ANADOLU UNIVERSITY JOURNAL OF EDUCATION FACULTY
ISSN: 2602-2249
Volume:6 Issue:1 2022 JANUARY

irtDemo R Paketi: Madde Tepki Kuramında Tahmin, Puanlama ve Çok Boyutluluk için Pedagojik Amaçlı Etkileşimli Web Uygulamaları

irtDemo R Package: Pedagogical Interactive Web Applications for Estimation, Scoring, and Multi Dimensionality in Item Response Theory

Metin BULUS[1] & Wes BONIFAY[2]

Article Type[3]: Technical Brief

Application Date: 12.04.2021 **Accepted Date:** 03.01.2022

To Cite This Article: Bulus, M., & Bonifay, W. (2022). irtDemo R Package: Pedagogical Interactive Shiny Web Applications for Estimation, Scoring, and Multi Dimensionality in Item Response Theory. *Anadolu Üniversitesi Eğitim Fakültesi Dergisi (AUJEF), 6*(1), 92-108.

ÖZ: Temel düzeyde ancak oldukça karmaşık istatistiksel teorilerin anlaşılması, temellerinde yatan denklemleri ve teoriyi anlamak için yardımcı etkileşimli teknolojik araçlar gerektirebilir. Bu çalışmada, tahmin, puanlama ve çok boyutluluk gibi bazı temel ancak karmaşık madde tepki kuramı kavramlarını göstermek veya keşfetmek için interaktif web uygulamaları koleksiyonu sunulmuştur. İnteraktif web uygulamaları *shiny* R paketi kullanılarak oluşturulmuştur. Kullanıcılar bu uygulamalara hem *irtDemo* R paketinden hem de bu çalışmada verilen linkleri kullanarak erişebilirler. Bu uygulamaların, gelişmiş ölçme konularıyla ilgilenen uygulayıcılara ve araştırmacılara başlangıç için bir avantaj sağlayacağını düşünülmektedir.

Anahtar sözcükler: madde tepki kuramı, çok boyutlu madde tepki kuramı, maksimum olabilirlik kestirimi, beklenen sonsal yetenek kestirimi, maksimum sonsal yetenek kestirimi

ABSTRACT: Comprehension of foundational but fairly complex statistical theories may require assistive interactive tools to understand underlying equations and theory. We provide a collection of interactive web applications to demonstrate or explore some of the fundamental yet complex item response theory concepts such as estimation, scoring and multidimensionality. Interactive web applications were developed via *shiny* R package. Users can access to these applications through *irtDemo* R package or links provided in this article. We hope that these applications give a head-start to emerging practitioners and researchers interested in advanced measurement topics.

Keywords: item response theory, multidimensional item response theory, maximum likelihood estimation, expected a posteriori ability estimation, maximum a posteriori ability estimation

[1] Lecturer, PhD, Adiyaman University, bulusmetin@gmail.com, ORCID: 0000-0003-4348-6322 (Corresponding author)
[2] Assoc. Prof., PhD, University of Missouri, bonifayw@missouri.edu, ORCID: 0000-0003-3853-7607
[3] An earlier version of this draft was submitted to the National Council on Measurement Education 2018 Annual Meeting held in New York with the title "*Pedagogical Applications for Estimation and Scoring in Item Response Theory*". This study was supported by the University of Missouri – Columbia.

1. INTRODUCTION

Recent research has demonstrated the effectiveness of computer-based activities in helping students to better comprehend statistical concepts (Braun, White, & Craig, 2014). Further, Garfield and Ben-Zvi (2007) assert that knowledge retention and student enjoyment will increase when traditional statistical instruction is supplemented with interactive applications that allow for exploration of statistical results. While statistical concepts can be explored by directly manipulating R code and observing the results, this process can be cumbersome and confusing to students, especially those who are new to the R environment. The `shiny` R package (Chang et al., 2017) converts complex R code into a simple interface that allows users to interact with and explore various concepts by varying parameters and visualizing output. Thus, the `shiny` R package (Chang et al., 2017) is a pedagogical tool that has great potential for elevating student knowledge, especially regarding topics as dense and advanced as item response theory (IRT).

The aim of this study is to introduce theoretical underpinnings for fundamental yet complex IRT concepts and provide pedagogical web applications for graduate students and scholars. We introduce five such applications that are intended to demonstrate and visualize the concepts of estimation, scoring, and multidimensionality. These applications were developed for classroom instruction using the R statistical software program (R Core Team, 2019) and the `shiny` R package (Chang et al., 2017). Users can access to these applications using `irtDemo` R package (Bulus & Bonifay, 2016) with the simple R interface or RStudio. Alternatively, applications can also be accessed through links provided in this article.

