steps. First, we modify the conditioning in the GFQ (Equation 9) to exclude $\mathbf{A}^\star$ and $\mathbf{Z}^\star$ values that lead to infinite extremal points along $\mathbf{d}$, and show that the modified fiducial distribution satisfies a Bernstein-von Mises theorem (Lemma S.2). Because the argument closely resembles that used in ? (?), details are omitted here. Next, we claim that the modification made to the conditioning set is minor, and thus conclude that the fiducial distribution with the particular selection rule follows the Bernstein-von Mises theorem as well.

Let $\tau_{jk}(\boldsymbol{\theta}_j, \mathbf{z}_i) = \alpha_{jk} + \boldsymbol{\beta}_j{}^\top \mathbf{z}_i$ be the linear regression on the latent variable, and set

[1] With probability one, a unique point (possibly infinity) is selected as the farthest along $\mathbf{d}$; whenever there are ties, a more rigorous treatment resulting in a unique vertex can be found in Appendix A of Hannig (2013).

$\alpha_{j0} = \infty$ and $\alpha_{jK_j} = -\infty$ by convention. With the help of these notations, we simplify the IRF (Equation S.1) to

$$f_j(\boldsymbol{\theta}_j, k|\mathbf{z}_i) = \Psi(\tau_{jk}(\boldsymbol{\theta}_j, \mathbf{z}_i)) - \Psi(\tau_{j,k+1}(\boldsymbol{\theta}_j, \mathbf{z}_i)), \tag{S.4}$$

and the set inverse function (Equation S.3) to

$$Q_{ij}(y_{ij}, a_{ij}, z_i) = \{\boldsymbol{\theta}_j \in \mathbb{R}^{q_j} : \tau_{j,y_{ij}+1}(\boldsymbol{\theta}_j, \mathbf{z}_i) < a_{ij} \leq \tau_{j,y_{ij}}(\boldsymbol{\theta}_j, \mathbf{z}_i)\}. \tag{S.5}$$

For each item j, let $\mathbf{i}_j = (i_{js})_{s=1}^{q_j}$ be a size-q_j sub-sample of observations in which i_{js}'s are sorted, e.g., in an ascending order[2]. Also let $\mathbf{k}_j = (k_{js})_{s=1}^{q_j}$ be a binary vector of length q_j, each element of which $k_{js} = 0$ or 1 indicates whether the right half-space $Q^{(0)}_{i_{js}j}(y_{i_{js}j}, a_{i_{js}j}, \mathbf{z}_{i_{js}}) = \{\boldsymbol{\theta}_j : a_{i_{js}j} \leq \tau_{jy_{i_{js}j}}(\boldsymbol{\theta}_j, \mathbf{z}_{i_{js}})\}$ or the left half-space $Q^{(1)}_{i_{js}j}(y_{i_{js}j}, a_{i_{js}j}, \mathbf{z}_{i_{js}}) = \{\boldsymbol{\theta}_j : \tau_{j,y_{i_{js}j}+1}(\boldsymbol{\theta}_j, \mathbf{z}_{i_{js}}) < a_{i_{js}j}\}$ is selected for each i_{js}. Each vertex of the possibly unbounded $\mathbb{R}^{q_j}$-polyhedron $Q_j(\mathbf{y}_{(j)}, \mathbf{a}_{(j)}, \mathbf{z})$ residing in the interior of Θ is the solution of a set of q_j linear equations of form $a_{i_{js}j} = \tau_{j,y_{i_{js}j}+k_{js}}(\boldsymbol{\theta}_j, \mathbf{z}_{i_{js}})$, contributed from the q_j observations from $\mathbf{i}_j$ and some suitable choices of left/right bounds $\mathbf{k}_j$. Notationally, we denote such a vertex by $\mathbf{v}_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)}, \mathbf{a}_{\mathbf{i}_j(j)}, \mathbf{z}_{\mathbf{i}_j})$, in which $\mathbf{y}_{\mathbf{i}_j(j)} = (y_{i_{js}j})_{s=1}^{q_j}$ collects the responses to item j for observations $\mathbf{i}_j$, and $\mathbf{a}_{\mathbf{i}_j(j)} = (a_{i_{js}j})_{s=1}^{q_j}$ and $\mathbf{z}_{\mathbf{i}_j} = (\mathbf{z}_{i_{js}})_{s=1}^{q_j}$ are the corresponding logistic and normal variates.

As a side note, if an observation contributes both the left and right bounds when determining a single vertex, then the vertex must be on the boundary of Θ_j determined by the order constraint of two adjacent intercepts, i.e., $\partial\Theta_j = \{\boldsymbol{\theta} : \alpha_{jk} = \alpha_{j,k+1}$ for some $k\}$. In large samples, however, this almost never happens. In fact, if there exists more than one endorsement to a response category k of item j, e.g., $y_{1j} = y_{2j} = k$, then $\alpha_{j,k+1} <$

[2]We sort the indices to avoid repeatedly counting permutations of $\mathbf{i}$ when summing across all possible subsamples.

$\min\{A^\star_{1j} - \boldsymbol{\beta}_j{}^\top \mathbf{Z}^\star_1, A^\star_{2j} - \boldsymbol{\beta}_j{}^\top \mathbf{Z}^\star_2\} \le \max\{A^\star_{1j} - \boldsymbol{\beta}_j{}^\top \mathbf{Z}^\star_1, A^\star_{2j} - \boldsymbol{\beta}_j{}^\top \mathbf{Z}^\star_2\} \le \alpha_{jk}$, with a strict inequality attained almost surely by the continuous nature of the logistic and normal variates. As long as the data-generating values of the item parameters are in the interior of the parameter space, all the response patterns happen with a positive probability, and thus the set inverse is eventually bounded away from $\partial\Theta$ with probability one. In addition, only a small fraction of $\mathbf{i}_j$ and $\mathbf{k}_j$ combinations are vertex-determining: It only happens when the q_j boundary hyperplanes of the selected half-spaces produce a non-singular linear system.

The selection rule considered in this proof identifies the extremal point of the set inverse along a fixed direction $\mathbf{d} = (\mathbf{d}_j)_{j=1}^m$ in the parameter space Θ. Denote the extremal point of each polyhedron $Q_j(\mathbf{y}_{(j)}, \mathbf{a}_{(j)}, \mathbf{z})$ along $\mathbf{d}_j$ by $\mathbf{v}_{\mathbf{d}_j}(\mathbf{y}_{(j)}, \mathbf{a}_{(j)}, \mathbf{z})$. $\mathbf{v}_{\mathbf{d}_j}(\mathbf{y}_{(j)}, \mathbf{a}_{(j)}, \mathbf{z})$ is either finite or infinite. Whenever all the coordinates are finite, $\mathbf{v}_{\mathbf{d}_j}(\mathbf{y}_{(j)}, \mathbf{a}_{(j)}, \mathbf{z})$ must be a vertex $\mathbf{v}_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)}, \mathbf{a}_{\mathbf{i}_j(j)}, \mathbf{z}_{\mathbf{i}_j})$ indexed by some $\mathbf{i}_j$ and $\mathbf{k}_j$.

Pooling across all m items in the test, write $\mathbf{i} = (\mathbf{i}_j)_{j=1}^m$, $\mathbf{k} = (\mathbf{k}_j)_{j=1}^m$, $\mathbf{a}_\mathbf{i} = (\mathbf{a}_{\mathbf{i}_j(j)})_{j=1}^m$, and $\mathbf{z}_\mathbf{i} = (\mathbf{z}_{\mathbf{i}_j})_{j=1}^m$. Also write $\mathbf{v}_{\mathbf{ik}}(\mathbf{y}_\mathbf{i}, \mathbf{a}_\mathbf{i}, \mathbf{z}_\mathbf{i}) = (\mathbf{v}_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)}, \mathbf{a}_{\mathbf{i}_j(j)}, \mathbf{z}_{\mathbf{i}_j}))_{j=1}^m$ as the collection of all vertices determined by $\mathbf{i}$ and $\mathbf{k}$, and $\mathbf{v}_\mathbf{d}(\mathbf{y}, \mathbf{a}, \mathbf{z}) = (\mathbf{v}_{\mathbf{d}_j}(\mathbf{y}_{(j)}, \mathbf{a}_{(j)}, \mathbf{z}))_{j=1}^m$ be the collection of all polyhedral extrema along $\mathbf{d}$. To avoid dealing with the extended Euclidean space including infinity, we next propose a modification to the fiducial distribution, for which a Bernstein-von Mises theorem can be proved.

B.1 A Bernstein von-Mises theorem for a modified fiducial distribution

An additional conditioning on $\|\mathbf{v}_\mathbf{d}(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star)\|_\infty < \infty$ is introduced to concentrate on finite extremal points along the pre-specified direction $\mathbf{d}$, in which $\|\cdot\|_\infty$ stands for the L^∞ norm. In particular, let $\tilde{g}_n^\mathbf{d}(\boldsymbol{\theta}|\mathbf{y})$ be the density of the following conditional distribution:

$$\mathbf{v}_\mathbf{d}(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) \mid \{Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) \neq \emptyset, \|\mathbf{v}_\mathbf{d}(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star)\|_\infty < \infty\}. \tag{S.6}$$

We now derive the exact formula of $\tilde{g}_n^{\mathbf{d}}(\boldsymbol{\theta}|\mathbf{y})$.

Lemma S.1. For a fixed data set $\mathbf{y}$, assume that every response category of every item is selected by at least two respondents. The modified fiducial distribution (Equation S.6) is absolutely continuous with density[3]

$$\tilde{g}_n^{\mathbf{d}}(\boldsymbol{\theta}|\mathbf{y}) \propto \sum_{\mathbf{i},\mathbf{k}} \int_{\mathbb{R}^{nr}} \prod_{j=1}^{m} \Bigg\{ \mathbb{I}_{\{\mathbf{v}_{\mathbf{d}}=\mathbf{v}_{\mathbf{i}_j\mathbf{k}_j}|D_{\mathbf{i}_j\mathbf{k}_j}\}}(\mathbf{z}_{\mathbf{i}_j}) d_{\mathbf{i}_j\mathbf{k}_j}(\boldsymbol{\theta}_j, \mathbf{z}_{\mathbf{i}_j}) \cdot \prod_{s=1}^{q_j} \psi(\tau_{j,y_{i_{js}j}+k_{js}}(\boldsymbol{\theta}_j, \mathbf{z}_i)) \cdot \prod_{i\notin \mathbf{i}_j} f_j(\boldsymbol{\theta}_j, y_{ij}|\mathbf{z}_i) \Bigg\} d\Phi(\mathbf{z}), \tag{S.7}$$

in which $d_{\mathbf{i}_j\mathbf{k}_j}(\boldsymbol{\theta}_j, \mathbf{z}_{\mathbf{i}_j}) = \left|\det(\partial\tau_{j,y_{i_{js}j}+k_{js}}(\boldsymbol{\theta}_j, \mathbf{z}_{i_{js}})/\partial\boldsymbol{\theta}_j)_{s=1}^{q_j}\right|$ is the Jacobian determinant corresponding to the selected sub-sample $\mathbf{i}_j$, $\Phi(\cdot)$ denotes a standard normal probability measure[4], and $\psi(x) = e^x/(1+e^x)^2$ is the standard logistic density function. In addition, $D_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)})$ denotes the set on the space of $\mathbf{a}_{(j)}$ and $\mathbf{z}$ such that $\mathbf{v}_{\mathbf{i}_j,\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)}, \mathbf{a}_{\mathbf{i}_j(j)}, \mathbf{z}_{\mathbf{i}_j})$ is interior, and the indicator function identifies $\mathbf{z}_{\mathbf{i}_j}$ values such that $\mathbf{v}_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)}, \mathbf{a}_{\mathbf{i}_j(j)}, \mathbf{z}_{\mathbf{i}_j})$ is extremal along direction $\mathbf{d}_j$ provided $\mathbf{a}_{(j)}$ and $\mathbf{z}$ are in $D_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)})$.

Proof. By the law of total probability, we have

$$\begin{aligned} &P\left\{\mathbf{v}_{\mathbf{d}}(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) \le \boldsymbol{\theta}, Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) \neq \emptyset, \|\mathbf{v}_{\mathbf{d}}(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star)\|_\infty < \infty\right\} \\ =& \sum_{\mathbf{i},\mathbf{k}} P\left\{\mathbf{v}_{\mathbf{ik}}(\mathbf{y}_{\mathbf{i}}, \mathbf{A}^\star_{\mathbf{i}}, \mathbf{Z}^\star_{\mathbf{i}}) \le \boldsymbol{\theta}, \mathbf{v}_{\mathbf{d}}(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) = \mathbf{v}_{\mathbf{ik}}(\mathbf{y}_{\mathbf{i}}, \mathbf{A}^\star_{\mathbf{i}}, \mathbf{Z}^\star_{\mathbf{i}}), D_{\mathbf{ik}}(\mathbf{y})\right\}, \end{aligned} \tag{S.8}$$

in which $D_{\mathbf{ik}}(\mathbf{y}) = \bigcap_{j=1}^m D_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)})$ denotes the event that $\mathbf{V}^\star_{\mathbf{ik}}$ is interior. The remaining task is to derive each summand on the RHS of Equation S.8 and then differentiate it with

[3]With a slight abuse of notation, we write $i \in \mathbf{i}_j$ ($i \notin \mathbf{i}_j$) indicating that i is (is not) a component of $\mathbf{i}_j$, and $i \in \mathbf{i}$ ($i \notin \mathbf{i}$) indicating that $i \in \mathbf{i}_j$ ($i \notin \mathbf{i}_j$) for some j (for all j). In addition, integrating with respect to the probability measure of $\mathbf{a}_{\mathbf{i}}$ or $\mathbf{z}_{\mathbf{i}}$ means integrating over the random variates corresponding to the unique observations in sub-sample $\mathbf{i}$.

