Bayesian Item Response Model when Performance is Affected by Test Anxiety

Cibele Queiroz da-Silva and Antonio Eduardo Gomes
Departmento de Estatística
Instituto de Ciências Exatas - Universidade de Brasília
Campus Universitário Darcy Ribeiro - Asa Norte
70910-900 - Brasília - DF - Brazil, cibeleqs@unb.br

ABSTRACT
We develop a Bayesian binary Item Response Model (IRM), which we denote as Test Anxiety Model (TAM), for estimating the proficiency scores when individuals might experience test anxiety. We consider order restricted item parameters conditionally to the examinees’ reported emotional state at the testing session. We consider three test anxiety levels: calm, anxious and very anxious. Using simulated data we show that taking into account test anxiety levels in an IRM help us to obtain fair proficiency estimates as opposed to the ones obtained with the two parameter logistic IRM (2PM) by Birnbaum (1968). For the 2PM, the proficiency estimates tend to be positively biased for both calm and anxious examinees.

1 Introduction

There is an extensive concern about test anxiety. When we look for pages related to this issue on the Internet, we find about 4,980,000 sites, many of them bringing counseling for dealing with the problem.

In 2005, the Education Testing Services (ETS) (online at www.ets.org/assessments), made public at the Internet a document entitled “Reducing Test Anxiety - A Guide for Praxis Test Takers”. This document describes some mental and physical symptoms of anxiety. Mental symptoms include mental blank-out, difficulty concentrating, negative thoughts and knowing the answers after the test, but not while taking it. Physical symptoms include nausea, cramps, faintness, sweating and increased breathing rate. The document also points out to the test takers some practical tips about good study habits and strategies to cope with test anxiety. (see www.ets.org/Media/Tests/PRAXIS/pdf/01361anxiety.pdf).

When Item Response Theory estimation of examinees’ proficiencies presents group related bias, either for or against a particular group, we say that the items are affected by differential item functioning (DIF) (see Hanson, 1998). It is possible that even when two examinees have exactly the same proficiency level, their estimated performances may be quite different, and those differences be caused by the group they belong to (eg. men or women).

IRT modelling of proficiencies accounting for test anxiety can be viewed as a DIF analysis. In this article we focus on a Bayesian extension of the two-parameter logistic Item
Response Model (2PM) (Birnbaum, 1968). We propose a methodology that allows for differential test anxiety levels in the estimation of the examinee’s proficiency scores. Our model acknowledges the test taker’s emotional stress and its consequential temporary cognitive interferences. The Bayesian approach offers powerful and flexible procedures in item response theory setting, especially due to the growing complexity of the item response theory models.

Using simulated data we show that taking into account test anxiety levels in an item response model help us to obtain fair proficiency estimates as opposed to the ones obtained with the 2PM model. For the 2PM, the proficiency estimates tend to be positively biased for both calm and anxious examinees.

2 The test anxiety model - TAM

Suppose that the examinees that will take a test are asked to complete a questionnaire at the testing session. The aim of this questionnaire is to assess the examinee’s anxiety level just before the test. Suppose that \( G = 3 \) increasingly ordered test anxiety levels are obtained after compiling and analysing the mentioned questionnaire: calm, anxious and very anxious.

In our model the conditional probability of the examinee \( j \) answering item \( i \) correctly, when his ability is \( \theta_j \) and his test anxiety level is \( g \) will be modeled as

\[
P_{ijg} = \frac{1}{1 + \exp \left[-\beta_2i\theta_j + \beta_1gi\right]} \tag{1}
\]

where \( \beta_2i \) is the discrimination parameter for item \( i \) while \( \beta_1gi \) is a location parameter (intercept) for item \( i \), for examinees with test anxiety level \( g \). It is reasonable to suppose that \( P_{ijg} \geq P_{ijv} \), for \( v > g \). That is, for the examinee \( j \), the higher the test anxiety level \( g \) observed, the smaller the probability of answering item \( i \) correctly.

