GENERALIZED FIDUCIAL INFERENCE FOR BINARY LOGISTIC ITEM RESPONSE MODELS

YANG LIU

UNIVERSITY OF CALIFORNIA, MERCED

JAN HANNIG

THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

Generalized fiducial inference (GFI) has been proposed as an alternative to likelihood-based and Bayesian inference in mainstream statistics. Confidence intervals (CIs) can be constructed from a fiducial distribution on the parameter space in a fashion similar to those used with a Bayesian posterior distribution. However, no prior distribution needs to be specified, which renders GFI more suitable when no *a priori* information about model parameters is available. In the current paper, we apply GFI to a family of binary logistic item response theory models, which includes the two-parameter logistic (2PL), bifactor and exploratory item factor models as special cases. Asymptotic properties of the resulting fiducial distribution are discussed. Random draws from the fiducial distribution can be obtained by the proposed Markov chain Monte Carlo sampling algorithm. We investigate the finite-sample performance of our fiducial percentile CI and two commonly used Wald-type CIs associated with maximum likelihood (ML) estimation via Monte Carlo simulation. The use of GFI in high-dimensional exploratory item factor analysis was illustrated by the analysis of a set of the Eysenck Personality Questionnaire data.

Key words: generalized fiducial inference, confidence interval, Markov chain Monte Carlo, item response theory, two-parameter logistic model, exploratory item factor analysis.

1. Introduction

Item response theory (IRT) refers to a collection of latent variable models and statistical methods that has been widely used for describing the underlying structure of survey questionnaires or standardized tests in psychological and educational research. In the current work, we focus on logistic IRT models for dichotomously scored items, e.g., questions with "yes/no" response options in an attitude survey, or multiple-choice questions with a single correct answer in an aptitude test. In particular, the binary response to each item in the test is modeled as a logistic regression on one or more latent variables, each of which represents some latent construct we intend to measure.

Maximum likelihood (ML) has been the gold-standard estimation method for IRT models. The ML estimates can be numerically found by expectation-maximization (EM; e.g., Bock & Aitkin, 1981) or Newton-type (e.g., Bock & Lieberman, 1970; Haberman, 2013) algorithms. The likelihood function of IRT models usually involves an intractable integration over the space of latent variables. Were the dimensionality of latent variables low, simple tensor-product Gaussian/rectangular quadrature suffices to approximate the integral. As the dimensionality increases, however, the naive quadrature representation suffers from the well-known "curse of dimensionality" that the total number of quadrature points grows exponentially fast. Adaptive quadrature (e.g., Schilling & Bock, 2005; Haberman, 2006), which re-scales the quadrature grid for each observed response pattern at each iteration based on the current parameter estimates,

Correspondence should be made to Yang Liu, School of Social Sciences, Humanities and Arts, University of California, Merced, 5200 North Lake Rd, Merced, CA 95343, USA. Email: yliu85@ucmerced.edu

is able to attain the same accuracy using much fewer points, and thus fares more efficient in high-dimensional problems. Alternatively, the integral can be approximated by Markov chain Monte Carlo (MCMC) techniques, which results in stochastic variants of EM- or Newton-type algorithms (e.g., Meng & Schilling, 1996; Cai, 2010a, b) that are also suitable for models with a large number of latent variables.

Bayesian inference (e.g., Albert, 1992; Patz & Junker, 1999; Edwards, 2010) based on sampling from the posterior distribution of model parameters has become popular in recent years, partly because of the enhanced computing power and availability of user-friendly software. Bayesian estimation circumvents the evaluation of the likelihood, and thus remains feasible in models with high-dimensional latent traits. However, it should be used with caution since specifying appropriate prior distributions and tuning the sampling algorithm require extraordinary statistical expertise. Even though the asymptotic optimality of Bayesian posteriors can be guaranteed by the Bernstein–von Mises theorem (e.g., Le Cam & Yang, 2000), erroneous results may be seen in real applications due to improperly chosen prior distributions or ill-behaved samplers.

Confidence intervals (CIs) convey information about the sampling variability, and should always be reported in company with point estimates. The most widely used interval estimator associated with the ML estimation of IRT models is the Wald-type CI, defined as the point estimate plus or minus the standard error multiplied by the proper normal quantile that matches the nominal coverage level. The standard error computation for IRT model parameters was discussed by, e.g., Cai (2008) and Yuan, Cheng, and Patton (2014). Caveats on the use of Wald-type CIs, due to the reliance on a quadratic approximation of the log-likelihood, have been raised in the statistical literature (e.g., Neale & Miller, 1997): For instance, they are not invariant under nonlinear transformations, may cover values beyond the boundary of the parameter space, and may have unsatisfying small-sample behaviors. As pointed out by a referee, CIs obtained by inverting the likelihood ratio or score test may have a better finite-sample performance. Those methods, however, have not yet been available in the IRT literature. Moreover, they require fitting the model multiple times for each parameter, which is computationally intensive, and thus may not be suitable for multidimensional models.

Bayesian inference is more flexible in terms of quantifying the sampling error. For a certain reparameterization of the model, converting accordingly each Monte Carlo sample from the original posterior yields an approximation to the transformed posterior, from which credible intervals can be constructed by taking, for example, the equi-tailed region. In finite samples, however, dissimilar interval estimates may be resulted from different prior configurations, and preferring one solution over others reduces in essence to the subtle question of prior selection.

In summary, the extant likelihood-based and Bayesian inference methods for IRT parameters both have their merits and deficiencies. In this paper, we aim at developing a comprehensive estimation and inference framework that is able to (a) deal with high-dimensional latent traits, (b) facilitate interval estimation for transformations of parameters, and (c) avoid as much subjectivity and ambiguity as possible in application. Generalized fiducial inference (GFI; Hannig, 2009, 2013), a new variant of Fisher's fiducial inference, is believed to achieve most, if not all, the aforementioned desiderata. In this article, we apply the GFI to a family of binary logistic IRT models; in particular, a fiducial distribution of item intercepts and slopes is derived. The resulting fiducial distribution is closely approximated by a Bayesian posterior with a data-dependent prior, which is shown to satisfy a Bernstein–von Mises-type asymptotic normality. A Markov chain Monte Carlo (MCMC) algorithm is proposed to obtain samples from the fiducial distribution, which can be subsequently used for constructing CIs. Using simulated data, we evaluate the comparative performance of the fiducial percentile CI against two types of ML Wald CIs in terms of empirical coverage and length. An real data example is provided in the end illustrating the use of GFI for exploratory item factor analysis.

2. Theory

2.1. Generalized Fiducial Inference

The origin of fiducial inference can be traced to Fisher (1930, 1933, 1935). To redress what he regarded as a "fallacy" of Bayesian inference that uninformative/flat priors are specified when such *a priori* information is indeed absent, Fisher invented a fiducial argument to transfer to the parameter space a prior-free probability distribution, namely, the fiducial distribution, which can be used for inferential purposes in ways that resemble the use of a Bayesian posterior. However, he failed to provide an unambiguous interpretation of the fiducial probability, and some of the claimed properties of the fiducial distribution could not be established (Zabell et al., 1992). As a result, fiducial inference has been considered Fisher's "one great failure" (Zabell et al., 1992), and largely eschewed by mainstream statisticians. Recently, from roots in the theory of structural inference (Fraser, 1968), Dempster–Shafer calculus (e.g., Dempster, 1968, 2008; Shafer, 1976), and generalized confidence intervals (Weerahandi, 1993), the re-formulated generalized fiducial inference (GFI; Hannig, 2009, 2013) was brought back to spotlight. GFI is a completely general framework adaptable to various parametric models, and usually has justified asymptotic frequentist properties under mild regularity conditions.

We illustrate the idea of fiducial inference using a simple example. Consider a normal location model $Y \sim \mathcal{N}(\theta, 1)$ with parameter $\theta \in \mathbb{R}$. When $\theta = \theta_0$ is known, data can be generated by $Y = \theta_0 + U$ in which $U \sim \mathcal{N}(0, 1)$. Conversely, we may want to make inference about θ_0 after observing $Y = y$.[1] Under most circumstances, we are not able to identify the data generating $U = u_0$ satisfying $y = \theta_0 + u_0$; otherwise, θ_0 can be obtained trivially by $\theta_0 = y - u_0$. Despite the fact that the exact recovery of θ_0 via u_0 is not viable, the quantity $y - u$ corresponds to the θ value that is needed to reproduce the observed data y for any fixed u. If we replace the fixed u by an independent and identically distributed (i.i.d.) copy of the data generating U, denoted $U^\star$, then the distribution of $y - U^\star$, referred to as a fiducial distribution of θ, gauges how plausible each θ value may reproduce y. The definition of a fiducial probability does not require any information other than the model and observed data, different from the definition of a Bayesian posterior probability in which prior information is indispensable. Because of its dependence on fixed data, the fiducial probability is also not the confidence probability in the usual frequentist sense.

We now introduce the theory of GFI. For a family of parametric models indexed by some parameter space Θ, the data-generating equation (DGE) expresses the data $\mathbf{Y}$ as a composition of parameters $\boldsymbol{\theta} \in \Theta$ and random components $\mathbf{U}$ with parameter-free distributions:

$$\mathbf{Y} = \mathbf{g}(\boldsymbol{\theta}, \mathbf{U}). \tag{1}$$

As the name suggests, the DGE characterizes how data are generated from the model; in the normal location example, the DGE is $Y = \theta + U$. When the parameters $\boldsymbol{\theta}$ are considered known, data $\mathbf{Y}$ can be obtained using Eq. 1 after sampling the random components $\mathbf{U}$ from their known distributions. For the reversal, i.e., making inference about θ, the data $\mathbf{Y} = \mathbf{y}$ are considered fixed and known, and Eq. 1 is regarded as an implicit function expressing parameters by the data and random components. A distribution on the parameter space is then implicitly determined by Eq. 1, transferred from the known distributions of $\mathbf{U}$. Properly explicating this relationship, in other words, "solving" $\boldsymbol{\theta}$ form the DGE, leads to a fiducial distribution that can be used for making inference about parameters $\boldsymbol{\theta}$. The same role-switching of data and parameters can be found in the duality between the likelihood and density functions, which is fundamental in likelihood-based inference. Here, applying the same idea to the DGE yields a probabilistic quantification regarding

---

[1] In the sequel, lowercase letters are routinely used for realizations of random variables.

which $\boldsymbol{\theta}$ in Θ is the truth, which is more intuitive than the deterministic quantification provided by the likelihood function.

Define the set inverse of the DGE:

$$Q(\mathbf{y}, \mathbf{u}) = \{\boldsymbol{\theta} \in \Theta : \mathbf{g}(\boldsymbol{\theta}, \mathbf{u}) = \mathbf{y}\}, \tag{2}$$

which contains all possible solutions to Eq. 1 given fixed $\mathbf{y}$ and $\mathbf{u}$. In the normal location example, the set inverse is a singleton set $\{y - u\}$; the unique element defines unambiguously a fiducial distribution, i.e., $y - U^\star$. In general, however, Eq. 2 may contain more than one element for some combinations of $\mathbf{y}$ and $\mathbf{u}$, and may be empty for others. More involved arguments are needed to define a fiducial distribution rigorously.

Infinitely many solutions to the DGE may co-exist. An example is the Bernoulli family $Y \sim \text{Bernoulli}(\theta), 0 \le \theta \le 1$. Its DGE can be expressed as $Y = \mathbb{I}\{U \le \theta\}$, in which $\mathbb{I}(\cdot)$ denotes the indicator function, and $U \sim \text{Uniform}(0, 1)$. Given $y \in \{0, 1\}$ and $0 \le u \le 1$ as realizations of Y and U, the corresponding set inverse is one of the two intervals partitioned by the value of u (see Hannig, 2009, Example 6): $u \le \theta \le 1$ if $y = 1$, and $0 \le \theta < u$ if $y = 0$. In fact, models for discrete data often yield non-singleton set inverse $Q(\mathbf{y}, \mathbf{u})$; therein favoring one element over others for the purpose of defining a fiducial distribution for $\boldsymbol{\theta}$ cannot be decided from the values of $\mathbf{y}$ and $\mathbf{u}$ *per se*. Hence, we need a user-defined rule that uniquely identifies an element from the set inverse: e.g., randomly selecting a point from $Q(\mathbf{y}, \mathbf{u})$.

It is also possible that the set determined by Eq. 2 is empty. For example, let $Y_1 = \mathbb{I}\{U_1 \le \theta\}$ and $Y_2 = \mathbb{I}\{U_2 \le \theta\}$ be two observations from Bernoulli(θ). If we observe $y_1 = 1$ and $y_2 = 0$, then the joint set inverse is $[u_1, 1] \cap [0, u_2)$, which is empty whenever $u_1 \ge u_2$. An empty set inverse $Q(\mathbf{y}, \mathbf{u})$ implies that no parameter value is able to recover $\mathbf{y}$ combined with the particular $\mathbf{u}$. Because the model is assumed to be correctly specified, intuitively it means that this $\mathbf{u}$ value is not helpful to the inference of $\boldsymbol{\theta}$ and should be discarded. One natural resolution is to concentrate on the set of $\mathbf{u}$ such that Eq. 2 is non-empty.

Following these heuristics, we define a fiducial distribution as

$$\mathbf{v}(Q(\mathbf{y}, \mathbf{U}^\star)) \mid \{Q(\mathbf{y}, \mathbf{U}^\star) \ne \emptyset\}, \tag{3}$$

in which $\mathbf{U}^\star$ is an i.i.d. copy of the data generating $\mathbf{U}$, and $\mathbf{v}(\cdot)$ denotes a selection rule. A random variable having the distribution determined by Eq. 3 is referred to as a generalized fiducial quantity (GFQ).

Two sources of non-uniqueness are inherent in Eq. 3. First, there may exist different DGEs for the same model, and thus different fiducial distributions (e.g., Hannig, 2013, Example 5.1). In our IRT application, we focus our attention on a specific DGE that has been widely used in practice for generating item response data; future studies are encouraged to investigate other possibilities. Second, when the set inverse consists of more than one point, different selection rules lead to different fiducial distributions. Hannig (2013) proved for a general class of models without random effects that the diameter of the set inverse converges to zero at a fast rate (of order $1/n$), which implies that the impact of selection rules is asymptotically negligible. Simulation studies suggest that the sizes of the polytopes produced by the proposed sampler (introduced in Sect. 2.4) are always tiny when the sample size is large. Thus, we conjecture that a higher-order convergence result similar to Hannig's result holds in our case as well. Theorem 2 presented in the current work is a initial step towards this direction, in which we establish that the size of the set inverse under the empirical Bayesian approximation to the fiducial distribution is of order $1/n$ for unidimensional models ($r = 1$).

2.2. GFI for Binary Logistic IRT Models

Next, we consider a family of binary logistic IRT models, and derive a GFQ for item parameters. Let a person i's response to a binary item j, $Y_{ij} = y_{ij} \in \{0, 1\}$, be modeled by the following conditional likelihood (also known as the item response function):

$$f_j(\boldsymbol{\theta}_j, y_{ij}|\mathbf{z}_i) = P\{Y_{ij} = y_{ij}|\mathbf{Z}_i = \mathbf{z}_i\} = \frac{1}{1 + e^{(-1)^{y_{ij}}\tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i)}}, \quad \tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i) = \alpha_j + \boldsymbol{\beta}_j^{\top}\mathbf{z}_i, \quad (4)$$

in which $\mathbf{Z}_i = (Z_{id})_{d=1}^r \in \mathbb{R}^r$ are the latent variables. In Eq. 4, α_j denotes the item intercept, and $\boldsymbol{\beta}_j$ denotes the r item slopes. We assume that the intercept is always freely estimated, but some slopes must be fixed for model identification. We denote all q_j free parameters that calibrate item j by $\boldsymbol{\theta}_j$, and write $\tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i)$ as the usual linear regression on the latent variable to highlight its dependence on $\boldsymbol{\theta}_j$. In addition, we restrict consideration to the case that $\mathbf{Z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_r)$, in which $\mathbf{I}_r$ is an r-dimensional identity matrix; that is, no correlation component is estimated among the latent variables. This general setup encompasses the two-parameter logistic (2PL), bifactor, and exploratory item factor analysis models, but not the independent-cluster or the general two-tier models; future research works are encouraged to extend the current framework to a broader class of IRT models.

Y_{ij} is a Bernoulli random variable with success probability given by $f_j(\boldsymbol{\theta}_j, 1|\mathbf{z}_i)$. The DGE of Y_{ij} has the following form:

$$Y_{ij} = \mathbb{I}\left\{U_{ij} \leq f_j(\boldsymbol{\theta}_j, 1|\mathbf{z}_i)\right\} = \mathbb{I}\{\text{logit}(U_{ij}) \leq \tau_j(\boldsymbol{\theta}_j, \mathbf{Z}_i)\} = \mathbb{I}\left\{A_{ij} \leq \tau_j(\boldsymbol{\theta}_j, \mathbf{Z}_i)\right\}, \quad (5)$$

in which $U_{ij} \sim \text{Uniform}(0, 1)$ independent of $\mathbf{Z}_i$, and $A_{ij} = \text{logit}(U_{ij}) \sim \text{Logistic}(0, 1)$. In Eq. 5, the free components of α_j and $\boldsymbol{\beta}_j$, i.e., $\boldsymbol{\theta}_j$, can be identified as parameters $\boldsymbol{\theta}$ in Eq. 1; A_{ij} and $\mathbf{Z}_i$ are the random components with parameter-free distributions; they jointly correspond to $\mathbf{U}$ in Eq. 1. The set inverse of Eq. 5 becomes

$$Q_{ij}(y_{ij}, a_{ij}, \mathbf{z}_i) = \{\boldsymbol{\theta}_j \in \mathbb{R}^{q_j} : a_{ij} \leq \tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i), \text{ if } y_{ij} = 1;$$
$$a_{ij} > \tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i), \text{ if } y_{ij} = 0\}. \quad (6)$$

The geometric representation of Eq. 6 is a half-space, i.e., one half of the Euclidean space $\mathbb{R}^{q_j}$ with the partitioning determined by the affine hyperplane $a_{ij} = \tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i)$; a graphical illustration using a 2PL item is shown in the left panel of Figure 1.

Now consider an $n \times m$ binary response data matrix, denoted $\mathbf{Y} = (Y_{ij})_{i=1}^n {}_{j=1}^m$, in which n is the sample size, m is the test length, and each Y_{ij} is generated from a version of Eq. 5. It is assumed that the n individual response patterns, denoted $\mathbf{Y}_i = (Y_{ij})_{j=1}^m$, $i = 1, \ldots, n$, are i.i.d, and that for each observation i, Y_{ij}, $j = 1, \ldots, m$, are independent conditional on $\mathbf{Z}_i$. This implies the independence of the corresponding logistic and normal variates, denoted $\mathbf{A} = (A_{ij})_{i=1}^n {}_{j=1}^m$ and $\mathbf{Z} = (\mathbf{Z}_i)_{i=1}^n$, respectively. The set inverse for the DGE of $\mathbf{Y}$ can be written as

$$Q(\mathbf{y}, \mathbf{a}, \mathbf{z}) = \underset{j=1}{\overset{m}{\times}} \bigcap_{i=1}^n Q_{ij}(y_{ij}, a_{ij}, \mathbf{z}_i), \quad (7)$$

in which we write $\times$ for the Cartesian product. For each j, we take the intersection for the reason that the set inverse by definition should include $\boldsymbol{\theta}_j$ values that are consistent with all individual
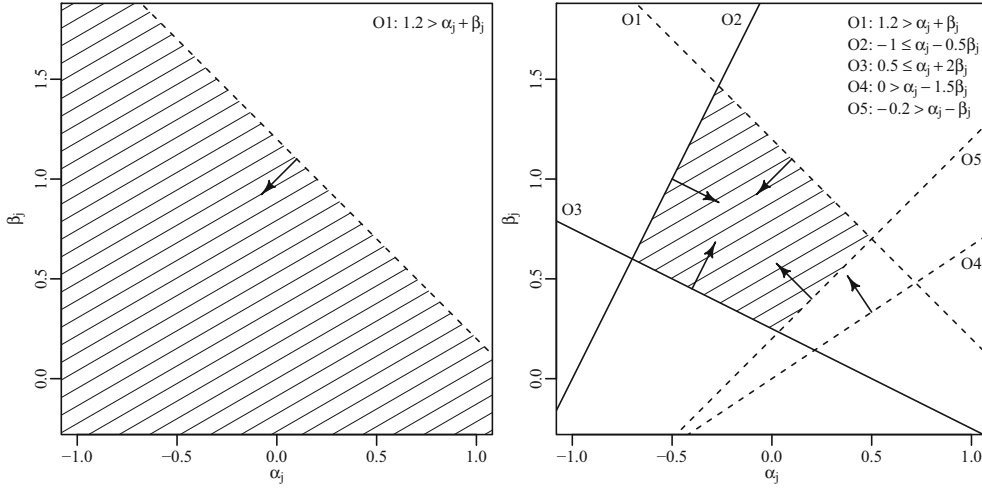
FIGURE 1.