[4]Here, the dimensionality of the random variable is suppressed for succinctness.

respect to $\boldsymbol{\theta}$.

Consider a single item j first. Set $\mathbf{v}_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)}, \mathbf{a}_{\mathbf{i}_j(j)}, \mathbf{z}_{\mathbf{i}_j}) = \boldsymbol{\theta}_j'$. Event $D_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)})$ requires that $\tau_{j,y_{i_{js}j}+k_{js}}(\boldsymbol{\theta}_j', \mathbf{z}_{i_{js}}) = a_{i_{js}j}$ for the selected half-spaces indexed by $(\mathbf{i}_j, \mathbf{k}_j)$, and that $\boldsymbol{\theta}_j'$ should not conflict with the remaining half-spaces: i.e., $\tau_{j,y_{ij}+1}(\boldsymbol{\theta}_j', \mathbf{z}_i) < a_{ij} \le \tau_{j,y_{ij}}(\boldsymbol{\theta}_j', \mathbf{z}_i)$ for all $i \notin \mathbf{i}_j$.

Given $D_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)})$, $\mathbf{v}_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)}, \mathbf{a}_{\mathbf{i}_j(j)}, \mathbf{z}_{\mathbf{i}_j})$ is extremal along $\mathbf{d}$ if and only if the normal cone of the polyhedral cone $\bigcap_{s=1}^{q_j} Q_{i_{js}j}^{(k_{js})}(y_{i_{js}j}, a_{i_{js}j}, \mathbf{z}_{i_{js}})$ with respect to $\mathbf{v}_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)}, \mathbf{a}_{\mathbf{i}_j(j)}, \mathbf{z}_{\mathbf{i}_j})$ contains a ray along the direction $\mathbf{d}_j$, i.e., for all $\boldsymbol{\theta}_j \in \bigcap_{s=1}^{q_j} Q_{i_{js}j}^{(k_{js})}(y_{i_{js}j}, a_{i_{js}j}, \mathbf{z}_{i_{js}})$, $[\boldsymbol{\theta}_j - \mathbf{v}_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)}, \mathbf{a}_{\mathbf{i}_j(j)}, \mathbf{z}_{\mathbf{i}_j})]^\top \mathbf{d}_j < 0$. For fixed $\mathbf{y}_{\mathbf{i}_j(j)}$ and $\mathbf{k}_j$, the latter condition only depends on the normal variates of the selected subsample $\mathbf{i}_j$, i.e., $\mathbf{z}_{\mathbf{i}_j}$, because the normal cone is spanned by the normal vectors of the selected hyperplanes, which has either ± 1 or 0 (determined by $\mathbf{y}_\mathbf{i}$ and $\mathbf{k}$) on the coordinates for intercepts and $\pm \mathbf{z}_{\mathbf{i}_j}$ on the coordinates for slopes.

Now we proceed to deriving the exact formula for each summand of Equation S.8. Fix $\mathbf{i}_j$ and $\mathbf{k}_j$, and condition on $\mathbf{Z}^\star = \mathbf{z}$. Using the argument articulated in the previous two paragraphs, we can write

$$
\begin{aligned}
&P\big\{\mathbf{v}_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)}, \mathbf{A}^\star_{\mathbf{i}_j(j)}, \mathbf{Z}^\star_{\mathbf{i}_j}) \le \boldsymbol{\theta}_j, \mathbf{v}_{\mathbf{d}_j}(\mathbf{y}_{(j)}, \mathbf{A}^\star_{(j)}, \mathbf{Z}^\star) = \mathbf{v}_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)}, \mathbf{A}^\star_{\mathbf{i}_j(j)}, \mathbf{Z}^\star_{\mathbf{i}_j}),\\
&\qquad D_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)}) \mid \mathbf{Z}^\star = \mathbf{z}\big\}\\
&= \int_{\boldsymbol{\theta}_j' \le \boldsymbol{\theta}_j} \mathbb{I}_{\{\mathbf{v}_\mathbf{d} = \mathbf{v}_{\mathbf{i}_j\mathbf{k}_j} | D_{\mathbf{i}_j\mathbf{k}_j}\}}(\mathbf{z}_{\mathbf{i}_j}) \cdot d_{\mathbf{i}_j\mathbf{k}_j}(\boldsymbol{\theta}_j', \mathbf{z}_{\mathbf{i}_j}) \cdot \prod_{s=1}^{q_j} \psi(\tau_{j,y_{i_{js}j}+k_{js}}(\boldsymbol{\theta}_j', \mathbf{z}_{i_{js}}))\\
&\quad \cdot \prod_{i \notin \mathbf{i}_j} \left[\int_{\tau_{j,y_{ij}+1}(\boldsymbol{\theta}_j', \mathbf{z}_i)}^{\tau_{j,y_{ij}}(\boldsymbol{\theta}_j', \mathbf{z}_i)} \psi(a_{ij}) da_{ij}\right] d\boldsymbol{\theta}'\\
&= \int_{\boldsymbol{\theta}_j' \le \boldsymbol{\theta}_j} \mathbb{I}_{\{\mathbf{v}_\mathbf{d} = \mathbf{v}_{\mathbf{i}_j\mathbf{k}_j} | D_{\mathbf{i}_j\mathbf{k}_j}\}}(\mathbf{z}_{\mathbf{i}_j}) d_{\mathbf{i}_j\mathbf{k}_j}(\boldsymbol{\theta}_j', \mathbf{z}_{\mathbf{i}_j}) \cdot \prod_{s=1}^{q_j} \psi(\tau_{j,y_{i_{js}j}+k_{js}}(\boldsymbol{\theta}_j', \mathbf{z}_{i_{js}})) \cdot \prod_{i \notin \mathbf{i}_j} f_j(\boldsymbol{\theta}_j', k | \mathbf{z}_i) d\boldsymbol{\theta}_j'
\end{aligned} \tag{S.9}
$$

In the first equality of Equation S.9, the Jacobian determinant and the first product are due to the change of variables from $\mathbf{a}_{\mathbf{i}_j(j)}$ to $\boldsymbol{\theta}'_j$, the product of integrals correspond to inequalities to be satisfied by the unselected half-spaces, and the indicator function decides whether the fixed $\mathbf{z}_{\mathbf{i}_j}$ values render the vertex $\mathbf{v}_{\mathbf{i}_j\mathbf{k}_j}$ the extremal point along $\mathbf{d}_j$. Note that a singular linear system $\tau_{j,y_{i_{js}j}+k_{js}}(\boldsymbol{\theta}'_j, \mathbf{z}_{i_{js}}) = a_{i_{js}j}$, $s = 1, \ldots, q_j$, is implied by certain $(\mathbf{i}_j, \mathbf{k}_j)$ pairs, which further results in a zero Jacobian determinant regardless of the $\mathbf{z}_{\mathbf{i}_j}$ values. Also, the vertices determined by some response patterns $y_{\mathbf{i}_j(j)}$ combined with $\mathbf{k}_j$ cannot be extremal along $\mathbf{d}_j$, and hence the indicator function equals to zero for all $\mathbf{z}_{\mathbf{i}_j}$ values. In those cases, the corresponding summand simply vanishes.

Due to the conditional independence assumption,

$$
\begin{aligned}
&P\left\{\mathbf{v}_{\mathbf{ik}}(\mathbf{y}_{\mathbf{i}}, \mathbf{A}^\star_{\mathbf{i}}, \mathbf{Z}^\star_{\mathbf{i}}) \le \boldsymbol{\theta}, \mathbf{v}_{\mathbf{d}}(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) = \mathbf{v}_{\mathbf{ik}}(\mathbf{y}_{\mathbf{i}}, \mathbf{A}^\star_{\mathbf{i}}, \mathbf{Z}^\star_{\mathbf{i}}), D_{\mathbf{ik}}(\mathbf{y})\right\} \\
&= \int_{\mathbb{R}^{nr}} \prod_{j=1}^{m} P\Big\{\mathbf{v}_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)}, \mathbf{A}^\star_{\mathbf{i}_j(j)}, \mathbf{Z}^\star_{\mathbf{i}_j}) \le \boldsymbol{\theta}_j, \mathbf{v}_{\mathbf{d}_j}(\mathbf{y}_{(j)}, \mathbf{A}^\star_{(j)}, \mathbf{Z}^\star) = \mathbf{v}_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)}, \mathbf{A}^\star_{\mathbf{i}_j(j)}, \mathbf{Z}^\star_{\mathbf{i}_j}), \\
&\qquad\qquad D_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)}) \mid \mathbf{Z}^\star = \mathbf{z}\Big\} d\Phi(\mathbf{z}).
\end{aligned} \tag{S.10}
$$

Equation S.7 is established by substituting Equation S.9 into Equations S.10 and S.8, and differentiating with respect to $\boldsymbol{\theta}$. □

There are only finitely many combinations of $\mathbf{y}_{\mathbf{i}}$ and $\mathbf{k}$, and the sub-samples $\mathbf{i}$ and $\mathbf{i}'$ such that $\mathbf{y}_{\mathbf{i}} = \mathbf{y}_{\mathbf{i}'}$ are exchangeable. As a consequence, we can replace the sum over $\mathbf{i}$ and $\mathbf{k}$ in Equation S.7, in which the number of summands grows as n increases, by a finite sum over $\mathbf{y}_{\mathbf{i}}$ and $\mathbf{k}$:

$$
\tilde{g}^{\mathbf{d}}_n(\boldsymbol{\theta}|\mathbf{y}) \propto G_n \left[\sum_{\mathbf{y}_{\mathbf{i}},\mathbf{k}} p_n(\mathbf{y}_{\mathbf{i}}) b^{\mathbf{d}}_{\mathbf{y}_{\mathbf{i}}\mathbf{k}}(\boldsymbol{\theta})\right] f_n(\boldsymbol{\theta}, \mathbf{y}). \tag{S.11}
$$

In Equation S.11, G_n is the total number of $\mathbf{i}$ and $\mathbf{k}$ combinations, and $p_n(\mathbf{y}_{\mathbf{i}})$ is the observed proportion of pattern $\mathbf{y}_{\mathbf{i}}$. By the standard theory of U-statistics, $p_n(\mathbf{y}_{\mathbf{i}})$ converges to the expected proportion of $\mathbf{y_i}$ in $P_{\boldsymbol{\theta}_0}$-probability. $f_n(\boldsymbol{\theta},\mathbf{y})$ is the sample likelihood function

$$f_n(\boldsymbol{\theta},\mathbf{y}) = \int \prod_{i=1}^{n}\prod_{j=1}^{m} f_j(\boldsymbol{\theta}_j, y_{ij}|\mathbf{z}_i)d\Phi(\mathbf{z}) \tag{S.12}$$

and

$$b^{\mathbf{d}}_{\mathbf{y_i k}}(\boldsymbol{\theta}) = \left\{ \int \prod_{j=1}^{m} \left[\mathbb{I}_{\{\mathbf{v_d}=\mathbf{v}_{\mathbf{i}_j\mathbf{k}_j}|D_{\mathbf{i}_j\mathbf{k}_j}(\mathbf{y}_{\mathbf{i}_j(j)})\}}(\mathbf{z}_{\mathbf{i}_j}) d_{\mathbf{i}_j\mathbf{k}_j}(\boldsymbol{\theta}_j,\mathbf{z}_{\mathbf{i}_j}) \prod_{s=1}^{q_j} \psi(\tau_{j,y_{i_{js}j}+k_{js}}(\boldsymbol{\theta}_j,\mathbf{z}_i)) \cdot \prod_{i\in\mathbf{i}\backslash\mathbf{i}_j} f_j(\boldsymbol{\theta}_j,k|\mathbf{z}_i)\right] d\Phi(\mathbf{z_i}) \right\} \Bigg/ \left[\int \prod_{i\in\mathbf{i}}\prod_{j=1}^{m} f_j(\boldsymbol{\theta}_j, y_{ij}|\mathbf{z}_i)d\Phi(\mathbf{z_i})\right], \tag{S.13}$$

in which $i \in \mathbf{i}\backslash\mathbf{i}_j$ means $i \in \mathbf{i}$ but $i \notin \mathbf{i}_j$. As we have discussed in the proof of Lemma S.1, $b^{\mathbf{d}}_{\mathbf{y_i k}}(\boldsymbol{\theta})$ can be 0, because: a) The linear system determined by $\mathbf{y_i}$ and $\mathbf{k}$ can be singular and thus the Jacobian determinant is zero for almost surely all $\mathbf{z_i}$; and b) the vertices determined by some $\mathbf{y_i}$ and $\mathbf{k}$ cannot be extremal along $\mathbf{d}$, no matter what the $\mathbf{z_i}$ values are. When the sample size is large enough, all $\mathbf{y_i}$ patterns emerge with probability 1, provided the true $\boldsymbol{\theta}_0$ is in the interior of the parameter space; therefore, the modified fiducial density (Equation S.7) is not degenerate for large enough n.

