CHILEAN JOURNAL OF STATISTICS

Edited by Víctor Leiva and Carolina Marchant

A free open access journal indexed by

Volume 12 Number 1
April 2021
ISSN: 0718-7912 (print)
ISSN: 0718-7920 (online)

Published by the
Chilean Statistical Society
SOCHE
SOCIEDAD CHILENA DE ESTADÍSTICA

Aims

The Chilean Journal of Statistics (ChJS) is an official publication of the Chilean Statistical Society (`www.soche.cl`). The ChJS takes the place of *Revista de la Sociedad Chilena de Estadística*, which was published from 1984 to 2000.

The ChJS covers a broad range of topics in statistics, as well as in artificial intelligence, big data, data science, and machine learning, focused mainly on research articles. However, review, survey, and teaching papers, as well as material for statistical discussion, could be also published exceptionally. Each paper published in the ChJS must consider, in addition to its theoretical and/or methodological novelty, simulations for validating its novel theoretical and/or methodological proposal, as well as an illustration/application with real data.

The ChJS editorial board plans to publish one volume per year, with two issues in each volume. On some occasions, certain events or topics may be published in one or more special issues prepared by a guest editor.

Chilean Journal of Statistics

Volume 12, Number 1
April 2021

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

Contents

Chilean Journal of Statistics
Vol. 12, No. 1, April 2021, 103–122

Bayesian Modeling
Research Paper

Bayesian analysis of an item response model with an AEP-based link function

Francisco J. Ariza-Hernandez[1] and Eduardo Gutiérrez-Peña[2]

[1]Facultad de Matemáticas, Universidad Autónoma de Guerrero, UAGro, Mexico,
[2]Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, IIMAS, UNAM, Mexico,
(Received: 17 February 2021 · Accepted in final form: 26 April 2021)

Abstract

We consider a robust Bayesian approach to the analysis of item response models, using the inverse of an asymmetric exponential power cumulative distribution function as a link function. This provides greater flexibility with respect to classic link functions such as the probit and the logit. We conduct a simulation study to evaluate the performance of our model. In order to draw samples from the posterior distribution of the parameters, we implement a Markov chain Monte Carlo scheme by means of the JAGS software. We also implement a posterior predictive model-checking method to assess the fit and relative performance of the various submodels. Finally, we provide a real-data example to illustrate the modeling approach proposed.

Keywords: Asymmetric exponential power (AEP) distribution · Generalized linear model · JAGS and R software · Rasch model · Sample-based inference.

Mathematics Subject Classification: Primary 62F15 · Secondary 62J12.

1. Introduction

Item response data come from applying a test to a set of individuals. The test is composed of a number of items. These tests are used extensively in schools, industry, and government, and for various purposes (see Baker and Kim, 2004; van der Linden and Hambleton, 1997; Fox, 2010). There is a very extensive literature about of the item response models, its development, description, and applications goes back to Lord (1952, 1980), who established the basis of item response theory (IRT), also called modern test theory.

Traditionally, frequentist analyses have been used in IRT. Recently, however, the Bayesian approach become very attractive for modeling item response data; (see Ghosh et al., 2000; Béguin and Glas, 2001; Bazán et al., 2006; Fox, 2010; Azevedo et al., 2011, 2012; Matteucci et al., 2012). This approach allows one to incorporate additional information to the analysis and provides powerful estimation methods based on simulated samples from posterior distributions.

Although item response modeling can be employed in more general contexts (see Reckase, 2009; Fu et al., 2009; Svetina, 2013; Bacci et al., 2014), and nonparametric settings (see Karabatsos, 2016; San Martin et al., 2011). In this paper, we focus on dichotomous response

*Corresponding author. Email: arizahfj@uagro.mx

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

data and unidimensional models with a continuous latent trait. The Rasch model (see Rasch, 1961) is by far the most popular model in this latter case. It is basically a logistic model, so the probit model is commonly used as an alternative.

Both the logit and probit link functions (corresponding to the standard logistic and the standard normal distributions, respectively) have traditionally been utilized when modeling dichotomous response data. Both of these link functions are symmetric. However, when the proportion of ones in the observed sample is very different from the proportion of zeros, or vice versa, the symmetric links commonly used may not be appropriate, as they may lead to misspecified models (see Chen et al., 1999a). This situation is not uncommon with item response experimental data.

The fit of item response models can be improved significantly by using asymmetric links. Several authors have worked with asymmetric links. Chen et al. (1999a) proposed a class models with skewed link to analyze binary data with covariates, Jiang et al. (2013) derived a new class of symmetric power link functions to model binary data and applied it to the Protea co-occurrence data. More recently, Durante (2019) proved that in the case of probit regression models which have Gaussian priors for the coefficients, the posterior belongs to the class of unified skew-normal distributions. Also, Naranjo et al. (2015) employed an asymmetric exponential power (AEP) distribution for the error of a linear regression model, and the inverse of the AEP cumulative distribution function (CDF) as a link function in a regression model for binary data, but not in the context of IRT.

Models with asymmetrical link functions have also been proposed in IRT settings. Samejima (2000) proposed a family of models called the logistic positive exponent family, which provides asymmetric item characteristic curves (ICCs). Bazán et al. (2006) introduced a skew-probit IRT link function based on the skew normal distribution, while Azevedo et al. (2011) used skew-normal distributions to model latent traits in an IRT two-parameter probit model under centered parameterizations. However, these models are not as flexible as the AEP distribution, which not only allows one to handle symmetry/asymmetry but also light/heavy tails.

In this paper, we build on the work of Zhu and Zinde-Walsh (2009) and Naranjo et al. (2015) to propose a Bayesian item response model based on the AEP distribution.

The outline of the paper is as follows. In Section 2, we briefly discuss the Rasch model and review the probability density and cumulative distribution functions of the AEP distribution. We describe the general model in Section 3. Then, in Section 4, we carry out Bayesian inferences on the parameters of interest via the just another Gibbs sampler (JAGS) software (see Plummer, 2017) within the `R` software (see R Core Team, 2020), and apply a posterior predictive model-checking method (see Sinharay et al., 2006) with the purpose of comparing various submodels. In Section 5, we present a simulation study and conducted to assess the performance of the Bayesian estimates. Also, a real-data example is given in this section to illustrate the AEP-based IRT model. Finally, Section 6 contains some concluding remarks.

2. Preliminaries

2.1 The Rasch model

We model the probability of the correct answer, p_{ik}, corresponding to i-th individual in the k-th item, as $p_{ik} = P(Y_{ik} = 1|\theta_i, a_k, b_k) = F(a_k\theta_i - b_k)$, for $i = 1, \ldots, N$ and $k = 1, \ldots, K$, where Y_{ik} is a random variable which takes the value of 1 if the i-th individual responds correctly to the k-th item and F is the CDF of a known parametric family. In the context of IRT, F is the ICC, $a_k > 0$ and $b_k \in \mathbb{R}$ are item parameters (called discrimination and difficulty parameters, respectively), and $\theta_i \in \mathbb{R}$ is the person parameter associated with the ability of individual i. The inverse of F is called the link function. The Rasch model is the simplest and most traditional model for p_{ik}. It is given by

$$p_{ik} = P(Y_{ik} = 1|\theta_i, b_k) = \frac{\exp(\theta_i - b_k)}{1 + \exp(\theta_i - b_k)}.$$

That is, the probability that the person i obtains a correct response to item k is a logistic function of the difference between the person's ability, θ_i, and difficulty of the item, b_k. Note that, if the person's ability is greater than the difficulty of the item, then the probability of success is higher in comparison with the probability of failure. A limitation of the Rasch model is that all items are assumed to discriminate between respondents in the same way (that is, $a_k = 1$ for all $k = 1, \ldots, K$); as a result, items only differ in item difficulty (see Fox, 2010). The probit model is another popular model for p_{ik}; it takes $p_{ik} = \Phi(\theta_i - b_k)$, where Φ is the standard normal CDF. The Rasch model can be approximated by a probit model by multiplying the parameter values by a scaling factor of 1.7.