We do not intend to provide an elaborate introduction to the theory underlying web applications. We briefly summarize key equations and introduce applications for five fundamental yet complex IRT concepts: maximum likelihood estimation (MLE) of person location in a Rasch model given item difficulties, MLE in the 2PL model, MLE in the 3PL Model, expected a posteriori (EAP) and maximum a posteriori (MAP) ability scoring, and multidimensional dichotomous IRT models. They are described below.

2. APPLICATIONS

Applications can be accessed through `irtDemo` R package. The package can be installed using the `install.packages("irtDemo")` command in the R environment. It should be loaded into the current R session using the `library(irtDemo)` command.

```
# Install the package
install.packages("irtDemo")
# Load into the current R session
library(irtDemo)
```

Application 1: Maximum Likelihood Estimation (MLE) of Person Location in a Rasch Model Given Item Difficulties

Conceptual Underpinnings

The theory and concepts underlying MLE of the person location in Rasch model is mostly drawn from De Ayala (2013), however a more elaborate explanation can be found in Baker and Kim (2004). In the Rasch model, for a person j with ability θ_j, probability of endorsing an item i with difficulty δ_i can be modeled as

$$P(x_{ij}|\theta_j, \delta_i) = \frac{e^{x_{ij}(\theta_j - \delta_i)}}{1 + e^{(\theta_j - \delta_i)}} \quad (1)$$

where x_{ij} is an indicator variable taking a value of 1 for correct endorsement and 0 for incorrect endorsement. There are often more than one item, therefore endorsing multiple item creates patterns such as $\underline{x}_j$ = 10110 which indicates that the jth person endorsed the firsts item correctly, the second item incorrectly and so on. The patterns observed in the data is often less than number of examines. In light of independence assumption, the likelihood of a given pattern $\underline{x}_j$ for a person with ability θ_j is

$$L(\underline{x}_j|\theta_j, \underline{\delta}) = \prod_{i=1}^{n} P_i^{x_{ij}}(1 - P_i)^{(1-x_{ij})} \quad (2)$$

where n is number of items, P_i is abbreviated version of $P(x_{ij}|\theta_j, \delta_i)$ for item i out of convenience, which was described in Equation 1. In this illustration, however, item difficulties are assumed to be known, therefore only the person location on the ability scale (θ_j) is maximized. The more the number of items the smaller the likelihood gets which poses problems for optimization because of miniscule increments in the likelihood function in Equation 2. Therefore, to avoid range restrictions, this function is log-transformed in terms of log-likelihood as

$$LL(\underline{x}_j|\theta_j, \underline{\delta}) = \sum_{i=1}^{n} [x_{ij} \log(P_i) + (1 - x_{ij}) \log(1 - P_i)] \quad (3)$$

In theory, we can try a finite set of plausible values to find the location of θ that maximizes $LL(\underline{x}_j|\theta_j, \underline{\delta})$ function in Equation 3, but this is impossible in practice because θ is continuous. Thus, there are infinite plausible values. There are two main shortages for this theoretical approach: (i) it is computationally intensive, (ii) we do not obtain standard error for θ_j. Therefore, in practice a software would focus on a feasible region via Tylor series expansion around a point. In more concrete terms, an iterative Newton-Raphson / Fisher scoring procedure is used to find the maximum of log-likelihood over the range of θ parameter space. Note that in this case there is a single equation at each iteration and a single unknown (θ_j), therefore Newton-Raphson equations can be constructed as

$$\hat{\theta}_{t+1} = \hat{\theta}_t - \frac{f'(\hat{\theta}_t)}{f''(\hat{\theta}_t)} \quad (4)$$

where $f(\hat{\theta}_t) = LL(\underline{x}_j|\theta_t, \underline{\delta})$, which was defined in Equation 3. Equation 4 means that if we have an initial guess (starting value) we can improve the guess by adding (or subtracting depending on the sign) the ratio of the slope (first derivative) to the change in slope (second derivative). As the algorithm approaches to the maximum this ratio becomes small, which means a small number is added or

subtracted. This iteration continues until a desired degree of accuracy, say below 0.0001. This dynamic ratio of the first derivative to the second derivative is called the step size. In the context of Rasch modeling, the first and second derivative of the log-likelihood function are

$$\frac{\partial LL(\underline{x}_j|\theta,\underline{\delta})}{\partial\theta} = r_j - \sum_{i=1}^{n} P_{ij}$$
$$\frac{\partial^2 LL(\underline{x}_j|\theta,\underline{\delta})}{\partial^2\theta} = -\sum_{i=1}^{n} P_{ij}(1-P_{ij}) \quad (5)$$

where r_j is observed score that is number of correctly endorsed items given response pattern for person j. Then, the Newton-Raphson equation becomes

$$\hat{\theta}_{t+1} = \hat{\theta}_t - \frac{r_j - \sum_{i=1}^{n} P_{ij}}{-\sum_{i=1}^{n} P_{ij}(1-P_{ij})} \quad (6)$$

Note that the nominator in the second terms in Equation 6 is observed (r_j) minus expected ($\sum_{i=1}^{n} P_{ij}$) score for person j, and the denominator is a function of the information in the scalar form as we are interested in only person location. From this information, we can get standard errors for θ_j, since minus reciprocal of the information is variance for θ_j. We want to continue iteration until the step size satisfies a pre-specified threshold, say 0.0001. Newton-Raphson iterations are expected to continue until difference between observed and expected person score is minimal, while we have highest information possible. This will be θ_j at which the $\underline{x}_j$ response pattern for the person is most likely to occur.