In both panels, the *horizontal axis* is intercept α_j, and the *vertical axis* is slope β_j. Each *straight line* gives the boundary condition in Eq. 6 for a particular observation: *Dashed lines* indicate strict inequalities, while *solid lines* indicate non-strict ones. The *arrow* perpendicular to each boundary points into the half-space. The *left plot* presents the set inverse function (the *shaded area*) for a single observation (Observation 1, abbreviated as O1). In the *right plot*, four additional observations (O2–O5) are included; the intersection of the five half-spaces is shown as the *shaded area*.

DGEs (Eq. 5). For easy reference, we introduce the notation $\mathbf{Y}_{(j)} = (Y_{ij})_{i=1}^{n}$ for all n responses to item j, and similarly $\mathbf{A}_{(j)} = (Y_{ij})_{i=1}^{n}$ for the corresponding logistic variates. Also, let

$$Q_j(\mathbf{y}_{(j)}, \mathbf{a}_{(j)}, \mathbf{z}) = \bigcap_{i=1}^{n} Q_{ij}(y_{ij}, a_{ij}, \mathbf{z}_i). \tag{8}$$

Geometrically, $Q_j(\mathbf{y}_{(j)}, \mathbf{a}_{(j)}, \mathbf{z})$ is an $\mathbb{R}^{q_j}$-polyhedron, whose faces are the boundaries of a selective collection of individual half-spaces (Eq. 5). For a single 2PL item, an illustration of Eq. 8 as an $\mathbb{R}^2$-polygon is given in the right panel of Figure 1. Because we assume that different items do not share parameters, the overall set inverse, i.e., Eq. 7, is finally obtained by taking the Cartesian product. In the sequel, we treat without loss of generality the set inverses (Eqs. 6–8) as the closure[2] of what we defined earlier; since the random components are continuous, all the stated properties that hold for the closure also apply to the interior with probability one.

Denote by $\mathbf{A}^{\star}$ and $\mathbf{Z}^{\star}$ i.i.d. copies of the random components $\mathbf{A}$ and $\mathbf{Z}$, respectively. A GFQ for item parameters can be constructed from the random set $Q(\mathbf{y}, \mathbf{A}^{\star}, \mathbf{Z}^{\star})$, provided the set is not empty. We remark that polyhedrons constituting the set inverse are unbounded with a positive probability for fixed n and $\mathbf{y}$: For example, when $n \leq q_j$, a non-empty polyhedron $Q_j(\mathbf{y}_{(j)}, \mathbf{a}_{(j)}, \mathbf{z})$ is certainly unbounded, because a bounded $\mathbb{R}^{q_j}$-polytope has at least $q_j + 1$ faces. For ease of exposition, we restrict ourselves to selection rules $\mathbf{v}(\cdot)$ returning finite values within the set inverse. Following the generic recipe (Eq. 3), a GFQ of item parameters is

$$\mathbf{v}(Q(\mathbf{y}, \mathbf{A}^{\star}, \mathbf{Z}^{\star})) \mid \left\{ Q(\mathbf{y}, \mathbf{A}^{\star}, \mathbf{Z}^{\star}) \neq \emptyset \right\}. \tag{9}$$

---

[2] For the set inverse function considered here, the closure amounts to the same polyhedrons with all the boundaries attained.

In the simulation study and the empirical example discussed later, we consider a selection rule that randomly (with equal probability) selects for each item an interior vertex of the corresponding polytope, which parallels Hannig's (2009) recommendation of non-informative and data-independent selection rules.

2.3. A Bernstein–von Mises Theorem

In Bayesian inference, the Bernstein–von Mises theorem describes the well-known phenomenon that a posterior distribution converges to a normal limit as the sample size increases; it is sometimes referred to as "the Bayesian central limit theorem" in fundamental text. To illustrate, we consider a one-parameter model with parameter θ; denote by θ_0 the true parameter that generates the observed data $\mathbf{y}$. Let $R(\mathbf{y})$ be a random variable that follows the posterior distribution of θ given observed data $\mathbf{y}$. The Bernstein–von Mises theorem implies that the distribution of $R(\mathbf{Y})$ approaches $\mathcal{N}(X, \sigma_0^2)$, in which $X \sim \mathcal{N}(\theta_0, \sigma_0^2)$, and σ_0^2 is the reciprocal of the sample Fisher information evaluated at θ_0. As a result, in large samples, a Bayesian credible interval has approximately the correct frequentist coverage. For instance, consider the one-sided credible interval $(-\infty, r_\alpha(\mathbf{y})]$ in which $r_\alpha(\mathbf{y})$ is the upper α quantile of $R(\mathbf{y})$. The normal approximation suggests $r_\alpha(\mathbf{Y}) \approx X + z_\alpha \sigma_0$, in which z_α is the upper α quantile of the standard normal distribution, so $P\{\theta_0 \leq r_\alpha(\mathbf{Y})\} \approx P\{\theta_0 \leq X + z_\alpha \sigma_0\} = 1 - \alpha$.

In this section, we establish a Bernstein–von Mises theorem for a posterior distribution derived from a data-dependent prior, which amounts to approximating the conditioning set involved in the GFQ (Eq. 9) by a first-order inclusion-exclusion expansion. Some notation is introduced first. Suppose that i.i.d. item response data $\mathbf{Y} = (\mathbf{Y}_i)_{i=1}^n$ are generated from the same logistic IRT family as described in the previous section. Each observation $\mathbf{Y}_i$ is a multinomial random variable with the following probability mass/likelihood function:

$$f(\boldsymbol{\theta}, \mathbf{y}_i) = \int_{\mathbb{R}^r} \prod_{j=1}^m f_j(\boldsymbol{\theta}_j, y_{ij}|\mathbf{z}_i) d\Phi(\mathbf{z}_i). \tag{10}$$

Let $\mathbf{s}(\boldsymbol{\theta}, \mathbf{y}_i) = \partial \log f(\boldsymbol{\theta}, \mathbf{y}_i)/\partial \boldsymbol{\theta}$ be the single-observation score vector, and $\mathbf{H}(\boldsymbol{\theta}, \mathbf{y}_i) = \partial^2 \log f(\boldsymbol{\theta}, \mathbf{y}_i)/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$ be the single-observation Hessian matrix. Also define $\mathcal{I}(\boldsymbol{\theta}) = \text{Cov}_{\boldsymbol{\theta}}\,[\mathbf{s}(\boldsymbol{\theta}, \mathbf{Y}_i)]$ which is usually referred to as the Fisher information matrix. It can be verified by direct calculation that

$$E_{\boldsymbol{\theta}}\,[\mathbf{s}(\boldsymbol{\theta}, \mathbf{Y}_i)] = \mathbf{0}, \tag{11}$$

and

$$\mathcal{I}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[\mathbf{s}(\boldsymbol{\theta}, \mathbf{Y}_i)\mathbf{s}(\boldsymbol{\theta}, \mathbf{Y}_i)^\top\right] = -E_{\boldsymbol{\theta}}\,[\mathbf{H}(\boldsymbol{\theta}, \mathbf{Y}_i)]. \tag{12}$$

Let $\boldsymbol{\theta}_0$ be the true parameter value that generates $\mathbf{Y}$, and $\mathcal{I}_0$ be a short-hand notation for $\mathcal{I}(\boldsymbol{\theta}_0)$. Also define the (scaled) sample score function $\mathbf{S}_n = n^{-1/2} \sum_{i=1}^n \mathbf{s}(\boldsymbol{\theta}_0, \mathbf{Y}_i)$. By Eqs. 11 and 12, and the Central Limit Theorem,

$$\mathbf{S}_n \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathcal{I}_0). \tag{13}$$

It follows that $\mathcal{I}_0^{-1}\mathbf{S}_n \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathcal{I}_0^{-1})$.

Let $I = (I_j)_{j=1}^m$ be an m-tuple of index sets, in which each I_j indexes a size-q_j sub-sample, i.e., $I_j \subset \{1, \ldots, n\}$ and $|I_j| = q_j$. For each item j, the linear system $A_{ij}^\star = \tau_j(\boldsymbol{\theta}_j, \mathbf{Z}_i^\star), i \in I_j$, has a unique solution with probability one, denoted $\mathbf{V}_{I_j}$, which can potentially be an interior vertex of the random polytope $Q_j(\mathbf{y}_{(j)}, \mathbf{A}_{(j)}^\star, \mathbf{Z}^\star)$. Pooling across all items, I determines a potential extremal point of $Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star)$, denoted $\mathbf{V}_I = (\mathbf{V}_{I_j})_{j=1}^m$; there are in total $C_n = \prod_{j=1}^m \binom{n}{q_j}$ different choices of I. Let D_I be the event that I determines an extremal point of the non-empty

set inverse; the event $Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) \neq \emptyset$ used for conditioning in Eq. 9 is then equivalent to $\bigcup_I D_I$.[3] Conditioning on the union of multiple D_I's is not easy to manipulate, and thus the following approximation is resorted to. Define event $D(\mathbf{y})$ by the law of total probability that each sub-sample I has probability C_n^{-1} to be selected, and that $Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star)$ is non-empty given that the selected I forms an extremal point. It is clear that $P\{D(\mathbf{y})\} \propto \sum_I P\{D_I\}$. Let $\mathbf{R}(\mathbf{y})$ be a random variable that follows the distribution of the selected $\mathbf{V}_I$ conditional on $D(\mathbf{y})$. $\mathbf{R}(\mathbf{y})$ differs from the GFQ (Eq. 9) only in the conditioning event: $D(\mathbf{y})$ used in the approximation can be considered as a first-order approximation to $\cup_I D_I$ in the inclusion-exclusion formula:

$$P\{\bigcup_I D_I\} = \sum_I P\{D_I\} - \sum_{I \neq I'} P\{D_I \cap D_{I'}\} + \sum_{I \neq I' \neq I''} P\{D_I \cap D_{I'} \cap D_{I''}\} - \cdots . \quad (14)$$

The construction of the equi-probability mixture distribution is inspired by Hannig's (2009, Sect. 4.1) suggested implementation of the fiducial recipe for continuous data. We conjecture that the higher-order terms on the right-hand side of Eq. 14 do not affect the conditional distribution as the sample size grows, but leave the theoretical justification for future research.

A roadmap for our theoretical justification is summarized as follows. We first establish that the density of $\mathbf{R}(\mathbf{y})$ has a closed-form expression (Lemma 1) and satisfies the desirable result (Theorem 1). Next, it is proved for unidimensional models ($r = 1$) that the diameter of the set inverse goes to 0 at the rate $1/n$ (Theorem 2), faster than the rate $1/\sqrt{n}$ at which the distribution of $\mathbf{R}(\mathbf{Y})$ approaches its normal limit. This provides partial support for the observation that different selection rules tend to give converged inference about model parameters when the sample size is large enough.

The following lemma gives explicitly the density of $\mathbf{R}(\mathbf{y})$, which amounts to a posterior density defined by a data-dependent prior; detailed derivations can be found in Appendix 1.

Lemma 1. (Density) *Consider a test of m dichotomous items each of which is characterized by a version of Eq. 4. Let $\Theta \subset \mathbb{R}^q$, $q = \sum_{j=1}^m q_j$, be the parameter space comprising all free intercepts and slopes $\boldsymbol{\theta} = (\boldsymbol{\theta}_j)_{j=1}^m$. For ease of exposition, the fixed slopes are set to zero.[4] Given observed response data $\mathbf{y} = (\mathbf{y}_i)_{i=1}^n = (y_{ij})_{i=1}^n {}_{j=1}^m$, the density of $\mathbf{R}(\mathbf{y})$ can be written as*

$$g_n(\boldsymbol{\theta}|\mathbf{y}) \propto \sum_I \int_{\mathbb{R}^{nr}} d_I(\boldsymbol{\theta}, \mathbf{z}_I) \prod_{j=1}^m \left\{ \prod_{i \in I_j} \frac{e^{\tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i)}}{[1 + e^{\tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i)}]^2} \prod_{i \in I_j^c} f_j(\boldsymbol{\theta}_j, y_{ij}|\mathbf{z}_i) \right\} d\Phi(\mathbf{z}). \quad (15)$$

In Eq. 15, Φ denotes the probability measure of $\mathcal{N}(\mathbf{0}, \mathbf{I}_{nr})$,[5] and $d_I(\boldsymbol{\theta}, \mathbf{z}_I) = \prod_{j=1}^m \left| \det(\partial \tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i) / \partial \boldsymbol{\theta}_j)_{i \in I_j} \right|$ gives a Jacobian determinant, in which $\mathbf{z}_I = (\mathbf{z}_i)_{i \in I}$.[6]

Remark 1. The connection to Bayesian inference can be seen from Eq. 15. Rewrite Eq. 15 by splitting the integral into two parts—one for $\mathbf{z}_I$, and the other for $\mathbf{z}_{I^c}$:

---

[3] Both V_I and D_I depend on the observed data $\mathbf{y}$; the dependency is omitted from the expressions for conciseness.

[4] In practice, slopes might be fixed at values other than zero. The theoretical properties discussed in the current work still apply after subtracting the inner product of those fixed slopes and the corresponding normal variates from A_{ij}'s and substituting its distribution function for the standard logistic density and distribution functions.

[5] For ease of notation, we use Φ to denote the probability measure corresponding to a standard normal distribution of arbitrary dimensionality. By default, the dimensionality is determined by the quantity in the parenthesis that follows.

[6] $i \in I$ means $i \in I_j$ for some j, with a slight abuse of notation. As a general notation, we put index set in subscript to denote the corresponding elements.

$$g_n(\boldsymbol{\theta}|\mathbf{y}) \propto \sum_I \int d_I(\boldsymbol{\theta}, \mathbf{z}_i) \prod_{i \in I} \left\{ \prod_{j \in J_i} \frac{e^{\tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i)}}{\left[1 + e^{\tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i)}\right]^2} \prod_{j \notin J_i} f_j(\boldsymbol{\theta}_j, y_{ij}|\mathbf{z}_i) \right\} d\Phi(\mathbf{z}_I)$$
$$\cdot \int \prod_{i \in I^c} \prod_{j=1}^m f_j(\boldsymbol{\theta}_j, y_{ij}|\mathbf{z}_i) d\Phi(\mathbf{z}_{I^c}), \tag{16}$$

in which $J_i = \{j : i \in I_j\}$ for $i \in I$. Note that the second line of Eq. 16 is the marginal likelihood function of the observations in I^c. We can multiply and divide the right-hand side of Eq. 16 by the likelihood of the vertex-determining observations I, and then simplify it to

$$g_n(\boldsymbol{\theta}|\mathbf{y}) \propto b_n(\boldsymbol{\theta}, \mathbf{y}) f_n(\boldsymbol{\theta}, \mathbf{y}). \tag{17}$$

In Eq. 17,

$$f_n(\boldsymbol{\theta}, \mathbf{y}) = \int \prod_{i=1}^n \prod_{j=1}^m f_j(\boldsymbol{\theta}_j, y_{ij}|\mathbf{z}_i) d\Phi(\mathbf{z}) \tag{18}$$

denotes the complete sample likelihood, and

$$b_n(\boldsymbol{\theta}, \mathbf{y}) = \sum_I \int d_I(\boldsymbol{\theta}, \mathbf{z}_I) \prod_{i \in I} \left\{ \prod_{j \in J_i} \frac{e^{\tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i)}}{\left[1 + e^{\tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i)}\right]^2} \prod_{j \notin J_i} f_j(\boldsymbol{\theta}_j, y_{ij}|\mathbf{z}_i) \right\} d\Phi(\mathbf{z}_I)$$
$$\Big/ \int \prod_{i \in I} \prod_{j=1}^m f_j(\boldsymbol{\theta}_j, y_{ij}|\mathbf{z}_i) d\Phi(\mathbf{z}_I) \tag{19}$$

is a function of both the item parameters and data. Therefore, the density of $\mathbf{R}(\mathbf{y})$ can be conceived as the (empirical) Bayesian posterior computed from the data-dependent prior proportional to Eq. 19.

It can be straightforwardly shown that the density expressed by Eq. 15 (or equivalently Eq. 17) satisfies a Bernstein–von Mises theorem. The proof is relegated to Appendix 2, which is similar to Ghosh and Ramamoorthi's (2003, Theorem 1.4.2) proof of a Bayesian Bernstein–von Mises theorem.

Theorem 1. (Bernstein–von Mises) *Suppose that item response data $\mathbf{Y} = (\mathbf{Y}_i)_{i=1}^n$ are i.i.d. with probability mass function $f(\boldsymbol{\theta}_0, \mathbf{y}_i)$. Let $\Theta \subset \mathbb{R}^q$ be the parameter space as usual. Assume that*

(i) *$m \geq r + 1$;*
(ii) *For all $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$ such that $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$, $f_{\boldsymbol{\theta}} \neq f_{\boldsymbol{\theta}'}$ for some response pattern;*
(iii) *$\boldsymbol{\theta}_0$ is at the interior of Θ;*
(iv) *The Fisher information matrix $\mathcal{I}_0$ is positive definite.*

Let $\bar{g}_n(\mathbf{h}|\mathbf{y}) = g_n(\boldsymbol{\theta}_0 + \mathbf{h}/\sqrt{n}|\mathbf{y})/\sqrt{n}$ be the density of $\sqrt{n}[\mathbf{R}(\mathbf{y}) - \boldsymbol{\theta}_0]$, H_n be the correspondingly rescaled parameter space, and $\phi_{\mathcal{I}_0^{-1}\mathbf{S}_n, \mathcal{I}_0^{-1}}$ be the density of $\mathcal{N}(\mathcal{I}_0^{-1}\mathbf{S}_n, \mathcal{I}_0^{-1})$. Then,

$$\int_{H_n} \left| \bar{g}_n(\mathbf{h}|\mathbf{Y}) - \phi_{\mathcal{I}_0^{-1}\mathbf{S}_n, \mathcal{I}_0^{-1}}(\mathbf{h}) \right| d\mathbf{h} \xrightarrow{P_{\boldsymbol{\theta}_0}} 0, \tag{20}$$

in which $P_{\boldsymbol{\theta}_0}$ denotes the probability measure of $\mathbf{Y}$ under the true parameter values $\boldsymbol{\theta}_0$, and $\xrightarrow{P_{\boldsymbol{\theta}_0}}$ means converges in probability under the true model.

Remark 2. Assumptions (ii) to (iv) are standard regularity conditions for establishing the asymptotic optimality of the ML estimator. (i) guarantees the existence of some neighborhood of $\boldsymbol{\theta}_0$ such that for $\boldsymbol{\theta}$ outside the likelihood ratio $f_n(\boldsymbol{\theta}, \mathbf{Y})/f_n(\boldsymbol{\theta}_0, \mathbf{Y})$ converges uniformly to zero in probability, which functions similarly to Assumption (v) in Ghosh and Ramamoorthi (2003).