2.2 The AEP distribution

The probability density function (PDF) of the rescaled AEP distribution, proposed by Zhu and Zinde-Walsh (2009), is stated as

$$\tilde{f}_{\mathrm{AEP}}(x|\mu, \tilde{\sigma}, \alpha, \delta_1, \delta_2) = \begin{cases} \dfrac{1}{\tilde{\sigma}} \exp\left\{ - \left| \dfrac{x-\mu}{\alpha\tilde{\sigma}/\Gamma(1+1/\delta_1)} \right|^{\delta_1} \right\}, & \text{if } x \leq \mu; \\ \dfrac{1}{\tilde{\sigma}} \exp\left\{ - \left| \dfrac{x-\mu}{(1-\alpha)\tilde{\sigma}/\Gamma(1+1/\delta_2)} \right|^{\delta_2} \right\}, & \text{if } x > \mu; \end{cases} \tag{1}$$

where $\mu \in \mathbb{R}$ is the location parameter, $\tilde{\sigma} > 0$ is the scale parameter, $\alpha \in (0, 1)$ is the skewness parameter, and δ_1 and δ_2 are the left- and right-tail parameters, respectively ($\delta_1 > 0, \delta_2 > 0$). For convenience, we consider a reparametrization of the scale parameter in Equation (1), $\tilde{\sigma} = \sqrt{2\pi}\sigma$, so that

$$f_{\mathrm{AEP}}(x|\mu, \sigma, \alpha, \delta_1, \delta_2) = \tilde{f}_{\mathrm{AEP}}(x|\mu, \sqrt{2\pi}\sigma, \alpha, \delta_1, \delta_2).$$

With this parametrization, the density function of the AEP distribution is given by

$$f_{\mathrm{AEP}}(x|\mu, \sigma, \alpha, \delta_1, \delta_2) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left\{ - \left| \dfrac{x-\mu}{\sqrt{2\pi}\alpha\sigma/\Gamma(1+1/\delta_1)} \right|^{\delta_1} \right\}, & \text{if } x \leq \mu; \\ \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left\{ - \left| \dfrac{x-\mu}{\sqrt{2\pi}(1-\alpha)\sigma/\Gamma(1+1/\delta_2)} \right|^{\delta_2} \right\}, & \text{if } x > \mu. \end{cases} \tag{2}$$

We use $X \sim \mathrm{AEP}(\mu, \sigma, \alpha, \delta_1, \delta_2)$ to denote Equation (2). Important properties of the AEP distribution have been discussed in the literature (see Zhu and Zinde-Walsh, 2009; Naranjo et al., 2015). If $\alpha = 1/2$ and $\delta_1 = \delta_2$, the distribution is symmetric. An important special case is when $\delta_1 = \delta_2 = 2$ and $\alpha = 1/2$, in which case Equation (2) is the $\mathrm{N}(\mu, \sigma^2)$ distribution. In Figure 1, we show the PDF and CDF of the AEP distribution in Equation (2) for a range of parameter values.

For the standard version of the AEP distribution ($\mu = 0$, $\sigma = 1$), the CDF can be expressed as

$$F_{\mathrm{AEP}}(x|\alpha,\delta_1,\delta_2) = \begin{cases} \alpha\left[1 - G\left(\left(\dfrac{x}{\sqrt{2\pi}\alpha/\Gamma(1+1/\delta_1)}\right)^{\delta_1}; \dfrac{1}{\delta_1}, 1\right)\right], & \text{if } x \le 0; \\ \alpha + (1-\alpha)G\left(\left(\dfrac{x}{\sqrt{2\pi}(1-\alpha)/\Gamma(1+1/\delta_2)}\right)^{\delta_2}; \dfrac{1}{\delta_2}, 1\right), & \text{if } x > 0; \end{cases} \tag{3}$$

where $G(v;\gamma,\beta)$ is the gamma CDF given by

$$G(v;\gamma,\beta) = \frac{1}{\Gamma(\gamma)\beta^{\gamma}} \int_0^v t^{\gamma-1} \exp\{-t/\beta\}\, \mathrm{d}t.$$

The proof of Equation (3) is given in Appendix A.

3. An AEP-based generalized linear model for binary data

3.1 Model specification

The IRT model based on the AEP distribution is formally defined as follows. Let Y_{ik} be a random variable representing the response of the i-th individual to the k-th item. This response variable is discrete, taking only two possible values. We define $Y_{ik} = 1$ if the i-th individual's response to the k-th item is correct and $Y_{ik} = 0$ for an incorrect response. Then, we have

$$Y_{ik}|\theta_i, a_k, b_k, \alpha_k, \delta_{1k}, \delta_{2k} \sim \mathrm{Bern}(p_{ik}), \tag{4}$$

where $\mathrm{Bern}(p_{ik})$ denotes the Bernoulli distribution, for $i = 1, \ldots, N$, $k = 1, \ldots, K$, and p_{ik} is given by

$$\begin{aligned} p_{ik} &= P(Y_{ik} = 1|\theta_i, a_k, b_k, \alpha_k, \delta_{1k}, \delta_{2k}) \\ &= F_{\mathrm{AEP}}(a_k\theta_i - b_k \,|\, \alpha_k, \delta_{1k}, \delta_{2k}). \end{aligned} \tag{5}$$

This represents the conditional probability that the i-th individual, with ability θ_i, responds correctly to the k-th item with discrimination parameter a_k and difficulty parameter b_k. The quantities α_k, δ_{1k} and δ_{2k} are the AEP parameters defined in Equation (2). This model assumes that a change in the probability of a specified response is described by the ICC in Equation (5), and that the responses to a pairs of items are statistically independent given the latent variable θ. The probability of success is modeled as a function of person, item and AEP parameters. Note that, for $\alpha = 0.5$ and $\delta_1 = \delta_2 = 2$, the Equation (5) reduces to the probit model.

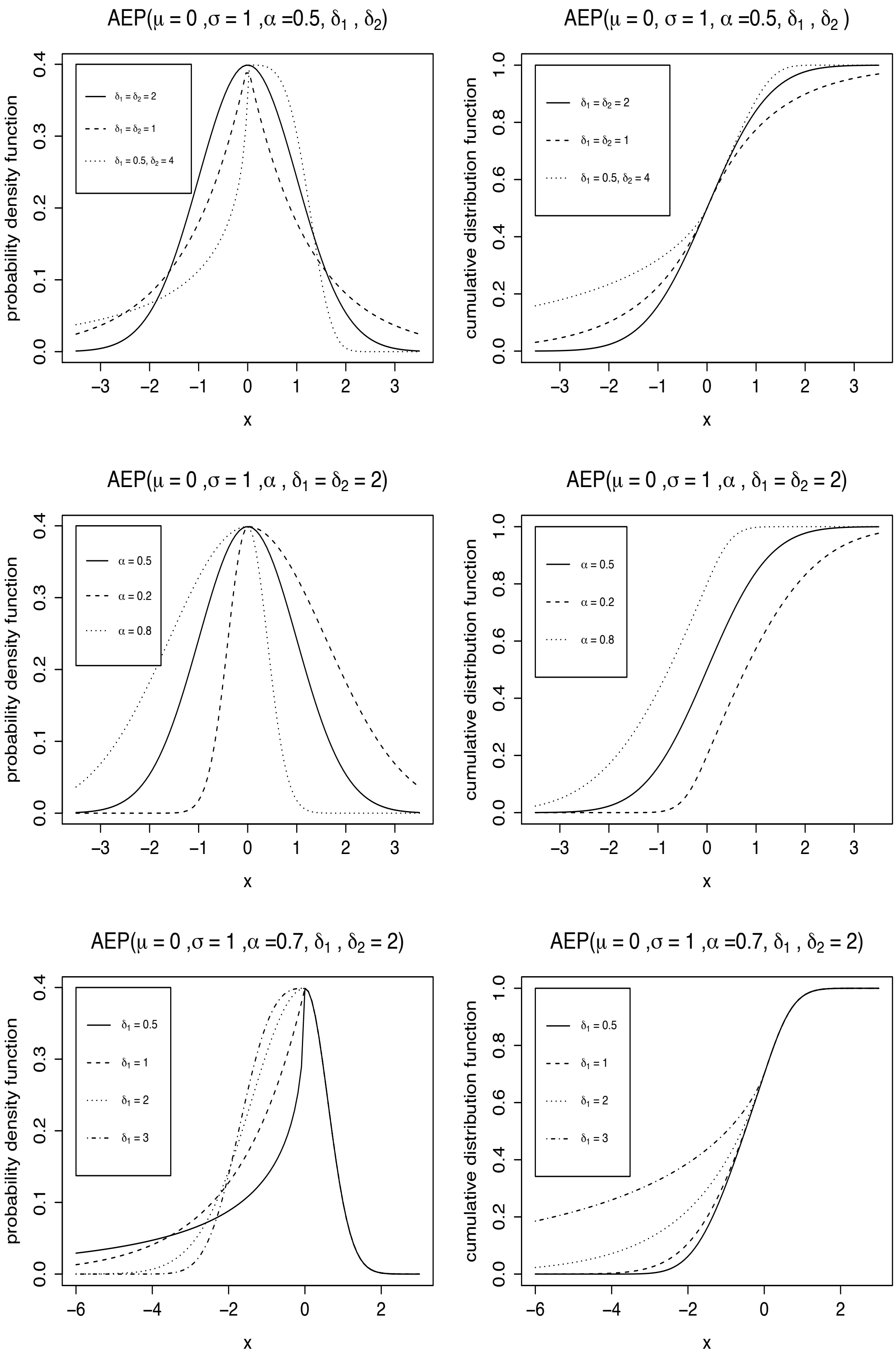

Figure 1. PDFs and CDFs of the AEP distribution for different values of the parameters.