Application

This application helps users to demonstrate, explore, and visualize the concept of maximum likelihood estimation (MLE) in Rasch models. This is a topic that is central to IRT but the non-interactive, equation-based explanations found in textbooks are often daunting to students and other users who are new to statistical estimation. By using the intuitive slider controls rather than typing in syntax, the user will be able to: visualize the concepts of convergence and divergence as the estimation algorithm successfully locates or fails to locate the person ability (θ) parameter; and explore the influence of unreasonable starting values and aberrant scores on maximum likelihood estimation.

Users can run Application 1 via running the R command below. They can also access the same application via https://irtdemo.shinyapps.io/mlest/.

```
# Maximum likelihood estimation
irtDemo("mle")
```

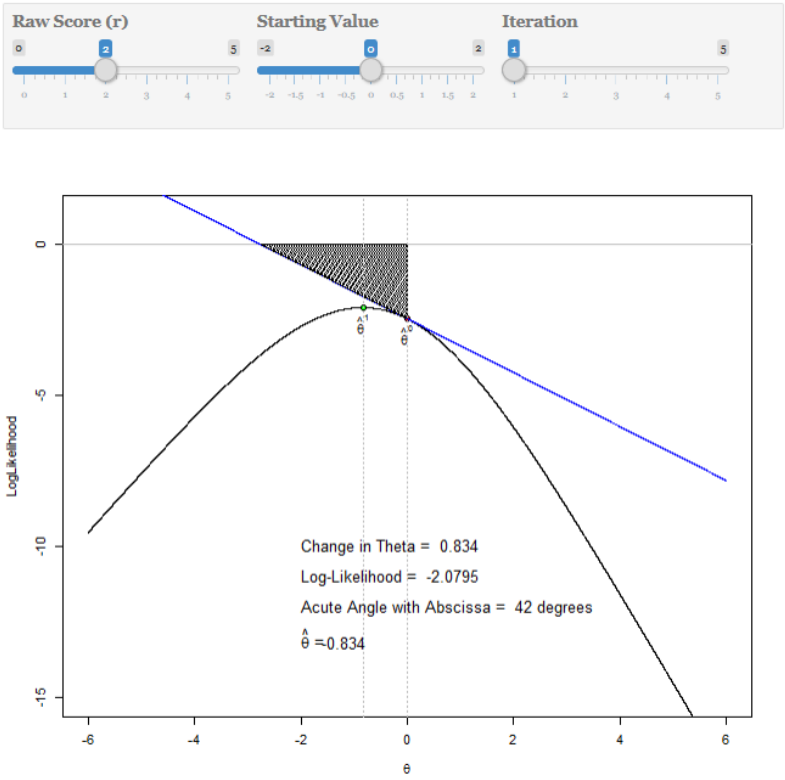

Figure 1: *Maximum likelihood estimation in the Rasch model*

To illustrate MLE in Rasch models, this application replicates the analysis presented in de Ayala (2009, p. 22), wherein the person location (θ_j) parameter is estimated given the item locations of a five item math test. Using slider controls, users can manipulate three inputs and observe various estimation concepts as well as the pitfalls inherent in MLE. The first input is the respondent's raw score, the second input is the starting value for the estimation algorithm, and the third input is the number of estimation iterations. The change in θ_j between iterations, the log-likelihood statistics, the acute angle with the abscissa, and the estimated person ability for that particular iteration are superimposed on the plot to reinforce the aforementioned concepts. Students can be asked to address the following questions:

Which raw score require more iterations?

Can you find a case where the algorithm diverges?

Can you find a case where the algorithm converges?

What happens to the change in θ_j when number of iterations increase?

What happens to convergence when estimating extreme scores such as zero (0)?

What happens when starting values deviate from the actual solution unreasonably?