The term $\left[\sum_{\mathbf{y_i},\mathbf{k}} p_n(\mathbf{y_i}) b^{\mathbf{d}}_{\mathbf{y_i k}}(\boldsymbol{\theta})\right]$ in Equation S.11 functions as a data-dependent prior, and the proof of Theorem 1 in ? (?) can be straightforwardly extended to establish the following Bernstein-von Mises result for the modified fiducial density.

Lemma S.2. Under the assumptions of Theorem 1,

$$\int_{H_n} \left| \bar{\tilde{g}}^{\mathbf{d}}_n(\mathbf{h}|\mathbf{Y}) - \phi_{\boldsymbol{\mathcal{I}}_0^{-1}\mathbf{s}_n, \boldsymbol{\mathcal{I}}_0^{-1}}(\mathbf{h}) \right| d\mathbf{h} \overset{P_{\boldsymbol{\theta}_0}}{\to} 0, \tag{S.14}$$

in which $\bar{\tilde{g}}^{\mathbf{d}}_n(\mathbf{h}|\mathbf{y}) = \tilde{g}^{\mathbf{d}}_n(\boldsymbol{\theta}_0 + \mathbf{h}/\sqrt{n}|\mathbf{y})/\sqrt{n}$.

B.2 The extra conditioning

The remaining task is to show that the extra conditioning on $\|\mathbf{v_d}(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star)\|_\infty < \infty$ does not change the fiducial distribution substantially. In particular, we want to establish

$$\frac{P\{Q(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star) \neq \emptyset, \|\mathbf{v_d}(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star)\|_\infty < \infty\}}{P\{Q(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star) \neq \emptyset\}} \to 1 \tag{S.15}$$

in $P_{\boldsymbol{\theta}_0}$-probability[5]. Because $\|\mathbf{v_d}(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star)\|_\infty = \infty$ implies that $Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star)$ is unbounded, we have

$$\begin{aligned}1 \geq & \frac{P\{Q(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star) \neq \emptyset, \|\mathbf{v_d}(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star)\|_\infty < \infty\}}{P\{Q(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star) \neq \emptyset\}} \\ \geq & \frac{P\{Q(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star) \neq \emptyset, \|\mathbf{v_d}(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star)\|_\infty < \infty\}}{P\{Q(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star) \neq \emptyset, \|\mathbf{v_d}(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star)\|_\infty < \infty\} + P\{Q(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star) \neq \emptyset \text{ but unbounded}\}}.\end{aligned} \tag{S.16}$$

Thus it suffices to show the second ratio in Equation S.16 converges to 1 in $P_{\boldsymbol{\theta}_0}$-probability.

We express the numerator of Equations S.15 and S.16, which is also the first term in the denominator of Equation S.16, as

$$\begin{aligned}& P\{Q(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star) \neq \emptyset, \|\mathbf{v_d}(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star)\|_\infty < \infty\} \\ =& \sum_{\mathbf{i},\mathbf{k}} P\{\mathbf{v_d}(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star) = \mathbf{v_{ik}}(\mathbf{Y_i}, \mathbf{A_i^\star}, \mathbf{Z_i^\star}), D_{\mathbf{ik}}(\mathbf{Y})\} \\ =& G_n \sum_{\mathbf{y_i},\mathbf{k}} p_n(\mathbf{y_i}) P\{\mathbf{v_d}(\mathbf{Y_{y_i}}, \mathbf{A}^\star, \mathbf{Z}^\star) = \mathbf{v_{ik}}(\mathbf{y_i}, \mathbf{A_i^\star}, \mathbf{Z_i^\star}), D_{\mathbf{ik}}(\mathbf{Y_{y_i}})\},\end{aligned} \tag{S.17}$$

using the exchangeability of sub-samples with the same response pattern. The notation $\mathbf{Y_{y_i}} = (\mathbf{y_i}, \mathbf{Y_{i^c}}^\top)^\top$ is used to denote a data matrix in which the responses for the selected

[5]Notationally, we use $P\{\cdot\}$ for probability calculations with respect to $\mathbf{A}^\star$ and $\mathbf{Z}^\star$, and $P_{\boldsymbol{\theta}_0}$ for the probability measure of $\mathbf{Y}$.

sub-sample $\mathbf{i}$ are fixed to $\mathbf{y_i}$ while the remaining responses $\mathbf{Y}_{\mathbf{i}^c}$, $\mathbf{i}^c = \{i : i \notin \mathbf{i}\}$, are deemed random[6]. $D_{\mathbf{ik}}(\mathbf{y})$ denotes the event that $\mathbf{v_{ik}}(\mathbf{y_i}, \mathbf{A}^\star_\mathbf{i}, \mathbf{Z}^\star_\mathbf{i})$ is interior, which has been explicated in the proof of Lemma S.1. In Equation S.17, the summand can be zero. To see this, write

$$P\{\mathbf{v_d}(\mathbf{Y}_{\mathbf{y_i}}, \mathbf{A}^\star, \mathbf{Z}^\star) = \mathbf{v_{ik}}(\mathbf{y_i}, \mathbf{A}^\star_\mathbf{i}, \mathbf{Z}^\star_\mathbf{i}), D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y_i}})\} = \int b^{\mathbf{d}}_{\mathbf{y_i k}}(\boldsymbol{\theta}) f_n(\boldsymbol{\theta}, \mathbf{Y}_{\mathbf{y_i}}) d\boldsymbol{\theta}. \tag{S.18}$$

Note that Equation S.18 is zero (almost surely) if and only if $b^{\mathbf{d}}_{\mathbf{y_i k}}(\boldsymbol{\theta})$ is zero for almost every $\boldsymbol{\theta}$, which is resulted from either a vanished indicator function that selects the extremal vertex or a degenerate linear system with a zero Jacobian determinant (see Equation S.13).

The second term in the denominator of Equation S.16 can be further bounded by:

$$\begin{aligned}
&P\{Q(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star) \neq \emptyset \text{ but unbounded}\} \\
=& P\left\{Q(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star) \text{ is unbounded}, \bigcup_{\mathbf{i},\mathbf{k}} D_{\mathbf{ik}}(\mathbf{Y})\right\} \\
\leq& G_n \sum_{\mathbf{y_i},\mathbf{k}} p_n(\mathbf{y_i}) P\{Q(\mathbf{Y}_{\mathbf{y_i}}, \mathbf{A}^\star, \mathbf{Z}^\star) \text{ is unbounded}, D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y_i}})\}.
\end{aligned} \tag{S.19}$$

Some terms in the last summation of Equation S.19 can be zero as well, when the combinations of $\mathbf{y_i}$ and $\mathbf{k}$ fail to produce vertices. In fact, $P\{D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y_i}})\}$ can be expressed as an integral in a form similar to Equation S.18:

$$P\{D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y_i}})\} = \int b_{\mathbf{y_i k}}(\boldsymbol{\theta}) f_n(\boldsymbol{\theta}, \mathbf{Y}_{\mathbf{y_i}}) d\boldsymbol{\theta}, \tag{S.20}$$

in which $b_{\mathbf{y_i k}}(\boldsymbol{\theta})$ is almost the same as $b^{\mathbf{d}}_{\mathbf{y_i k}}(\boldsymbol{\theta})$ (Equation S.13) with the exception that the indicator functions selecting extremal vertices are dropped. $P\{D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y_i}})\}$ is then zero (almost surely) if and only if $b_{\mathbf{y_i k}}(\boldsymbol{\theta})$ is constantly zero, which is caused by a singular Jacobian.

[6] Again by the exchangeability of sub-samples with the same response pattern, we assume without loss of generality that $\mathbf{i}$ always indexes observations at the beginning of the sample.

Because the sums in Equations S.17 and S.19 are over finitely many terms, it further suffices to show that for any $\mathbf{y}_\mathbf{i}$ and $\mathbf{k}$ such that $P\{D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y}_\mathbf{i}})\} > 0$ with $P_{\boldsymbol{\theta}_0}$-probability 1, and any $\mathbf{y}_{\mathbf{i}'}$ and $\mathbf{k}'$ such that $P\{\mathbf{v}_\mathbf{d}(\mathbf{Y}_{\mathbf{y}_{\mathbf{i}'}}, \mathbf{A}^\star, \mathbf{Z}^\star) = \mathbf{v}_{\mathbf{k}'}(\mathbf{y}_{\mathbf{i}'}, \mathbf{A}^\star_{\mathbf{i}'}, \mathbf{Z}^\star_{\mathbf{i}'}), D_{\mathbf{i}'\mathbf{k}'}(\mathbf{Y}_{\mathbf{y}_{\mathbf{i}'}})\} > 0$ with $P_{\boldsymbol{\theta}_0}$-probability 1,

$$
\begin{aligned}
&\frac{P\{Q(\mathbf{Y}_{\mathbf{y}_\mathbf{i}}, \mathbf{A}^\star, \mathbf{Z}^\star) \text{ is unbounded}, D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y}_\mathbf{i}})\}}{P\{\mathbf{v}_\mathbf{d}(\mathbf{Y}_{\mathbf{y}_{\mathbf{i}'}}, \mathbf{A}^\star, \mathbf{Z}^\star) = \mathbf{v}_{\mathbf{k}'}(\mathbf{y}_{\mathbf{i}'}, \mathbf{A}^\star_{\mathbf{i}'}, \mathbf{Z}^\star_{\mathbf{i}'}), D_{\mathbf{i}'\mathbf{k}'}(\mathbf{Y}_{\mathbf{y}_{\mathbf{i}'}})\}} \\
&= \frac{P\{Q(\mathbf{Y}_{\mathbf{y}_\mathbf{i}}, \mathbf{A}^\star, \mathbf{Z}^\star) \text{ is unbounded}, D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y}_\mathbf{i}})\}}{P\{D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y}_\mathbf{i}})\}} \cdot \frac{P\{D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y}_\mathbf{i}})\}}{P\{\mathbf{v}_\mathbf{d}(\mathbf{Y}_{\mathbf{y}_{\mathbf{i}'}}, \mathbf{A}^\star, \mathbf{Z}^\star) = \mathbf{v}_{\mathbf{k}'}(\mathbf{y}_{\mathbf{i}'}, \mathbf{A}^\star_{\mathbf{i}'}, \mathbf{Z}^\star_{\mathbf{i}'}), D_{\mathbf{i}'\mathbf{k}'}(\mathbf{Y}_{\mathbf{y}_{\mathbf{i}'}})\}} \\
&\to 0
\end{aligned}
\tag{S.21}
$$

in $P_{\boldsymbol{\theta}_0}$-probability. The proof of Equation S.21 is presented as the following two lemmas. Assuming that all the assumptions of Theorem 1 hold, we show that the first ratio in the second line of Equation S.21 converges to 0 in $P_{\boldsymbol{\theta}_0}$-probability (Lemma S.3), and that the second ratio is bounded with a $P_{\boldsymbol{\theta}_0}$-probability approaching 1 (Lemma S.4).

Lemma S.3. Fix $\varepsilon > 0$. For any $\mathbf{y}_\mathbf{i}$ and $\mathbf{k}$ such that $P\{D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y}_\mathbf{i}})\} > 0$ with $P_{\boldsymbol{\theta}_0}$-probability 1,

$$
P\{Q(\mathbf{Y}_{\mathbf{y}_\mathbf{i}}, \mathbf{A}^\star, \mathbf{Z}^\star) \text{ is unbounded} \mid D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y}_\mathbf{i}})\} \to 0 \tag{S.22}
$$

in $P_{\boldsymbol{\theta}_0}$-probability.