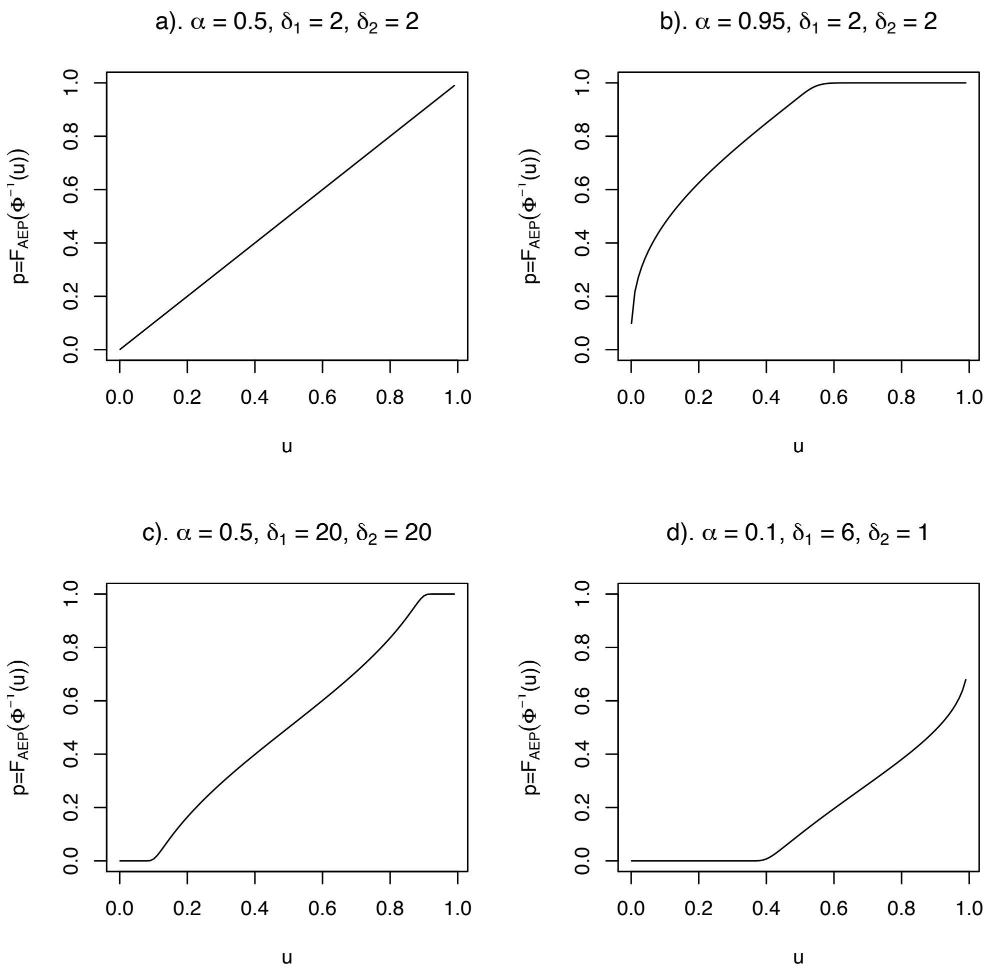

Figure 2. Comparison of AEP-based and probit link functions.

3.2 Likelihood function

Let $\boldsymbol{y} = (y_{11}, \ldots, y_{NK})^\top$ denote the observed item response data. Then, the likelihood function for the AEP-IRT model is stated as

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\eta}; \boldsymbol{y}) &= \prod_{i=1}^{N}\prod_{k=1}^{K} p_{ik}^{y_{ik}}(1-p_{ik})^{1-y_{ik}} \\ &= \prod_{i=1}^{N}\prod_{k=1}^{K} [F_{\text{AEP}}(m_{ik} \,|\, \boldsymbol{\eta})]^{y_{ik}}[1 - F_{\text{AEP}}(m_{ik} \,|\, \boldsymbol{\eta})]^{1-y_{ik}}, \end{aligned}$$

where $m_{ik} = a_k\theta_i - b_k$, $i = 1, \ldots, N$, $k = 1, \ldots, K$; $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$, $\boldsymbol{\xi} = (\boldsymbol{a}, \boldsymbol{b})$ and $\boldsymbol{\eta} = (\boldsymbol{\alpha}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$, with $\boldsymbol{a} = (a_1, \ldots, a_K)$, $\boldsymbol{b} = (b_1, \ldots, b_K)$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$, $\boldsymbol{\delta}_1 = (\delta_{1k}, \ldots, \delta_{1k})$ and $\boldsymbol{\delta}_2 = (\delta_{2k}, \ldots, \delta_{2k})$.

Note that the model proposed here is described in terms of N ability parameters, K discrimination parameters, K difficulty parameters, K skewness parameters and K pairs of tail parameters. Hence, it has a total of N person parameters and $5K$ unknown parameters. This model is overparameterized. In fact, for two different sets of parameter values the model may give the same success probabilities and so the model may be unidentifiable. For example, the linear predictor in Equation (5) can be written as $a_k\theta_i - b_k = a_k(10\theta_i -$

$50)/10 - (b_k - 50a_k/10) = a_k^*\theta_i^* - b_k^*$; that is, the model with a_k, b_k, θ_i is the same as with a_k^*, b_k^*, θ_i^*. Thus, the parameters cannot be uniquely estimated, unless certain constraints are imposed. From the Bayesian viewpoint, this problem may be solved by specifying suitable priors for the parameters of interest (see Chen et al., 2003; Matteucci et al., 2012; Naranjo et al., 2015).

As pointed out in Section 1, when the proportion of ones in the observed sample is very different from the proportion of zeros, or vice versa, the symmetric links commonly used may not be appropriate. To visualize the flexibility of the AEP-based link function with respect to the probit link function, in Figure 2 we plot $F_{\text{AEP}}(\Phi^{-1}(u)|\,\alpha, \delta_1, \delta_2)$ over the interval $(0, 1)$ for selected values of α, δ_1 and δ_2.

4. Bayesian inference

4.1 Prior distribution

In this paper we use a Bayesian approach to make statistical inference about the parameters of interest. In this setting, the parameters are regarded as random variables and have prior distributions that reflect the uncertainty about their true values before observing the data. Several authors have suggested informative as well as noninformative prior distributions for the item para-meters; for example, lognormal priors for the discrimination parameters and a normal prior for difficulty parameters (see Albert, 1992; Patz and Junker, 1999; Rupp et al., 2004; Fox and Glas, 2001; Matteucci et al., 2012; Bazán et al., 2006). Ghosh et al. (2000) pointed out that, with noninformative priors, posterior distributions for item and person parameters may be improper when the sum of the binary responses for an item or person takes its minimum or maximum possible value. However, they prove that under certain conditions the joint posterior distribution is proper.