Similarly to the estimation paradigm for person location (θ_j), we can find item difficulties (δ_i) at which log-likelihood of a pattern is maximized given the vector of examinee location estimates ($\underline{\theta}$). In reality, both are unknown so the algorithm take turns via fixing one of the item parameter or person parameter at the provisional value (often starting values for the first iteration) until the log-likelihood function is maximized jointly. This two-stage estimation is referred to as joint maximum likelihood

estimation (Birnbaum, 1968). Although this may work well with Rasch models, it tends to be problematic with smaller samples and complex models with two or more item parameters to be optimized. There are other more complex estimation paradigms such as conditional maximum likelihood estimation (Andersen, 1972), marginal maximum likelihood estimation (Bock and Atkin, 1981; Bock and Lieberman, 1970), Bayesian estimation which are described in Baker & Kim (2004). Recently for more complex models Metropolis-Hastings Robbins-Monro algorithm has been proposed by Cai (2010). Explanation of these estimation paradigms are beyond the scope of this illustration, however for more information readers are referred to references.

Application 2: Maximum Likelihood Estimation in the 2PL Model

Conceptual Underpinnings

Below we summarize the theory and concepts underlying the Application 2 which are mostly drawn from de Ayala (2013). In the 2PL model, for a person j with ability θ_j probability of endorsing an item i with location (difficulty) δ_i, and item discrimination α_i can be modeled as

$$P(x_{ij}|\theta_j, \delta_i, \alpha_i) = \frac{e^{x_{ij}\alpha_i(\theta_j - \delta_i)}}{1 + e^{\alpha_i(\theta_j - \delta_i)}} \tag{7}$$

where x_{ij} is an indicator variable taking a value of 1 for correct endorsement and 0 for incorrect endorsement for examinee j and item i, α_i is item discrimination for item i and provides information as to how well an item differentiates examinees with different abilities.

Similar to the Rasch model in Application 1, in light of independence assumption, the likelihood of a pattern $\underline{x}_j$ given person ability θ_j is

$$L(\underline{x}_j|\theta_j, \underline{\delta}, \underline{\alpha}) = \prod_{i=1}^{n} P_i^{x_{ij}}(1 - P_i)^{(1-x_{ij})} \tag{8}$$

where n is number of items, P_i is abbreviated version of $P(x_{ij}|\theta_j, \delta_i, \alpha_i)$ for convenience. Contrary to Application 1, in this illustration, examinees' location are assumed to be known, therefore item location on the ability scale δ_i and discrimination α_i are simultaneously maximized. For similar reasons in Application 1, that is, to avoid range restrictions this function is log-transformed as

$$LL(\underline{x}_j|\theta_j, \underline{\delta}, \underline{\alpha}) = \sum_{i=1}^{n} [x_{ij} \log(P_i) + (1 - x_{ij}) \log(1 - P_i)] \tag{9}$$

To keep matters simple, we do not go in to details of Newton-Raphson equations. However, one may come to realize that Newton-Raphson equations consist of two equations with partial derivatives with respect to δ and α parameters.

Application

This application is the same as Application 1, though it allows users to investigate a slightly more complex IRT model: the two-parameter logistic (2PL) model. This application will help the user to grasp the complexity of estimating one additional parameter (i.e., the discrimination parameter) as compared to the Rasch model; understand that the 2PL MLEs are found by searching across a 3-dimensional surface rather than a line; and explore the influence of item discrimination on the estimation of item location.

Users can run Application 2 via running the R command below. They can also access the same application via https://irtdemo.shinyapps.io/est2pl/.

```
# Maximum likelihood estimation in 2PL model
irtDemo("est2pl")
```