Remark 3. As remarked in van der Vaart (2000, Sect. 10.2), the alternative "centering sequence" $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, in which $\hat{\boldsymbol{\theta}}$ is the ML estimator, can be used in place of $\mathcal{I}_0^{-1}\mathbf{S}_n$ in Eq. 20, because the the latter is a local linear approximation of the former at the true parameter values $\boldsymbol{\theta}_0$ and the two are asymptotically equivalent.

When $r = 1$, we are able to control the diameter of the set inverse by an $O_p(n^{-1})$ term (Theorem 2). Because the rate of convergence in Theorem 1 is of order $1/\sqrt{n}$, the same convergence result also holds for all other points selected from the set inverse given that each sub-sample I serves equally likely as the selected extremal point. The proof is provided in Appendix 3.

Theorem 2. *Suppose that Assumptions (i)–(iv) of Theorem 1 hold. Consider $r = 1$. For any $K > 0$, define*

$$\rho_K(\mathbf{y}) = P\{diam Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) > K/n \mid D(\mathbf{y})\}, \tag{21}$$

Then, for each $\varepsilon > 0$,

$$P_{\boldsymbol{\theta}_0}\{\exists K, N > 0: \ \rho_K(\mathbf{Y}) < \varepsilon, \ \forall n > N\} \to 1. \tag{22}$$

Remark 4. A majority of the proof (Appendix 3) is extensible to multidimensional models (i.e., $r > 1$), except for the last part that involves a case enumeration.

2.4. A Markov Chain Monte Carlo Algorithm

Next, we introduce an MCMC algorithm to sample from the fiducial distribution (Eq. 9). Our main task is to sample $\mathbf{A}^\star$ and $\mathbf{Z}^\star$ such that the set inverse $Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star)$ is non-empty. We solve this high-dimensional truncated sampling problem by a Gibbs sampler, which consists of two types of conditional sampling steps, one for $A_{ij}^\star$ and the other for $Z_{id}^\star$. After initialization, our algorithm sequentially draws each random component from its conditional distribution given the latest values of the rest. By the standard theory for Gibbs samplers, the generated Markov Chain converges to the joint distribution of $\mathbf{A}^\star$ and $\mathbf{Z}^\star$ conditional on $Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) \neq \emptyset$. Following the update of each random component, the interior polyhedrons are rebuilt accordingly. After each MCMC cycle, one extremal point of the set inverse is selected and recorded as an instance of the GFQ. Next, we discuss the two Gibbs sampling steps, the choice of starting values, and some tuning details of the algorithm.

Conditional sampling of $A_{ij}^\star$. Fix i and j. The goal of this step is to obtain an update of $A_{ij}^\star$ such that the resulting new half-space has a non-empty intersection with the interior polyhedron determined by all current realizations of the random components except for those of the ith observation. Notationally, we use superscript 0 to highlight the dependency solely on the current values of the random components, and superscript 1 the involvement of the updated one. Let $\mathbf{y}_{-i(j)} = (y_{kj})_{k \neq i}$, $\mathbf{a}_{-i(j)}^0 = (a_{kj}^0)_{k \neq i}$, and $\mathbf{z}_{-i}^0 = (\mathbf{z}_k^0)_{k \neq i}$. Any valid update of $A_{ij}^\star$, denoted a_{ij}^1, should satisfy the following condition:

$$Q_{ij}(y_{ij}, a_{ij}^1, \mathbf{z}_i^0) \cap Q_j(\mathbf{y}_{-i(j)}, \mathbf{a}_{-i(j)}^0, \mathbf{z}_{-i}^0) \neq \emptyset \Leftrightarrow \begin{cases} a_{ij}^1 \leq \max_{\boldsymbol{\theta}_j \in \mathcal{V}_{-ij}^0} \tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i^0), & \text{if } y_{ij} = 1; \\ a_{ij}^1 \geq \min_{\boldsymbol{\theta}_j \in \mathcal{V}_{-ij}^0} \tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i^0), & \text{if } y_{ij} = 0. \end{cases} \tag{23}$$
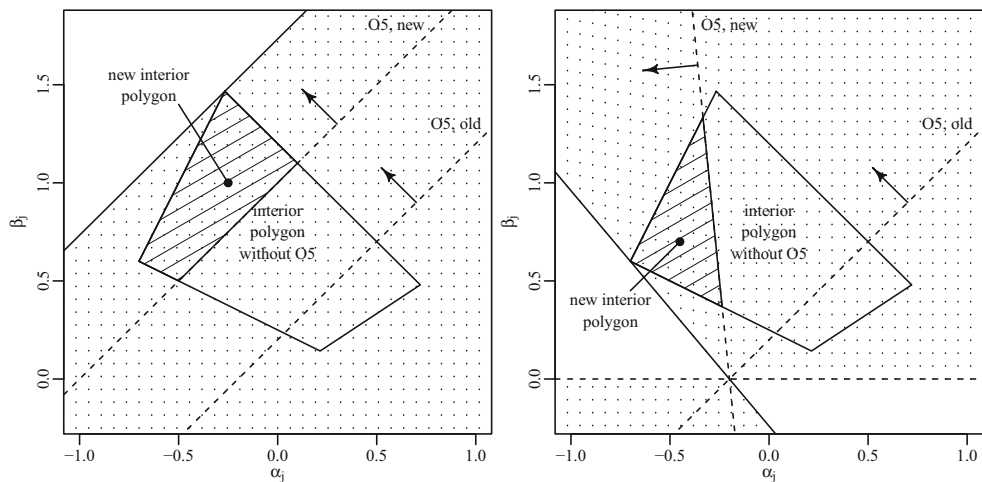
FIGURE 2.

The *left panel* illustrates updating $A_{5j}^{\star}$ for Observation 5 (O5) given values of other random components. The updated half-space (O5, new) is parallel to the old O5. The *dot-filled area* is the feasible region for the updated half-space to have non-empty intersection with the interior polygon without old O5. The *shaded area* shows the updated *polygon*. Similarly, the *right panel* illustrates the conditional sampling of $Z_5^{\star}$. This time both the old and new O5 pass through the same point on $\beta_j = 0$.

in which $\mathcal{V}_{-ij}^0$ denotes the collection of interior vertices of $Q_j(\mathbf{y}_{-i(j)}, \mathbf{a}_{-i(j)}^0, \mathbf{z}_{-i}^0)$. Equation 23 follows from the fact that the left-hand side intersection, i.e., the updated interior polyhedron for item j, is non-empty if and only if at least one point in $Q_j(\mathbf{y}_{-i(j)}, \mathbf{a}_{-i(j)}^0, \mathbf{z}_{-i}^0)$ satisfies the inequality posed by $Q_{ij}(y_{ij}, a_{ij}^1, \mathbf{z}_i^0)$; due to convexity, it suffices to require at least one vertex of the polyhedron $Q_j(\mathbf{y}_{-i(j)}, \mathbf{a}_{-i(j)}^0, \mathbf{z}_{-i}^0)$ satisfying the inequality. Therefore, we sample $A_{ij}^{\star} = a_{ij}^1$ from Logistic(0, 1) truncated from above by $\max_{\boldsymbol{\theta}_j \in \mathcal{V}_{-ij}^0} \tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i^0)$ when $y_{ij} = 1$, and from below by $\min_{\boldsymbol{\theta}_j \in \mathcal{V}_{-ij}^0} \tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i^0)$ when $y_{ij} = 0$. A graphical illustration of Step 1 using a 2PL item can be found in the left panel of Figure 2.

Conditional sampling of $Z_{id}^{\star}$. Fix i and d. The goal of this step is to sample $Z_{id}^{\star}$ from a suitably truncated standard normal distribution ensuring for all items that the updated interior polyhedrons are not empty. Let $\mathbf{z}_i^d = (z_{i1}^0 \cdots z_{i,d-1}^0 z_{id}^1 z_{i,d+1}^0 \cdots z_{ir}^0)^\top$. For each item j, the updated z_{id}^1 should satisfy:

$$Q_{ij}(y_{ij}, a_{ij}^0, \mathbf{z}_i^d) \cap Q_j(\mathbf{y}_{-i(j)}, \mathbf{a}_{-i(j)}^0, \mathbf{z}_{-i}^0) \neq \emptyset \Leftrightarrow \begin{cases} a_{ij}^0 \leq \max_{\boldsymbol{\theta}_j \in \mathcal{V}_{-ij}^0} \tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i^d), \text{ if } y_{ij} = 1; \\ a_{ij}^0 \geq \min_{\boldsymbol{\theta}_j \in \mathcal{V}_{-ij}^0} \tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i^d), \text{ if } y_{ij} = 0. \end{cases}$$
(24)

Pooling across all items, we express the desired truncation of this sampling step as

$$z_{id}^1 \in \bigcap_{j=1}^{m} \bigcup_{\boldsymbol{\theta}_j \in \mathcal{V}_{-ij}^0} \{a_{ij}^0 \leq \tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i^d), \text{ if } y_{ij} = 1;$$
$$a_{ij}^0 \geq \tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i^d), \text{ if } y_{ij} = 0.\}$$
(25)

The geometric object implied by Eq. 25 can be a finite interval, an infinite interval, or a disjoint union of intervals. An example using a single 2PL item can be found in the right panel of Figure 2.

Starting values. A non-empty set $Q(\mathbf{y}, \mathbf{a}^0, \mathbf{z}^0)$ is required to initialize our Gibbs sampler, which can be constructed from some suitable starting values of the parameters and random components. Suppose that some initial guess of the parameter values $\boldsymbol{\theta}^0 = (\theta_j^0)_{j=1}^m$ and the latent variable values $\mathbf{z}^0$ are available. For each i and j, we could execute the Gibbs sampling step of $A_{ij}^\star$ to obtain starting values a_{ij}^0 assuming that the interior polytope has only one vertex $\boldsymbol{\theta}_j^0$; that is, we sample $A_{ij}^\star$ from Logistic(0, 1) truncated from above by $\tau_j(\boldsymbol{\theta}_j^0, \mathbf{z}^0)$ if $y_{ij} = 1$, and truncated from below by the same quantity if $y_{ij} = 0$. It is clear that the resulting set inverse function is non-empty, because it contains at least some neighborhood of $\boldsymbol{\theta}^0$.

In practice, conveniently computable parameter estimates, such as various weighted least square methods based on tetrachoric correlations (e.g., Muthén, 1978; Gunsjö, 1994), can be used as $\boldsymbol{\theta}^0$; alternatively, one could use naive starting values such as 0 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 simply from a standard normal distribution. From our experience, the generated Markov chain often appears stationary after several thousand iterations, and the final results are not affected by the choice of starting points.

Additional tuning of the sampler. To simplify the sampling algorithm, we restrict the item parameters to a compact set $[-M, M]^q$ with $M > 0$ being some pre-specified large number. Hence, the set inverse $Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star)$ always comprises closed polytopes which can be efficiently represented by their vertices. The results are not significantly affected by the choice of M provided the sample size is large enough, in which case the generated polyhedrons usually have small diameters and thus are unlikely to attain the arbitrary bounding box.

In small samples, however, polytopes attaining the bounding box emerge every now and then, resulting from unbounded polyhedrons hard-truncated to the arbitrary bound. Consequently, the marginal fiducial distribution for the associated item parameters can be heavy-tailed; it leaves visible "spikes" on the trace plots, and yields less efficient interval estimators. To resolve this, we propose an extra tuning operation based on the observation that unbounded polyhedrons are typically resulted from lacking lower/upper bounds for the slope parameters. For fixed item j and dimension d on which item j loads, the single-entry set inverse (Eq. 6) imposes an upper bound for the corresponding slope parameter if $y_{ij} = 1$ and $Z_{id}^\star < 0$, or $y_{ij} = 0$ and $Z_{id}^\star > 0$; a lower bound is imposed otherwise. Some combinations of y_{ij} and $Z_{id}^\star$ rarely occur under certain data-generating models,[7] which may lead to a shortage, if not a sheer absence, of bounds on one side. A natural workaround is to modify the set inverse $Q_{ij}(y_{ij}, a_{ij}, \mathbf{z}_i)$ to give each slope both lower and upper bounds; in particular, we define

$$Q_{ij}^M(y_{ij}, a_{ij}, \mathbf{z}_i) = \{\boldsymbol{\theta}_j \in \mathbb{R}^{q_j} : -M + \boldsymbol{\beta}_j^\top \mathbf{z}_i \le a_{ij} \le \alpha_j + \boldsymbol{\beta}_j^\top \mathbf{z}_i, \text{ if } y_{ij} = 1;$$
$$\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{z}_i < a_{ij} \le M + \boldsymbol{\beta}_j^\top \mathbf{z}_i, \text{ if } y_{ij} = 0\}, \qquad (26)$$

which, for fixed a_{ij} and $\mathbf{z}_i$, approaches Eq. 6 as M increases. Pilot studies suggest that replacing Eq. 6 by Eq. 26 with parameter bound $M = 20$ in the construction of a fiducial distribution significantly relieves the problems caused by the heavy-tailedness. In practice, we do not expect item parameters to go beyond this value as well.

---

[7] For example, if a 2PL item is moderately difficult but highly discriminating, then observing either a correct response with a negative $Z_i^\star$ or an incorrect response with a positive $Z_i^\star$ is unlikely; therefore, the generated set inverse functions may not have an upper bound for the corresponding slope parameter.

TABLE 1.
True item parameter values in the Monte Carlo simulation.

Item	Communality	Skewness	Loading λ_j	Threshold τ_j	Slope β_j	Intercept α_j	Difficulty δ_j
1	Low	No	0.32	0.00	0.57	0.00	0.00
2	Low	Low	0.32	0.50	0.57	−0.90	1.58
3	Low	High	0.32	1.00	0.57	−1.79	3.16
4	Medium	No	0.71	0.00	1.70	0.00	0.00
5	Medium	Low	0.71	−0.50	1.70	1.20	−0.71
6	Medium	High	0.71	−1.00	1.70	2.40	−1.41
7	High	No	0.95	0.00	5.10	0.00	0.00
8	High	Low	0.95	0.50	5.10	−2.69	0.53
9	High	High	0.95	1.00	5.10	−5.38	1.05

3. Simulation Study

We report next a comparative evaluation of fiducial and ML Wald-type interval estimators via Monte Carlo simulations. Nine-item tests ($m = 9$) and two sample size conditions ($n = 100$ and 500) were considered. Under each condition, 500 data sets were simulated. Apart from the intercepts and slopes in the original parameterization of the model, three additional parameters are of interest to us. The item difficulty parameter,

$$\delta_j = -\alpha_j/\beta_j, \tag{27}$$

gauges the latent variable level at which a correct response is produced with 50 % chance. The loading λ_j and threshold τ_j are standardizations of slope and intercept, respectively:

$$\lambda_j = \frac{\beta_j/1.7}{\sqrt{1 + (\beta_j/1.7)^2}}, \tag{28}$$

and

$$\tau_j = \delta_j \lambda_j. \tag{29}$$

They are defined on a standardized scale pertaining to the notion of explained variance (communality), which is the preferred metric in the literature of item factor analysis (e.g., Wirth & Edwards, 2007). The true item parameters, tabulated in Table 1, were determined by two factors: (a) $\lambda_j^2 = 0.1, 0.5, 0.9$, representing low, medium, and high communality, and (b) $|\tau_j| = 0, 0.5, 1$, representing no, low, and high skewness.

We implemented the previously discussed Gibbs sampler (Sect. 2.4) in Fortran. We set 0 as the starting value for intercepts, and 1 for slopes; $\mathbf{z}^0$ were generated from the standard normal distribution, and $\mathbf{a}^0$ were generated by running the Gibbs sampling step once, as described in the previous section. For each simulated dataset, we ran 60000 MCMC cycles, and burned in the first 10000 to remove the influence of starting values; 5000 draws were then extracted by applying a thinning interval of 10, from which equi-tailed percentile CIs (FID) were obtained. The parameter bound M is set to 20.

The ML estimates of item parameters were found by the Bock-Aitkin EM algorithm using M*plus* 7.0 (Muthén & Muthén, 1998-2012). The integral in the response pattern likelihood function

TABLE 2.
Empirical coverage and median length of CIs ($n = 100$).

Item	CI type	Empirical coverage					Median length				
		λ_j	τ_j	β_j	α_j	δ_j	λ_j	τ_j	β_j	α_j	δ_j
1	MWH	0.93	0.95	0.96	0.95	1.00	0.50	0.47	**0.99**	**0.85**	**1.67**
	MWO	0.94	0.97	0.96	0.97	1.00	0.53	0.51	1.07	0.92	1.80
	FID	0.94	0.96	0.94	0.96	0.98	**0.50**	**0.47**	1.02	0.85	3.09
2	MWH	0.93	0.96	0.94	0.95	**0.91**	0.53	0.51	**1.06**	**0.94**	**3.00**
	MWO	0.94	0.98	0.97	0.98	**0.92**	0.57	0.55	1.16	1.02	3.33
	FID	0.94	0.96	0.94	0.95	0.97	**0.52**	**0.50**	1.10	0.94	6.73
3	MWH	**0.93**	0.96	0.96	0.97	**0.89**	0.64	0.63	**1.32**	**1.25**	**6.41**
	MWO	0.95	0.97	0.98	0.97	**0.90**	0.69	0.69	1.45	1.35	7.18
	FID	0.95	0.96	0.95	0.95	0.98	**0.64**	**0.62**	1.39	1.27	30.08
4	MWH	**0.93**	0.94	0.96	0.95	0.95	**0.36**	0.49	**1.79**	**1.21**	**0.70**
	MWO	0.95	0.95	0.97	0.96	0.96	0.40	0.52	1.99	1.28	0.74
	FID	0.94	0.94	0.94	0.94	0.94	0.37	**0.49**	1.94	1.25	0.72
5	MWH	**0.91**	0.96	0.97	0.99	0.95	**0.39**	0.52	**1.99**	**1.46**	**0.82**
	MWO	0.94	0.97	0.98	0.99	0.96	0.45	0.55	2.25	1.60	0.86
	FID	0.94	0.95	0.94	0.96	0.96	0.40	**0.51**	2.21	1.55	0.89
6	MWH	**0.88**	0.96	0.95	0.97	**0.91**	**0.48**	0.61	**2.56**	**2.37**	**1.29**
	MWO	0.95	0.97	0.96	0.98	0.93	0.58	0.65	2.91	2.69	1.42
	FID	0.94	0.96	0.94	0.96	0.97	0.49	**0.61**	3.03	2.79	1.58
7	MWH	**0.82**	**0.93**	**0.93**	0.97	**0.93**	0.17	**0.49**	17.15	3.48	**0.51**
	MWO	0.95	0.94	0.97	1.00	0.95	0.26	0.51	25.13	4.18	0.53
	FID	1.00	0.95	1.00	0.95	0.95	**0.16**	0.49	**5.76**	**2.87**	0.53
8	MWH	**0.84**	0.94	**0.92**	**0.92**	0.95	0.16	0.52	12.99	6.29	**0.55**
	MWO	0.97	0.95	0.95	0.95	0.95	0.27	0.54	18.45	8.60	0.58
	FID	0.99	0.96	0.99	0.98	0.95	**0.15**	**0.52**	**5.76**	**3.87**	0.57
9	MWH	**0.77**	0.94	**0.87**	**0.88**	0.96	0.21	0.60	20.18	15.80	**0.69**
	MWO	0.94	0.96	0.94	0.95	0.97	0.39	0.64	40.42	38.52	0.78
	FID	0.98	0.95	0.98	0.99	0.94	**0.20**	**0.60**	**6.28**	**6.59**	0.73

Bold coverage probabilities are values less than 0.931, the lower limit of a normal-approximation 95 % CI for the nominal level 0.95 across 500 replications. The shortest median length among three CIs for each parameter is also shown in bold font. Non-positive definite Hessian matrices were obtained for 33 out of 500 simulated data sets; those replications were excluded when calculating summary statistics for MWH intervals.