Here, we assume the item parameters to be exchangeable. We also assume monotonicity of the ICC, which is satisfied when the discrimination parameter is restricted to be positive. Thus, we assume the following prior distribution for the item parameters

$$(a_k, b_k) \sim \mathrm{N}(\boldsymbol{\mu}_\xi, \boldsymbol{\Sigma}_\xi)\, I_A(a_k),$$

where $A = \{a \in \mathbb{R} : a > 0\}$, $k = 1, \ldots, K$, and I_A is the indicator function of the set A. Note that this prior is not conjugate for the observed likelihood. A typical prior for person parameters assumes that the individual are chosen randomly from an unknown population, where each individual has the same probability of being chosen. Individuals are also assumed to be sampled independently, so we assume that

$$\theta_i \sim \mathrm{N}(\mu_\theta, \sigma_\theta), \quad i = 1, \ldots, N,$$

where $\mu_\theta, \sigma_\theta$ are known parameters. These priors have been suggested by others authors (see Bazán et al., 2006; Sinharay et al., 2006; Fox, 2010; Matteucci et al., 2012). In generalized linear models, some authors have proposed an elicitation scheme for a class of informative prior distributions for the regression parameters based on historical data (see Chen et al., 1999b, 2003). Naranjo et al. (2015) proposed some alternative prior distributions for the AEP parameters, which have the advantage of allowing one to derive the full conditional distributions required for a Gibbs sampler. With some adjustments, the Jeffreys prior distribution can be computed from the Fisher information matrix given by Zhu and Zinde-Walsh (2009).

Assuming prior independence of the parameters, we can write the joint prior distribution as

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\eta}) &= p(\boldsymbol{\theta})\, p(\boldsymbol{\xi})\, p(\boldsymbol{\eta}) \\ &= \prod_{i=1}^{N} p(\theta_i) \left\{ \prod_{k=1}^{K} p(a_k)\, p(b_k)\, p(\alpha_k)\, p(\delta_{1k})\, p(\delta_{2k}) \right\}. \end{aligned}$$

4.2 Posterior sampling

By the Bayes theorem, the posterior distribution of the parameters of interest is established as

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\eta} | \boldsymbol{y}) &= L(\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\eta}; \boldsymbol{y})\, p(\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\eta}) / p(\boldsymbol{y}) \\ &\propto \prod_{i=1}^{N} \prod_{k=1}^{K} \left\{ [F_{\mathrm{AEP}}(m_{ik} \,|\, \boldsymbol{\eta})]^{y_{ik}} [1 - F_{\mathrm{AEP}}(m_{ik} \,|\, \boldsymbol{\eta})]^{1-y_{ik}} \right. \\ &\qquad \left. \times\; p(\theta_i)\, p(a_k)\, p(b_k)\, p(\alpha_k)\, p(\delta_{1k})\, p(\delta_{2k}) \right\}. \end{aligned} \tag{6}$$

Note that the joint posterior distribution is analytically intractable and thus obtaining the marginal posterior densities of the parameters is not an easy task; however, samples from Equation (6) can be obtained using Markov chain Monte Carlo (MCMC) techniques. The most common MCMC methods are the Gibbs sampling (see Gelfand and Smith, 1990; Casella and George, 1992) and the Metropolis-Hastings (see Metropolis et al., 1953; Chib and Greenberg, 1995). Currently, many of the MCMC algorithms have been already implemented in computer programs, such as, `WinBUGS` (see Spiegelhalter et al., 2003), `JAGS` (see Plummer, 2017) and Stan (see Stan Development Team, 2014). All of these software packages provide programs for Bayesian modeling through posterior simulation given a specified model and data. In particular, `JAGS` provides several samplers and attempts to use the most efficient one to update the parameters of the model at each iteration. The `R` packages named `R2WinBUGS`, `R2jags` and `rstan` allow one to run `WinBUGS`, `JAGS` and `Stan` from within `R`, respectively. There are several `R` packages for IRT. Choi and Asilkalkan (2019) presentes a summary of the `IRT` package that have been developed over the last decade. In this paper, we utilize `JAGS` within `R` to obtain samples from the posterior distributions of interest (see Appendix B).

4.3 A posterior predictive model-checking method

The posterior predictive model-checking (PPMC) method is a popular Bayesian model-checking tool, has a strong theoretical basis, and can provide graphical or numerical summaries about the model fit (or lack thereof). For IRT models, Sinharay et al. (2006) presented an extensive explanation of the PPMC method and discuss different discrepancy measures to detect various violations to model assumptions. Azevedo et al. (2012) developed Bayesian methods for the multiple-group IRT model, including an estimation method based on MCMC and different posterior predictive assessment tools. The idea of PPMC is to generate replicate data sets by simulating from the posterior predictive distribution, and then compare these simulated samples with the observed data. If the replicated data and the observed data differ systematically, it is an indication of a potential model misfit.

The choice of discrepancy measure is crucial in the application of the PPMC method. In this paper, we used the Observed Score Distribution (OSD) as the discrepancy measure, which has been employed by Béguin and Glas (2001). This discrepancy measure is given by

$$\text{OSD} = \sum_k \frac{[NC_k - \text{E}(NC_k)]^2}{\text{E}(NC_k)}, \tag{7}$$

where NC_k denotes the number of examinees getting exactly k correct items, and $\text{E}(NC_k)$ is the expected value of NC_k under the model, for $k = 0, 1, \ldots, K$.

In addition, here we propose an alternative discrepancy measure based on the Kullback-Leibler divergence between the "true" model and an "approximate" model, stated as

$$D_{\text{KL}}(\pi||\widetilde{\pi}) = \sum_k \pi_k \log\left(\frac{\pi_k}{\widetilde{\pi}_k}\right), \tag{8}$$

where $\pi_k = \text{E}(NC_k)/N$ and $\widetilde{\pi}_k = NC_k/N$.

In order to assess the fit of the IRT model to a given data set, we can repeat the following steps a large number of times:

(1) Generate a draw of the parameters of interest from the posterior distribution given by Equation (6).
(2) Obtain a data set from the model given in Equations (4)-(5), using the parameters drawn in the previous step.
(3) Compute the values of the predictive and realized discrepancy measures given in Equations (7) or (8), utilizing the data set drawn in the previous step.

With the predictive and realized discrepancy measures, we can creates plots to assess the fit of the IRT model.

5. Numerical applications

5.1 Simulation study

We carried out a simulation study to assess the performance of the Bayesian estimators of the parameters of interest. The procedure was applied to each of several combinations of data-generating and fitted models. Table 1 shows the cases we considered, that is:

- The AEP-III model is based on Equations (4)–(5). This model can describe both symmetry/asymmetry and light/heavy tails separately for each item.
- In the AEP-II model, the tails of the AEP distribution for each item are described by means of the parameters δ_{1k} and δ_{2k}, while α_k is held fixed at $\alpha_k = 0.5$.
- In the AEP-I model, the symmetry/asymmetry for each item is formulated by means of the skewness parameter α_k, while δ_{1k} and δ_{2k} are held fixed at $\delta_{1k} = \delta_{2k} = 2$.

Note that the AEP-I and AEP-II models are both particular cases of the AEP-III model.

Table 1. Cases examined in the simulation study.

Fitted	Data-generating model			
model	AEP-I	AEP-II	AEP-III	Probit
AEP-I	•			•
AEP-II		•		•
AEP-III			•	•
Probit	•	•	•	•

We now describe the simulation study:

(1) We simulated $B = 100$ data sets from each data-generating model (see below for details).

(2) For each simulated sample, we obtained Bayesian estimators of the parameters of interest, both for model given in Equations (4)–(5) and for the two-parameter probit model.

   a) We calculated the Bayes estimators as the sample mean from Equation (6) using `JAGS` within `R`. We employed two chains, each with 26,000 iterations, with a burn-in of 1000 iterations and a thinning rate of 50, so we kept a total of 500 iterations to make inferences about the parameters of interest.
   The analysis of each sample took around 1.74 minutes on a computer with a 4 GHz Intel Core i7 processor and 32GB of RAM.

   b) We calculated 95% credible intervals for each parameter; these intervals are based on the 2.5-th quantile and the 97.5-th quantile of the corresponding posterior sample.

(3) From these B samples, we computed the mean squared error (MSE) of the estimators as

$$\widehat{\mathrm{MSE}} = s^2(\theta) + \widehat{B}^2(\theta),$$

where $s^2(\theta) = \sum_{i=1}^{B}(\theta_i^B - \bar{\theta}_B)^2/(B-1)$ is the sample variance of the Bayes estimators, $\widehat{B}(\theta) = \bar{\theta}_B - \theta$ is the bias, $\bar{\theta}_B = \sum_{i=1}^{B} \theta_i^B/B$, and θ_i^B is the Bayes estimator corresponding to the i-th sample.

(4) Finally, we computed the coverage of the corresponding credible intervals.

To perform this study, we used `R` together with the `R2jags` package (see Su and Yajima, 2020). Our simulated data sets consist of $N = 100$ individuals and $K = 3, 5, 10, 20$ items. The true values of the parameters utilized to generate the data sets were varying according to the Table 2.

Table 2. Parameter values for the simulation study.