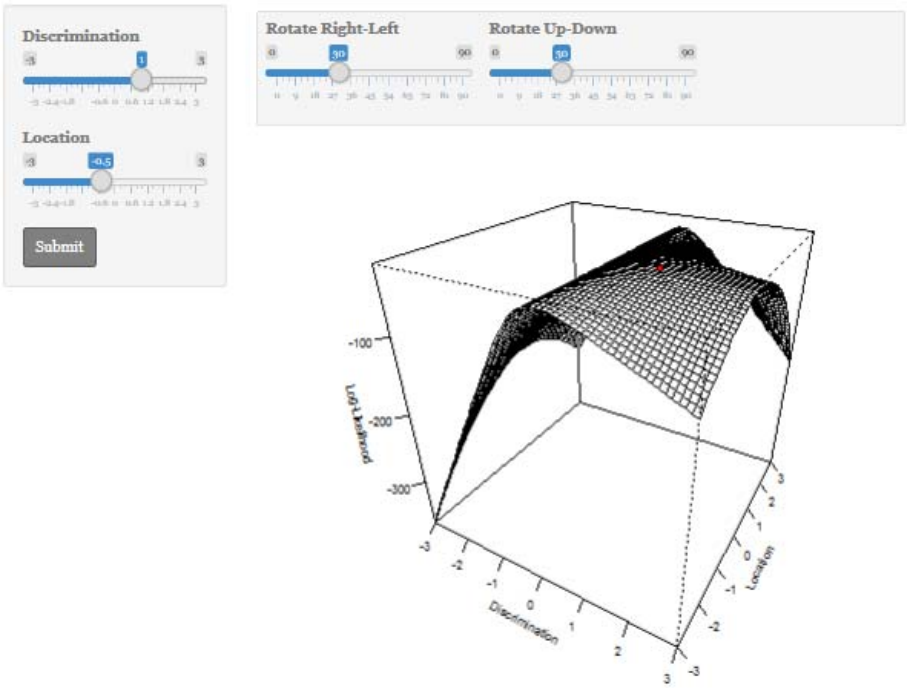

Figure 2: *Maximum likelihood estimation in 2PL model*

The 2PL MLE application simulates 50 responses to an item with discrimination and location parameters determined by the user. Two additional inputs enable the user to change the orientation of the 3D plot to get a better perspective of the estimated item parameters at which the log-likelihood function in Equation 9 is maximized. To keep matters simple, in this application we do not go into details of Newton-Raphson equations. However, one may come to realize that Newton-Raphson equations consist of two equations with partial derivatives with respect to δ and α parameters. In the application we only use empirical finite values to draw the 3D plot. Unlike Application 1, the search would have followed a funnel-like zigzag beginning from starting values and narrowing down to the point of maximum.

Application 3: Maximum Likelihood Estimation in the 3PL Model

Conceptual Underpinnings

In the 3PL model, for a person j with ability θ_j probability of endorsing an item i with location (difficulty) δ_i, and item discrimination α_i, and guessing parameter c_i can be modeled as

$$P(x_{ij}|\theta_j, \delta_i, \alpha_i, c_i) = c_i + \frac{e^{x_{ij}\alpha_i(\theta_j - \delta_i)} - c_i}{1 + e^{\alpha_i(\theta_j - \delta_i)}} \tag{10}$$

where x_{ij} is an indicator variable taking a value of 1 for correct endorsement and 0 for incorrect endorsement for examinee j and item i. Similar to Application 1 and 2 the likelihood of a pattern $\underline{x}_j$ for a person with ability θ_j in light of independence assumption is

$$L(\underline{x}_j|\theta_j, \underline{\delta}, \underline{\alpha}, \underline{c}) = \prod_{i=1}^{n} P_i^{x_{ij}}(1 - P_i)^{(1-x_{ij})} \tag{11}$$

where n is number of items, P_i is abbreviated version of $P(x_{ij}|\theta_j, \delta_i, \alpha_i, c_i)$ for convenience and was defined in Equation 10. Similar to Application 2, in this illustration examinees' location are assumed to be known, therefore item location on the ability scale δ_i, discrimination α_i, and guessing parameter c_i are simultaneously is maximized. Log-likelihood takes the form of

$$LL(\underline{x}_j|\theta_j, \underline{\delta}, \underline{\alpha}, \underline{c}) = \sum_{i=1}^{n}[x_{ij}\log(p_i) + (1 - x_{ij})\log(1 - p_i)] \tag{12}$$

In this application, to keep matters simple, we do not go in to details of Newton-Raphson equations. However, one should get the intuition that in this case the Newton-Raphson equations consist of three partial derivatives with respect to δ, α and c parameters. In the application we only use empirical finite values to draw the 3D plot, layers representing the 4th dimension.

Application

Application 3 aids in understanding estimation of the three-parameter logistic (3PL) model (Birnbaum, 1968), one of the most common, yet complex model used in practice today for the analysis of dichotomous response data. In particular, this application will help the user to comprehend the complexity of estimating an additional parameter – the lower asymptote, or "guessing" parameter – as compared to the 2PL model; and realize that 3PL MLEs are searched through a four-dimensional space.

Users can run Application 3 via running the R command below. They can also access the same application via https://irtdemo.shinyapps.io/est3pl/

```
# Maximum likelihood estimation in 3PL model
irtDemo("est2pl")
```