Proof. Fix $\delta > 0$ such that the δ-ball around $\boldsymbol{\theta}_0$, i.e., $C^\delta_{\boldsymbol{\theta}_0} = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le \delta\}$, is contained in the interior of the parameter space. The random quantity on the left-hand side (LHS) of

Equation S.22 has the following bound:

$$\begin{aligned}
&P\{Q(\mathbf{Y}_{\mathbf{y_i}}, \mathbf{A}^\star, \mathbf{Z}^\star) \text{ is unbounded} \mid D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y_i}})\} \\
=\ &P\{Q(\mathbf{Y}_{\mathbf{y_i}}, \mathbf{A}^\star, \mathbf{Z}^\star) \text{ is unbounded}, \mathbf{v}_{\mathbf{ik}}(\mathbf{y_i}, \mathbf{A}_\mathbf{i}^\star, \mathbf{Z}_\mathbf{i}^\star) \in C^\delta_{\boldsymbol{\theta}_0} \mid D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y_i}})\} \\
&+ P\{Q(\mathbf{Y}_{\mathbf{y_i}}, \mathbf{A}^\star, \mathbf{Z}^\star) \text{ is unbounded}, \mathbf{v}_{\mathbf{ik}}(\mathbf{y_i}, \mathbf{A}_\mathbf{i}^\star, \mathbf{Z}_\mathbf{i}^\star) \notin C^\delta_{\boldsymbol{\theta}_0} \mid D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y_i}})\} \\
\leq\ &P\{Q(\mathbf{Y}_{\mathbf{y_i}}, \mathbf{A}^\star, \mathbf{Z}^\star) \text{ is unbounded}, \mathbf{v}_{\mathbf{ik}}(\mathbf{y_i}, \mathbf{A}_\mathbf{i}^\star, \mathbf{Z}_\mathbf{i}^\star) \in C^\delta_{\boldsymbol{\theta}_0} \mid D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y_i}})\} \\
&+ P\{\mathbf{v}_{\mathbf{ik}}(\mathbf{y_i}, \mathbf{A}_\mathbf{i}^\star, \mathbf{Z}_\mathbf{i}^\star) \notin C^\delta_{\boldsymbol{\theta}_0} \mid D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y_i}})\} \qquad \text{(S.23)}
\end{aligned}$$

The second term in the last line of Equation S.23 converges to 0 in $P_{\boldsymbol{\theta}_0}$-probability, because a Bernstein-von Mises theorem, which further implies consistency, can be established for the density proportional to $b_{\mathbf{y_i k}}(\boldsymbol{\theta}) f_n(\boldsymbol{\theta}, \mathbf{y})$ using similar techniques described in ? (?). Also let

$$u_{\mathbf{ik}}(\boldsymbol{\theta}, \mathbf{y}) = P\{Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) \text{ is unbounded} \mid \mathbf{v}_{\mathbf{ik}}(\mathbf{y_i}, \mathbf{A}_\mathbf{i}^\star, \mathbf{Z}_\mathbf{i}^\star) = \boldsymbol{\theta}, D_{\mathbf{ik}}(\mathbf{y})\}. \qquad \text{(S.24)}$$

The first term in the last line of Equation S.23 can be bounded by a constant multiple of $\int_{C^\delta_{\boldsymbol{\theta}_0}} u_{\mathbf{ik}}(\boldsymbol{\theta}, \mathbf{Y}_{\mathbf{y_i}}) d\boldsymbol{\theta}$ due to the boundedness of $C^\delta_{\boldsymbol{\theta}_0}$. Thus, it suffices to show that

$$P\left\{\int_{C^\delta_{\boldsymbol{\theta}_0}} u_{\mathbf{ik}}(\boldsymbol{\theta}, \mathbf{Y}_{\mathbf{y_i}}) d\boldsymbol{\theta}\right\} \to 0 \qquad \text{(S.25)}$$

in $P_{\boldsymbol{\theta}_0}$-probability.

Fix a $\boldsymbol{\theta} \in C^\delta_{\boldsymbol{\theta}_0}$. Let H_{jk} be a size-$[2(r_j + 1)]$ subsample such that $y_{ij} = k - 1$ for exactly $r_j + 1$ respondents therein and $y_{ij} = k$ for the rest. Let H be a union of disjoint H_{jk}'s across all $j = 1, \ldots, m$, and $k = 1, \ldots, K_j - 1$. Subsamples of form H are referred to as *eligible* in the sequel, in the sense that each H may determine a collection of bounded polytopes (one for each H_{jk}) that contains $\boldsymbol{\theta}$ with a positive probability. For example, a sufficient condition for H to determine such a bounded polytope is to require that a) for each H_{jk}, the outward normal vectors corresponding to the left half-spaces of $\{i \in H_{jk} : y_{ij} = k-1\}$ and the right half-spaces of $\{i \in H_{jk} : y_{ij} = k\}$ fall respectively in the $2(r_j+1)$ orthants on the sub-parameter-space of α_{jk} and the free slopes in $\boldsymbol{\beta}_j$, and that b) conditional on the normal variates, the logistic variates are chosen such that $\boldsymbol{\theta} \in \cap_{i \in H_{jk}} Q_{ij}(y_{ij}, a_{ij}, \mathbf{z}_i)$. a) guarantees boundedness, which is a consequence of an extension of the (iii) $\Rightarrow$ (i) part of Lemma 4 in ? (?)[7]; in the meantime, b) ensures the inclusion of $\boldsymbol{\theta}$ in the interior of the resulting intersection.

Without loss of generality, assume $\mathbf{i}$ consists of the observations at the beginning of the sample. Define $t_n(\mathbf{y})$ as the number of disjoint eligible sub-samples in $\mathbf{i}^c = \{i : i \notin \mathbf{i}\}$; the sub-samples are denoted by $H_1, \ldots, H_{t_n(\mathbf{y})}$. Given a sample path $\mathbf{y}$ such that $t_n(\mathbf{y}) \to \infty$ as $n \to \infty$, we have

$$P\left\{\limsup_{t_n(\mathbf{y})\to\infty}\{H_{t_n(\mathbf{y})} \text{ determines bounded polytopes}\} \mid \mathbf{v}_{\mathbf{ik}}(\mathbf{y}_{\mathbf{i}}, \mathbf{A}^\star_{\mathbf{i}}, \mathbf{Z}^\star_{\mathbf{i}}) = \boldsymbol{\theta}, D_{\mathbf{ik}}(\mathbf{y})\right\} = 1 \tag{S.26}$$

by the second Borel-Cantelli Lemma; the independence of events is resulted from our selection of non-overlapping sub-samples. Moreover, if $H_{t_n(\mathbf{y})}$ determines bounded polytopes, then

[7] In the Lemma 4 of ? (?), the number of half-spaces is fixed at 1 plus the dimension of those half-spaces; however, the proof of the (iii) $\Rightarrow$ (i) part therein also applies to larger collections of half-spaces.

$Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star)$ is certainly bounded; therefore, we have

$$\begin{aligned}
1 &= P\left\{\limsup_{t_n(\mathbf{y})\to\infty}\{H_{t_n(\mathbf{y})} \text{ determines bounded polytopes}\} \mid \mathbf{v}_{\mathbf{ik}}(\mathbf{y}_\mathbf{i}, \mathbf{A}^\star_\mathbf{i}, \mathbf{Z}^\star_\mathbf{i}) = \boldsymbol{\theta}, D_{\mathbf{ik}}(\mathbf{y})\right\}\\
&\leq P\left\{\limsup_{t_n(\mathbf{y})\to\infty}\{Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) \text{ is bounded}\} \mid \mathbf{v}_{\mathbf{ik}}(\mathbf{y}_\mathbf{i}, \mathbf{A}^\star_\mathbf{i}, \mathbf{Z}^\star_\mathbf{i}) = \boldsymbol{\theta}, D_{\mathbf{ik}}(\mathbf{y})\right\}\\
&= P\left\{\lim_{t_n(\mathbf{y})\to\infty}\{Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) \text{ is bounded}\} \mid \mathbf{v}_{\mathbf{ik}}(\mathbf{y}_\mathbf{i}, \mathbf{A}^\star_\mathbf{i}, \mathbf{Z}^\star_\mathbf{i}) = \boldsymbol{\theta}, D_{\mathbf{ik}}(\mathbf{y})\right\}\\
&\leq \lim_{t_n(\mathbf{y})\to\infty} P\left\{Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) \text{ is bounded} \mid \mathbf{v}_{\mathbf{ik}}(\mathbf{y}_\mathbf{i}, \mathbf{A}^\star_\mathbf{i}, \mathbf{Z}^\star_\mathbf{i}) = \boldsymbol{\theta}, D_{\mathbf{ik}}(\mathbf{y})\right\},
\end{aligned} \tag{S.27}$$

which implies $u_{\mathbf{ik}}(\boldsymbol{\theta}, \mathbf{y})$ converges to 0 along the subsequence $t_n(\mathbf{y})$. In Equation S.27, the third line follows from the fact that events $\{Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star)$ is bounded$\}$ form a monotonically non-decreasing sequence as $n \to \infty$, and the last inequality is obtained by applying Fatou's Lemma. The convergence result holds for every $\boldsymbol{\theta} \in C^\delta_{\boldsymbol{\theta}_0}$ along the same subsequence, and thus for the integral because $C^\delta_{\boldsymbol{\theta}_0}$ is a bounded set. Since the true parameters $\boldsymbol{\theta}_0$ are assumed to lie in the interior of the parameter space, all response pattern probabilities are positive. Therefore, $t_n(\mathbf{Y}) \to \infty$ in $P_{\boldsymbol{\theta}_0}$-probability as $n \to \infty$, from which the desired result follows.

□

Lemma S.4. Fix $\mathbf{y}_\mathbf{i}$ and $\mathbf{k}$ such that $P\{D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y}_\mathbf{i}})\} > 0$ with $P_{\boldsymbol{\theta}_0}$-probability 1, and $\mathbf{y}_{\mathbf{i}'}$ and $\mathbf{k}'$ such that $P\{\mathbf{v}_\mathbf{d}(\mathbf{Y}_{\mathbf{y}_{\mathbf{i}'}}, \mathbf{A}^\star, \mathbf{Z}^\star) = \mathbf{v}_{\mathbf{k}'}(\mathbf{y}_{\mathbf{i}'}, \mathbf{A}^\star_{\mathbf{i}'}, \mathbf{Z}^\star_{\mathbf{i}'}), D_{\mathbf{i}'\mathbf{k}'}(\mathbf{Y}_{\mathbf{y}_{\mathbf{i}'}})\} > 0$ with $P_{\boldsymbol{\theta}_0}$-probability 1. There exists an $L > 0$ such that

$$P_{\boldsymbol{\theta}_0}\left\{\frac{P\{D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y}_\mathbf{i}})\}}{P\{\mathbf{v}_\mathbf{d}(\mathbf{Y}_{\mathbf{y}_{\mathbf{i}'}}, \mathbf{A}^\star, \mathbf{Z}^\star) = \mathbf{v}_{\mathbf{k}'}(\mathbf{y}_{\mathbf{i}'}, \mathbf{A}^\star_{\mathbf{i}'}, \mathbf{Z}^\star_{\mathbf{i}'}), D_{\mathbf{i}'\mathbf{k}'}(\mathbf{Y}_{\mathbf{y}_{\mathbf{i}'}})\}} \leq L\right\} \to 1. \tag{S.28}$$

Proof. By Equations S.18 and S.20, the ratio in Equation S.28 can be expressed as

$$\frac{P\{D_{\mathbf{ik}}(\mathbf{Y}_{\mathbf{y_i}})\}}{P\{\mathbf{v_d}(\mathbf{Y}_{\mathbf{y_{i'}}}, \mathbf{A}^\star, \mathbf{Z}^\star) = \mathbf{v}_{\mathbf{k}'}(\mathbf{y}_{\mathbf{i}'}, \mathbf{A}^\star_{\mathbf{i}'}, \mathbf{Z}^\star_{\mathbf{i}'}), D_{\mathbf{i}'\mathbf{k}'}(\mathbf{Y}_{\mathbf{y_{i'}}})\}} = \frac{\int b_{\mathbf{y_i k}}(\boldsymbol{\theta}, \mathbf{y_i}) f_n(\boldsymbol{\theta}, \mathbf{Y}_{\mathbf{y_i}}) d\boldsymbol{\theta}}{\int b^{\mathbf{d}}_{\mathbf{y_{i'} k}}(\boldsymbol{\theta}, \mathbf{y_{i'}}) f_n(\boldsymbol{\theta}, \mathbf{Y}_{\mathbf{y_{i'}}}) d\boldsymbol{\theta}}. \tag{S.29}$$

The density proportional to the numerator integrand satisfies a Bernstein-von Mises theorem. As a consequence,

$$\begin{aligned}\int_{C^\delta_{\boldsymbol{\theta}_0}} b_{\mathbf{y_i k}}(\boldsymbol{\theta}, \mathbf{y_i}) f_n(\boldsymbol{\theta}, \mathbf{Y}_{\mathbf{y_i}}) d\boldsymbol{\theta} &\le \int b_{\mathbf{y_i k}}(\boldsymbol{\theta}, \mathbf{y_i}) f_n(\boldsymbol{\theta}, \mathbf{Y}_{\mathbf{y_i}}) d\boldsymbol{\theta} \\ &= (1+\epsilon_n) \int_{C^\delta_{\boldsymbol{\theta}_0}} b_{\mathbf{y_i k}}(\boldsymbol{\theta}, \mathbf{y_i}) f_n(\boldsymbol{\theta}, \mathbf{Y}_{\mathbf{y_i}}) d\boldsymbol{\theta},\end{aligned} \tag{S.30}$$

for any δ such that $C^\delta_{\boldsymbol{\theta}_0}$ is contained in the interior of the parameter space, in which $\epsilon_n \downarrow 0$. Since $b_{\mathbf{y_i k}}(\boldsymbol{\theta}, \mathbf{y_i})$ is a continuous function of $\boldsymbol{\theta}$ (see Equation S.13), it is bounded on $C^\delta_{\boldsymbol{\theta}_0}$.