Parameters	from	to
a_k	0.5	2.0
b_k	-2.0	2.0
α_k	0.1	0.9
δ_{1k}	0.5	4.0
δ_{2k}	0.5	4.0

We assumed the following priors for the discrimination and difficulty parameters:

$$a_k \sim \mathrm{N}(1,1)I(a_k > 0); \quad b_k \sim \mathrm{N}(0,1), \quad k = 1, \ldots, K,$$

while, for the AEP parameters, we took the priors used by Naranjo et al. (2015), namely,

$$\alpha_k \sim \mathrm{Beta}(1,1); \quad \delta_{1k} \sim \mathrm{Gamma}(1,1) \quad \text{and} \quad \delta_{2k} \sim \mathrm{Gamma}(1,1), \quad k = 1, \ldots, K.$$

In Tables 3 and 4, we show the estimated MSE and coverage for the cases considered in Table 1: AEP-I versus Probit, AEP-II versus Probit, AEP-III versus Probit, and Probit versus AEP-x, (x = I,II,II). Generally speaking, the AEP models outperform the probit model, especially AEP-III and AEP-II.

When we generated simulated data from an AEP-x model, and fitted both the corresponding AEP-x and the probit models, we observed that in all cases the coverage of the credible intervals for the item parameter was close to 100% for the AEP-x models. In contrast, the coverage obtained for the credible intervals of the discrimination and difficulty parameters of the probit model was much lower, even reaching zero in some cases. In general, the estimated MSE and bias are lower for the AEP-x models than for the probit model (see Table 3). Also, when we simulated data from the probit model, and fitted both the AEP-x and the probit models, we observed that both the coverage and the MSE are very similar for all models and close to 100%. As expected, all three AEP models fit the data generated from the probit model reasonably well (Table 4).

5.2 A real-data example

Next, an example is given to illustrate the Bayesian item response modeling approach proposed in this paper. We consider a data set previously analyzed by Fox (2010), which consists of 200 eighth-grade students that are subjected to a mathematics test with 5 items. The data set contains the responses of the examinees, where 1 indicates a correct answer and 0 an incorrect answer. We assume that the five items measure a unidimensional ability represented by θ, which is a continuous latent variable that takes values on the real line. We estimate the item parameters of both the probit IRT model and the AEP-IRT model using the MCMC methodology described above. This example was also implemented utilizing the `JAGS` package within `R` (see Appendix B for details).

The probability of a correct response by examinee i to item k is modeled by the following item response models:

(Probit) $P(Y_{ik} = 1|\theta_i, a_k, b_k) = \Phi(a_k\theta_i - b_k)$,
(AEP-I) $P(Y_{ik} = 1|\theta_i, a_k, b_k, \alpha_k) = F_{\text{AEP}}(a_k\theta_i - b_k|\alpha_k)$,
(AEP-II) $P(Y_{ik} = 1|\theta_i, a_k, b_k, \delta_{1k}, \delta_{2k}) = F_{\text{AEP}}(a_k\theta_i - b_k|\delta_{1k}, \delta_{2k})$,
(AEP-III) $P(Y_{ik} = 1|\theta_i, a_k, b_k, \alpha_k, \delta_{1k}, \delta_{2k}) = F_{\text{AEP}}(a_k\theta_i - b_k|\alpha_k, \delta_{1k}, \delta_{2k})$,

for $i = 1, \ldots, 200$ and $k = 1, \ldots, 5$, where F_{AEP} is given in Equation (3), and Φ is the standard normal CDF.

The prior distributions used were as follows: for all models, we employed a $\text{N}(0, 1)$ distribution for the difficulty parameters, while a truncated normal distribution, $\text{N}(1, 1)I(a_k > 0)$, was utilized for the discrimination parameters. These values of the hyperparameters indicate a moderate level of discrimination and average level of difficulty. Assuming that the individuals are sampled independently from the population, we specified a $\text{N}(0, 1)$ for the ability parameters of all of the models. This restriction identifies the two-parameter item response model (see Fox, 2010). Finally, for the AEP parameters we took the priors used by Naranjo et al. (2015). That is, $\alpha_k \sim \text{Beta}(1, 1)$, $\delta_{1k} \sim \text{Gamma}(1, 1)$ and $\delta_{2k} \sim \text{Gamma}(1, 1)$, for $k = 1, \ldots, 5$.

For each model, we employed two chains, each with 26000 iterations, and the first 1000 were discarded, taking a thinning rate of 50. Thus, 1000 posterior samples were used to obtain the summary statistics about the parameters of interest. Standard convergence diagnostics were carried out. To mention a few, the value of Gelman-Rubin $\hat{R}$ was close to 1 for each parameter of interest and for all the models we considered. Also, the Geweke diagnostics were calculated and showed evidence of convergence.

Table 3. Estimated MSE | Bias | (coverage).

<table>
<tr><th>Parameters</th><th>AEP-I</th><th>Probit</th><th>AEP-II</th><th>Probit</th><th>AEP-III</th><th>Probit</th></tr>
<tr><td>a_1</td><td>0.08 |-0.005| (1.0)</td><td>0.08 |0.02 | (1.0)</td><td>0.12 |-0.16| (1.0)</td><td>0.17 |0.27| (0.70)</td><td>0.018 |-0.10| (1.0)</td><td>0.148 |0.36| (1.0)</td></tr>
<tr><td>a_2</td><td>0.17 |-0.27| (0.96)</td><td>0.15 |-0.23| (0.96)</td><td>0.54 |-0.68| (0.92)</td><td>0.34 |0.27| (0.90)</td><td>0.161|-0.33 | (1.0)</td><td>0.220|-0.33| (1.0)</td></tr>
<tr><td>a_3</td><td>0.08 |-0.05| (1.0)</td><td>0.09 |0.07| (1.0)</td><td>0.04 |-0.03| (1.0)</td><td>0.14 |-0.46| (0.96)</td><td>0.043|-0.18| (1.0)</td><td>0.187 | 0.36|(0.96)</td></tr>
<tr><td>b_1</td><td>0.53 |-0.72| (1.0)</td><td>0.30 |-0.56| (0.33)</td><td>0.15 |-0.09| (1.0)</td><td>0.08 |0.23| (0.71)</td><td>0.011|0.09| (1.0)</td><td>0.944| 0.95| (0.0)</td></tr>
<tr><td>b_2</td><td>0.15 |0.38| (1.0)</td><td>0.54 |0.73| (0.0)</td><td>0.05 |-0.01| (1.0)</td><td>0.05 |0.07| (0.60)</td><td>0.001|-0.01| (1.0)</td><td>0.031| 0.11| (1.0)</td></tr>
<tr><td>b_3</td><td>1.25 |1.11| (0.90)</td><td>1.59 |-1.25| (0.0)</td><td>0.04 |0.06| (1.0)</td><td>0.03 |-0.04| (0.10)</td><td>0.012|-0.08| (1.0)</td><td>0.355|0.57| (0.41)</td></tr>
<tr><td>α_1</td><td>0.07 |-0.27| (1.0)</td><td>-</td><td>-</td><td>-</td><td>0.095 |-0.22|(1.0)</td><td>-</td></tr>
<tr><td>α_2</td><td>0.005 |0.06| (1.0)</td><td>-</td><td>-</td><td>-</td><td>0.015|0.07| (1.0)</td><td>-</td></tr>
<tr><td>α_3</td><td>0.12 |0.34| (1.0)</td><td>-</td><td>-</td><td>-</td><td>0.264| 0.45| (1.0)</td><td>-</td></tr>
<tr><td>δ_{11}</td><td>-</td><td>-</td><td>0.46 |0.21| (0.96)</td><td>-</td><td>0.354|-0.12| (1.0)</td><td>-</td></tr>
<tr><td>δ_{12}</td><td>-</td><td>-</td><td>2.04 |0.26| (1.0)</td><td>-</td><td>0.794|-0.52| (1.0)</td><td>-</td></tr>
<tr><td>δ_{13}</td><td>-</td><td>-</td><td>4.84 |-2.04| (0.84)</td><td>-</td><td>0.447| 0.27| (0.84)</td><td>-</td></tr>
<tr><td>δ_{21}</td><td>-</td><td>-</td><td>4.48 |-2.63| (0.68)</td><td>-</td><td>5.644 |2.32|(0.68)</td><td>-</td></tr>
<tr><td>δ_{22}</td><td>-</td><td>-</td><td>1.61 |-0.13|(1.0)</td><td>-</td><td>0.469 |-0.38|(1.0)</td><td>-</td></tr>
<tr><td>δ_{23}</td><td>-</td><td>-</td><td>0.51 |-0.48| (1.0)</td><td>-</td><td>0.324|0.38| (1.0)</td><td>-</td></tr>
</table>

Table 4. Probit versus AEP: Estimated MSE ⟨coverage⟩.