(Eq.10) was approximated using 49 equally spaced rectangular quadrature points from -5 to 5. We adopted the software's default convergence criteria, maximum number of iterations, and starting values. Two types of Wald CIs were computed from the two commonly used sample estimates of the Fisher information matrix: the Hessian form (MWH; in M*plus*, estimator = ML) and the outer-product form (MWO; estimator = MLF). The Delta-method standard errors were used for transformed parameters.

The empirical coverage and median length of CIs are two main criteria for comparison. Intervals having coverage probabilities greater than or equal to the nominal level (95 % in the current work) and short lengths are preferred. Whenever a trade-off between coverage and length is observed, we always prioritize coverage over length. The results are tabulated in Tables 2 and 3 for the two sample size conditions, respectively.

As expected, the difference among the three candidate CIs is more salient in the small-sample condition ($n = 100$); in large samples (e.g., $n = 500$ in the current study), the three methods are

TABLE 3.
Empirical coverage and median length of CIs ($n = 500$).

Item	CI type	Empirical coverage					Median length				
		λ_j	τ_j	β_j	α_j	δ_j	λ_j	τ_j	β_j	α_j	δ_j
1	MWH	0.94	0.96	0.95	0.96	0.97	0.22	0.21	0.44	0.38	**0.67**
	MWO	0.95	0.95	0.95	0.96	0.97	0.23	0.21	0.45	0.38	0.68
	FID	0.95	0.95	0.95	0.95	0.95	**0.22**	**0.21**	**0.44**	**0.37**	0.72
2	MWH	0.96	0.95	0.95	0.96	0.94	0.24	0.23	0.47	0.42	**1.40**
	MWO	0.96	0.96	0.96	0.96	0.94	0.24	0.23	0.48	0.42	1.42
	FID	0.95	0.95	0.95	0.96	0.96	**0.24**	**0.22**	**0.47**	**0.41**	1.61
3	MWH	0.95	0.96	0.95	0.96	**0.92**	0.29	0.28	0.58	0.55	**2.85**
	MWO	0.95	0.97	0.96	0.96	**0.92**	0.29	0.29	0.59	0.55	2.90
	FID	0.95	0.96	0.95	0.96	0.97	**0.28**	**0.28**	**0.58**	**0.54**	3.62
4	MWH	0.95	**0.93**	0.94	**0.93**	0.93	**0.16**	0.22	0.79	0.53	**0.31**
	MWO	0.95	0.93	0.94	0.93	0.94	0.17	0.22	0.80	0.53	0.31
	FID	0.94	0.94	0.94	0.94	0.94	0.16	**0.22**	**0.79**	**0.52**	0.31
5	MWH	**0.93**	0.95	0.94	0.95	0.95	**0.18**	0.23	**0.87**	0.64	**0.37**
	MWO	0.93	0.95	0.95	0.96	0.95	0.18	0.23	0.88	0.65	0.38
	FID	0.94	0.95	0.94	0.95	0.94	0.18	**0.23**	0.87	**0.63**	0.38
6	MWH	**0.92**	0.95	0.95	0.95	0.94	**0.22**	0.27	**1.06**	**0.98**	**0.58**
	MWO	0.94	0.95	0.96	0.96	0.95	0.22	0.27	1.08	1.01	0.59
	FID	0.93	0.94	0.93	0.94	0.95	0.22	**0.27**	1.07	0.99	0.60
7	MWH	**0.93**	0.95	**0.93**	0.96	0.95	0.08	0.22	4.49	1.22	**0.23**
	MWO	0.94	0.95	0.94	0.96	0.95	0.09	0.22	4.80	1.23	0.23
	FID	0.97	0.95	0.97	0.95	0.95	**0.07**	**0.22**	**2.93**	**1.15**	0.23
8	MWH	0.95	0.95	0.96	0.95	0.96	0.08	0.23	4.19	2.30	0.25
	MWO	0.95	0.95	0.96	0.95	0.96	0.08	0.23	4.45	2.40	0.25
	FID	0.99	0.95	0.99	0.97	0.95	**0.06**	**0.23**	**2.88**	**1.80**	**0.25**
9	MWH	0.94	0.94	**0.92**	**0.91**	0.94	0.10	0.27	5.33	5.03	0.31
	MWO	0.95	0.94	0.95	0.94	0.93	0.11	0.27	5.95	5.79	0.31
	FID	0.97	0.94	0.97	0.97	**0.93**	**0.08**	**0.26**	**3.19**	**3.21**	**0.31**

Bold coverage probabilities are values less than 0.931, the lower limit of a normal-approximation 95 % CI for the nominal level 0.95 across 500 replications. The shortest median length among three CIs for each parameter is also shown in bold font. Non-positive definite Hessian matrices were obtained for 9 out of 500 simulated datasets; those replications were excluded when calculating summary statistics for MWH intervals.

more comparable in accordance with the asymptotic theory. Hence, we only discuss the results for $n = 100$ (Table 2) here.

For the original parameterization, MWO and FID always exhibit well-calibrated coverage, with FID being uniformly more efficient (i.e., having shorter lengths) than MWO across all items. In contrast, MWH significantly under-covers for large slopes (items 7–9), and skewed intercepts when combined with large slopes (items 8 and 9); what is worse, MWH also tends to be much wider than FID for those parameters. For low and medium communality items (items 1–6), however, MWH is the most reliable and efficient choice for slope and intercept parameters, trailed by FID with slightly less desirable lengths.

The coverage of MWH for loading parameters decreases substantially as the true value increases; for high communality items (items 7–9), its empirical coverage can be even lower than 80 %. This may be construed as the failure of normal approximation when the true parameters are closed to the boundary (here, 1 is the upper bound for the loading parameter), due to a skewed sampling distribution of the ML estimate. Having lengths comparable to MWH on

average, FID, however, is able to maintain well-controlled coverage; moreover, for large loading parameters (items 7–9), FID achieves the highest empirical coverage with the shortest median length. For threshold parameters, all three candidate methods show acceptable coverage; MWO is less favorable than MWH and FID, because it always yields wider intervals.

Both MWH and MWO are subject to insufficient coverage for non-zero difficulty parameters in low communality items (items 2 and 3). When a small slope co-occurs with a somewhat large intercept, the difficulty parameter tends to be large, i.e., close to infinity, which may lead to a non-normal sampling distribution of the ML estimate, and consequently the poor performance of normal-approximation intervals. In the meantime, the coverage of FID is not affected by extreme difficulty values for low communality items, compensated by excessive lengths (for item 3, FID is almost 5 times as wide as MWH).

In summary, FID, although not always the most efficient interval estimator, is always reliable in terms of coverage for all five parameterizations. The gold-standard method MWH is liberal when the ML estimates have non-normal sampling distributions, which is likely to happen for extreme parameters in small samples. The alternative MWO approach often yields conservative intervals that are adequate in coverage but typically wider than the corresponding FID and MWH ones.

4. Empirical Example

In this section, we apply the proposed GFI to an exploratory item factor analysis (EIFA) problem. The dataset being analyzed is the UK female normative sample data of the revised Eysenck Personality Questionnaire (EPQ-R; Eysenck, Eysenck, & Barrett, 1985). We are grateful to Dr. Paul Barrett for granting us access to the data. This questionnaire was originally designed to measure three dimensions of individual differences: extraversion (E), neuroticism (N), and psychoticism (P). In this analysis, we only use the 12 short form items from each subscale, so there are 36 items ($m = 36$) in total. The sample size is $n = 824$, after all incomplete cases deleted.

In EIFA, substantive researchers are more interested in the multi-factor structure of the scale, and the strength of each test item being associated with each factor. In this sense, the standardized loading-threshold parameterization is more helpful, because it is on a scale that eases the computation of variance/covariance of test items explained by the factors. In addition, analytic rotations of factor loadings (see Browne, 2001, for a review) are often applied to obtain more interpretable patterns of item factor dependency. The goal of this analysis is to obtain CIs for rotated factor loadings and the inter-factor correlations.

An r-dimensional ($r > 1$) EIFA model can be parameterized by Eq. 4: For each of the first $r - 1$ items, indexed by $j = 1, \ldots, r - 1$, the last j slopes are fixed to 0; for the remaining items, all slopes are freely estimated. Then, unrotated factor loadings for each item were computed as a non-linear transformation of the slopes:

$$\lambda_j = \frac{\beta_j/1.7}{\sqrt{1 + \beta_j^\top \beta_j/1.7^2}}, \tag{30}$$

which is a generalization of Eq. 28. The Crawford–Ferguson Quartimax criterion (Crawford & Ferguson, 1970) was minimized to obtain rotated factor loadings and inter-factor correlations, which leads to an implicit non-linear transformation of the unrotated loadings that does not have a closed-form expression. Implicit differentiation is required to compute the Delta-method standard errors for the rotated ML solutions, which has been described by Jennrich (1973).

Using GFI, however, we can easily approximate the fiducial distribution of rotated solutions by applying the transformation given by Eq. 30 and then the rotation routine to each Monte Carlo sample from the marginal fiducial distribution of slopes. We tuned the Gibbs sampler similarly as described in the simulation study, with the exception that the weighted least square solution (`estimator = WLSMV`) produced by M*plus* 7.0 (Muthén & Muthén, 1998-2012) and the corresponding factor score estimates were used as starting values $\boldsymbol{\theta}^0$ and $\mathbf{z}^0$, respectively, in order to accelerate the convergence of the generated Markov chain. The R package `GPArotation` (Bernaards & Jennrich, 2005) was used to perform analytic rotation. Note that the directions and the order of factors are not identified for the rotated solutions; a matching procedure was applied to establish a uniform orientation across all MCMC iterations, similar to that described by Asparouhov and Muthén (2012) in the context of Bayesian EIFA.

We fitted three-, four-, and five- factor EIFA models to the EPQ-R data; for succinctness, only the five-factor solution is reported here. The high dimensionality and the relatively small sample size render likelihood-based estimation and inference very challenging; with GFI, however, we are still able to obtain substantively meaningful results. Fiducial medians and 95 % equi-tailed percentile CIs for rotated loadings and inter-factor correlations are shown in Figures 3 and 4, respectively.

When multiple CIs for selected parameters are reported, Benjamini and Yekutieli (2005) recommended a general procedure that controls the false coverage-statement rate (FCR), i.e., the expected proportion of the selected parameters not covered by the constructed CIs. Here, we are interested in testing whether the rotated loadings and inter-factor correlations are 0. Thus, we computed for each parameter the empirical two-sided p-values for the corresponding test, selected R significant parameters by the Benjamini–Hochberg step-up procedure (Benjamini and Hochberg,1995) at nominal level 0.05, and then construct $100(1-0.05R/\tilde{q})$ % CIs for all loading and correlation parameters, in which $\tilde{q} = rm + r(r-1)/2 = 190$ is the total number of parameters being tested for the five-factor EIFA.

The psychoticism items dominate the first factor in the five-factor EIFA. Factor 2 and 3 yield a further decomposition of the extraversion subscale. The separation of the two factors is driven by two locally dependent pairs of items (e.g., Liu & Thissen, 2012): Factor 2 is led by two "party" items, i.e., "Can you easily get some life into a rather dull party? (E51)" and "Can you get a party going? (E78)"; the two items loaded the highest on factor 3 are "Are you a talkative person? (E6)" and "Are you mostly quiet when you are with other people? (E47)", both related to loquaciousness. The remaining extraversion items are moderately cross-loaded on both factors. The correlation between factor 2 and 3 is about 0.5, which is the highest among all factors. Meanwhile, the neuroticism items are split into halves (factor 4 and 5). After examining the item stems, we conclude that factor 5 is mainly indicated by the mood-related items in the neuroticism subscale, e.g., "Does your mood often go up and down? (N3)" and "Do you often feel 'fed-up'? (N26)". Factor 4, on the other hand, is defined by the items related to worrying and nerves. In addition, extraversion (factors 2 and 3) is nearly uncorrelated with the emotion-related neuroticism (factor 5), but negatively correlated with the emotion-free one (factor 4).

To qualify the fiducial solution, we conducted a 100-replication bootstrap simulation: Data sets were generated from a five-dimensional model using the point estimates of the rotated factor loadings and inter-factor correlations as the true values. The empirical coverage accumulated across the 100 resamples is also included in Figures 3 and 4. It is observed that for almost all loading and correlation parameters, the empirical coverage of the fiducial percentile interval is close to the nominal level (coverage frequency $>$90). But for some moderately high loadings and the largest inter-factor correlation, the fiducial interval is too liberal. We conclude that in general the fiducial intervals obtained in the current example can be trusted; however, we need further investigations on those problematic cases to better understand the behavior of GFI in EIFA.
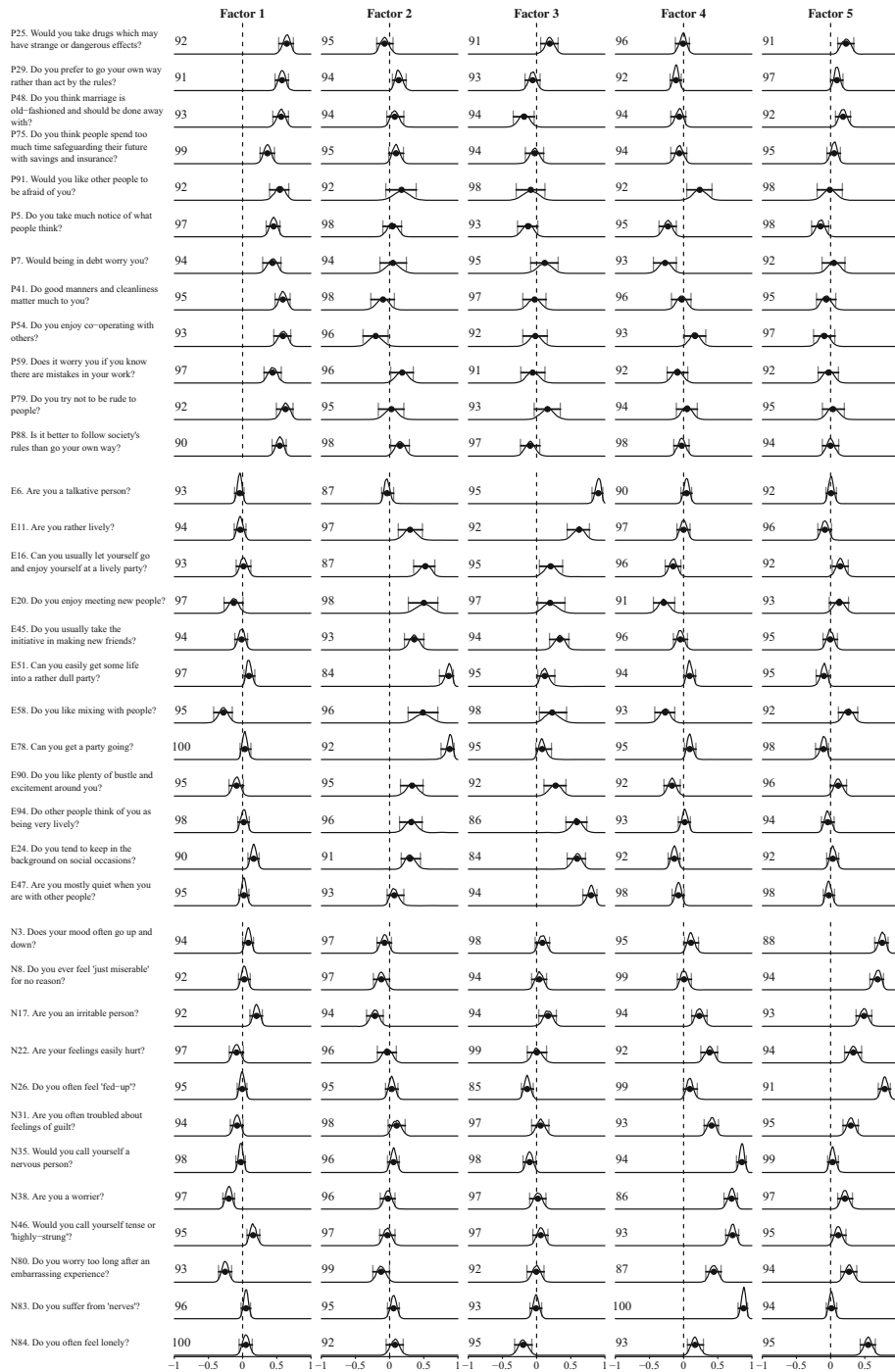
FIGURE 3.

Point estimates and 95 % CIs for rotated factor loadings in the five-factor model. The tabular layout has a *row* for each item; item stems are listed in the *leftmost column*. The following *five columns* correspond to the five factors. Within each cell, the estimated fiducial density is shown in the background. Superimposed are the fiducial median (shown as *dots*), and the 95 % fiducial equi-tailed percentile CIs. The 0 point on the factor loading scale is highlighted by the *vertical dashed lines*. For each parameter, the empirical coverage frequency in the bootstrap simulation is also included.
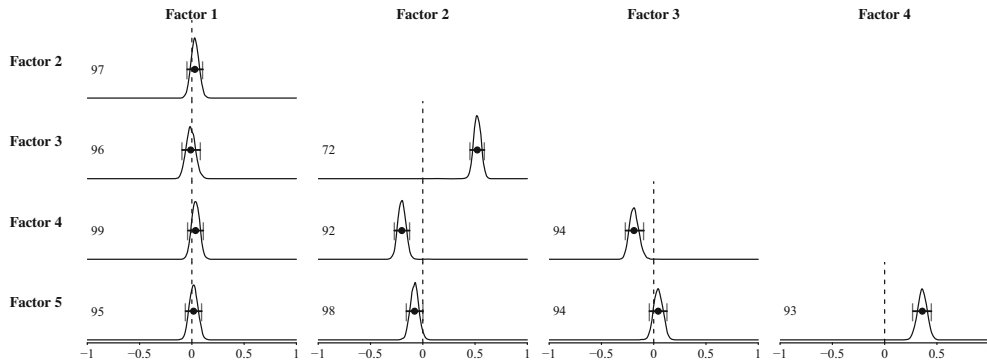
FIGURE 4.

Point estimates and 95 % CIs for inter-factor correlations in the five-factor model. The tabular layout resembles a correlation matrix. Within each cell, the estimated fiducial density is shown in the background. Superimposed are the fiducial median (shown as *dots*), and the 95 % fiducial equi-tailed percentile CIs. The 0 point on the correlation scale is highlighted by the *vertical dashed lines*. For each parameter, the empirical coverage frequency in the bootstrap simulation is also included.

5. Discussion and Conclusion

In the current research, GFI is employed to address interval estimation problems for a family of binary logistic IRT models. We derive a fiducial distribution for item parameters, prove a Bernstein–von Mises theorem analogous to the well-known version for Bayesian posteriors, and implement an efficient MCMC sampler to fit the model. It has been observed in the simulation study that the fiducial percentile CI outperforms the commonly used ML Wald-type CIs when the sample size is small and the generating parameters are extreme. In addition, as shown in the EIFA example, GFI offers great flexibility and reliable performance when interval estimation is desired for complex transformations of parameters. All these render GFI a promising statistical tool catering to the gaining popularity of item response models in psychological and educational testing.

As pointed out by a referee, good coverage coupled with short width of CIs often translates to small mean squared error of the corresponding point estimates; in this regard, we observed from pilot simulations that the fiducial median can be less biased and variable than the ML estimate when the sample size is small, in line with the empirical coverage and length results reported in Sect. 3. However, since the improvement is often outweighed by the large sampling variability, it is highly recommended to rely on CIs, rather than point estimates, when interpreting model parameters in small-sample calibrations. Even in large-scale educational testing, the usefulness of CIs is likely underestimated. Operational researchers in educational assessment programs tend to only pay heed to point estimates, because their pool of respondents is often large. Indeed, when the sample size is large enough, ML, fiducial, and Bayesian fittings should not be dissimilar because of their asymptotic equivalence. What is often ignored, however, is the trade-off between the sample size and model complexity in determining the amount of sampling variability of point estimates, the degree of which is largely unknown in practice until CIs are calculated. Therefore, we believe that methods producing high-quality CIs, such as GFI, deserve more attention than they receive at the moment.

There are limitations and extensions of the current study that remain to be addressed by future research.