Because $b^{\mathbf{d}}_{\mathbf{y_i k}}(\boldsymbol{\theta}, \mathbf{y_i})$ is also bounded on $C^\delta_{\boldsymbol{\theta}_0}$ by continuity, it follows that

$$\frac{\int b_{\mathbf{y_i k}}(\boldsymbol{\theta}, \mathbf{y_i}) f_n(\boldsymbol{\theta}, \mathbf{Y}_{\mathbf{y_i}}) d\boldsymbol{\theta}}{\int b^{\mathbf{d}}_{\mathbf{y_{i'} k}}(\boldsymbol{\theta}, \mathbf{y_{i'}}) f_n(\boldsymbol{\theta}, \mathbf{Y}_{\mathbf{y_{i'}}}) d\boldsymbol{\theta}} \le \frac{\sup_{\boldsymbol{\theta} \in C^\delta_{\boldsymbol{\theta}_0}} b_{\mathbf{y_i k}}(\boldsymbol{\theta}, \mathbf{y_i})}{\inf_{\boldsymbol{\theta} \in C^\delta_{\boldsymbol{\theta}_0}} b^{\mathbf{d}}_{\mathbf{y_{i'} k}}(\boldsymbol{\theta}, \mathbf{y_{i'}})} (1+\epsilon_n). \tag{S.31}$$

Note that the infimum in the denominator of Equation S.31 is bounded away from 0, because the selection of $\mathbf{y_{i'}}$ and $\mathbf{k}'$ guarantees that an extremal point along $\mathbf{d}$ is produced with a positive probability. This concludes the proof of Lemma S.4. □

Appendix C: Conditional sampling of $A^\star_{ij}$

Fix observation i and item j. The goal of this step is to obtain an updated $A^\star_{ij}$ such that the implied half-spaces have a non-empty intersection with the interior polytope determined by all but the ith observations evaluated at the current values of the corresponding random components; the latter is readily available from Line 5 of Algorithm 1 in ? (?). Here, we only describe the case when a middle category on the response scale is selected: i.e.,

$0 < y_{ij} < K_j - 1$. The workaround we implement to reduce the impact of a long-tailed fiducial distribution (see the later discussion in Appendix G) amounts to augmenting the actual response scale with two "phantom" extreme categories that have no endorsement in the observed data. This effectively turns every actual response option into a middle category, and thus the discussion here suffices.

For notational convenience, we use a superscript 0 to highlight the dependency solely on the current values of random components at this particular sampling step, and a superscript 1 to denote the involvement of the updated one. Let $\mathbf{a}^0_{-i(j)}$ and $\mathbf{z}^0_{-i}$ be the current values of the logistic and normal variates without the ith observation, $\mathbf{a}^1_{(j)}$ be the logistic variates including the updated ith component, and $\mathbf{y}_{-i(j)} = (y_{i'j})_{i' \neq i}$ be the corresponding item responses to the same item. The updated value a^1_{ij} should yield a non-empty polytope: i.e.,

$$Q_j(\mathbf{y}_{(j)}, \mathbf{a}^1_{(j)}, \mathbf{z}^0) = Q_j(\mathbf{y}_{-i(j)}, \mathbf{a}^0_{-i(j)}, \mathbf{z}^0) \cap Q_{ij}(y_{ij}, a^1_{ij}, \mathbf{z}^0) \neq \emptyset. \tag{S.32}$$

Denoted V^0_{-ij} the collection of vertices of $Q_j(\mathbf{y}_{-i(j)}, \mathbf{a}^0_{-i(j)}, \mathbf{z}^0)$. Due to convexity, Equation S.32 is further identical to the existence of at least one element of V^0_{-ij} being consistent with each of the two updated half-spaces: i.e.,

$$a^1_{ij} > \tau_{j,y_{ij}+1}(\boldsymbol{\theta}_j, \mathbf{z}^0_i) \tag{S.33}$$

for some $\boldsymbol{\theta}_j \in V^0_{-ij}$, and

$$a^1_{ij} \leq \tau_{jy_{ij}}(\boldsymbol{\theta}_j, \mathbf{z}^0_i), \tag{S.34}$$

for some $\boldsymbol{\theta}_j \in V^0_{-ij}$ as well. It follows that $A^\star_{ij} = a^1_{ij}$ should be generated from a standard logistic distribution truncated to $[\min_{\boldsymbol{\theta}_j \in V^0_{-ij}} \tau_{j,y_{ij}+1}(\boldsymbol{\theta}_j, \mathbf{z}^0_i), \max_{\boldsymbol{\theta}_j \in V^0_{-ij}} \tau_{jy_{ij}}(\boldsymbol{\theta}_j, \mathbf{z}^0_i)]$. An implementation of the updating step is described in Algorithm S.1.

Algorithm S.1 Updating $A_{ij}^{\star}$

1: set $m = \infty$ and $M = -\infty$
2: **for** $\boldsymbol{\theta}_j \in V_{-ij}^0$ **do**
3: compute $m_1 = \tau_{j,y_{ij}+1}(\boldsymbol{\theta}_j, \mathbf{z}_i^0)$
4: **if** $m_1 < m$ **then**
5: $m = m_1$
6: **end if**
7: compute $m_2 = \tau_{jy_{ij}}(\boldsymbol{\theta}_j, \mathbf{z}_i^0)$
8: **if** $m_2 > M$ **then**
9: $M = m_2$
10: **end if**
11: **end for**
12: generate $A_{ij}^{\star} = a_{ij}^1$ from the logistic distribution truncated to $[m, M]$

Samples from truncated logistic distributions (Line 12) are obtained by an implementation of the slice sampler (Neal, 2003), which is by itself an MCMC algorithm; five cycles are performed for each call of the sampler, which appears to behave well in a pilot study. We also found that slice sampling outperforms the inverse cumulative distribution function (cdf) approach when the truncation bounds are extreme.

Appendix D: Conditional sampling of $Z_{id}^{\star}$

The conditional sampling of $Z_{id}^{\star}$ is slightly more involved than that of $A_{ij}^{\star}$, because a single $Z_{id}^{\star}$ may be associated with multiple items. $Z_{id}^{\star} = z_{id}^1$ should be sampled from a suitably truncated standard normal distribution ensuring for each associated item that the updated interior polytope is not empty.

Fix i and d. Let $\mathbf{z}_{i,-d}^0 = (z_{ie}^0)_{e \neq d}$ be the current values of all but the dth dimension of the normal variates, and $\boldsymbol{\theta}_{j,-d}$ be the item parameters without the dth slope. Also write

$$\tau_{jk}^d(\boldsymbol{\theta}_{j,-d}, \mathbf{z}_{i,-d}^0) = \alpha_{jk} + \sum_{e \neq d} \beta_{je} z_{ie}^0. \tag{S.35}$$

For all items j loading on dimension d, the updated value z_{id}^1 should satisfy

$$\beta_{jd} z_{id}^1 < a_{ij}^0 - \tau_{j,y_{ij}+1}^d(\boldsymbol{\theta}_{j,-d}, \mathbf{z}_{i,-d}^0) \tag{S.36}$$

for some $\boldsymbol{\theta}_j \in V_{-ij}^0$, and

$$\beta_{jd} z_{id}^1 \geq a_{ij}^0 - \tau_{jy_{ij}}^d(\boldsymbol{\theta}_{j,-d}, \mathbf{z}_{i,-d}^0) \tag{S.37}$$

for some $\boldsymbol{\theta}_j \in V_{-ij}^0$ as well. Let J_d be the collection of items that are associated with $Z_{id}^\star$; equations S.36 and S.37 together yield the desirable truncation:

$$\begin{aligned} Z_{id}^\star \in \bigcap_{j \in J_d} \Bigg[&\Bigg(\bigcup_{\boldsymbol{\theta}_j \in V_{-ij}^0} \{z_{id}^1 : \beta_{jd} z_{id}^1 < a_{ij}^0 - \tau_{j,y_{ij}+1}^d(\boldsymbol{\theta}_{j,-d}, \mathbf{z}_{i,-d}^0)\} \Bigg) \\ &\cap \Bigg(\bigcup_{\boldsymbol{\theta}_j \in V_{-ij}^0} \{z_{id}^1 : \beta_{jd} z_{id}^1 \geq a_{ij}^0 - \tau_{jy_{ij}}^d(\boldsymbol{\theta}_{j,-d}, \mathbf{z}_{i,-d}^0)\} \Bigg) \Bigg]. \end{aligned} \tag{S.38}$$

Both equations S.36 and S.37 define one-sided intervals for z_{id}^1, the direction of which is contingent upon the sign of β_{jd} for each vertex in V_{-ij}^0. As a consequence, Equation S.38 might be an interval or a union of disjoint intervals. The foregoing updating mechanism is summarized as Algorithm S.2.

Again, the technique of slice sampling is used in Line 23 of Algorithm S.2. As mentioned earlier, the truncation T can be either a bounded interval, or a disjoint union of bounded intervals. In the latter case, the sampling is done in three steps: a) computing probabilities of the intervals under a standard normal distribution and normalizing to a total sum of one; b) randomly selecting an interval with probabilities computed in step a); c) slice sampling on the selected interval.

Algorithm S.2 Updating $Z_{id}^{\star}$

set $T = (-\infty, \infty)$
for items $j = 1, \ldots, m$ **do**
  if β_{jd} is fixed to 0 **then**
    cycle the item loop
  else
    set $T_j = \emptyset$
    for $\boldsymbol{\theta}_j \in V_{-ij}^0$ **do**
      if $\beta_{jd} = 0$ **then**
        cycle the vertex loop
      else
        compute $m_1 = [a_{ij}^0 - \tau_{jy_{ij}}^d(\boldsymbol{\theta}_{j,-d}, \mathbf{z}_{i,-d}^0)]/\beta_{jd}$
        compute $m_2 = [a_{ij}^0 - \tau_{j,y_{ij}+1}^d(\boldsymbol{\theta}_{j,-d}, \mathbf{z}_{i,-d}^0)]/\beta_{jd}$
        if $\beta_{jd} > 0$ **then**
          update $T_j = T_j \cup [m_1, m_2]$
        else
          update $T_j = T_j \cup [m_2, m_1]$
        end if
      end if
    end for
  end if
  update $T = T \cap T_j$
end for
generate $Z_{id}^{\star} = z_{id}^1$ from the standard normal distribution truncated to T

Appendix E: Updating interior polytopes

Inside the observation loop of Algorithm 1, all interior polytopes need to be renewed after the logistic and normal variates are updated. Geometrically, it amounts to cutting the old polytope formed by the rest of the observations, i.e., $Q_j(\mathbf{y}_{-i(j)}, \mathbf{a}_{-i(j)}^0, \mathbf{z}^0)$, by the two new half-spaces $\tau_{j,y_{ij}+1}(\boldsymbol{\theta}_j, \mathbf{z}_i^1) < a_{ij}^1$ and $\tau_{jy_{ij}}(\boldsymbol{\theta}_j, \mathbf{z}_i^1) \geq a_{ij}^1$; the resulting intersection is certainly non-empty due to the truncation enforced for $A_{ij}^{\star}$'s and $Z_{id}^{\star}$'s.

The updating algorithm requires an effective representation of the $\mathbb{R}^{q_j}$-polytope $Q_j(\mathbf{y}_{-i(j)}, \mathbf{a}_{-i(j)}^0, \mathbf{z}^0)$ for each item j. It is well-known that a convex polytope is uniquely determined by its vertices; so we need to record V_{-ij}^0. With a slight abuse of notation, we now consider V_{-ij}^0 as a set of doublets $\mathcal{V}_j = (\boldsymbol{\theta}_j, \mathbf{i}_j)$, in which $\mathbf{i}_j$ indexes the observations that are used to solve for

$\boldsymbol{\theta}_j$. If a half-space, say $\tau_{jy_{ij}}(\boldsymbol{\theta}_j, \mathbf{z}_i^1) \geq a_{ij}^1$, is known to cut the polytope, vertices in V_{-ij}^0 can be partitioned into two groups by whether or not they are consistent with the cutting half-space: Those satisfying $\tau_{jy_{ij}}(\boldsymbol{\theta}_j, \mathbf{z}_i^1) \geq a_{ij}^1$ are still *feasible*, and the rest become infeasible. In addition to vertices, we also track the edges of the old polytope, denoted E_{-ij}^0. Each edge connects a pair of vertices sharing $q_j - 1$ observations, i.e., $\mathcal{E}_j = (\mathcal{V}_j, \mathcal{V}_j')$, in which $\mathcal{V}_j = (\boldsymbol{\theta}_j, \mathbf{i}_j)$, $\mathcal{V}' = (\boldsymbol{\theta}_j', \mathbf{i}_j')$, and $|\mathbf{i}_j \cap \mathbf{i}_j'| = q_j - 1$; in other words, the shared observations determines this particular edge. One advantage of keeping the edges is that new vertices introduced by the cutting half-space can be easily obtained: An edge together with the cutting hyperplane $\tau_{jy_{ij}}(\boldsymbol{\theta}_j, \mathbf{z}_i^1) = a_{ij}^1$ produce a vertex if and only if the edge connects a feasible-infeasible pair of vertices, provided the resulting linear system is non-singular. In addition, the vertex and edge lists need to be updated; entries that are no longer feasible should be removed, and the new ones produced by the cutting half-space should be appended. A pseudo-code of the polytope-cutting procedure is provided as Algorithm S.3; in Line 12 of Algorithm 1, two executions of Algorithm S.3 are needed for the left and right half-spaces corresponding to a single observation, respectively.