Parameters	Probit	AEP-I	AEP-II	AEP-III
a_1	0.16 ⟨0.86⟩	0.16 ⟨0.95⟩	0.23 ⟨1.0⟩	0.14 ⟨1.0⟩
a_2	0.15 ⟨1.0⟩	0.15 ⟨1.0⟩	0.27 ⟨1.0⟩	0.03 ⟨1.0⟩
a_3	0.18 ⟨1.0⟩	0.17 ⟨1.0⟩	0.06 ⟨1.0⟩	0.16 ⟨1.0⟩
b_1	0.09 ⟨0.90⟩	0.11 ⟨1.0⟩	0.06 ⟨1.0⟩	0.47 ⟨1.0⟩
b_2	0.05 ⟨1.0⟩	0.01 ⟨1.0⟩	0.03 ⟨1.0⟩	0.02 ⟨1.0⟩
b_3	0.08 ⟨0.90⟩	0.39 ⟨1.0⟩	0.16 ⟨1.0⟩	0.51⟨1.0⟩

Table 5 shows the parameter estimates for each of the fitted models. Posterior means are used as point estimates of the parameters of interest. The deviance information criterion –DIC– (see Spiegelhalter et al., 2002) for each model was obtained too, and was utilized to assess the fit of the various models.

Table 5. Parameter estimation for Fox (2010) data set.

Parameters	Probit	AEP-I	AEP-II	AEP-III
a_1	1.54	1.52	1.84	1.65
a_2	0.90	0.88	1.78	1.54
a_3	0.66	0.70	1.50	1.32
a_4	0.91	0.89	1.47	1.48
a_5	0.46	0.49	1.61	1.22
b_1	-0.27	-0.04	-0.26	-0.02
b_2	-0.79	-0.25	-0.52	-0.05
b_3	-0.11	0.05	0.31	0.55
b_4	-0.73	-0.35	-0.83	-0.19
b_5	-0.42	-0.18	-0.02	0.53
α_1	–	0.58	–	0.50
α_2	–	0.69	–	0.61
α_3	–	0.56	–	0.51
α_4	–	0.63	–	0.62
α_5	–	0.59	–	0.61
δ_{11}	–	–	3.80	1.36
δ_{12}	–	–	0.18	0.75
δ_{13}	–	–	0.76	0.56
δ_{14}	–	–	1.03	1.26
δ_{15}	–	–	0.10	0.67
δ_{21}	–	–	4.36	1.15
δ_{22}	–	–	5.68	1.61
δ_{23}	–	–	3.79	1.70
δ_{24}	–	–	3.50	1.38
δ_{25}	–	–	3.67	1.53
DIC	1332.9	1303.0	1463.6	2235.1

We observe that item five discriminates poorly in the AEP-I and Probit models, with values less than 1, except for item 1. Item 1 is the most discriminative item for all fitted models. The average estimated discrimination level is 0.89 with the AEP-I IRT model, which is slightly smaller than the prior mean. The posterior means for the skewness parameters show that the marginal posterior densities are non-symmetric and slightly positively skewed (see the AEP-III and AEP-I models). For the estimated difficulty parameter, we can observe that the item with higher values for difficulty is item 3, whereas the less difficult items are 2 and 4. The proportion of correct responses for each of the five items are estimated as 56%, 73%, 54%, 71%, and 65%, respectively. In Table 5, we also observe that, for the AEP-I IRT model, we have a lower value of DIC with respect to the probit, AEP-III and AEP-II IRT models. This criterion indicates that the AEP-I IRT model is preferable with respect to the other models considered.

We applied the PPMC method to assess the fit of each of the IRT models. We calculated the values of discrepancy measure OSD and compared the observed and predicted score distribution through plots of the discrepancy measures. Figure 3 and Figure 4 show the OSD and KL discrepancy measures, respectively. The PPMC method provides graphical evidence that the Probit and AEP-I models cannot adequately explain the observed score

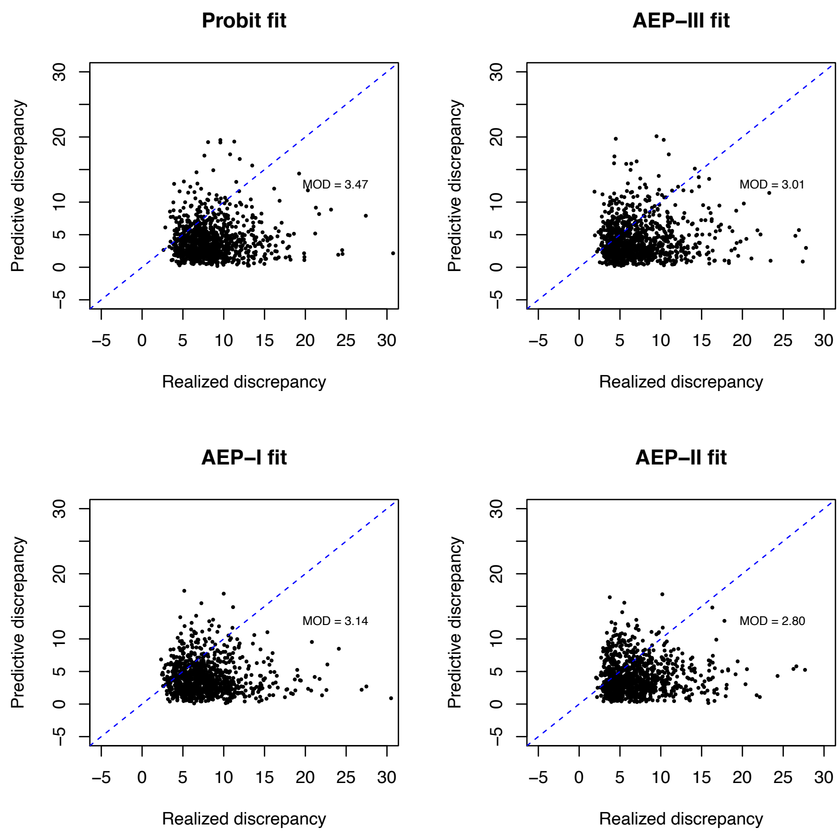

Figure 3. PPMC based on the OSD discrepancy.

distribution of the actual dataset, even though the AEP-I model seems preferable under the DIC criterion. As a quick numerical summary of the plots, we also calculated the average orthogonal distances from the 45-degree line to the points given by the realized and predictive OSDs. We used this mean orthogonal distance (MOD) to provide a quantitative measure of the fit (included in Figures 3 and 4). A large value of the MOD suggests that the model does not adequately capture the features of the data.

Neither the results in Figure 3 nor those in Figure 4 are in agreement with the results obtained using the DIC (see Table 5). The DIC has been criticized on several grounds (see Spiegelhalter et al., 2014). In this particular application, the PPMC procedure seems to yield better results.

6. Concluding remarks

In this paper, we have proposed the use of link functions based on the asymmetric exponential power distribution to model item response data. These link functions provide great flexibility to model a wide range of item characteristic curve shapes and include the symmetric probit model as a special case. The resulting model can handle both symmetry/asymmetry and light/heavy tails at the same time.

In contrast with traditional approaches to IRT modeling, the Bayesian approach has a number of advantages. For one thing, inferences based on posterior simulations are both more flexible and relatively easy to implement in `JAGS` within the `R` software. Also, the possible lack of identifiability in the general IRT model may be tackled using suitable prior distributions.