First, the Bernstein–von Mises theorem (Theorem 1) is only established for an approximation of the fiducial distribution, which is a limitation of the current work. Although similar constructions of the empirical Bayesian approximation have been considered "fiducial" by some authors (e.g.,

Hannig, 2009), its exact relation with GFI is yet to be demonstrated. In addition, an extension of Theorem 2 to multidimensional models ($r > 1$) should be pursued. As the latent dimensionality increases, we cannot discuss all the possible cases as we do in the last part of the unidimensional proof (see Appendix 3). More intricate arguments involving the high-dimensional Euclidean geometry are expected to replace the current case-enumerating one.

Second, the current study is more theoretically oriented, and the simulation study is more of an illustration than a demonstration of the proposed GFI. Carefully designed large-scale simulation studies should be conducted to evaluate all existing frequentist and Bayesian inference methods in the context of more involved multidimensional IRT models. It is particularly of interest to compare GFI to a number of stochastic variants of the EM algorithm (e.g., Cai, 2010a, b) for ML estimation, and to "less informative" Bayesian methods using flat priors. Apart from standard criteria of parameter recovery, practical matters, such as the computational time and convergence of the Markov chain, are subject to comparison as well.

Third, the reliable performance of GFI observed under the combination of a small sample size and extreme parameter values prompts our speculation that in general GFI is able to handle nearly unidentified models properly. For example, as brought up by a referee, the guessing parameter in a three-parameter logistic (3PL) model (Birnbaum, 1968) is difficult to estimate. In those cases, the log-likelihood function is flat, and finding its mode can be challenging. Adding a prior distribution ameliorates the numerical condition; however, since the contribution from the likelihood is very little, the performance of a Bayesian estimator is almost completely contingent upon how good the prior is. Meanwhile, GFI may produce wide CIs if there is indeed barely any information contained in the data. Yet those wide CIs at least have trustworthy coverage, from which sound statistical inferences can be made.

Finally, the usefulness of GFI for other inferential purposes, such as goodness of testing and test scoring, should be explored. It has been identified that the quality of asymptotic covariance matrix estimates plays an important role in determining the performance of various quadratic form goodness of fit statistics (e.g., Cai, 2008; Liu & Maydeu-Olivares, 2013). Suitable co-variation estimates of the fiducial distribution are natural candidates to this end, and their performance should be examined and compared to existing approaches such as the inverse outer-product/Hessian information estimators. As an alternative, a fiducial analogy of Bayesian posterior predictive checks (Rubin, 1984) can be easily programmed; theoretical inquisition and empirical evaluation can be pursued as a distinctive line of research to validate its use. As for test scoring, a Monte Carlo sample from the marginal distribution of $\mathbf{Z}^\star$ given $Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) \neq \emptyset$, an incidental product of the sampling algorithm, can be used to calibrate the latent traits for each observation. Consistency of individual latent score estimates in some proper sense is anticipated.

Appendix 1: Proof of Lemma 1

Let $\mathbf{V}$ be the random variable that equals to one of the C_n potential extremal points with equal probability unconditionally, i.e., $P\{\mathbf{V} = \mathbf{V}_I\} = C_n^{-1}$. It follows that

$$P\left\{\mathbf{V} \le \boldsymbol{\theta}, Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) \ne \emptyset\right\} = C_n^{-1} \sum_I P\left\{\mathbf{V}_I \le \boldsymbol{\theta}, D_I\right\}. \tag{31}$$

The remaining task is to derive each summand on the right-hand side (RHS) of Eq. 31 and then differentiate it with respect to $\boldsymbol{\theta}$.

Consider a single item j first. Recall that $\mathbf{V}_{I_j}$ is the potential vertex determined by sub-sample I_j. When $\mathbf{V}_{I_j} = \boldsymbol{\theta}'_j$ and serves as an interior vertex of $Q_j(\mathbf{y}_{(j)}, \mathbf{A}^\star_{(j)}, \mathbf{Z}^\star)$, it means that $\tau_j(\boldsymbol{\theta}'_j, \mathbf{Z}^\star_i) = A^\star_{ij}$ for all $i \in I_j$, and that $\boldsymbol{\theta}'_j$ should not conflict with the half-spaces of the other observations: i.e., for all $i \in I_j^c$, $A^\star_{ij} \le \tau_j(\boldsymbol{\theta}'_j, \mathbf{Z}^\star_i)$ if $y_{ij} = 1$, and $A^\star_{ij} > \tau_j(\boldsymbol{\theta}'_j, \mathbf{Z}^\star_i)$ if $y_{ij} = 0$. Thus, conditional on $\mathbf{Z}^\star = \mathbf{z}$, we have

$$
\begin{aligned}
&P\left\{\mathbf{V}_{I_j} \le \boldsymbol{\theta}_j, Q_j(\mathbf{y}_{(j)}, \mathbf{A}^\star_{(j)}, \mathbf{Z}^\star) \ne \emptyset \mid \mathbf{V} = \mathbf{V}_I, \mathbf{Z}^\star = \mathbf{z}\right\} \\
&= \int_{\boldsymbol{\theta}'_j \le \boldsymbol{\theta}_j} \left|\det(\partial \tau_j(\boldsymbol{\theta}_j, \mathbf{z}_i)/\partial \boldsymbol{\theta}_j)_{i \in I_j}\right| \prod_{i \in I_j} \frac{e^{\tau_j(\boldsymbol{\theta}'_j, \mathbf{z}_i)}}{[1 + e^{\tau_j(\boldsymbol{\theta}'_j, \mathbf{z}_i)}]^2} \prod_{i \in I_j^c} f_j(\boldsymbol{\theta}'_j, y_{ij}|\mathbf{z}_i) d\boldsymbol{\theta}'_j, \quad (32)
\end{aligned}
$$

in which the determinant and the first product are due to the change of variables from $(A^\star_{ij})_{i \in I_j}$ to $\mathbf{V}_{I_j}$ (the standard logistic density $\psi(x) = e^x/(1 + e^x)^2$), and the second product corresponds to the logistic probabilities of those inequalities that the other observations should satisfy. Due to the conditional independence assumption,

$$
\begin{aligned}
&P\left\{\mathbf{V}_I \le \boldsymbol{\theta}, Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) \ne \emptyset \mid \mathbf{V} = \mathbf{V}_I\right\} \\
&= \int_{\mathbb{R}^{nr}} \prod_{j=1}^m P\left\{\mathbf{V}_{I_j} \le \boldsymbol{\theta}_j, Q(\mathbf{y}_{(j)}, \mathbf{A}^\star_{(j)}, \mathbf{Z}^\star) \ne \emptyset \mid \mathbf{V} = \mathbf{V}_I, \mathbf{Z}^\star = \mathbf{z}\right\} d\Phi(\mathbf{z}) \\
&= \int_{\mathbb{R}^{nr}} \int_{\boldsymbol{\theta}' \le \boldsymbol{\theta}} d_I(\boldsymbol{\theta}', \mathbf{z}_I) \prod_{j=1}^m \left\{\prod_{i \in I_j} \frac{e^{\tau_j(\boldsymbol{\theta}'_j, \mathbf{z}_i)}}{[1 + e^{\tau_j(\boldsymbol{\theta}'_j, \mathbf{z}_i)}]^2} \prod_{i \in I_j^c} f_j(\boldsymbol{\theta}'_j, y_{ij}|\mathbf{z}_i)\right\} d\boldsymbol{\theta}' d\Phi(\mathbf{z}). \quad (33)
\end{aligned}
$$

Equation 15 is established by substituting Eq. 33 back into the RHS of Eq. 31, switching the order of integrals, and differentiating with respective to $\boldsymbol{\theta}$.

Appendix 1: Proof of Theorem 1

We start from re-expressing the density of $\mathbf{R}(\mathbf{y})$, i.e., Eq. 15. Note that the summands of Eq. 15 corresponding to I and I', $I \ne I'$, are the same whenever $\mathbf{y}_I = \mathbf{y}_{I'}$; hence, the (outer) sum over index sets I therein can be reduced to a finite sum over sub-sample response patterns $\mathbf{y}_I$. Note that $\bigcup_{j=1}^m I_j$ has at least $\max_j q_j$ and at most $\sum_{j=1}^m q_j$ elements. Let $G_n = \binom{n}{\sum_{j=1}^m q_j}$ be the total number of size-$\sum_{j=1}^m q_j$ sub-samples. Also let $p_n(\mathbf{y}_I) = G_n^{-1} \sum_I \mathbb{I}\{\mathbf{Y}_I = \mathbf{y}_I\}$. By the standard theory of U-statistics, $p_n(\mathbf{y}_I) \overset{P_{\boldsymbol{\theta}_0}}{\to} \pi_0(\mathbf{y}_I)$, in which $\pi_0(\mathbf{y}_I)$ is determined by the data-generating parameter values $\boldsymbol{\theta}_0$, and $\pi_0(\mathbf{y}_I) = 0$ if $|I| < \sum_{j=1}^m q_j$. Then, the density can be written as

$$g_n(\boldsymbol{\theta}|\mathbf{y}) \propto b_n(\boldsymbol{\theta}, \mathbf{y}) f_n(\boldsymbol{\theta}, \mathbf{y}). \tag{34}$$

In Eq. 34, $f_n(\boldsymbol{\theta}, \mathbf{y})$ is the sample likelihood, and

$$b_n(\boldsymbol{\theta}, \mathbf{y}) = G_n \sum_{\mathbf{y}_I} p_n(\mathbf{y}_I) b_{\mathbf{y}_I}(\boldsymbol{\theta}), \tag{35}$$

in which

$$b_{\mathbf{y}_I}(\boldsymbol{\theta}) = \int d_I(\boldsymbol{\theta}, \mathbf{z}_I) \prod_{i \in I} \left\{ \prod_{j \in J_i} \frac{e^{\tau_j(\theta_j, \mathbf{z}_i)}}{[1 + e^{\tau_j(\theta_j, \mathbf{z}_i)}]^2} \prod_{j \notin J_i} f_j(\theta_j, y_{ij}|\mathbf{z}_i) \right\} \mathrm{d}\Phi(\mathbf{z}_I)$$
$$\bigg/ \int \prod_{i \in I} \prod_{j=1}^{m} f_j(\theta_j, y_{ij}|\mathbf{z}_i) \mathrm{d}\Phi(\mathbf{z}_I). \tag{36}$$

Equation 35 is a repetition of Eq. 19. Also let

$$a_n(\boldsymbol{\theta}, \mathbf{y}) = f_n(\boldsymbol{\theta}, \mathbf{y}) b_n(\boldsymbol{\theta}, \mathbf{y}) \tag{37}$$

be the RHS of Eq. 34.

Next, we consider the local parameter $\mathbf{h} = \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$. Some short-hand notation is introduced for conciseness: Let $b_{n,\mathbf{h}} = b_n(\boldsymbol{\theta}_0 + \mathbf{h}/\sqrt{n}, \mathbf{y})/G_n$, $a_{n,\mathbf{h}} = a_n(\boldsymbol{\theta}_0 + \mathbf{h}/\sqrt{n}, \mathbf{y})$, and $f_{n,\mathbf{h}} = f_n(\boldsymbol{\theta}_0 + \mathbf{h}/\sqrt{n}, \mathbf{y})/\sqrt{n}$; also let $b_0 = \sum_{\mathbf{y}_I} \pi_0(\mathbf{y}_I) b_{\mathbf{y}_I}(\boldsymbol{\theta}_0)$, $a_{n,0} = a_n(\boldsymbol{\theta}_0, \mathbf{y})$, and $f_{n,0} = f_n(\boldsymbol{\theta}_0, \mathbf{y})$. Using this new notation, the corresponding density of the local parameter can be written as

$$\bar{g}_n(\mathbf{h}|\mathbf{y}) \propto a_{n,\mathbf{h}} = b_{n,\mathbf{h}} f_{n,\mathbf{h}}. \tag{38}$$

For each $\mathbf{y}_I$, $b_{\mathbf{y}_I}(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ (it is in fact differentiable). In addition, we know that $p_n(\mathbf{y}_I) \to \pi_0(\mathbf{y}_I)$ in $P_{\boldsymbol{\theta}_0}$-probability. Consequently, $b_{n,\mathbf{h}} \to b_0$ in $P_{\boldsymbol{\theta}_0}$-probability.

We also consider the Taylor series expansion of $\log f_{n,\mathbf{h}}$ at the true parameter $\boldsymbol{\theta}_0$:

$$\log \frac{f_{n,\mathbf{h}}}{f_{n,0}} = \mathbf{h}^\top \mathbf{S}_n + \frac{1}{2n} \sum_{i=1}^{n} \mathbf{h}^\top \mathbf{H}(\boldsymbol{\theta}_0, \mathbf{y}_i) \mathbf{h} + R_{n,\mathbf{h}}. \tag{39}$$

Here, some comments are made for each term of Eq. 39. (a) The sequence $\{\mathbf{S}_n\}$ is tight by the convergence result given by Eq. 13; hence, for each $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset \mathbb{R}^q$ such that $P(K_\varepsilon) > 1 - \varepsilon$ and $\mathbf{S}_n \in K_\varepsilon$ for all n. If we restrict the consideration to K_ε, then the first term of Eq. 39 is bounded for each $\mathbf{h}$. (b) By the (Uniform) Law of Large Numbers, the second term converges to $\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}$ in probability (the convergence is uniform for $\mathbf{h}$ in compact sets). (c) The remainder term has the following form:

$$R_{n,\mathbf{h}} = \sum_{i=1}^{n} \sum_{|\mathbf{t}|=3} \frac{f^{(\mathbf{t})}(\bar{\boldsymbol{\theta}}, \mathbf{Y}_i)}{\mathbf{t}!} \left(\frac{\mathbf{h}}{\sqrt{n}} \right)^{\mathbf{t}}. \tag{40}$$

In Eq. 40, $\mathbf{t} = (t_1, \dots, t_q)$ is a q-tuple of non-negative integers serving as a multi-index: $|\mathbf{t}| = \sum_{s=1}^{q} t_s$, $\mathbf{h}^{\mathbf{t}} = h_1^{t_1} \cdots h_q^{t_q}$, $\mathbf{t}! = \frac{q!}{t_1! \cdots t_q!}$, and $f^{(\mathbf{t})} = \frac{\partial^{|\mathbf{t}|} f}{\partial^{t_1} \theta_1 \cdots \partial^{t_q} \theta_q}$, where $h_1, \dots, h_q$ and $\theta_1, \dots, \theta_q$ are the coordinates of $\mathbf{h}$ and $\boldsymbol{\theta}$, respectively. $\bar{\boldsymbol{\theta}}$ lies between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_0 + \mathbf{h}/\sqrt{n}$.

Now we proceed to the proof of Theorem 1, i.e., Eq. 20. By an argument similar to Ghosh and Ramamoorthi (2003), it suffices to show for each $\varepsilon > 0$ that

$$\int_{H_n} \left| \frac{a_{n,\mathbf{h}}}{f_{n,0}} - b_0 e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2} \mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} \right| d\mathbf{h} \xrightarrow{P_{\boldsymbol{\theta}_0}} 0 \tag{41}$$

To see this, let $D_n = \int_{H_n} a_{n,\mathbf{h}} d\mathbf{h}/f_{n,0}$. The left-hand side (LHS) of Eq. 20 can be bounded by

$$D_n^{-1} \int_{H_n} \left| \frac{a_{n,\mathbf{h}}}{f_{n,0}} - b_0 e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} \right| d\mathbf{h} + \int_{H_n} \left| D_n^{-1} b_0 e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} - \phi_{\mathcal{I}_0^{-1}\mathbf{S}_n, \mathcal{I}_0^{-1}}(\mathbf{h}) \right| d\mathbf{h}.$$

(42)

Notice that Eq. 41 implies $|D_n - b_0 \int_{H_n} e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} d\mathbf{h}| \overset{P_{\boldsymbol{\theta}_0}}{\to} 0$. We also know that

$$D e^{\frac{1}{2}\mathbf{S}_n^\top \mathcal{I}_0^{-1}\mathbf{S}_n} \leq \int_{H_n} e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} d\mathbf{h} \leq D' e^{\frac{1}{2}\mathbf{S}_n^\top \mathcal{I}_0^{-1}\mathbf{S}_n},$$

(43)

for some suitable constants D and D', because the local parameter space H_n satisfies $\Theta - \boldsymbol{\theta}_0 \subset H_n \subset \mathbb{R}^q$. It follows that D_n^{-1} is $O_p(1)$, and that the first integral in Eq. 42 converges to zero in probability. Further let $T_{1,n}$ be the integral in Eq. 43, and $T_{2,n} = |D_n^{-1}b_0 - T_{1,n}^{-1}|$; then, the second integral of Eq. 42 can be written as $T_{1,n}T_{2,n}$. The sequence $\{T_{1,n}\}$ is tight by Eq. 43, so for each $\eta > 0$, there exists an L_η such that $P(T_{1,n} \leq L_\eta) > 1 - \eta$ for all n. Moreover, $T_{2,n} \overset{P_{\boldsymbol{\theta}_0}}{\to} 0$ by Eq. 41. Fix $\varepsilon, \eta > 0$, we have

$$P(T_{1,n}T_{2,n} > \varepsilon) \leq P(T_{1,n}T_{2,n} > \varepsilon, T_{1,n} \leq L_\eta) + P(T_{1,n} > L_\eta) \leq P(T_{2,n} > \varepsilon/L_\eta) + \eta,$$

(44)

which can be made less than 2η for large enough n. Therefore, $T_{1,n}T_{2,n} \overset{P_{\boldsymbol{\theta}_0}}{\to} 0$. Because both integrals in Eq. 42 converge to 0 in probability, Eq. 20 is established.

For the remaining part of the proof, we partition the domain of integration of Eq. 20 into four regions (for n large enough), and establish the desired convergence on each part. The four regions are as follows:

$$A_{1,n} = \{\mathbf{h} : \|\mathbf{h}\| < B\log n\} \cap H_n, \text{ for some large number } B > 0;$$
$$A_{2,n} = \{\mathbf{h} : B\log n \leq \|\mathbf{h}\| < \delta\sqrt{n}\} \cap H_n, \text{ for some small number } \delta > 0;$$
$$A_{3,n} = \{\mathbf{h} : \delta\sqrt{n} \leq \|\mathbf{h}\| \leq B'\sqrt{n}\} \cap H_n, \text{ for another large number } B' > 0;$$
$$A_{4,n} = \{\mathbf{h} : \|\mathbf{h}\| > B'\sqrt{n}\} \cap H_n.$$

In terms of the constants, we first choose δ and B to ensure the convergence on $A_{2,n}$. The convergence on $A_{1,n}$ holds for any $B > 0$, so it also holds for the particular B that we select. Then we consider region $A_{4,n}$ and select B'. Finally, we show the integral convergences for $\mathbf{h}/\sqrt{n}$ in any compact sets excluding $\mathbf{0}$, from which the convergence on $A_{3,n}$ follows.