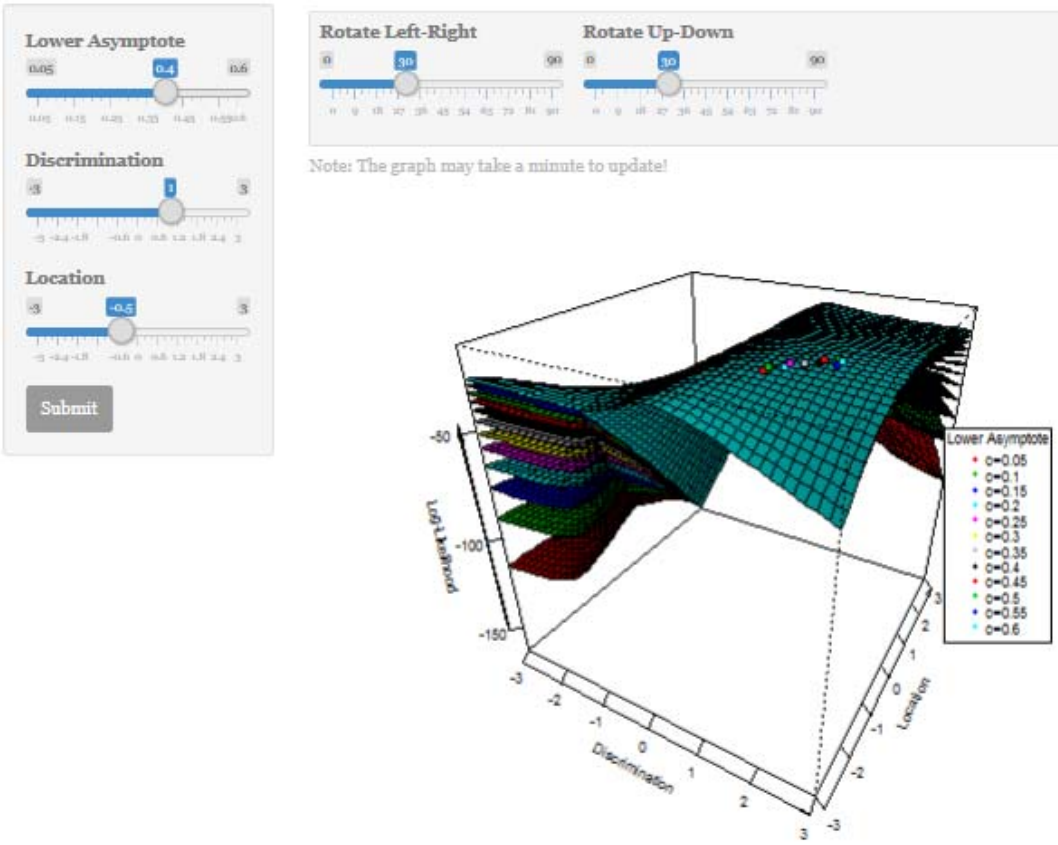

Figure 3: *Maximum likelihood estimation in 3PL model*

Application 3 simulates 50 responses to an item with discrimination, location, and lower asymptote parameters determined by the user. The legend has several plausible values for lower asymptote, so that the user can compare the lower asymptote specified in the slider to the plausible values in the legend. Each plausible value is maximized with respect to discrimination and location. In other words, each surface plot is the feasible region for a given plausible lower asymptote. As in the 2PL MLE application, the user also has the ability to change the orientation of the perspective plot and thereby gain a better understanding of the estimated item parameters at which the log-likelihood function in Equation 12 is maximized in the 3PL model.

Application 4: EAP and MAP Scoring

Conceptual Underpinnings

The likelihood function was defined in Application 1, however in this case the likelihood is estimated via quadrature along the continuum of the ability scale. Thus, the function takes the form of

$$L(\underline{x}_j|q,\underline{\delta}) = \prod_{i=1}^{n} P_i^{x_{ij}}(1-P_i)^{(1-x_{ij})} \tag{13}$$

where q is the quadrature point on the ability scale. As we need to find likelihood along the continuum of the ability scale, the likelihood has to be estimated for each plausible value on the ability scale. This

requires integration. In practice, this is conducted via numerical integration, as such, ability scale is divided uniformly by a pre-determined number of points called quadrature. Then the likelihood function in Equation 13 is computed for each quadrature point and summed across. As the number of quadrature points increases, the likelihood distribution is represented more correctly. However, there is a trade-off between correct representation of likelihood function and computational burden. Often quadrature points around 40 is sufficient for simple models, and is specified as 41 by default in this illustration. The prior information is taken into consideration via multiplying likelihood function by prior distribution weights as

$$f(\hat{\theta}_i) = \sum_q q \; L(\underline{x}_j|q, \underline{\delta})A(q) \tag{14}$$

where $f(\hat{\theta}_i)$ is estimated posterior distribution density for examinee i, $A(q)$ is calculated form prior density by dividing probability of each quadrature point with the overall probability of all quadrature points. The EAP can be found as

$$\hat{\theta}_i = \frac{\sum_q q \; L(\underline{x}_j|q, \underline{\delta})A(q)}{L(\underline{x}_j|q, \underline{\delta})A(q)} \tag{15}$$

and MAP is the value on the posterior distribution that has the maximum $f(\hat{\theta}_i)$ defined in Equation 14. Again as we do not know the examinee's location, we work on the full candidates of values for the examinee location along the continuum. Plots are drawn using finite values, and to make matters simple MAP estimates are obtained by brute force, however the rationale of the MLE in Application 1 applies to this section as well.