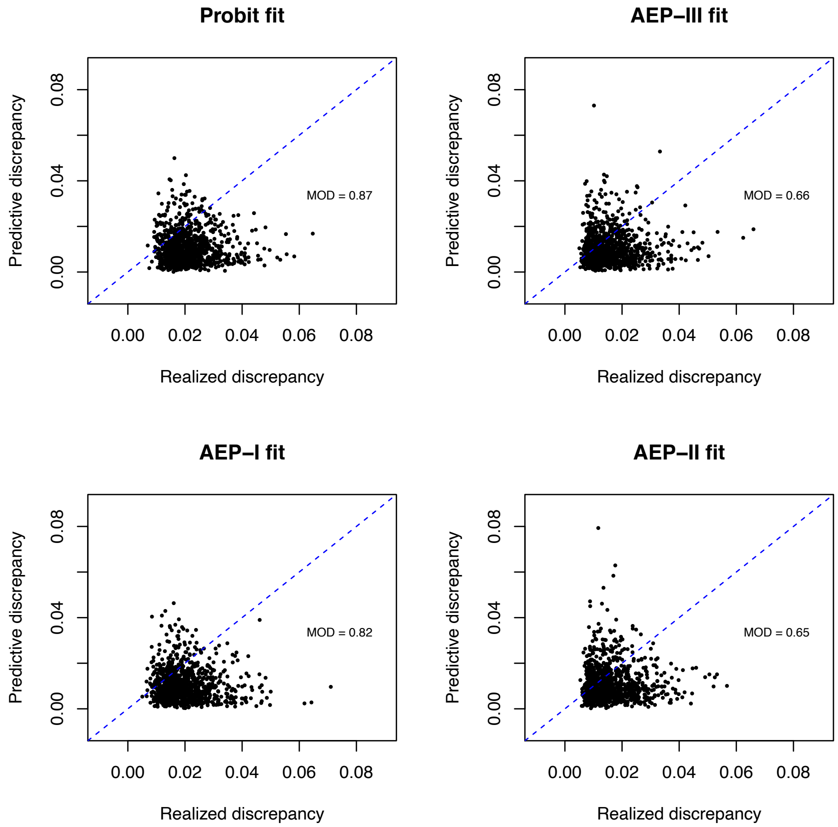

Figure 4. PPMC based on the D_{KL} discrepancy.

Our simulation study shows that the general IRT AEP-based model and the corresponding Bayesian estimates perform well. Our results also suggests that the DIC does not provide a good measure of model fit in our setting, perhaps because it is not based on a proper predictive criterion. By contrast, the posterior predictive model-checking procedure used here provides a nice graphical summary and, together with the mean orthogonal distance, provides a better way of comparing models. Moreover, in the real data example the Kullback-Leibler discrepancy proposed here seems to outperform the OSD discrepancy.

Acknowledgements

The authors would like to thank an Associate Editor and two anonymous reviewers for valuable comments which helped us improve the presentation of the paper. The first author was supported by a postdoctoral research fellowship from CONACYT, Mexico. The work of the second author was partially supported by the Sistema Nacional de Investigadores (CONACYT), Mexico.

Appendix A

Consider the standard version ($\mu = 0$, $\sigma = 1$) of the AEP PDF given in Equation (2) stated as

$$f_{\mathrm{AEP}}(x|\alpha,\delta_1,\delta_2) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} \exp\left\{ -\left| \dfrac{x}{\sqrt{2\pi}\alpha/\Gamma(1+1/\delta_1)} \right|^{\delta_1} \right\}, & \text{if } x \le 0; \\ \dfrac{1}{\sqrt{2\pi}} \exp\left\{ -\left| \dfrac{x}{\sqrt{2\pi}(1-\alpha)/\Gamma(1+1/\delta_2)} \right|^{\delta_2} \right\}, & \text{if } x > 0. \end{cases}$$

The CDF of the AEP distribution is expressed by

- For $x \le 0$,

$$\begin{aligned} F(x|\alpha,\delta_1,\delta_2) &= \int_{-\infty}^{x} f_{\mathrm{AEP}}(z|\alpha,\delta_1,\delta_2)\,\mathrm{d}z \\ &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\left| \frac{x}{\sqrt{2\pi}\alpha/\Gamma(1+1/\delta_1)} \right|^{\delta_1} \right\} \mathrm{d}z. \end{aligned}$$

Now, making the change of variable

$$t = \left(\frac{|z|}{\sqrt{2\pi}\alpha/\Gamma(1+1/\delta_1)} \right)^{\delta_1}, \qquad dt = \delta_1 |z|^{\delta_1 - 1} \left(\frac{\Gamma(1+1/\delta_1)}{\sqrt{2\pi}\alpha} \right)^{\delta_1} dz = -\delta_1\, t z^{-1}\, dz,$$

we have

$$dz = -\frac{\sqrt{2\pi}\,\alpha\, t^{1/\delta_1 - 1}}{\Gamma(1/\delta_1)}\, dt.$$

Note also that $t \to +\infty$ as $z \to -\infty$, while $t \to t_1(x) \equiv \left(\frac{|x|}{\sqrt{2\pi}\alpha/\Gamma(1+1/\delta_1)} \right)^{\delta_1}$ as $z \to x$. Hence,

$$\begin{aligned} F(x|\alpha,\delta_1,\delta_2) &= \alpha \int_{t_1(x)}^{\infty} \frac{t^{1/\delta_1 - 1} \exp(-t)}{\Gamma(1/\delta_1)}\, \mathrm{d}t \\ &= \alpha \left(1 - \int_0^{t_1(x)} \frac{t^{1/\delta_1 - 1} \exp(-t)}{\Gamma(1/\delta_1)}\, \mathrm{d}t \right) \\ &= \alpha \left[1 - G\left(t_1(x); \frac{1}{\delta_1}, 1 \right) \right]; \quad x \le 0, \end{aligned} \tag{9}$$

where $G()$ denotes the gamma CDF.

- For $x > 0$,

$$
\begin{aligned}
F(x|\alpha,\delta_1,\delta_2) &= \int_{-\infty}^{x} f_{\mathrm{AEP}}(z|\alpha,\delta_1,\delta_2)\,\mathrm{d}z \\
&= \int_{-\infty}^{0} f_{\mathrm{AEP}}(z|\alpha,\delta_1,\delta_2)\,dz + \int_{0}^{x} f_{\mathrm{AEP}}(z|\alpha,\delta_1,\delta_2)\,\mathrm{d}z \\
&= \alpha + \int_{0}^{x} f_{\mathrm{AEP}}(z|\alpha,\delta_1,\delta_2)\,\mathrm{d}z \qquad \text{by Equation (9)}
\end{aligned}
$$

  Similarly to the previous case, making the change of variable

$$
t = \left(\frac{|z|}{\sqrt{2\pi}(1-\alpha)/\Gamma(1+1/\delta_2)}\right)^{\delta_2}, \qquad dt = \delta_2 z^{\delta_2-1}\left(\frac{\Gamma(1+1/\delta_2)}{\sqrt{2\pi}(1-\alpha)}\right)^{\delta_2} dz,
$$

  we have

$$
\begin{aligned}
F(x|\alpha,\delta_1,\delta_2) &= \alpha + (1-\alpha)\int_{0}^{t_2(x)} \frac{t^{1/\delta_2-1}\exp(-t)}{\Gamma(1/\delta_2)}\,\mathrm{d}t \\
&= \alpha + (1-\alpha)G\left(t_2(x);\frac{1}{\delta_2},1\right); \quad x > 0,
\end{aligned}
$$

  where

$$
t_2(x) = \left(\frac{|x|}{\sqrt{2\pi}(1-\alpha)/\Gamma(1+1/\delta_2)}\right)^{\delta_2}.
$$

Appendix B. JAGS implementation

The proposed models were all implemented in `JAGS` using the `R2jags` package to fit the models and to perform convergence diagnostics right within `R`. Here we use the data set of Section 5.2 to illustrate the implementation of our model in `JAGS`.

