SEMIPARAMETRIC IRT MODELS FOR NON-NORMAL LATENT TRAITS

Sally Paganin [1]Department of Biostatistics, Harvard School of Public Health,
Harvard University (e-mail: spaganin@hsph.harvard.edu)

ABSTRACT: Item Response Theory models are widely used in many domains of applications to analyze questionnaires data, scaling categorical data into continuous construct. Interpretable inference is often obtained relying on a set of assumptions for the latent constructs, as for example normality for the unknown subject-specific latent traits. This assumption can often be unrealistic and lead to biased results, hence we consider more flexible models using Bayesian nonparametric mixtures for the individual latent traits. We study several identifiability constraints, and compare inferential results and different Markov chain Monte Carlo strategies for posterior sampling.

KEYWORDS: 2PL, Bayesian nonparametrics, Dirichlet Process, MCMC, NIMBLE.

1 IRT models for binary responses

Let y_{ij} denote the answer of an individual j to item i for $j = 1,\dots,N$ and $i = 1,\dots,I$, with $y_{ij} = 1$, when the answer is correct and 0 otherwise. Typically, different individuals are assumed to work independently, while responses from the same individuals are assumed independent conditional to the latent trait (local independence assumption). Hence each answer y_{ij}, conditionally to the latent parameters, is assumed to be a realization of a Bernoulli distribution, and the probability of a correct response is typically modeled via logistic regression.

2 Semiparametric 2PL models

In the two-parameter logistic (2PL) model, the conditional probability of a correct response is modeled as

$$\Pr(y_{ij} = 1|\lambda_i,\beta_i,\eta_j) = \frac{\exp\{\lambda_i(\eta_j - \beta_i)\}}{1 = \exp\{\lambda_i(\eta_j - \beta_i)\}}, i = 1,\dots,I, \quad j = 1,\dots,N. \quad (1)$$

where η_j represents the health status, or more in general latent trait, of the j-th individual, while β_i and λ_i encode item characteristics. The parameter $\lambda_i > 0$ is often referred to as *discrimination*, while β_i is called *difficulty* because for any fixed η_j the probability of a correct response to item i is decreasing in β_i. When $\lambda_i = 1$ for all $i = 1, \dots, I$, the model in **??** reduces to the one-parameter logistic (1PL) model. Often, conditional log-odds in **??** are reparametrized as $\lambda_i \eta_i + \gamma_i$, with $\gamma_i = -\lambda_i \times \beta_i$. Sometimes this is reffered to as slope-intercept parameterization as opposed to the IRT parameterization in considered traditionally for interpretation.

Traditional literature assumes that $\eta_j \sim \mathcal{N}(0,1)$ for $j = 1, \dots, N$, but there are situations in which such assumption can be too restrictive. We can extend the model in **??** to describe more flexible latent trait distributions using a Dirichlet Process (DP) mixture of normal distributions

$$\begin{aligned} \eta_j | G \sim G, \quad G \sim DP(\alpha, G_0), \\ G_0 \equiv \mathcal{N}(0, \sigma_0^2) \times \text{InvGamma}(\nu_1, \nu_2) \end{aligned} \tag{2}$$

where α is the concentration parameter and G_0 the base measure. Alternative representations of the DP are known as the Chinese Restaurant Process (CRP) **?** or the truncated stick-breaking (SB) **?**.

3 Model estimation

Estimation of the model parameters is carried out in the Bayesian framework via MCMC methods, using NIMBLE **?**, a R software for hierarchical models. The NIMBLE system provides a suite of different sampling algorithms along with the possibility to code user-defined samplers. We compare results from the parametric and semiparametric 2PL model, using NIMBLE's default sampling configuration, that mixes conjugate samplers with adaptive Metropolis Hastings algorithm.

Typically parameters of the 2PL model are not identifiable, so constraints are either included in the model or one can post-process posterior samples to meet the constraints. This last approach is typical of parameter-expanded algorithms, which embed targeted models in a larger specification. We found this last option to be the most efficient in terms on both MCMC mixing and time.

In traditional literature on parametric 2PL model, identification is obtained constraining the discrimination parameters λ_i, for $i = 1, \dots, I$ to be positive, when the latent trait distribution is assumed to be a standard normal. Since we are relaxing the normal assumption on the latent traits, we considered sum-to-zero constraints on the item parameters, i.e. $\sum_i \beta_i = 0$, $\sum_i \log(\lambda_i) = 0$.

4 Inferential results

We compare inferential results via simulation. We simulate data from two different scenarios changing the distribution generating the latent traits. We simulate responses from $N = 3,000$ individuals to $I = 20$ binary items. Values for the discrimination parameters $\{\lambda_i\}_{i=1}^{20}$ are sampled from a Uniform distribution over the interval $(0.5, 2)$, while values for difficulty parameters $\{\beta_i\}_{i=1}^{20}$ are sampled from a Normal distribution with mean zero and variance 2.

In particular, we considered two different generating distribution for the latent traits. A unimodal scenario, where η_j are i.i.d. draws from a $\mathcal{N}(0,1)$ and a multimodal scenario where

$$\eta_j \sim 0.4 \times \mathcal{N}(-3,1) + 0.2 \times \mathcal{N}(-2,4) + 0.4 \times \mathcal{N}(2,1). \tag{3}$$

We chose moderately vague priors for the item parameters, $\beta_i \sim \mathcal{N}(0,3)$ and $\log(\lambda_i) \sim \mathcal{N}(0.5, 0.5)$. In the parametric model, η_js are assumed to follow $\mathcal{N}(0,1)$, while for DP we choose $G_0 \equiv \mathcal{N}(0,3) \times InvGamma(1.01, 2.01)$. We run the MCMC for $50,000$ iterations using a 10% burn-in of 5000 iterations, and check traceplots for convergence.

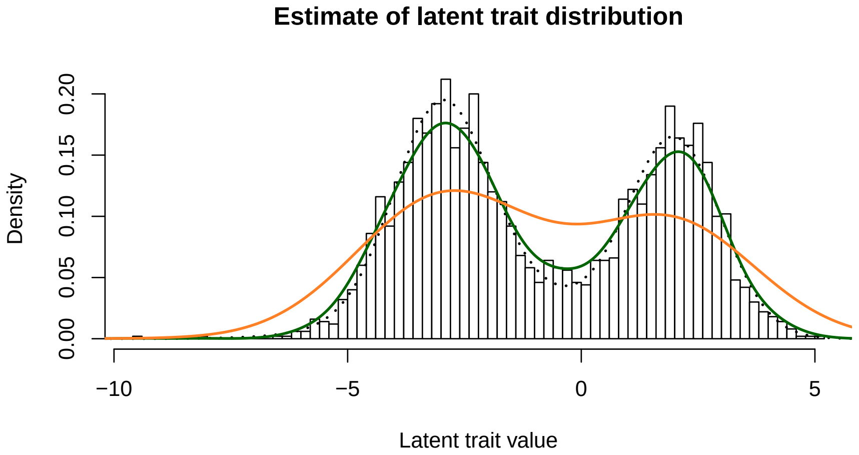

Figure 1. *Comparison of the latent trait density estimates, using a parametric 2PL model (orange line) and a semiparametric 2PL model (green line). The dotted black lines indicate the true distribution in (**??**).*

Figure 1 compares density estimates of the latent trait distribution from the parametric and semiparametric models, computed taking the posterior means of the η_js. It can be noticed that the parametric model leads to a flat distribution because of the underlying normal assumption, while the semiparametric specification recover the true density structure. Better estimation of the latent abilities helps to avoid bias in inference, for example when estimating item parameters or item characteristics curves (ICC).