Application

Application 4 demonstrates Expected a Posteriori (EAP, Bock & Mislevy, 1982) and Maximum a Posteriori (MAP, Birnbaum, 1969) scoring techniques in the Rasch model. This is illustrated by replicating the analysis presented in de Ayala (2009, p. 79), in which the person location is estimated given the item locations of the five math items. In particular, this application allows users to: understand how the posterior distribution is obtained; differentiate between EAP and MAP estimation; realize that the ability for extreme responses can be estimated unlike with ML estimation in Application 1; visualize the concept of quadrature as it is used in IRT estimation; observe that MAP scores are more sensitive to characteristics of the prior distribution and the number of quadrature points; understand that increasing the number of quadrature points will result in a smoother posterior and thus more precise EAP and MAP estimates.

Users can run Application 4 via running the R command below. They can also access the same application via https://irtdemo.shinyapps.io/eapmap/.

```
# Expected a posteriori and
# maximum a posteriori ability estimation
irtDemo("eapmap")
```

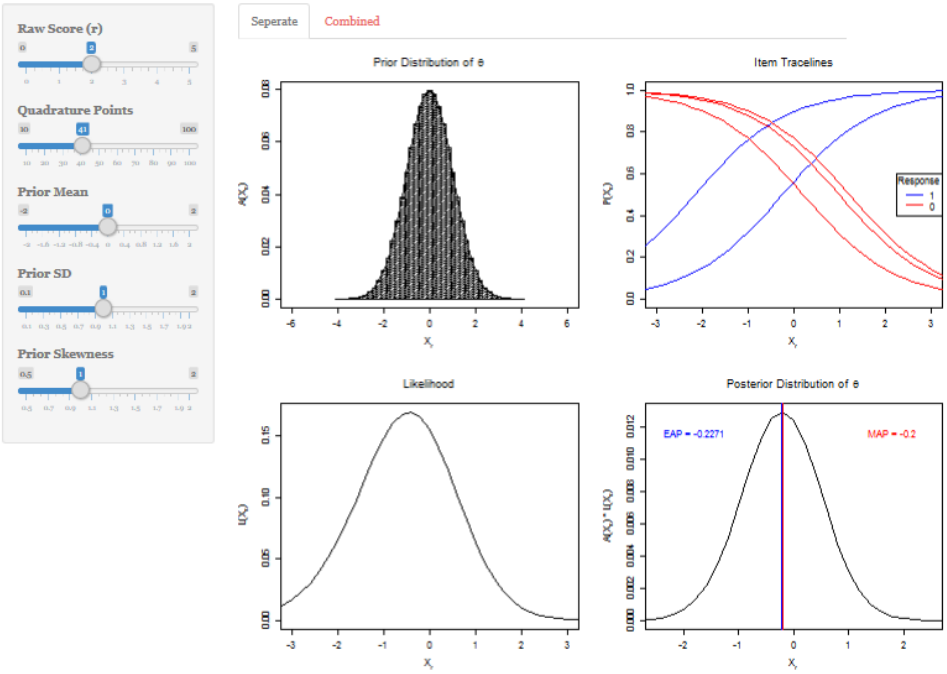

Figure 4: *Scoring (with separate plots)*

The plot in Figure 4 has four panels. The first panel on the upper-left corner is the prior distribution of the θ divided into quadrature intervals for computational ease. The larger number of quadrature points the smoother the prior distribution of θ. The location and shape of this distribution depends on the prior mean, prior standard deviation (SD) and prior skewness (values that can be manipulated by the user using slider controls). The second plot on the upper-right corner demonstrates items tracelines, which are probability of endorsing the given category (1 or 0) for each item in the raw score vector (e.g. 00011, this value can be manipulated using slider control on the left panel). The third plot on the lower-left corner demonstrates likelihood function given the raw score pattern, and the fourth plot on the lower-right corner is the posterior distribution of θ, which is the product of the prior distribution and the likelihood function. EAP and MAP estimates are shown on the third plot.