(1) **Packages**. Load the required `R` packages:

```
library(R2jags)
library(coda)
library(lattice)
library(R2WinBUGS)
library(rjags)
```

(2) **Data.** Read the data from the working directory:

```
setwd("my directory")
cito<-matrix(read.table(file="cito.txt", sep=","),200,5,byrow=T)
N<-dim(cito)[1]; K<-dim(cito)[2]
cito.data <- list("cito","N","K")
```

(3) **The model.** Write the model in BUGS code and save it as “cito.model.jags” in the working directory.

```
model{
    for (i in 1:N){
         for(k in 1:K){
             p[i,k]<-phi(a[k]*theta[i]-b[k])
             Y[i,k] ~ dbern(p[i,k])
                    }
      theta[i] ~ dnorm(0,1)
               }
    for (i in 1:K){
      a[k] ~ dnorm(1,1)T(0,)
      b[k] ~ dnorm(0,1)
           }
     }
```

(4) **Parameters.** Define the parameters of interest:

```
cito.params<-c("a","b")
```

(5) **Initial values.** Define the starting values for the MCMC runs:

```
cito.inits<-function()
  {
     list("a"=c(0.5,0.5,0.5,0.5,0.5), "b"=c(0,0,0,0,0))
  }
```

Alternatively, specify separate starting values for each chain:

```
units1<-list("a"=c(0.1,0.1,0.1,0.1,0.1), "b"=c(-4,-4,-4,-4,-4))
units2<-list("a"=c(3,3,3,3,3), "b"=c(4,4,4,4,4))
cito.inits2<-list(units1, units2)
```

(6) **Fit.** Fit the model in `JAGS`:

```
set.seed(123)
fit.cito<-jags(data=cito.data, inits = cito.inits2, parameters.to.save=
         cito.params, n.chains =2, n.iter = 9000, n.burnin=1000,
                  model.file="cito.model.jags")
    print(fit.cito)
```

(7) **Diagnostic.** Convert the model output into an MCMC object in order to have access to several convergence diagnostics:

```
cito.mcmc<-as.mcmc(cito,fit)
xyplot(cito.mcmc,layout=c(2,6), aspect="fill")
densityplot(cito.mcmc)
autocorr.plot(cito.mcmc)
elman.plot(cito.mcmc)
geweke.diag(cito.mcmc)
geweke.plot(cito.mcmc)
raftery.diag(cito.mcmc)
raftery.plot(cito.mcmc)
heidel.diag(cito.mcmc).
```

References

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

Azevedo, C.L., Andrade, D.F., and Fox, J.P., 2012. A Bayesian generalized multiple group IRT model with model-fit assessment tools. Computational Statistics and Data Analysis, 56, 4399–4412.

Azevedo, C.L., Bolfarine, H., and Andrade, D.F., 2011. Bayesian inference for skew-normal IRT model under the centred parametrization. Computational Statistics and Data Analysis, 55, 353–365.

Bacci, S., Bartolucci, F., and Gnaldi, M., 2014. A class of multidimensional latent class IRT models for ordinal polytomous item responses. Communications in Statistics: Theory and Methods, 43, 4:787–800.

Baker, F. and Kim, S., 2004. Item Response Theory: Parameter Estimation Techniques. Statistics: A Series of Textbooks and Monographs. Taylor and Francis, New York.

Bazán, J.L., Branco, M.D., and Bolfarine, H., 2006. A skew item response model. Bayesian Analysis, 1, 861–892.

Béguin, A.A. and Glas, C.A.W., 2001. MCMC estimation and some model-fit analysis of multidimensional IRT models. Psychometrika, 66, 541–562.

Casella, G. and George, E.I., 1992. Explaining the Gibbs sampler. The American Statistician, 46, 167–174.

Chen, M.H., Dey, D.K., and Shao, Q.M., 1999a. A new skewed link model for dichotomous quantal response data. Journal of the American Statistical Association, 94, 1172–1186.

Chen, M.H., Ibrahim, J.G., Shao, Q.M., and Weiss, R.E., 2003. Prior elicitation for model selection and estimation in generalized linear mixed models. Journal of Statistical Planning and Inference, 111, 57–76.

Chen, M.H., Ibrahim, J.G., and Yiannoutsos, C., 1999b. Prior elicitation, variable selection and Bayesian computation for logistic regression models. Journal of the Royal Statistical Society B, 61, 223–243.

Chib, S. and Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. The American Statistician, 49:327–335.

Choi, Y.J. and Asilkalkan, A., 2019. R packages for item response theory analysis: Descriptions and features. Measurement: Interdisciplinary Research and Perspective, 17, 168–175.

Durante, D., 2019. Conjugate bayes for probit regression via unified skew-normal distributions. Biometrika, 106, 765–779.

Fox, J.P., 2010. Bayesian Item Response Modeling. Theory and Applications. Springer.

Fox, J.P. and Glas, C.A., 2001. Bayesian estimation of a multilevel IRT model using Gibbs sampling. Psychometrika, 66, 271–288.

Fu, Z.H., Tao, J., and Shi, N., 2009. Bayesian estimation in the multidimensional three-parameter logistic model. Journal of Statistical Computation and Simulation, 79, 819–835.

Gelfand, A.E. and Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. Journal of American Statistical Association, 85, 398–409.

Ghosh, M., Ghosh, A., Chen, M.H., and Agresti, A., 2000. Noninformative priors for one-parameter item response models. Journal of Statistical Planning and Inference, 88, 99–115.

Jiang, X., Dey, D.K., Prunier, R., Wilson, A.M., and Holsinger, K.E., 2013. A new class of the flexible link function with application to species co-occurrense in cape floristic region. The Annals of Applied Statistics, 7, 2180–2204.

Karabatsos, G., 2016. Bayesian nonparametric response models. In Handbook of Item Response Theory. Volume I: Models. (van der Linden, W.J., editor). CRC Press, Boca Raton, FL, US, pp. 323–336.

Lord, F.M., 1952. A theory of test scores. Psychometric Monograph, 7, 84.

Lord, F.M., 1980. Applications of Item Response Theory to Practical Testing Problems. Lawrence Erlbaum Associates, NJ, US.

Matteucci, M., Mignani, S., and Veldkamp, B.P., 2012. Prior distributions for item parameters in IRT models. Communications in Statistics: Theory and Methods, 41, 2944–2958.

Metropolis, N., Rosenbluth, A.W., Rosenbulth, M.N., Teller, A.H., and Teller, E., 1953. Equations of state calculations by fast computing machines. Journal of Chemical Physics, 21, 1087–1091.

Naranjo, L., Pérez, C., and Martín, J., 2015. Bayesian analysis of some models that use the asymmetric exponential power distribution. Statistics and Computing, 25, 497–514.

Patz, R.J. and Junker, B.W., 1999. A straightforward approach to Markov chain Monte Carlo methods for item response models. Journal of Educational and Behavioral Statistics, 24, 146–178.

Plummer, M., 2017. JAGS Version 4.3.0 user Manual.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rasch, G., 1961. On General Laws and the Meaning of Measurement in Psychology. University of California Press, Berkeley, California, US.

Reckase, M.D., 2009. Multidimensional Item Response Theory. Springer, New York.

Rupp, A.A., Dey, D.K., and Zumbo, B.D., 2004. To Bayes or not to Bayes, from whether to when: Applications of Bayesian methodology to modeling. Structural Equation Modeling: A Multidisciplinary Journal, 11, 424–451.

Samejima, F., 2000. Logistic positive exponent family of models: Virtue of asymmetric item characteristic curve. Psychometrika, 65, 319–335.

San Martin, E., Jara, A., Rolin, J.M., and Mouchart, M., 2011. On the Bayesian nonparametric generalization of IRT-type models. Psychometrika, 76, 385–409.

Sinharay, S., Johnson, M.S., and Stern, H.S., 2006. Posterior predictive assessment of item theory models. Applied Psychological Measurement, 30, 298–321.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and van der Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society B, 64, 583–639.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and van der Linde, A., 2014. The deviance information criterion: 12 years on. Journal of the Royal Statistical Society B, 76, 485–493.

Spiegelhalter, D.J., Thomas, A., Best, N.G., and Lunn, D., 2003. WinBUGS Version 1.4 User Manual.

Stan Development Team, 2014. Stan Modeling Language: User's Guide and Reference Manual. Version 2.2.0.

Su, Y.S. and Yajima, M., 2020. R2jags: Using R to Run 'JAGS'. R package version 0.6-1.

Svetina, D., 2013. Assessing dimensionality of noncompensatory multidimensional item response theory with complex structures. Educational and Psychological Measurement, 73, 312–338.

van der Linden, W.J. and Hambleton, R.K., 1997. Handbook of Modern Item Response Theory. Springer, New York.

Zhu, D. and Zinde-Walsh, V., 2009. Properties and estimation of asymmetric exponential power distribution. Journal of Econometrics, 148, 86–99.

Information for authors

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF electronic version of the manuscript in a free format to the Editors-in-Chief of the ChJS (E-mail: `chilean.journal.of.statistics@gmail.com`). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a"cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

Preparation of accepted manuscripts

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at `http://soche.cl/chjs/files/ChJS.zip`. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at `http://www.ams.org/mathscinet/msc/`. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., 1994. Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., 1997. Local influence. In Kotz, S., Read, C.B., and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 1., Wiley, New York, pp. 380-385.

Rukhin, A.L., 2009. Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters, 79, 1004-1007.

Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Peña, D., and Pankratz, A.E., 2000. Outliers in multivariate time series. Biometrika, 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

Copyright