Region $A_{2,n}$ Because the likelihood function is three times continuously differentiable with respect to $\boldsymbol{\theta}$, and also because there are finitely many (i.e., $\prod_{j=1}^m K_j$) individual patterns of $\mathbf{y}_i$, the remainder term (Eq. 40) of the Taylor expansion (Eq. 39) has the following bound for each $\delta > 0$ and $\|\mathbf{h}\| \leq \delta\sqrt{n}$:

$$|R_{n,\mathbf{h}}| \leq M(\delta)\frac{\|\mathbf{h}\|^3}{n^{3/2}} \leq M(\delta)\delta^3,$$

(45)

as a result of the multinomial theorem and the Cauchy–Schwarz inequality, in which $M(\delta)$ is a constant multiple of $|\max_{|\mathbf{t}|=3, \mathbf{y}_i} \sup_{\|\boldsymbol{\theta}-\theta_0\|\leq\delta} f^{(\mathbf{t})}(\boldsymbol{\theta}, \mathbf{y}_i)|$. Since $M(\delta) \downarrow$ as $\delta \downarrow 0$, Eq. 45 allows us to choose δ small enough such that $|R_{n,\mathbf{h}}| < \frac{1}{4}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}$ for all $\mathbf{h} \in A_{2,n}$. Then for such δ,

$$\int_{A_{2,n}} \left| \frac{a_{n,\mathbf{h}}}{f_{n,0}} - b_0 e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} \right| d\mathbf{h}$$

$$\leq \int_{A_{2,n}} \frac{a_{n,\mathbf{h}}}{f_{n,0}} d\mathbf{h} + \int_{A_{2,n}} b_0 e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} d\mathbf{h}$$

$$\leq \sup_{\mathbf{h} \in A_{2,n}} b_{n,\mathbf{h}} \int_{A_{2,n}} \frac{f_{n,\mathbf{h}}}{f_{n,0}} d\mathbf{h} + b_0 \int_{A_{2,n}} e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} d\mathbf{h}$$

$$\leq \left(\sup_{\mathbf{h} \in A_{2,n}} b_{n,\mathbf{h}} + b_0 \right) \int_{A_{2,n}} e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{4}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} d\mathbf{h} + o_p(1). \tag{46}$$

In the last line of Eq. 46, the parenthesized term is bounded due to the continuity of function $b_{\mathbf{y}_l}(\boldsymbol{\theta})$, the boundedness of set $A_{2,n}$, and our selection of δ. The $o_p(1)$ term comes from the uniform convergence of the second term in the Taylor expansion (Eq. 39). Also notice that

$$\int_{A_{2,n}} e^{-\frac{1}{4}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} d\mathbf{h} \leq C e^{-C'B \log n} (\delta \sqrt{n} - B \log n)^q \leq C'' n^{q/2 - C'B}, \tag{47}$$

where C, C', and C'' are constants. By selecting B large enough, Eq. 47 implies $\int_{A_{2,n}} e^{-\frac{1}{4}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} d\mathbf{h}$ $\to 0$. Finally, an argument using tightness similar to Eqs. 43 and 44 shows that the RHS of Eq. 46 converges to 0 in probability.

Region $A_{1,n}$ The convergence on $A_{1,n}$ can be established similarly. Fix an arbitrary $B > 0$. For the particular δ, we have selected

$$\sup_{\mathbf{h} \in A_{1,n}} |R_{n,\mathbf{h}}| \leq M(\delta) \sup_{\mathbf{h} \in A_{1,n}} \frac{\|\mathbf{h}\|^3}{n^{3/2}} \leq M(\delta) B^3 \frac{\log^3 n}{n^{3/2}} = o(1), \tag{48}$$

in which $M(\delta)$ is the same as in Eq. 45. Then

$$\int_{A_{1,n}} \left| \frac{a_{n,\mathbf{h}}}{f_{n,0}} - b_0 e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} \right| d\mathbf{h} \leq \int_{A_{1,n}} b_{n,\mathbf{h}} \left| e^{R_{n,\mathbf{h}}} - 1 \right| e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} d\mathbf{h}$$

$$+ \int_{A_{1,n}} |b_{n,\mathbf{h}} - b_0| e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} d\mathbf{h} + o_p(1). \tag{49}$$

In Eq. 49, the $o_p(1)$ term is again due to the uniform convergence of the second term in Eq. 39. Eq. 48 implies that $\sup_{\mathbf{h} \in A_{1,n}} |e^{R_{n,\mathbf{h}}} - 1| \to 0$; together with $b_{n,\mathbf{h}} \overset{P_{\boldsymbol{\theta}_0}}{\to} b_0$ and the boundedness of $A_{1,n}$, both integrals on the RHS of Eq. 49 converges to 0 in probability (the tightness argument needs to be used again).

Region $A_{4,n}$ Assume for a moment that there exists a large number B' such that

$$\sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > B'} \min_{\mathbf{y}_i} f(\boldsymbol{\theta}, \mathbf{y}_i) < f(\boldsymbol{\theta}_0, \mathbf{y}_i^\circ)^{2/f(\boldsymbol{\theta}_0, \mathbf{y}_i^\circ)}, \tag{50}$$

in which $\mathbf{y}_i^\circ$ is the least plausible individual response pattern under $P_{\boldsymbol{\theta}_0}$. Also write $p_n(\mathbf{y}_i^\circ)$ be the observed proportion of $\mathbf{y}_i$; $p_n(\mathbf{y}_i^\circ) \overset{P_{\boldsymbol{\theta}_0}}{\to} f(\boldsymbol{\theta}_0, \mathbf{y}_i^\circ)$. Then, on region $A_{4,n}$ defined by such a B',

$$P\left\{\frac{\min_{y_i} f(\boldsymbol{\theta}, \mathbf{y}_i)^{p_n(\mathbf{y}_i^\circ)}}{f(\boldsymbol{\theta}_0, \mathbf{y}_i^\circ)} < 1\right\} \geq P\left\{p_n(\mathbf{y}_i^\circ) > \frac{f(\boldsymbol{\theta}_0, \mathbf{y}_i^\circ)}{2}\right\} \to 1. \tag{51}$$

Therefore, we have

$$\frac{f_{n,\mathbf{h}}}{f_{n,0}} \leq \left[\frac{\min_{y_i} f(\boldsymbol{\theta}, \mathbf{y}_i)^{p_n(\mathbf{y}_i^\circ)}}{f(\boldsymbol{\theta}_0, \mathbf{y}_i^\circ)}\right]^n \leq \rho^n + o_p(1) \tag{52}$$

for some $0 < \rho < 1$. Also note that this likelihood ratio bound is not affected if finitely many observations are removed from $f_{n,\mathbf{h}}$, which is the case after dividing by the denominator of each summand of $b_{n,\mathbf{h}}$. As a result,

$$\int_{A_{4,n}} \left|\frac{a_{n,\mathbf{h}}}{f_{n,0}} - b_0 e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}}\right| d\mathbf{h} \leq \int_{A_{4,n}} \frac{b_{n,\mathbf{h}} f_{n,\mathbf{h}}}{f_{n,0}} d\mathbf{h} + b_0 \int_{A_{4,n}} e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} d\mathbf{h}$$

$$\leq K\rho^n + b_0 \int_{A_{4,n}} e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \mathcal{I}_0 \mathbf{h}} d\mathbf{h} + o_p(1), \tag{53}$$

in which K is a constant. Equation 53 results from the fact that: (a) The numerator of Eq. 36 is integrable with respect to the Lebesgue measure on the parameter space, which contributes to the constant K; (b) $p_n(\mathbf{y}_I) \overset{P_{\boldsymbol{\theta}_0}}{\to} \pi_0(\mathbf{y}_I)$, so the latter also contributes to K, while the difference of the two is merged into the $o_p(1)$ term. The second term on the RHS of Eq. 53 converges to zero by a similar tightness argument and the tail estimate of a multivariate normal distribution. These altogether show that the LHS of Eq. 53 converges to zero in probability.

Now we prove the result stated by Eq. 50; we denote the RHS of Eq. 50 by η.

First, consider the parameter subspace of α_j and $\boldsymbol{\beta}_j$ for each j. Let $L_j = \|(\alpha_j\ \boldsymbol{\beta}_j^\top)^\top\|$, and $\mathbf{d}_j = (\alpha_j\ \boldsymbol{\beta}_j^\top)^\top / L_j \in \mathbb{R}^{r+1}$ be a unit directional vector, in which the coordinates corresponding to fixed slopes are set to 0. Also introduce the partition $\mathbf{d}_j = (x_j\ \mathbf{e}_j^\top)^\top$ separating the direction of the intercept parameter, i.e., the first coordinate x_j, from those of the slopes. Then, we write

$$\tau_j(\boldsymbol{\theta}_j, \mathbf{Z}_i^\star) = \alpha_j + \boldsymbol{\beta}_j^\top \mathbf{Z}_i^\star = L_j(x_j + \mathbf{e}_j^\top \mathbf{Z}_i^\star), \tag{54}$$

in which $x_j + \mathbf{e}_j^\top \mathbf{Z}_i^\star \sim \mathcal{N}(x_j, 1 - x_j^2)$. For fixed $\mathbf{d}_j$, define $H_{\mathbf{d}_j}^\varepsilon(y) = \{\mathbf{z}_i \in \mathbb{R}^r : (-1)^y(x_j + \mathbf{e}_j^\top \mathbf{z}_i) \geq \varepsilon\}$ for $\varepsilon \geq 0$.

Now pool across multiple items. A direct consequence of Lemma 2, which is presented soon, is that $\mathbb{R}^r \subset \bigcup_{j=1}^{r+1} H_{\mathbf{d}_j}^0(y_{ij})$ for properly selected $(y_{ij})_{j=1}^{r+1}$ (recall that we assume $m > r$, so there are sufficient items). Then, for any $\varepsilon > 0$, the following bound can be established for the likelihood of an individual response pattern in which the first $r + 1$ items have the selected pattern $(y_{ij})_{j=1}^{r+1}$:

$$f(\boldsymbol{\theta}, \mathbf{y}_i) = \int_{\mathbb{R}^r} \prod_{j=1}^m f_j(\boldsymbol{\theta}_j, y_{ij}|\mathbf{z}_i) d\Phi(\mathbf{z}_i)$$

$$\leq \sum_{j=1}^{r+1} \int_{H_{\mathbf{d}_j}^0(y_{ij})} f_j(\boldsymbol{\theta}_j, y_{ij}|\mathbf{z}_i) d\Phi(\mathbf{z}_i)$$

$$\leq \sum_{j=1}^{r+1} \int_{H_{\mathbf{d}_j}^\varepsilon(y_{ij})} f_j(\boldsymbol{\theta}_j, y_{ij}|\mathbf{z}_i) d\Phi(\mathbf{z}_i) + \sum_{j=1}^{r+1} \Phi\{H_{\mathbf{d}_j}^0(y_{ij}) \backslash H_{\mathbf{d}_j}^\varepsilon(y_{ij})\}$$

$$\leq \sum_{j=1}^{r+1} \frac{1}{1+e^{\varepsilon L_j}} + \sum_{j=1}^{r+1} \Phi\{H_{\mathbf{d}_j}^0(y_{ij}) \backslash H_{\mathbf{d}_j}^\varepsilon(y_{ij})\} \tag{55}$$

In the last line of Eq. 55, each summand of the second term can be made smaller than $\frac{\eta}{2(r+1)}$ by choosing a proper ε; this result can be strengthened to hold uniformly for all directions $\mathbf{d}_j$ on $\mathbb{R}^{r+1}$, as a consequence of Lemma 3. In addition, since there are only finitely many intercept parameters, we can choose a large enough B' (i.e., $\boldsymbol{\theta}$ is sufficiently distant from $\boldsymbol{\theta}_0$) such that $\frac{1}{1+e^{\varepsilon L_j}} < \frac{\eta}{2(r+1)}$ for all j. Consequently, for each $\boldsymbol{\theta}$ satisfying $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > B'$, we are able to find an individual response pattern $\mathbf{y}_i$ such that the corresponding value of Eq. 55 can be bounded by the desired number η, which establishes the result stated by Eq. 50. The two lemmas required in the foregoing proof are presented next.

Lemma 2. *Consider a sequence of affine hyperplanes* $\{\mathbf{z} \in \mathbb{R}^r : \mathbf{a}_i^\top \mathbf{z} = b_i\}_{i=1}^k$. *Let half-space* H_i *be either* $\mathbf{a}_i^\top \mathbf{z} \geq b_i$ *or* $\mathbf{a}_i^\top \mathbf{z} \leq b_i$. *There exists some choice of* $\{H_i\}_{i=1}^k$ *such that* $\mathbb{R}^r \subset \bigcup_{i=1}^k H_i$, *if and only if* $\mathbf{a}_i$*'s are linearly dependent.*

Proof. ($\Leftarrow$) Suppose $\mathbf{a}_i$'s are linearly dependent. There exists an $\mathbf{a}_i$ that can be written as a non-trivial linear combination of the others. Without loss of generality, let $\mathbf{a}_1$ be such a vector:

$$\mathbf{a}_1 = \sum_{i=2}^k c_i \mathbf{a}_i, \tag{56}$$

in which at least one c_i is non-zero. If $\sum_{i=2}^k c_i b_i \geq b_1$, then for $i = 2, \ldots, k$ set $H_i = \{\mathbf{z} : \mathbf{a}_i^\top \mathbf{z} \geq b_i\}$ when $c_i \leq 0$ and $H_i = \{\mathbf{z} : \mathbf{a}_i^\top \mathbf{z} \leq b_i\}$ when $c_i > 0$. It follows that

$$\bigcap_{i=2}^k H_i^c \subset \left\{\mathbf{z} : \sum_{i=2}^k c_i \mathbf{a}_i^\top \mathbf{z} > \sum_{i=2}^k c_i b_i\right\} \subset \{\mathbf{z} : \mathbf{a}_1^\top \mathbf{z} \geq b_1\}. \tag{57}$$

By letting H_1 be the RHS of Eq. 57, we have $\mathbb{R}^r \subset \bigcap_{i=1}^k H_i$. A similar argument can be used to establish the statement when $\sum_{i=2}^k c_i b_i < b_1$.

($\Rightarrow$) Suppose the $\mathbf{a}_i$'s are linearly independent, which implies that the set of Eqs. $\{\mathbf{a}_i^\top \mathbf{z} = b_i\}_{i=1}^k$ has at least one solution, denoted $\mathbf{z}'$. Consider the k-dimensional subspace spanned by the coordinate system $\{\mathbf{a}_i\}_{i=1}^n$ with an origin at $\mathbf{z}'$. For each i, the half-space H_i corresponds to either the positive or negative side of vector $\mathbf{a}_i$, depending on the direction of the inequality. No matter how we choose the H_i's, there will be one out of 2^k "orthants" corresponding to $\bigcap_{i=1}^k H_i^c$ left uncovered, which proves the "only if" part. $\square$

Lemma 3. *Let* $Z_x \sim \mathcal{N}(x, 1-x^2)$ *be a one-parameter family of normal random variables with* $x \in [-1, 1]$. *Given any* $\eta \in (0, 1/2)$, *there exists an* $\varepsilon > 0$ *such that* $\sup_{x \in [-1,1]} P(|Z_x| \leq \varepsilon) < \eta$.

Proof. By symmetry, $\sup_{x \in [0,1]} P(|Z_x| \leq \varepsilon) = \sup_{x \in [-1,1]} P(|Z_x| \leq \varepsilon)$, so we only need to consider non-negative x's in the proof. Note that for all $\varepsilon \in [0, 1)$ and $x > \varepsilon$,

$$P(Z_x \leq \varepsilon) = \Phi\left(\frac{\varepsilon - x}{\sqrt{1-x^2}}\right) \downarrow 0, \tag{58}$$

as $x \uparrow 1$, due to the monotonicity of the functions involved. Now fix an $\eta \in (0, 1/2)$. Equation 58 implies there exists an $x' \in (1/2, 1)$ such that $P(Z_{x'} \leq 1/2) < \eta$. Then for all $x \in (x', 1]$ and $\varepsilon \in (0, 1/2]$, we have

$$P(|Z_x| \leq \varepsilon) \leq P(Z_x \leq \varepsilon) \leq P(Z_{x'} \leq \varepsilon) < \eta. \tag{59}$$

For $x \in [0, x']$, the variance of Z_x is bounded from below by $1 - x'^2$. We select ε' such that $P(|Z_{x'} - x'| \leq \varepsilon') < \eta$. Then by Anderson's inequality,

$$P(|Z_x| \leq \varepsilon') \leq P(|Z_x - x| \leq \varepsilon') \leq P(|Z_{x'} - x'| \leq \varepsilon') < \eta. \tag{60}$$

The statement follows by setting $\varepsilon = \min\{1/2, \varepsilon'\}$.

Region $A_{3,n}$ Let K_{-0} be any compact subset of Θ which is bounded away from $\boldsymbol{\theta}_0$. By a well-known application of Jensen's inequality:

$$E_{\boldsymbol{\theta}_0} \log \frac{f(\boldsymbol{\theta}, \mathbf{Y}_i)}{f(\boldsymbol{\theta}_0, \mathbf{Y}_i)} \leq \log E_{\boldsymbol{\theta}_0} \frac{f(\boldsymbol{\theta}, \mathbf{Y}_i)}{f(\boldsymbol{\theta}_0, \mathbf{Y}_i)} = 0. \tag{61}$$

In fact, the inequality in Eq. 61 is strict by the model identification assumption (ii) of Theorem 1. Because K_{-0} is compact, there exists a positive number κ such that

$$\sup_{\boldsymbol{\theta} \in K_{-0}} E_{\boldsymbol{\theta}_0} \log \frac{f(\boldsymbol{\theta}, \mathbf{Y}_i)}{f(\boldsymbol{\theta}_0, \mathbf{Y}_i)} < -\kappa, \tag{62}$$

by the continuity of the LHS function. Moreover, by the Uniform Law of Large Numbers,

$$\sup_{\boldsymbol{\theta} \in K_{-0}} \left| \frac{1}{n} \sum_{i=1}^{n} \log \frac{f(\boldsymbol{\theta}, \mathbf{Y}_i)}{f(\boldsymbol{\theta}_0, \mathbf{Y}_i)} - E_{\boldsymbol{\theta}_0} \log \frac{f(\boldsymbol{\theta}, \mathbf{Y}_i)}{f(\boldsymbol{\theta}_0, \mathbf{Y}_i)} \right| \stackrel{P_{\boldsymbol{\theta}_0}}{\to} 0. \tag{63}$$

Therefore, $\sup_{\boldsymbol{\theta} \in K_{-0}} \prod_{i=1}^{n} f(\boldsymbol{\theta}, \mathbf{Y}_i) / \prod_{i=1}^{n} f(\boldsymbol{\theta}_0, \mathbf{Y}_i) \stackrel{P_{\boldsymbol{\theta}_0}}{\to} 0$, which implies

$$\sup_{\mathbf{h} \in A_{3,n}} \frac{f_{n,\mathbf{h}}}{f_{n,0}} \stackrel{P_{\boldsymbol{\theta}_0}}{\to} 0, \tag{64}$$

because $\mathbf{h} \in A_{3,n}$ implies $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \in [\delta, B']$. It follows that

$$\int_{A_{3,n}} \left| \frac{b_{n,\mathbf{h}} f_{n,\mathbf{h}}}{f_{n,0}} - b_0 e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \boldsymbol{\mathcal{I}}_0 \mathbf{h}} \right| d\mathbf{h}$$

$$\leq \int_{A_{3,n}} \left| \frac{b_{n,\mathbf{h}} f_{n,\mathbf{h}}}{f_{n,0}} \right| d\mathbf{h} + b_0 \int_{A_{3,n}} e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \boldsymbol{\mathcal{I}}_0 \mathbf{h}} d\mathbf{h}$$

$$\leq \sup_{\mathbf{h} \in A_{3,n}} \left| \frac{f_{n,\mathbf{h}}}{f_{n,0}} \right| \int_{A_{3,n}} b_{n,\mathbf{h}} d\mathbf{h} + b_0 \int_{A_{3,n}} e^{\mathbf{h}^\top \mathbf{S}_n - \frac{1}{2}\mathbf{h}^\top \boldsymbol{\mathcal{I}}_0 \mathbf{h}} d\mathbf{h}.$$

$$\tag{65}$$

Equation 65 converges in probability to 0 due to the integrability of $b_{n,\mathbf{h}}$, the tail estimates of a multivariate normal distribution, and the tightness of $\mathbf{S}_n$. The proof is now complete.