Recording edges facilitates finding new vertices, i.e., Line 16-19 of Algorithm S.3; however, the algorithm may fail whenever the linear system determined by observations $(\mathbf{i}_j \cap \mathbf{i}_j') \cup \{i\}$ (Line 16) is singular. When $K_j > 2$, it could happen occasionally; it corresponds to the case that the new half-space cuts the polytope exactly along the edge. In theory, this loophole can be redressed by treating all vertices satisfying $\tau_{jy_{ij}}(\boldsymbol{\theta}_j, \mathbf{z}_i^1) - a_{ij}^1 = 0$ as infeasible; it follows that the edge being cut along is first removed in Line 14, and then added back in Line 24 with the new observation included in the index set. In the presence of numerical errors, a slacking constant $\varepsilon > 0$ should be used instead of the exact zero in practice; in the current implementation of the algorithm, ε is chosen to be 10^{-10}.

In terms of data structure, we recommend the use of linked lists (i.e., adjacent units are

Algorithm S.3 Cutting $Q_j(\mathbf{y}_{-i(j)}, \mathbf{a}^0_{-i(j)}, \mathbf{z}^0)$ by $\tau_{jy_{ij}}(\boldsymbol{\theta}_j, \mathbf{z}^1_i) \geq a^1_{ij}$

for $\mathcal{V}_j = (\boldsymbol{\theta}_j, \mathbf{i}_j) \in V^0_{-ij}$ **do** ▷ check feasibility
  if $\tau_{jy_{ij}}(\boldsymbol{\theta}_j, \mathbf{z}^1_i) \geq a^1_{ij}$ **then**
    cycle the vertex loop ($\mathcal{V}_j$ feasible)
  else
    remove $\mathcal{V}_j$ ($\mathcal{V}_j$ infeasible)
  end if
end for
if all vertices are feasible **then**
  terminate the program
end if
create an empty vertex list V^1_{ij} and an empty edge list E^1_{ij}
for $\mathcal{E}_j = (\mathcal{V}_j, \mathcal{V}'_j) \in E^0_{-ij}$ **do** ▷ obtain new vertices
  if both $\mathcal{V}_j$ and $\mathcal{V}'_j$ are feasible **then** cycle the edge loop
  else if neither $\mathcal{V}_j$ nor $\mathcal{V}'_j$ is feasible **then** remove $\mathcal{E}$
  else
    set $\mathbf{i}''_j = (\mathbf{i}_j \cap \mathbf{i}'_j) \cup \{i\}$
    calculate the new vertex determined by $\mathbf{i}''_j$, denoted $\boldsymbol{\theta}''_j$
    append $\mathcal{V}''_j = (\boldsymbol{\theta}''_j, \mathbf{i}''_j)$ to V^1_{ij}
    in $\mathcal{E}_j$, replace the infeasible vertex by $\mathcal{V}''_j$
  end if
end for
for $\mathcal{V}_j$, $\mathcal{V}'_j \in V^1_{ij}$ **do** ▷ obtain new edges
  if $|\mathbf{i}_j \cap \mathbf{i}'_j| = q_j - 1$ **then**
    add $(\mathcal{V}_j, \mathcal{V}'_j)$ to E^1_{ij}
  end if
end for
append V^1_{ij} to V^0_{-ij}
append E^1_{ij} to E^0_{-ij}

concatenated via pointers) instead of arrays (i.e., adjacent units are stored in consecutive memory locations) as containers of vertex and edge lists, for the reason that the former eases removal and addition of elements to arbitrary locations in the list, which appears frequently in Algorithm S.3.

Appendix F: Starting values

The proposed sampler requires initial values of the logistic and normal variates, denoted by $\mathbf{a}^{(0)}$ and $\mathbf{z}^{(0)}$, which imply a non-empty and bounded polytope $Q_j(\mathbf{y}_{(j)}, \mathbf{a}^{(0)}_{(j)}, \mathbf{z}^{(0)})$ for each j. There is certainly more than one way to achieve this. Our algorithm, described in this section, requires user-input of starting values $\boldsymbol{\theta}^{(0)}$ and factor score estimates $\mathbf{z}^{(0)}$. The logistic variates $\mathbf{a}^{(0)}$ are subsequently generated using Algorithm S.1, in which each interior polytope comprises only one vertex $\boldsymbol{\theta}^{(0)}_j$; the non-emptiness and boundedness of polytopes are ensured by the truncated sampling.

The boundedness requirement is unnecessary in theory; for each fixed $\mathbf{y}$ the polytope can be unbounded with a positive probability. However, the sampling algorithm, especially the polytope-updating part (Algorithm S.3), only applies to bounded cases. As a result, an arbitrarily specified initial bounding box is needed (similar configurations can be found in Cisewski and Hannig, 2012, and ?, ?). For the GRM, we define the following bounding box for $\boldsymbol{\theta}_j$:

$$
\begin{aligned}
&\alpha_{j1} \geq -M, \alpha_{j,K-1} \leq M \\
&\alpha_{jk} \geq \alpha_{j,k+1}, \text{ for all } k = 1, \ldots, K-2 \\
&-M \leq \beta_{jd} \leq M, \text{ for all } d = 1, \ldots, r.
\end{aligned} \tag{S.39}
$$

The parameter bound M is a tuning parameter of the sampling algorithm. We found in Section 4.1 that the choice of M (20 versus 200) does not have a substantial impact on the results. In general, we recommend using the value $M = 20$, which led to reliable and efficient point and interval estimates in various simulation studies we have conducted. Based on the foregoing discussion, we outline the starting value program as Algorithm S.4.

In practice, $\boldsymbol{\theta}^{(0)}$ can be provided by computationally economical limited information

Algorithm S.4 Starting values

for items $j = 1, \ldots, m$ **do**
  set V^0_{-ij} and E^0_{-ij} to represent the initial bounding box
  for observations $i = 1, \ldots, n$ **do**
    generate $A^\star_{ij} = a^{(0)}_{ij}$ from Logistic$(0, 1)$ truncated to $[\tau_{j,y_{ij}+1}(\boldsymbol{\theta}^{(0)}_j, \mathbf{z}^{(0)}_i),$ $\tau_{jy_{ij}}(\boldsymbol{\theta}^{(0)}_j, \mathbf{z}^{(0)}_i)]$
    Update the jth polytope (Algorithm S.3)
  end for
end for

estimators, such as various weighted least square methods based on polychoric correlations (e.g., Muthén, 1978; Gunsjö, 1994). Alternatively, one could use naive starting values such as ordered constants for intercepts and 1 for slopes. $\mathbf{z}^{(0)}$ can be generated from the conditional distribution of the latent variables given $\mathbf{y}$ and $\boldsymbol{\theta}^{(0)}$ or point estimates (e.g., EAP) derived from such distribution. $\mathbf{z}^{(0)}$ can also be generated from a standard normal distribution unconditionally. The naive starting values are indeed nowhere near the true item parameters and factor scores, nor the center of the fiducial distribution, but they work reasonably well in our Monte Carlo experiments. From our experience, the generated Markov chain appears stationary after about a thousand iterations, and the final results are not significantly affected by the choice of initial status.

Appendix G: Long-tailedness and a workaround

When the sample size is small relatively to the generating parameter values, the fiducial distribution tends to be long-tailed: It produces spikes on the trace plot (see the left panel of Figure S.1 for an illustration) which subsequently leads to excessively wide CIs and undue dependency of the inferential results on the size of the initial bounding box.

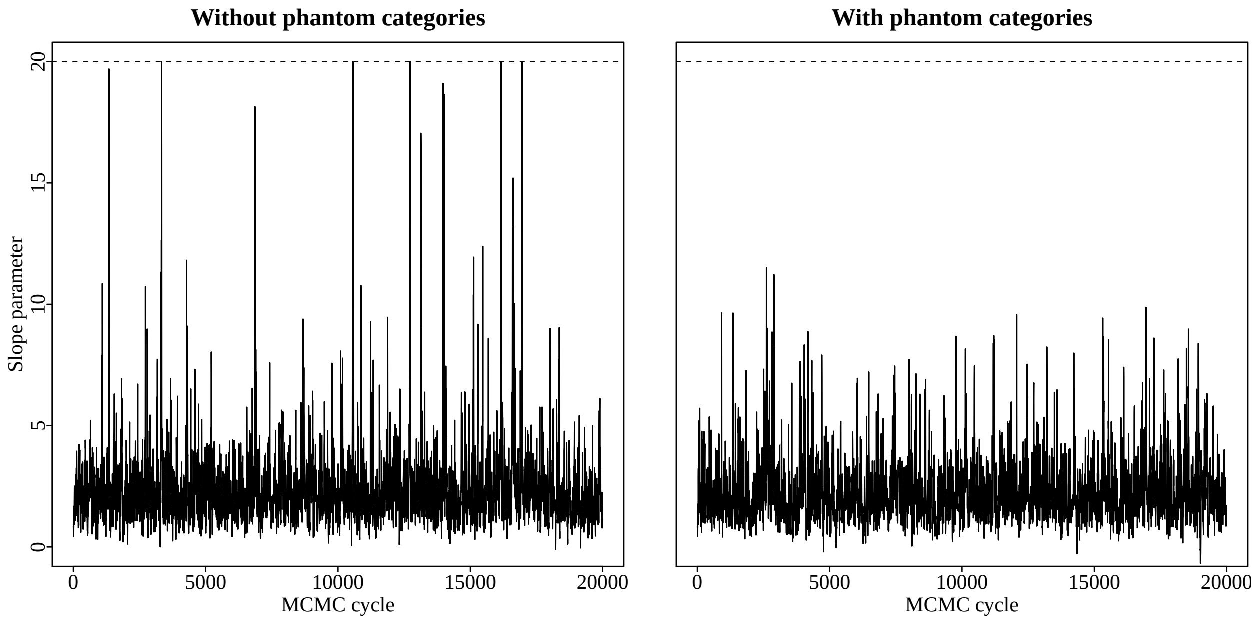

Figure S.1: Trace plot for a slope parameter before (left) and after (right) implementing the workaround. The data set used for illustration is composed of 50 observations and five 3-category items. The arbitrary bound M is set to 20, which is highlighted by the horizontal dashed lines.

? (?) proposed a workaround in the context of binary IRT modeling; their approach can be straightforwardly generalized to accommodate the ordinal case. In fact, the tuning operation has an interesting interpretation under the GRM. The idea is to introduce two "phantom" response categories outside the actual response scale (from 0 to $K_j - 1$), coded as $y_{ij} = -1$ and K_j, in company with two additional intercept parameters α_{j0} and α_{jK}. This extra configuration converts the actual extremal responses 0 and $K_j - 1$ into middle categories; therefore, the modified set inverse for Y_{ij} involves two-sided inequalities for all observable responses, i.e.,

$$\tilde{Q}_{ij}(y_{ij}, a_{ij}, \mathbf{z}_i) = \{\boldsymbol{\theta}_j \in \Theta_j : \alpha_{j,k+1} + \boldsymbol{\beta}_j{}^\top \mathbf{z}_i < a_{ij} \leq \alpha_{jk} + \boldsymbol{\beta}_j{}^\top \mathbf{z}_i\} \tag{S.40}$$

It follows that each observation provides both lower and upper bounds for each slope parameter. No endorsement of the phantom categories can be found in the observed data, so estimates of the extra intercepts are not meaningful. Moreover, freely estimating α_{j0} and α_{jK} increases the dimension of the parameter space, and results in longer computation time. Therefore, we fix $\alpha_{j0} = M$ and $\alpha_{jK} = -M$ in the current implementation, which has proved to reduce the influence of long-tailedness in our pilot investigation (see the right panel of Figure S.1).

Appendix H: Additional simulation results

More detailed comparisons among the five candidate interval estimators in the reported simulation studies (Sections 4.2 and 4.3) can be found in Figures S.2 to S.7, in which the coverage and median log length ratio (MLLR) results are summarized for each individual parameter.