For \( i \) and \( j \) fixed, when the \( P_{ijg} \)'s are arranged in increasing order, that is

\[
P_{ij1} \geq P_{ij2} \geq \ldots \geq P_{ijG},
\]

it implies that the \( \beta_1gi \)'s should be arranged in decreasing order,

\[
\beta_1i1 \leq \beta_1i2 \leq \ldots \leq \beta_1iG.
\]

Our model is formulated for binary items only and there exists an intercept parameter \( \beta_1gi \) for each observed level (group) \( g \) of an ordered test anxiety scale. The discrimination parameters \( \beta_2i \) are supposed to vary with the items but not with the test anxiety observed categories. Our model accounts for DIF by allowing for differential test anxiety levels in the estimation of the examinee’s proficiency scores.

2
The likelihood is given by
\[ L(D | \theta, \beta_1, \beta_2) = \prod_{j=1}^{n} \prod_{i=1}^{I} \prod_{g=1}^{G} P_{ijg}^{u_{ij}} v_{ijg} Q_{ijg}^{(1-u_{ij})v_{ijg}} \]
\[ = \prod_{j=1}^{n} \prod_{i=1}^{I} \prod_{g=1}^{G} [P_{ijg}(\theta_j, \beta_{1gi}, \beta_{2gi})]^{u_{ij}} v_{ijg} [Q_{ijg}(\theta_j, \beta_{1gi}, \beta_{2gi})]^{(1-u_{ij})v_{ijg}}. \tag{2} \]
under the constraint that for every fixed \( i \) and \( j \), \( P_{ij1} \geq P_{ij2} \geq \ldots \geq P_{ijG} \).

3 Bayesian test anxiety model

Following Soares et al. (2009), and also using some of their notation, in our model we also admit that the examinees are grouped into \( G \) groups of test anxiety and that
\[ \theta_j \sim N(\mu_{g(j)}, \sigma^2_{g(j)}). \]
In order to ensure the models’ identifiability, we define group 1 as a reference group and assume that \( \lambda_1 = (\mu_1, \sigma^2_1) = (0, 1) \). The other groups are described as focal groups, and their associated proficiency distribution parameters are represented by \( \lambda_g = (\mu_g, \sigma^2_g) \), \( g = 2, \ldots, G \) and will be estimated. Let \( \lambda = (\lambda_2, \ldots, \lambda_G) \).

Comparability among the estimated proficiencies for examinees of different groups is only possible when part of the items in the test has no DIF. That implies that
\[ \beta_{11i} = \beta_{12i} = \ldots = \beta_{1Gi}, \quad i \in I_A \subset \{1, \ldots, I\}. \]
That is, even when an examinee is under extreme test anxiety, he or she will not react differently to item \( i \) than another less anxious examinee. Such restriction also ensures convergence of the generated MCMC chains in the Bayesian analysis.

For expressing the joint posterior of \( \Psi = \{\Psi_1, \Psi_2\} \), where \( \Psi_1 = \{\theta, \lambda\} \) and \( \Psi_2 = \{\beta_1, \beta_2\} \), we will assume that the parameters are independent a priori and that \( D \) denotes the set of observed data. Therefore,
\[ \pi(\Psi | D) \propto L(D | \theta, \beta_1, \beta_2) \pi(\Psi) \]
\[ \propto L(D | \theta, \beta_1, \beta_2) \pi(\Psi_1) \pi(\Psi_2). \tag{3} \]

3.1 Prior specifications

Following Soares et al. (2009), we assume that the
\[ \pi(\Psi_1) = \prod_{j=1}^{n} \pi(\theta_j | \mu_{g(j)}, \sigma^2_{g(j)}) \prod_{g=2}^{G} \pi(\mu_g) \pi(\sigma^2_g), \]
3
where \( (\theta_j \mid \mu_{g(j)}, \sigma_{g(j)}^2) \sim N(\mu_{g(j)}, \sigma_{g(j)}^2), \mu_g \sim N(m_0, \sigma_0^2) \) and \( \sigma_g^2 \sim IG(\alpha_0, \beta_0) \), where \( IG \) denotes an inverse-gamma distribution. We also assume that \( \pi(\Psi_2) = \prod_{i=1}^{G} \prod_{g=1}^{G} \pi(\beta_{1gi})\pi(\beta_{2i}) \).