Appendix 3: Proof of Theorem 2

Recall that $\mathbf{V}$ has density $g_n(\boldsymbol{\theta}|\mathbf{y})$ conditional on $D(\mathbf{y})$. Take $\delta > 0$. For each fixed $\mathbf{y}$, $\rho_K(\mathbf{y})$, defined by Eq. 22, can be bounded by

$$
\begin{aligned}
\rho_K(\mathbf{y}) &= P\{\operatorname{diam} Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) > K/n \mid D(\mathbf{y})\} \\
&= P\{\operatorname{diam} Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) > K/n, \|\mathbf{V} - \boldsymbol{\theta}_0\| \le \delta \mid D(\mathbf{y})\} \\
&\quad + P\{\operatorname{diam} Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) > K/n, \|\mathbf{V} - \boldsymbol{\theta}_0\| > \delta \mid D(\mathbf{y})\} \\
&\le P\{\operatorname{diam} Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) > K/n \mid \|\mathbf{V} - \boldsymbol{\theta}_0\| \le \delta, D(\mathbf{y})\} \\
&\quad + P\{\|\mathbf{V} - \boldsymbol{\theta}_0\| > \delta \mid D(\mathbf{y})\}.
\end{aligned}
\tag{66}
$$

Theorem 1 implies that for $\mathbf{Y}$ generated from $P_{\boldsymbol{\theta}_0}$, $P\{\|\mathbf{V} - \boldsymbol{\theta}_0\| > \delta \mid D(\mathbf{Y})\}$, as a measurable function of $\mathbf{Y}$, converges to 0 in $P_{\boldsymbol{\theta}_0}$-probability: i.e.,

$$
P\{\|\mathbf{V} - \boldsymbol{\theta}_0\| > \delta \mid D(\mathbf{Y})\} = \int_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > \delta} g_n(\boldsymbol{\theta}|\mathbf{Y})d\boldsymbol{\theta} \xrightarrow{P_{\boldsymbol{\theta}_0}} 0.
\tag{67}
$$

Hence, we focus on the first term in Eq. 66. This term can be further bounded by

$$
\begin{aligned}
&P\{\operatorname{diam} Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) > K/n \mid \|\mathbf{V} - \boldsymbol{\theta}_0\| \le \delta, D(\mathbf{y})\} \\
&= \sum_I P\{\operatorname{diam} Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) > K/n \mid \|\mathbf{V} - \boldsymbol{\theta}_0\| \le \delta, \mathbf{V} = \mathbf{V}_I, D(\mathbf{y})\} \\
&\quad \cdot P\{\mathbf{V} = \mathbf{V}_I \mid \|\mathbf{V} - \boldsymbol{\theta}_0\| \le \delta, D(\mathbf{y})\} \\
&= \sum_{\mathbf{y}_I} P\{\operatorname{diam} Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) > K/n \mid \|\mathbf{V} - \boldsymbol{\theta}_0\| \le \delta, \mathbf{V} = \mathbf{V}_I, D(\mathbf{y})\} \\
&\quad \cdot \left(\sum_{I': \mathbf{y}_{I'} = \mathbf{y}_I} P\{\mathbf{V} = \mathbf{V}_{I'} \mid \|\mathbf{V} - \boldsymbol{\theta}_0\| \le \delta, D(\mathbf{y})\} \right) \\
&\le \sum_{\mathbf{y}_I} P\{\operatorname{diam} Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) > K/n \mid \|\mathbf{V}_I - \boldsymbol{\theta}_0\| \le \delta, \mathbf{V} = \mathbf{V}_I, D(\mathbf{y})\} \\
&= \sum_{\mathbf{y}_I} \int P\{\operatorname{diam} Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) > K/n \mid \|\mathbf{V}_I - \boldsymbol{\theta}_0\| \le \delta, \mathbf{V} = \mathbf{V}_I, D(\mathbf{y}), \mathbf{Z}_I^\star = \mathbf{z}_I\} d\Phi(\mathbf{z}_I).
\end{aligned}
\tag{68}
$$

The first sum over index sets I in the second line of Eq. 68 can be collapsed into a finite sum over all patterns of $\mathbf{y}_I$ in the third line, for the reason that sub-samples I and I' having the same response pattern $\mathbf{y}_I = \mathbf{y}_{I'}$ are exchangeable. Note that the event being conditioned on in the integrand of the last line of Eq. 68 happens with a positive probability almost surely under the probability measure of $\mathbf{Z}^\star$; to simplify notation, write $E_{\mathbf{y}_I}^\delta(\mathbf{z}_I) = \{\|\mathbf{V}_I - \boldsymbol{\theta}_0\| \le \delta, \mathbf{V} = \mathbf{V}_I, D(\mathbf{y}), \mathbf{Z}_I^\star = \mathbf{z}_I\}$ as that event. Because there are only finitely many patterns of $\mathbf{y}_I$, it suffices to prove that for each $\varepsilon > 0$ and some $\delta > 0$,

$$
P_{\boldsymbol{\theta}_0} \left\{ \exists K, N > 0 : \int P\{\operatorname{diam} Q(\mathbf{Y}, \mathbf{A}^\star, \mathbf{Z}^\star) > K/n \mid E_{\mathbf{y}_I}^\delta(\mathbf{z}_I)\} d\Phi(\mathbf{z}_I) < \varepsilon, \ \forall n > N \right\} \to 1.
\tag{69}
$$

So fix $\mathbf{y}_I$ and δ for the rest of the proof. Also note that conditional on $E_{\mathbf{y}_I}^{\delta}(\mathbf{z}_I)$, the remaining observations $i \notin I$ are independent.

To proceed, we sequentially project the set inverse $Q(\mathbf{y}, \mathbf{A}^{\star}, \mathbf{Z}^{\star})$ onto m subspaces, each of which is spanned by the q_j free parameters for item j. For each projection, we find a bounding random variable for its diameter; then, the sum of constructed bounds across all projections serves as a upper bound, up to a constant multiplier depending on the dimension of the parameter space, for the diameter of the set inverse. We prove the result stated in Eq. 69 with the diameter of $Q(\mathbf{y}, \mathbf{A}^{\star}, \mathbf{Z}^{\star})$ replaced by the constructed bound. In order to establish the desired property for the bounding variables, we allocate the rest observations (i.e., not in I) to each projection, and subsequently use the standard theory for order statistics of i.i.d. random variables. In particular, we rearrange those observations to fill a growing two-dimensional array indexed by a pair of indices s and j: The second dimension of the array, $j = 1, \ldots, m$, is filled first, then the first one; therefore, the first dimension indexed by $s = \lfloor i/m \rfloor, i = 1, \ldots, n$, grows as the sample size increases. Notationally, elements corresponding to an observation in the array are denoted by a subscript $[sj]$.[8]

Fix $\mathbf{V} = \mathbf{V}_I = \boldsymbol{\theta}$ for now. For each item j, let $\tilde{\boldsymbol{\beta}}_j$ be the collection of the r_j free slopes. Now intersecting the half-space of a new observation $[sj]$ in the two-dimensional array with those of observations I_j, the resulting intersection on the subspace of $\boldsymbol{\theta}_j$ can be either bounded (i.e., a simplex) or unbounded. The following lemma provides sufficient and necessary conditions for the (un)bounded case:

Lemma 4. *Consider $p + 1$ half-spaces: $H_i = \{\mathbf{x} \in \mathbb{R}^p : \mathbf{n}_i^{\top} \mathbf{x} \le b_i\}, i = 1, \ldots, p + 1$, in which $\mathbf{n}_i$'s are considered fixed. Then, the following statements are equivalent:*
(i) $\bigcap_{i=1}^{p+1} H_i$ *is bounded for all choices of b_i's, $i = 1, \ldots, p + 1$, such that the intersection is not empty;*
(ii) $\bigcap_{i=1}^{p+1} H_i$ *is a bounded simplex for some choices of b_i's, $i = 1, \ldots, p + 1$;*
(iii) *For all $\mathbf{c} \in \mathbb{R}^p$, there exists $i \in \{1, \ldots, p + 1\}$ such that $\mathbf{n}_i^{\top} \mathbf{c} > 0$;*
(iv) *There exists $i \in \{1, \ldots, p + 1\}$ such that $\mathbf{n}_j$'s, $j \ne i$, are linear independent, and that $\mathbf{n}_i = -\sum_{j \ne i} \gamma_j \mathbf{n}_j$ with $\gamma_j > 0$ for all $j \ne i$.*

Proof. (i) $\Rightarrow$ (ii). We can always make the intersection non-empty by choosing $b_i > 0$ for all $i = 1, \ldots, p + 1$. In this case, $\bigcap_{i=1}^{p+1} H_i$ must contain some neighborhood of $\mathbf{0}$. So (i) $\Rightarrow$ (ii) is trivial.

(ii) $\Rightarrow$ (iii). Fix b_i's, $i = 1, \ldots, p + 1$, such that $\bigcap_{i=1}^{p+1} H_i$ is a bounded simplex. Take $\mathbf{x}_0 \in \bigcap_{i=1}^{p+1} H_i$; i.e., $\mathbf{n}_i^{\top} \mathbf{x}_0 \le b_i$ for all $i = 1, \ldots, p + 1$. If there exists $\mathbf{c} \in \mathbb{R}^p$ such that $\mathbf{c}^{\top} \mathbf{n}_i \le 0$ for all i, then $\mathbf{n}_i^{\top} \mathbf{x}_0 + \lambda \mathbf{n}_i^{\top} \mathbf{c} \le b_i$ for all i and all $\lambda > 0$. This implies $\mathbf{x}_0 + \lambda \mathbf{c} \in \bigcap_{i=1}^{p+1} H_i$ for all $\lambda > 0$, which contradicts the boundedness.

(iii) $\Rightarrow$ (i). On each direction $\mathbf{c}$, choose i such that $\mathbf{n}_i^{\top} \mathbf{c} > 0$. For every possible value of the corresponding b_i, there exists some $\lambda_0 > 0$ such that for all $\lambda > \lambda_0$, $\mathbf{n}_i^{\top}(\lambda \mathbf{c}) > b_i$, i.e., $\lambda \mathbf{c}_i \notin H_i$. So $\bigcap_{i=1}^{p+1} H_i$ is always bounded.

(iii) $\Rightarrow$ (iv). Let C_i be the convex cone defined by all but the ith normal vectors. (iii) implies $-\mathbf{n}_i^{\top} \mathbf{c} < 0$ for all $\mathbf{c} \in C_i^N = \{\mathbf{c} : \mathbf{n}_i^{\top} \mathbf{c} \le 0, \text{ for all } j \ne i\}$, i.e., the normal cone (denoted by a superscript N) of C_i. Hence, $-\mathbf{n}_i \in (C_i^N)^N = C_i$.

(iv) $\Rightarrow$ (iii). For $\mathbf{c} \in C_i^N$, (iv) implies $\mathbf{n}_i^{\top} \mathbf{c} > 0$. For $\mathbf{c} \notin C_i^N$, there exists some $j \ne i$ such that $\mathbf{n}_j^{\top} \mathbf{c} > 0$. □

---

[8] For example, if $I = \{1, \ldots, \sum_{j=1}^m q_j\}$ is the first $\sum_{j=1}^m q_j$ observations in the sample, then $[sj]$ corresponds to the observation $i = \sum_{j=1}^m q_j + (s - 1)m + j$.
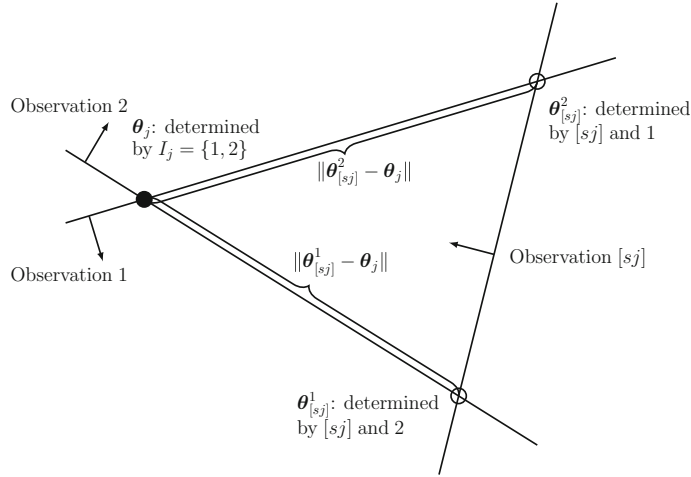
FIGURE 5.

Illustration of notation used in the proof of Theorem 2. Here, $r = 1$, and j is fixed. $I_j = \{1, 2\}$, which determines the fixed vertex $\boldsymbol{\theta}_j$ (shown as a *dot*). The line corresponding to the new observation $[sj]$ intersects with those of observations 1 and 2, respectively, and produces two new vertices $\boldsymbol{\theta}_{[sj]}^2$ and $\boldsymbol{\theta}_{[sj]}1$ (shown as *circles*). The sum of $\|\boldsymbol{\theta}_{[sj]}^1 - \boldsymbol{\theta}_j\|$ and $\|\boldsymbol{\theta}_{[sj]}^2 - \boldsymbol{\theta}_j\|$ gives an upper bound of the diameter of the plotted *triangle*.

Let $\tilde{\mathbf{z}}_{ij}$ be the elements of $\mathbf{z}_i$ associated with $\tilde{\boldsymbol{\beta}}_j$. For each $i \in I_j$, write $\mathbf{n}_{ij} = \omega_{ij}(1 \ \tilde{\mathbf{z}}_{ij}^\top)^\top$ as the normal vector of the corresponding $(r_j + 1)$-dimensional half-space, in which $\omega_{ij} = \pm 1$ is determined by the item response y_{ij}. Similar notation is defined for observations in the array: Let $\tilde{\mathbf{Z}}_{[sj]}^\star$ be the elements of $\mathbf{Z}_{[sj]}^\star$ associated with $\tilde{\boldsymbol{\beta}}_j$, and $\mathbf{N}_{[sj]}^\star = \omega_{[sj]}(1 \ \tilde{\mathbf{Z}}_{[sj]}^{\star\top})^\top$ be the corresponding (random) normal vector; the random variable $\omega_{[sj]} = \pm 1$ depends on this observation's response to item j, which is denoted $y_{[sj]}$ for simplicity. For each j, Lemma 4 implies that observation $[sj]$ produces a bounded intersection if there exist positive real numbers $\gamma_i, i \in I_j$, such that

$$\omega_{[sj]}\tilde{\mathbf{Z}}_{[sj]}^\star = -\sum_{i \in I_j} \gamma_i \omega_{ij} \tilde{\mathbf{z}}_{ij}, \tag{70}$$

and

$$\omega_{[sj]} = -\sum_{i \in I_j} \gamma_i \omega_{ij}. \tag{71}$$

Conditioning on $\mathbf{V}_j = \mathbf{V}_{I_j} = \boldsymbol{\theta}_j$, the intersection cannot be empty, which introduces a truncation to $A_{[sj]}^\star$, i.e., the associated logistic variate for observation $[sj]$ and item j:

$$(-1)^{y_{[sj]}}(A_{[sj]}^\star - \alpha_j - \tilde{\boldsymbol{\beta}}_j^\top \tilde{\mathbf{Z}}_{[sj]}^\star) \geq 0. \tag{72}$$

Fix j. When Eqs. 70 and 71 hold, let $\boldsymbol{\theta}_{[sj]}^i = (\alpha_j^i \ \boldsymbol{\beta}_j^{i\top})^\top, i \in I_j$, be the vertex on the subspace of $\boldsymbol{\theta}_j$ determined by observations $I_j \setminus \{i\}$ together with the new observation $[sj]$, which is random due to its dependency on $A_{[sj]}^\star$ and $\mathbf{Z}_{[sj]}^\star$. Also let $I_j^i = I_j \setminus \{i\}$ for $i \in I_j$, and treat $\tilde{\mathbf{z}}_{I_j^i} = (\tilde{\mathbf{z}}_{ij})_{i \in I_j^i}$ as an $r_j \times r_j$ matrix throughout this part of derivation. A geometric illustration of these notations for $r = 1$ is shown in Figure 5.

Applying the formula for inverting a partitioned matrix, we have

$$
\begin{pmatrix} \tilde{\mathbf{z}}_{I_j^i} & \mathbf{1} \\ \tilde{\mathbf{Z}}_{[sj]}^{\star\top} & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \tilde{\mathbf{z}}_{I_j^i}^{-1} + \dfrac{\tilde{\mathbf{z}}_{I_j^i}^{-1}\mathbf{1}\tilde{\mathbf{Z}}_{[sj]}^{\star\top}\tilde{\mathbf{z}}_{I_j^i}^{-1}}{1 - \tilde{\mathbf{Z}}_{[sj]}^{\star\top}\tilde{\mathbf{z}}_{I_j^i}^{-1}\mathbf{1}} & \dfrac{-\tilde{\mathbf{z}}_{I_j^i}^{-1}\mathbf{1}}{1 - \tilde{\mathbf{Z}}_{[sj]}^{\star\top}\tilde{\mathbf{z}}_{I_j^i}^{-1}\mathbf{1}} \\[6pt] \dfrac{-\tilde{\mathbf{Z}}_{[sj]}^{\star\top}\tilde{\mathbf{z}}_{I_j^i}^{-1}}{1 - \tilde{\mathbf{Z}}_{[sj]}^{\star\top}\tilde{\mathbf{z}}_{I_j^i}^{-1}\mathbf{1}} & \dfrac{1}{1 - \tilde{\mathbf{Z}}_{[sj]}^{\star\top}\tilde{\mathbf{z}}_{I_j^i}^{-1}\mathbf{1}} \end{pmatrix}. \tag{73}
$$

It follows that the elements of $\boldsymbol{\theta}_{[sj]}^i - \boldsymbol{\theta}_j$ can be expressed as following:

$$
\tilde{\boldsymbol{\beta}}_{[sj]}^i - \tilde{\boldsymbol{\beta}}_j = \frac{\tilde{\mathbf{z}}_{I_j^i}^{-1}\mathbf{1}(A_{[sj]}^\star - \tilde{\boldsymbol{\beta}}_j^\top \tilde{\mathbf{Z}}_{[sj]}^\star - \alpha_j)}{\tilde{\mathbf{Z}}_{[sj]}^{\star\top}\tilde{\mathbf{z}}_{I_j^i}^{-1}\mathbf{1} - 1}, \tag{74}
$$

and

$$
\alpha_{[sj]}^i - \alpha_j = -\tilde{\mathbf{z}}_{ij}^\top(\tilde{\boldsymbol{\beta}}_{[sj]}^i - \tilde{\boldsymbol{\beta}}_j), \text{ for all } i \in I_j^i. \tag{75}
$$

Define

$$
\bar{U}_{[sj]}^\star = \left[\sum_{i \in I_j} \frac{\left\| \tilde{\mathbf{z}}_{I_j^i}^{-1}\mathbf{1} \right\| \left(1 + \sum_{i' \in I_j^i} \|\tilde{\mathbf{z}}_{i'j}\|^2 \right)}{\left| \tilde{\mathbf{Z}}_{[sj]}^{\star\top}\tilde{\mathbf{z}}_{I_j^i}^{-1}\mathbf{1} - 1 \right|} \right] |A_{[sj]}^\star - \tilde{\boldsymbol{\beta}}_j^\top \tilde{\mathbf{Z}}_{[sj]}^\star - \alpha_j| \tag{76}
$$

If both Eqs. 70 and 71 are satisfied, the random variable defined by Eq. 76 gives an upper bound for $\|\boldsymbol{\theta}_{[sj]}^i - \boldsymbol{\theta}\|$. Also define

$$
U_{[sj]}^\star = \begin{cases} \bar{U}_{[sj]}^\star, & \text{if Eqs. 70 and 71 hold;} \\ \infty, & \text{otherwise.} \end{cases} \tag{77}
$$

which is a random variable that is defined on the extended real line.