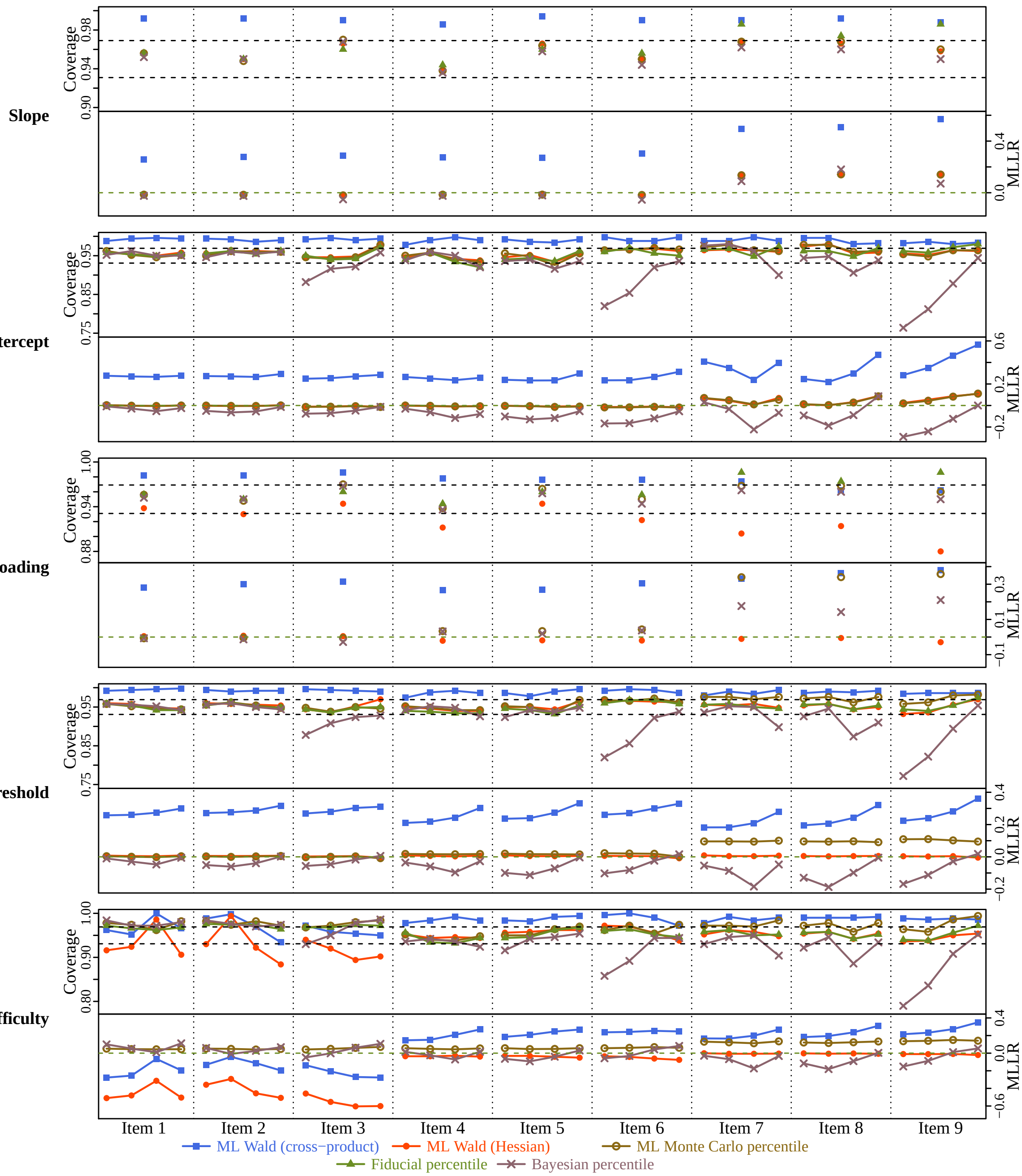

Figure S.2: Empirical coverages and median log length ratio (MLLR) of the four types of interval estimators (shown in different colors) under a unidimensional GRM. Here, the sample size $n = 100$ and the number of items $m = 9$. Each row corresponds to one type of parameter, in which coverage is plotted in the upper panel and median length relative to the fiducial percentile CI in the lower panel, and parameters belonging to different items are separated by vertical dotted lines. The two horizontal dashed lines on the upper panel gives a 95% normal-approximation confidence band for the nominal level 0.95. The horizontal dashed line on the lower panel indicates the median length of the fiducial percentile CI.

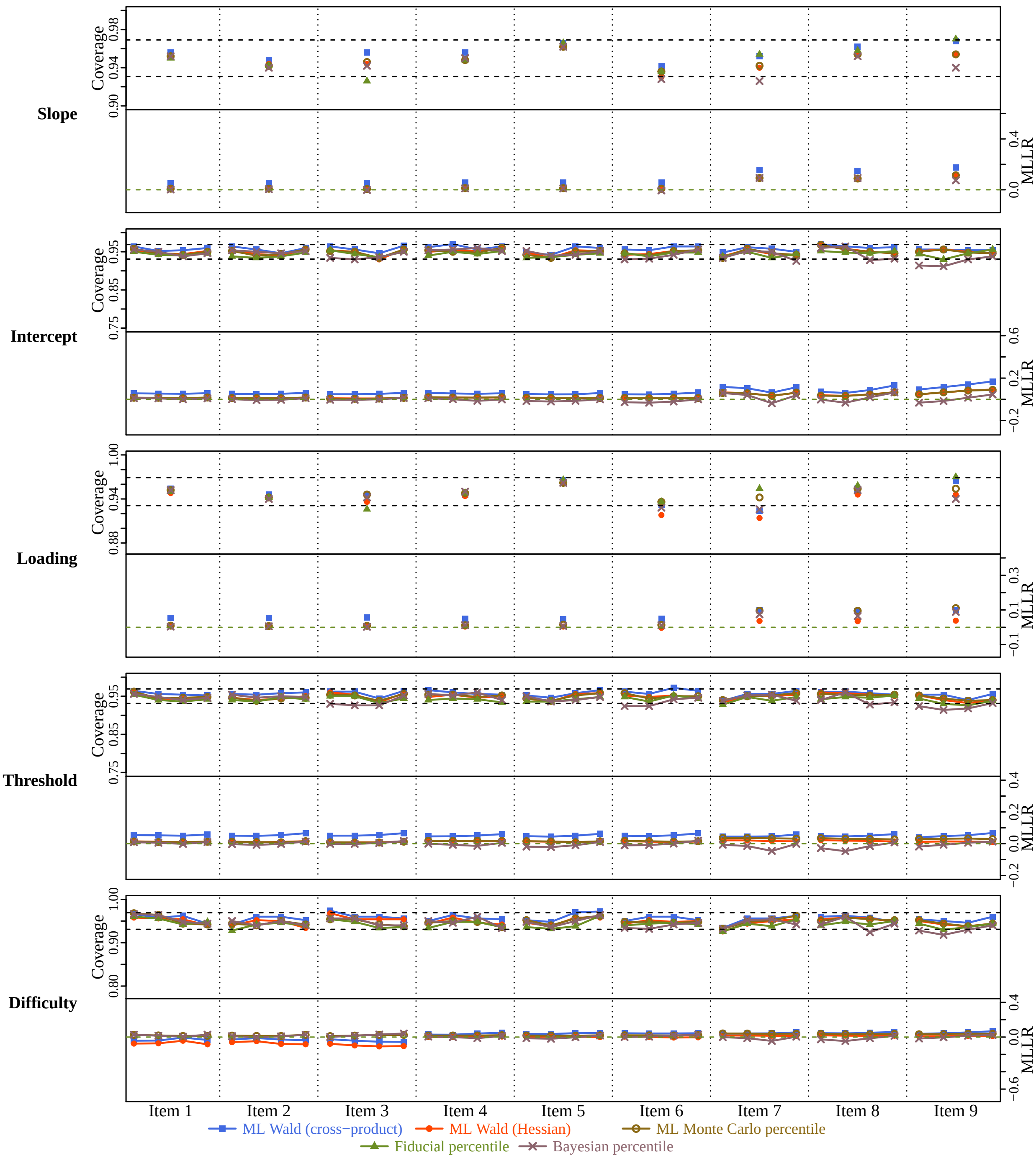

Figure S.3: Empirical coverages and median log length ratio (MLLR) of the four types of interval estimators (shown in different colors) under a unidimensional GRM. Here, the sample size $n = 500$ and the number of items $m = 9$. Each row corresponds to one type of parameter, in which coverage is plotted in the upper panel and median length relative to the fiducial percentile CI in the lower panel, and parameters belonging to different items are separated by vertical dotted lines. The two horizontal dashed lines on the upper panel gives a 95% normal-approximation confidence band for the nominal level 0.95. The horizontal dashed line on the lower panel indicates the median length of the fiducial percentile CI.

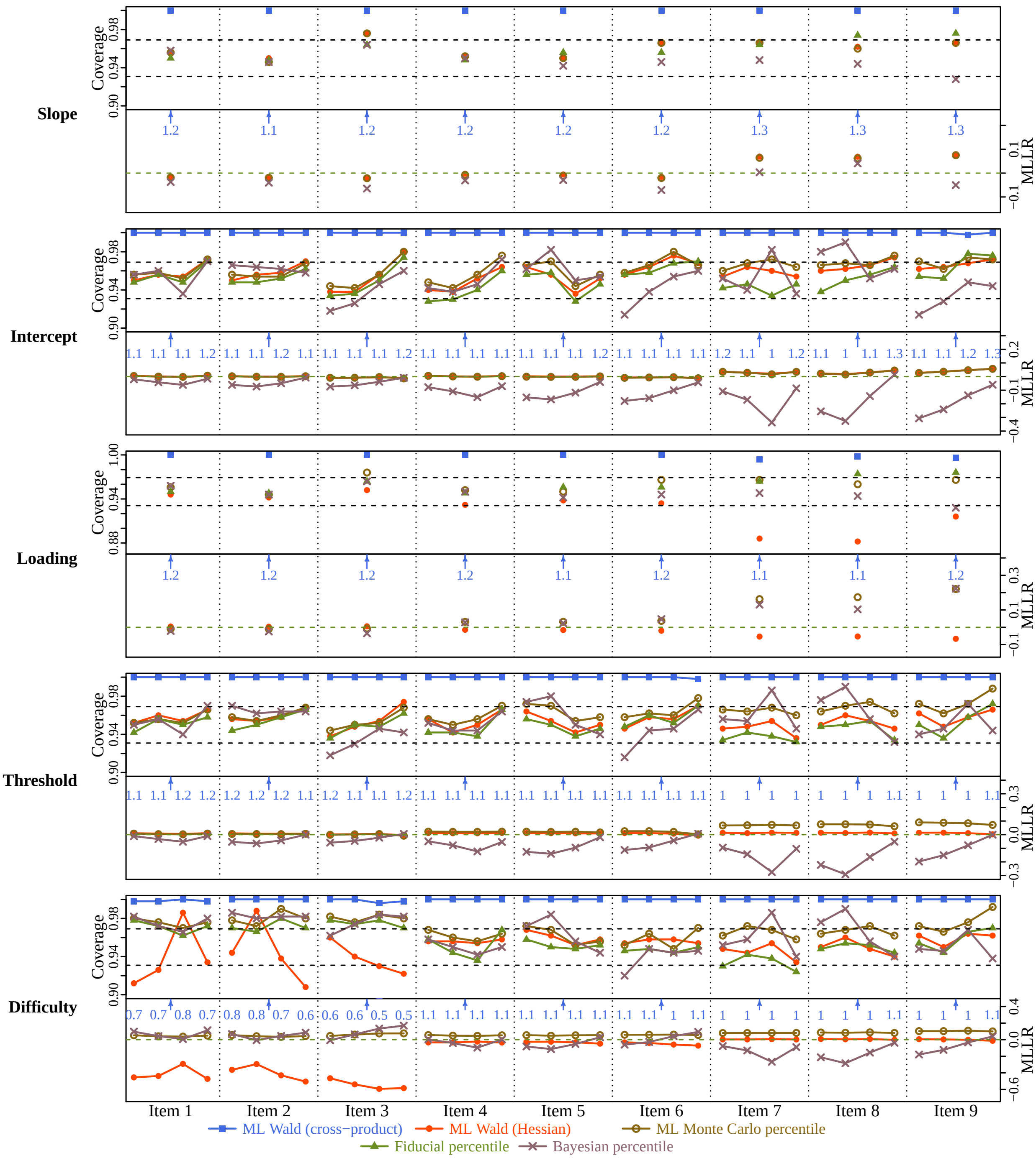

Figure S.4: Empirical coverages and median log length ratio (MLLR) of the four types of interval estimators (shown in different colors) under a unidimensional GRM. Here, the sample size $n = 100$ and the number of items $m = 18$. Each row corresponds to one type of parameter, in which coverage is plotted in the upper panel and median length relative to the fiducial percentile CI in the lower panel, and parameters belonging to different items are separated by vertical dotted lines. The two horizontal dashed lines on the upper panel gives a 95% normal-approximation confidence band for the nominal level 0.95. The horizontal dashed line on the lower panel indicates the median length of the fiducial percentile CI. The blue-colored numbers indicate the median length ratios for the cross-product-type Wald CI.