For the prior specification of \( \beta_{2i} \), we follow Patz and Junker (1999) and let the \( \beta_{2i} \)'s be IID lognormal(\( \mu, \sigma_{\beta}^2 \)). It can be shown that \( p(x) = \text{lognormal}(x \mid \mu, \sigma^2) \) implies that \( \frac{\ln(x) - \ln(\mu)}{\sigma} \sim N(0, 1) \). Therefore, the prior for \( \beta_{2i} \) is \( \beta_{2i} \sim \text{log-normal}(0, \sigma_{\beta}^2) \), \( i = 1, \ldots, I \).

For dealing with \( \beta_{1i} \), due to the mentioned imposed constraints, \( \beta_{11i} \leq \beta_{12i} \leq \ldots \leq \beta_{1Gi} \), for a given item \( i \), we follow Gelfand et al. (1992). Therefore, the prior for \( \beta_{1gi} \) is

\[
\beta_{1gi} \sim N(0, \sigma_{\beta_1}^2), \quad \text{i, 1, \ldots, I, and, } \quad g = 1, \ldots, G.
\]

### 4 Simulated data and Results

For simulating the data to be used in the analysis we considered \( G = 3 \) groups of increasing test anxiety levels, that we denote as calm, anxious and very anxious. We worked with 20 items, with 15 of them presenting differential response to the anxiety levels and the remaining ones were left as anchor items. The values of the \( \beta_{1gi} \) parameters used in the simulations are presented in column 3 of Tables 1 and 2. The discrimination parameters, \( \beta_{2i} \), were fixed at 1.0, 1.0, 1.5, 1.5, 2.0, 2.0, 1.0, 1.0, 1.5, 1.5, 2.0, 2.0, 1.0, 1.0, 1.5, 1.5, 2.0, 2.0, 1.0 and 1.0 for items 1 to 20, respectively. However, for brevity, we will only present analyses for the \( \beta_{1gi} \) parameters and for the proficiencies \( \theta_j \).

For each of the \( G = 3 \) groups, we considered \( n_g = 2,000 \) examinees, \( g = 1, 2, 3 \). Their proficiencies were simulated according to normal distributions as follows:

- for \( g = 1 \), the calm examinees, we considered that \( \theta_{1g} \sim N(0, 1) \), in order to ensure the models’ identifiability. We denote this group as the reference group;
- for \( g = 2 \), the nervous examinees, we considered that \( \theta_{2g} \sim N(-0.1, 1) \). We denote this group as the focal group I;
- for \( g = 3 \), the very nervous examinees, we considered that \( \theta_{3g} \sim N(-0.5, 1) \). We denote this group as the focal group II.

Conditioning on \( \theta_{1g} \), the data for 6,000 examinees on the 20 binary items were generated according to a Bernoulli distribution with success probability given by model (2.1):

\[
P_{ijg} = \frac{1}{1 + \exp[-\beta_{2i}\theta_{1g} + \beta_{1gi}]}, \quad i = 1, \ldots, 20, \quad j \in J(g) \text{ and } g = 1, 2, 3.
\]

Using simulated data we show that taking into account test anxiety levels in an IRM help us to obtain fair proficiency estimates as opposed to the ones obtained with the two parameter logistic IRM (2PM) by Birnbaum (1968). For the 2PM, the proficiency estimates tend to
be positively biased for both, calm and anxious examinees. That means that disregarding that test anxiety plays a role in the examinees’ performances can lead to unfair estimated proficiency scores, with the calm individuals tending to get better scores than the anxious ones, even when two such individuals share the very same proficiency.

**Bibliography**