Pooling across all observations in the array, we have

$$
\operatorname{diam} Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) \le C \sum_{j=1}^m \min_{t \le s} U_{[tj]}^\star, \tag{78}
$$

in which C is a constant determined by the dimension of the parameter space. It follows that

$$
\begin{aligned}
&\int P\{\operatorname{diam} Q(\mathbf{y}, \mathbf{A}^\star, \mathbf{Z}^\star) > K/n \mid E_{\mathbf{y}_I}^\delta(\mathbf{z}_I)\} d\Phi(\mathbf{z}_I) \\
&\le \int P\left\{ \sum_{j=1}^m \min_{t \le s} U_{[tj]}^\star > \frac{K}{Cn} \,\middle|\, E_{\mathbf{y}_I}^\delta(\mathbf{z}_I) \right\} d\Phi(\mathbf{z}_I) \\
&\le \sum_{j=1}^m \int P\left\{ \min_{t \le s} U_{[tj]}^\star > K'/n \,\middle|\, E_{\mathbf{y}_I}^\delta(\mathbf{z}_I) \right\} d\Phi(\mathbf{z}_{I_j}),
\end{aligned} \tag{79}
$$

in which $K' = \frac{K}{Cm}$. Now fix $\varepsilon, \delta > 0$. It suffices to prove for each summand of Eq. 79:

$$P_{\theta_0}\left\{\exists K', N > 0 : \int P\left\{\min_{t \le s} U^\star_{[tj]} > K'/n \mid E^\delta_{\mathbf{y}_I}(\mathbf{z}_I)\right\} d\Phi(\mathbf{z}_I) < \varepsilon, \ \forall n > N\right\} \to 1. \quad (80)$$

For each item j, define $T^k_{sj} = \{t : t \le s, y_{[tj]} = k\}$ for $k = 0$ and 1, respectively. We intend to prove that the sub-collections satisfy

$$|T^k_{sj}|/n \xrightarrow{P_{\theta_0}} \rho_k, \ \text{as } n \to \infty \text{ for some } 0 < \rho_k < 1. \quad (81)$$

In this case, we write $\varrho = \min\{\rho_0, \rho_1\}$. Within each sub-collection, $U^\star_{[tj]}$, $t \in T^k_{sj}$, are i.i.d. conditional on $E^\delta_{\mathbf{y}_I}(\mathbf{z}_I)$; let $\varphi_j(u, \boldsymbol{\theta}_j, \mathbf{y}_{I_j}, \tilde{\mathbf{z}}_{I_j}, k)$ be its corresponding (conditional) density. We also intend to find a set $B_j \subset \mathbb{R}^{r_j^2}$ such that $P\{\tilde{\mathbf{Z}}^\star_{I_j} \notin B_j\} < \varepsilon/2$, and also a $\kappa > 0$ such that for every $\tilde{\mathbf{z}}_{I_j} \in B_j$, there exists a particular $y_{[sj]} = k$, for which

$$\inf\left\{\varphi_j(u, \boldsymbol{\theta}_j, \mathbf{y}_{I_j}, \tilde{\mathbf{z}}_{I_j}, k) : 0 \le u \le \eta, \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le \delta\right\} \ge \kappa \quad (82)$$

for some $\eta > 0$. Assume for a moment that Eqs. 81 and 82 hold. Then we can construct a sequence of i.i.d. non-negative random variables $\{X_n\}$, whose density function is constantly equal to κ within $[0, \eta]$. By the Delta method and the standard result for i.i.d. uniform order statistics, $n \min_{i \le n} X_i \xrightarrow{d} W/\kappa$, in which $W \sim \text{Exp}(1)$. Fix K' such that $P\left(W/\kappa > K'\right) < \varepsilon/8$. By the Portmanteau Lemma, there exists an n_1 such that for all $n > n_1$, $P\{n \min_{i \le \lfloor \varrho n/2 \rfloor} X_i > K'\} \le P\{W/\kappa > K'\} + \varepsilon/8 \le \varepsilon/4$. Also take n_2 such that $K'/n_2 < \eta$, and n_3 such that $P\{\min_{k=0,1} |T^k_{sj}|/n < \varrho/2\} \le \varepsilon/4$. Thus, for every $\mathbf{z}_{I_j} \in B_j$, there exists $k = 0$ or 1 such that along the corresponding subsequence T^k_{sj}:

$$P\left\{\min_{t \le s} U^\star_{[tj]} > K'/n \mid E^\delta_{\mathbf{y}_I}(\mathbf{z}_I)\right\} \le P\left\{\min_{t \in T^k_{sj}} U^\star_{[tj]} > K'/n \mid E^\delta_{\mathbf{y}_I}(\mathbf{z}_I)\right\}$$

$$\le P\left\{\min_{i \le \lfloor \varrho n/2 \rfloor} X_i > K'/n\right\} + \varepsilon/4 \le \varepsilon/2 \quad (83)$$

for all $n > \max\{n_1, n_2, n_3\}$. It follows that for all these large n's,

$$\int P\left\{\min_{t \le s} U^\star_{[tj]} > K'/n \mid E^\delta_{\mathbf{y}_I}(\mathbf{z}_I)\right\} d\Phi(\mathbf{z}_{I_j})$$

$$\le \int_{B_j} P\left\{\min_{t \le s} U^\star_{[tj]} > K'/n \mid E^\delta_{\mathbf{y}_I}(\mathbf{z}_I)\right\} d\Phi(\mathbf{z}_{I_j}) + \varepsilon/2 \le \varepsilon, \quad (84)$$

This implies the intended results (Eq. 80).

When $\mathbf{Y}$ is considered random, in fact, the probability that Eq. 81 holds for both $k = 0$ and 1 goes to 1, because the data-generating parameter values $\boldsymbol{\theta}_0$ are assumed to be in the interior of the parameter space (and thus $\varrho > 0$ is determined solely by $\boldsymbol{\theta}_0$).

Let $\bar{\varphi}_j(u, \boldsymbol{\theta}_j, \tilde{\mathbf{z}}_{I_j})$ be the density of $\bar{U}_{[sj]}$ conditional on $E_{\mathbf{y}_I}^{\delta}(\mathbf{z}_I)$, and the event:

$$C_j(\mathbf{y}_{I_j}, \tilde{\mathbf{z}}_{I_j}, y_{[sj]}) = \{\omega_{[sj]}\tilde{\mathbf{Z}}_{[sj]}^{\star} = -\sum_{i \in I_j} \gamma_i \omega_{ij}\tilde{\mathbf{z}}_{ij},$$

$$\omega_{[sj]} = -\sum_{i \in I_j} \gamma_i \omega_{ij},$$

$$\gamma_i > 0 \text{ for all } i \in I_j\}. \tag{85}$$

Then, $\varphi_j(u, \boldsymbol{\theta}_j, \mathbf{y}_{I_j}, \tilde{\mathbf{z}}_{I_j}, y_{[sj]}) = \bar{\varphi}_j(u, \boldsymbol{\theta}_j, \tilde{\mathbf{z}}_{I_j})P\{C_j(\mathbf{y}_{I_j}, \tilde{\mathbf{z}}_{I_j}, y_{[sj]})|E_{\mathbf{y}_I}^{\delta}(\mathbf{z}_I)\}$. Next, we find proper lower bounds for the two parts on the RHS, respectively, which subsequently establishes Eq. 82.

First, fix a $y_{[sj]}$ ensuring Eqs. 70 and 71. For easy reference, let

$$\sigma_j(\tilde{\mathbf{z}}_{I_j}, \tilde{\mathbf{Z}}_{[sj]}^{\star}) = \sum_{i \in I_j} \frac{\left\|\tilde{\mathbf{z}}_{I_j^i}^{-1}\mathbf{1}\right\| \left(1 + \sum_{i' \in I_j^i} \|\tilde{\mathbf{z}}_{i'j}\|^2\right)}{\left|\tilde{\mathbf{Z}}_{[sj]}^{\star \top}\tilde{\mathbf{z}}_{I_j^i}^{-1}\mathbf{1} - 1\right|} \tag{86}$$

and

$$\mu_j(\boldsymbol{\theta}_j, \tilde{\mathbf{Z}}_{[sj]}^{\star}) = \tilde{\boldsymbol{\beta}}_j^{\top}\tilde{\mathbf{Z}}_{[sj]}^{\star} + \alpha_j. \tag{87}$$

Then we rewrite Eq. 76 as $\bar{U}_{[sj]} = \sigma_j(\tilde{\mathbf{z}}_{I_j}, \tilde{\mathbf{Z}}_{[sj]}^{\star})|A_{[sj]}^{\star} - \mu_j(\boldsymbol{\theta}_j, \tilde{\mathbf{Z}}_{[sj]}^{\star})|$, whose density function is

$$\bar{\varphi}_j(u, \boldsymbol{\theta}_j, \tilde{\mathbf{z}}_{I_j}) = \int \frac{\bar{\psi}(\mu_j(\boldsymbol{\theta}_j, \tilde{\mathbf{z}}_{[sj]}) + u/\sigma_j(\tilde{\mathbf{z}}_{I_j}, \tilde{\mathbf{z}}_{[sj]}))}{\sigma_j(\tilde{\mathbf{z}}_{I_j}, \tilde{\mathbf{z}}_{[sj]})}d\Phi(\tilde{\mathbf{z}}_{[sj]})$$

$$+ \int \frac{\bar{\psi}(\mu_j(\boldsymbol{\theta}_j, \tilde{\mathbf{z}}_{[sj]}) - u/\sigma_j(\tilde{\mathbf{z}}_{I_j}, \tilde{\mathbf{z}}_{[sj]}))}{\sigma_j(\tilde{\mathbf{z}}_{I_j}, \tilde{\mathbf{z}}_{[sj]})}d\Phi(\tilde{\mathbf{z}}_{[sj]}) \tag{88}$$

in which $\bar{\psi}(\cdot)$ is the standard logistic density conditional on Eq. 72. By the theory of multivariate normal random variables, we can find

$$B_j^1 = \{\tilde{\mathbf{z}}_{I_j} \in \mathbb{R}^{r_j^2} : \lambda \le \|\tilde{\mathbf{z}}_{ij}\| \le L, \text{ for all } i \in I_j\} \tag{89}$$

with properly defined λ and L such that $P\{\tilde{\mathbf{Z}}_{I_j}^{\star} \in B_j^1\} > 1 - \varepsilon/4$. Also for fixed $D' > 0$ and $D > \delta > 0$, define

$$G_j(\tilde{\mathbf{z}}_{I_j}) = \{\tilde{\mathbf{z}}_{[sj]} \in \mathbb{R}^{r_j} : \delta' \le \left|\tilde{\mathbf{z}}_{[sj]}^{\top}\tilde{\mathbf{z}}_{I_j^i}^{-1}\mathbf{1} - 1\right| \le D' \text{ for all } i \in I_j, \|\tilde{\mathbf{z}}_{[sj]}\| \le D\}. \tag{90}$$

Note that $\tilde{\mathbf{Z}}_{[sj]}^{\star \top}\tilde{\mathbf{z}}_{I_j^i}^{-1}\mathbf{1} - 1 \sim \mathcal{N}(-1, \mathbf{1}^{\top}\tilde{\mathbf{z}}_{I_j^i}^{-\top}\tilde{\mathbf{z}}_{I_j^i}^{-1}\mathbf{1})$, in which the variance is uniformly bounded from above and below for all $\tilde{\mathbf{z}}_{I_j} \in B_j^1$. It follows that

$$\inf_{\tilde{\mathbf{z}}_{I_j} \in B_j^1} P\{G_j(\tilde{\mathbf{z}}_{I_j})\} > 0. \tag{91}$$

Thus, by restricting the integrals on the RHS of Eq. 88 to $G_j(\tilde{\mathbf{z}}_{I_j})$, we are able to obtain an uniform lower bound of $\bar{\varphi}_j(u, \boldsymbol{\theta}_j, \tilde{\mathbf{z}}_{I_j})$ for all $\tilde{\mathbf{z}}_{I_j} \in B_j^1$.

Our final task is to find $B_j^2 \subset \mathbb{R}^{r_j^2}$ such that $P\{\tilde{\mathbf{Z}}_{I_j}^\star \in B_j^2\} > 1 - \varepsilon/4$, and that $P\{C_j(\mathbf{y}_{I_j}, \tilde{\mathbf{z}}_{I_j}, k) \mid E_{\mathbf{y}_I}^\delta(\mathbf{z}_I)\}$ has a uniform lower bound for all $\tilde{\mathbf{z}}_{I_j} \in B_j^2$. Here, we only prove the statement for $r = 1$, and we conjecture that an extended argument can be established for $r > 1$. When $r = 1$, $|I_j| = 2$; without loss of generality, let I_j be the first two observations. We fix j, and for simplicity denote the two normal vectors corresponding to the first two observations by $\mathbf{n}_1 = \omega_1(1\ z_1)^\top$ and $\mathbf{n}_2 = \omega_2(1\ z_2)^\top$, in which $\omega_1, \omega_2 = \pm 1$. We now discuss two cases with different combinations of ω_1 and ω_2; in either case, the joint probability of $Z_{[sj]}^\star = -\gamma_1\omega_1 z_1 - \gamma_2\omega_2 z_2$ and $1 = -\gamma_1\omega_1 - \gamma_2\omega_2$, $\gamma_1, \gamma_2 > 0$, is uniformly bounded from below for $(z_1\ z_2)^\top \in B_j^2 = \{(z_1\ z_2)^\top : |z_1 - z_2| \geq \eta, |z_1| \leq H, |z_2| \leq H\}$ with properly selected $\eta, H > 0$ such that $P\{\tilde{\mathbf{Z}}_{I_j}^\star \in B_j^2\} > 1 - \varepsilon/4$.

Case 1 $\omega_1 = 1$ *and* $\omega_2 = 1$. Choose $\omega_{[sj]} = -1$, which happens with positive probability provided the data-generating parameter values are in the interior of the parameter space. Then, $\mathbf{N}_{[sj]}^\star = -\gamma_1\mathbf{n}_1 - \gamma_2\mathbf{n}_2$ implies $\gamma_1 + \gamma_2 = 1$ and $Z_{[sj]}^\star = -\gamma_1 z_1 - \gamma_2 z_2$, i.e., $Z_{[sj]}^\star$ falls in the line segment between $-z_1$ and $-z_2$. For all $(z_1\ z_2)^\top \in B_j^2$, $P\{\min\{-z_1, -z_2\} \leq Z_{[sj]}^\star \leq \max\{-z_1, -z_2\}\} > \Phi(H) - \Phi(H + \eta)$.

Case 2 $\omega_1 = 1$ *and* $\omega_2 = -1$. In this case, the constraints are $\omega_{[sj]} = -\gamma_1 + \gamma_2$ and $\omega_{[sj]}Z_{[sj]}^\star = -\gamma_1 z_1 + \gamma_2 z_2$. If we choose $\omega_{[sj]} = 1$, then $\gamma_2 = 1 + \gamma_1$. It follows that $Z_{[sj]}^\star = \gamma_1(z_2 - z_1) + z_2$, which is greater than z_2 when $z_2 > z_1$ and less than z_2 when $z_2 < z_1$. Then, both $P\{Z_{[sj]}^\star < z_2\}$ and $P\{Z_{[sj]}^\star > z_2\}$ are uniformly greater than $1 - \Phi(L)$ for all $(z_1\ z_2)^\top \in B_j^2$. A similar argument applies to the case when $\omega_{[sj]} = -1$ is chosen.

The remaining combinations of ω_1 and ω_2 are reflections of the two cases having been discussed. Altogether we have shown that $P\{C_j(\mathbf{y}_{I_j}, \tilde{\mathbf{z}}_{I_j}) | E_{\mathbf{y}_I, \mathbf{k}_I}^\delta(\mathbf{z}_I)\}$ is uniformly bounded from below for $\tilde{\mathbf{z}}_{I_j} \in B_j^2$. Take $B_j = B_j^1 \cap B_j^2$; $P\{\tilde{\mathbf{Z}}_{I_j}^\star \in B_j\} > 1 - \varepsilon/2$. Then, the proof is now complete for $r = 1$.

References

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational and Behavioral Statistics*, *17*(3), 251–269.

Asparouhov, T., & Muthén, B. (2012). Comparison of computational methods for high dimensional item factor analysis. Unpublished manuscript retrieved from www.statmodel.com.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B: Methodological, 57*(1), 289–300

Benjamini, Y., & Yekutieli, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. *Journal of the American Statistical Association*, *100*(469), 71–81.

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, *65*, 676–696.

Birnbaum, A. (1968). Some latent train models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395–479). Reading, MA: Addison-Wesley.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*(4), 443–459.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika*, *35*(2), 179–197.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*(1), 111–150.

Cai, L. (2008). SEM of another flavour: Two new applications of the supplemented EM algorithm. *British Journal of Mathematical and Statistical Psychology*, *61*(2), 309–329.

Cai, L. (2010a). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika*, *75*(1), 33–57.

Cai, L. (2010b). Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*, *35*(3), 307–335.

Crawford, C. B., & Ferguson, G. A. (1970). A general rotation criterion and its use in orthogonal rotation. *Psychometrika*, *35*(3), 321–332.

Dempster, A. P. (1968). A generalization of bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, *20*(2), 205–247.

Dempster, A. P. (2008). The dempster-shafer calculus for statisticians. *International Journal of Approximate Reasoning*, *48*(2), 365–377.

Edwards, M. C. (2010). A Markov chain Monte Carlo approach to confirmatory item factor analysis. *Psychometrika*, *75*(3), 474–497.

Eysenck, S. B., Eysenck, H. J., & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and individual differences*, *6*(1), 21–29.

Fisher, R. A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society*, *26*, 528–535.

Fisher, R. A. (1933). The concepts of inverse probability and fiducial probability referring to unknown parameters. *Proceedings of the Royal Society of London. Series A*, *139*(838), 343–348.

Fisher, R. A. (1935). The fiducial argument in statistical inference. *Annals of Eugenics*, *6*(4), 391–398.

Fraser, D. A. S. (1968). *The structure of inference*. New York: John Wiley & Sons.

Ghosh, J. K., & Ramamoorthi, R. (2003). *Bayesian nonparametrics*. New York: Springer.

Gunsjö, A. (1994). *Faktoranalys av ordinala variabler. Studia statistica Upsaliensia*. Stockholm, Sweden: Acta Universitatis Upsaliensis.

Haberman, S. J. (2006). Adaptive quadrature for item response models. *ETS Research Report Series*, *2006*(2), 1–10.

Haberman, S. J. (2013). A general program for item-response analysis that employs the stabilized Newton-Raphson algorithm. *ETS Research Report Series*, *2013*(2), 1–98.

Hannig, J. (2009). On generalized fiducial inference. *Statistica Sinica*, *19*(2), 491.

Hannig, J. (2013). Generalized fiducial inference via discretization. *Statistica Sinica*, *23*(2), 489–514.

Jennrich, R. I. (1973). Standard errors for obliquely rotated factor loadings. *Psychometrika*, *38*(4), 593–604.

Le Cam, L., & Yang, G. L. (2000). *Asymptotics in Statistics: Some Basic Concepts. Springer Series in Statistics*. New York: Springer-Verlag.

Liu, Y., & Maydeu-Olivares, A. (2013). Identifying the source of misfit in item response theory models. *Multivariate Behavioral Research*. In press.

Liu, Y., & Thissen, D. (2012). Identifying local dependence with a score test statistic based on the bifactor logistic model. *Applied Psychological Measurement*, *36*(8), 670–688.

Meng, X.-L., & Schilling, S. (1996). Fitting full-information item factor models and an empirical investigation of bridge sampling. *Journal of the American Statistical Association*, *91*(435), 1254–1267.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, *43*(4), 551–560.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.

Neale, M. C., & Miller, M. B. (1997). The use of likelihood-based confidence intervals in genetic models. *Behavior genetics*, *27*(2), 113–120.

Patz, R. J., & Junker, B. W. (1999). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics*, *24*(4), 342–366.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applies statistician. The Annals of Statistics (pp. 1151–1172).

Schilling, S., & Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*, *70*(3), 533–555.

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press.

Van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge University Press.

Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association*, *88*(423), 899–905.

Wirth, R., & Edwards, M. C. (2007). Item factor analysis: current approaches and future directions. *Psychological methods*, *12*(1), 58.

Yuan, K.-H., Cheng, Y., & Patton, J. (2014). Information matrices and standard errors for MLEs of item parameters in IRT. *Psychometrika*, *79*(2), 232–254.

Zabell, S. L. (1992). R. A. Fisher and fiducial argument. *Statistical Science*, *7*(3), 369–387.