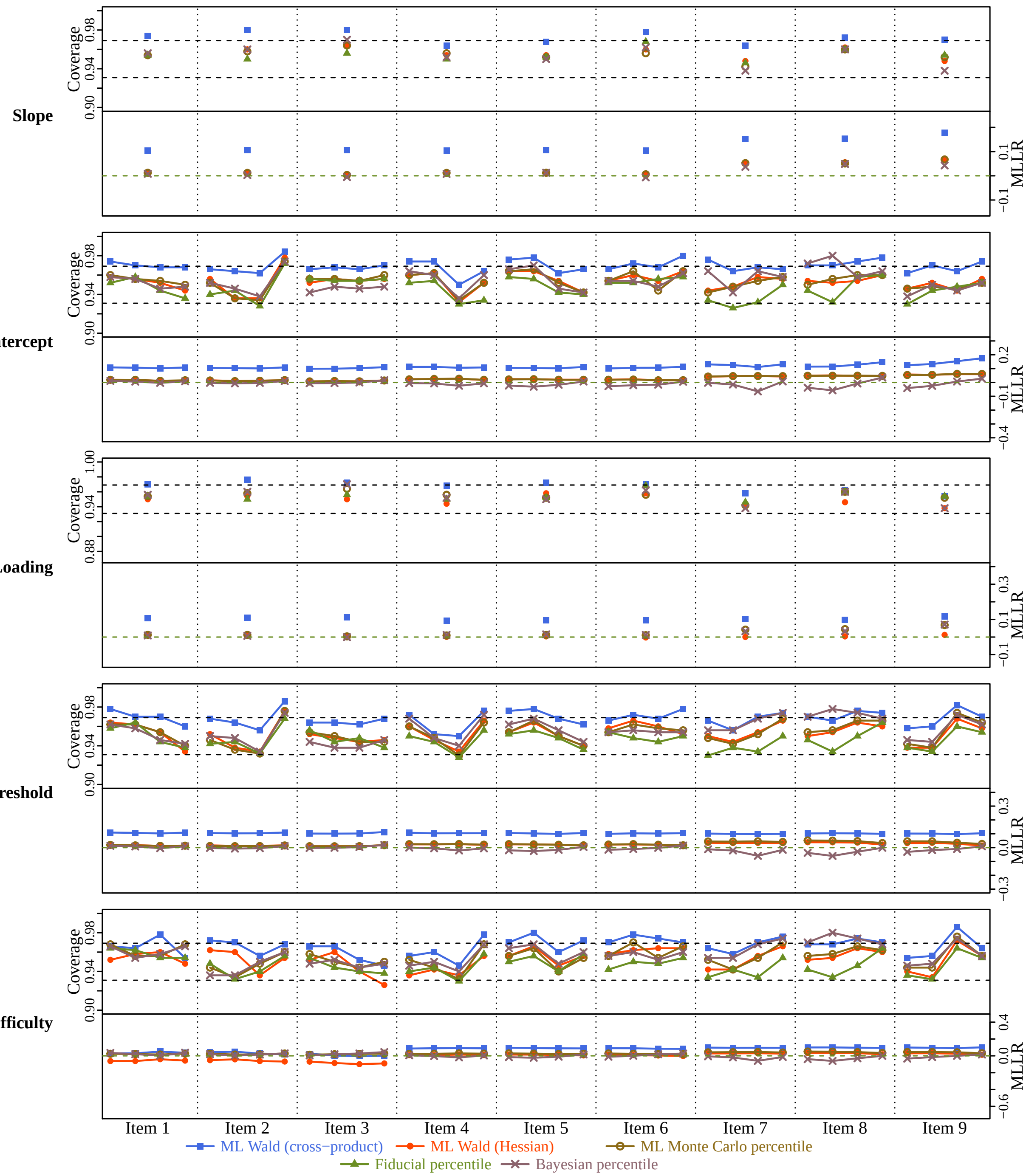

Figure S.5: Empirical coverages and median log length ratio (MLLR) of the four types of interval estimators (shown in different colors) under a unidimensional GRM. Here, the sample size $n = 500$ and the number of items $m = 18$. Each row corresponds to one type of parameter, in which coverage is plotted in the upper panel and median length relative to the fiducial percentile CI in the lower panel, and parameters belonging to different items are separated by vertical dotted lines. The two horizontal dashed lines on the upper panel gives a 95% normal-approximation confidence band for the nominal level 0.95. The horizontal dashed line on the lower panel indicates the median length of the fiducial percentile CI.

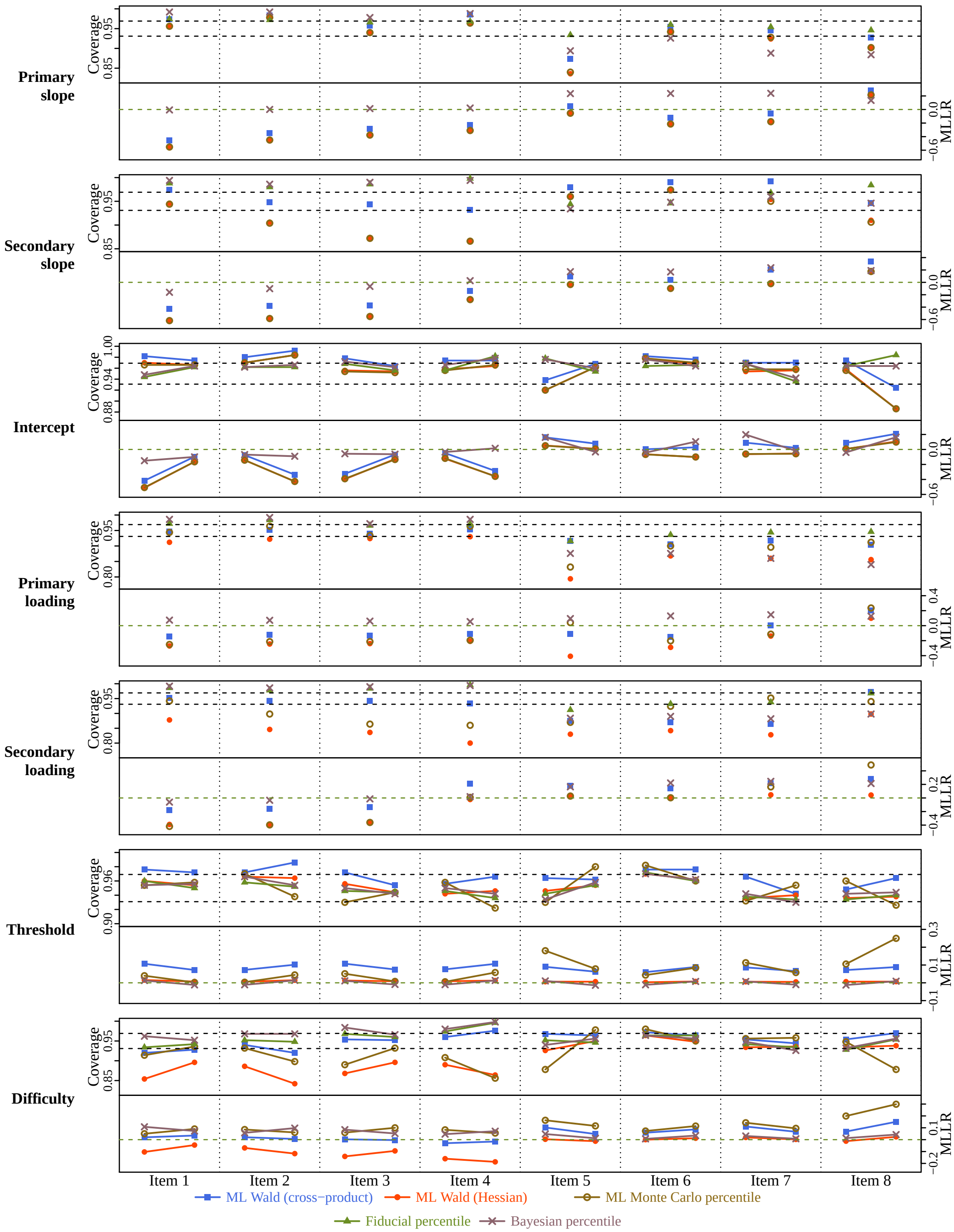

Figure S.6: Empirical coverages and median log length ratio (MLLR) of the five types of interval estimators (shown in different colors) under a bifactor GRM. Here, the sample size $n = 200$ and the number of items $m = 8$. Each row corresponds to one type of parameter, in which coverage is plotted in the upper panel and median length relative to the fiducial percentile CI in the lower panel, and parameters belonging to different items are separated by vertical dotted lines. The two horizontal dashed lines on the upper panel gives a 95% normal-approximation confidence band for the nominal level 0.95. The horizontal dashed line on the lower panel indicates the median length of the fiducial percentile CI.

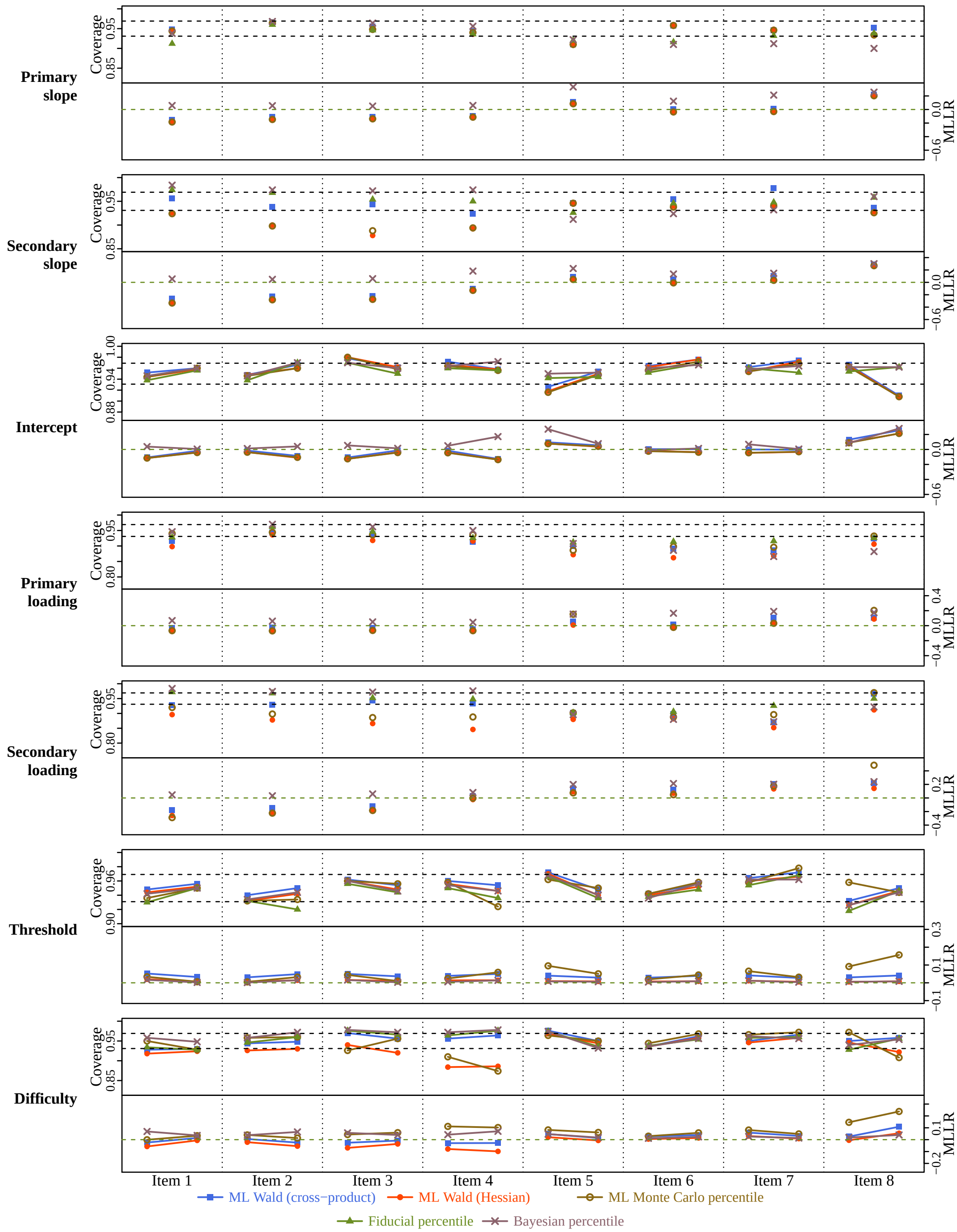

Figure S.7: Empirical coverages and median log length ratio (MLLR) of the five types of interval estimators (shown in different colors) under a bifactor GRM. Here, the sample size $n = 500$ and the number of items $m = 8$. Each row corresponds to one type of parameter, in which coverage is plotted in the upper panel and median length relative to the fiducial percentile CI in the lower panel, and parameters belonging to different items are separated by vertical dotted lines. The two horizontal dashed lines on the upper panel gives a 95% normal-approximation confidence band for the nominal level 0.95. The horizontal dashed line on the lower panel indicates the median length of the fiducial percentile CI.