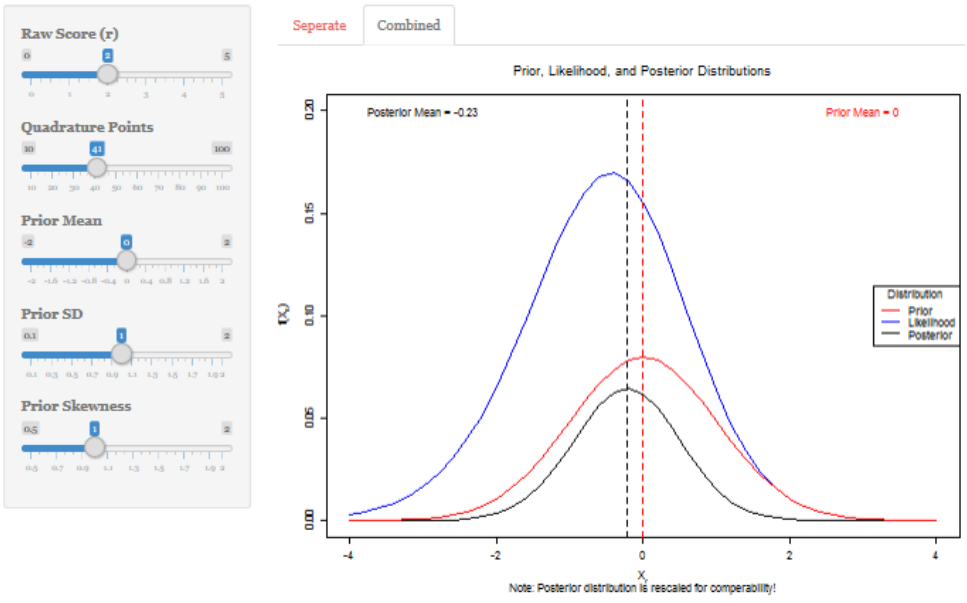

Figure 5: *Scoring (with combined plot)*

In another panel, the plot in Figure 5 juxtaposes prior, likelihood and posterior distributions. Posterior distribution is rescaled for comparability because product of prior and likelihood result in a relatively smaller distribution on the same scale. Prior and posterior means are shown the figure.

Application 5: Multidimensional Dichotomous IRT Models

Conceptual Underpinnings

In the previous application, it was assumed that to endorse an item correctly a test taker's ability level for a particular domain should be sufficient (unidimensional). There are cases where knowledge in more than one domain is required so that the test taker would endorse an item correctly. For example some math problems may require solid understanding of the text, in which the test taker's ability should be high both in math knowledge and reading comprehension. In this case the measurement model is said to be multidimensional. Considering multidimensional 4 PL (M4PL) model, for the compensatory model in general form, probability of endorsing an item j correctly for subject i given a set of ability levels for two dimensions $(\theta_{i1}, \theta_{i2})$, a set of item discriminations for each of the two dimensions $(\alpha_{j1}, \alpha_{j2})$, and the intercept (γ_j) is

$$P\left(x_{ij}=1\middle|\theta_{i1},\theta_{i2},\alpha_{j1},\alpha_{j2},\gamma_j\right)=c+(d-c)\frac{e^{D[(\alpha_{j1}\theta_{i1}+\alpha_{j2}\theta_{i2})+\gamma_j]}}{1+e^{D[(\alpha_{j1}\theta_{i1}+\alpha_{j2}\theta_{i2})+\gamma_j]}} \tag{16}$$

and for the non-compensatory (or partially compensatory) model the probability function is defined as

$$P\left(x_{ij}=1\middle|\theta_{i1},\theta_{i2},\alpha_{j1},\alpha_{j2},\gamma_j\right)=c+(d-c)\left(\frac{e^{D[\alpha_{j1}\theta_{i1}+\gamma_{j1}]}}{1+e^{D[\alpha_{j1}\theta_{i1}+\gamma_{j1}]}}\right)\left(\frac{e^{D[\alpha_{j2}\theta_{i2}+\gamma_{j2}]}}{1+e^{D[\alpha_{j2}\theta_{i2}+\gamma_{j2}]}}\right) \tag{17}$$

where c and d are lower and upper asymptotes respectively, D is constant which is 1 for logistic function and 1.702 for normal approximation, and the intercept is defined as a function of discrimination and difficulty parameters as $\gamma_j=-(\alpha_{j1}\delta_{j1}+\alpha_{j2}\delta_{j2})$ for the compensatory model, and $\gamma_{j1}=-(\alpha_{j1}\delta_{j1})$ and $\gamma_{j2}=-(\alpha_{j2}\delta_{j2})$ for the non-compensatory model. Note that when $c=0$ and $d=1$ the multidimensional model becomes M2PL model, when $c>0$ and $d=1$ it becomes M3PL, and finally when $c>0$ and $d<1$ it becomes M4PL.

Multidimensional difficulty index is defined as $\Delta_j=-\frac{\gamma_j}{A_j}$ where $A_j=\sqrt{\alpha_{j1}^2+\alpha_{j2}^2}$ and is multidimensional discrimination. The item direction (represented by an arrow) on the contour plot for the compensatory model is determined from the angle $\omega_j=\arccos\left(\alpha_{j1}/A_j\right)$. So the arrow representing a multidimensional item starts from the origin and ends in abscissa $(\Delta_j+A_j)\cos(\omega_j)$ and ordinate $(\Delta_j+A_j)\sin(\omega_j)$. This is not so straight forward for the non-compensatory models, so it is not shown on the plot. Information function is defined as $I\left(\theta_j\right)=A_j^2P(1-P)$ where P is short form of probability function in Equation 16 or 17.

Application

This application is not related to estimation or scoring but have been produced to channel a complex concept in IRT. The final application allows users to explore an item response surface (IRS) by manipulating hypothetical person locations and item discrimination and location parameters for each latent trait in a multidimensional item response theory (MIRT) (Bonifay, 2019; Reckase, 2009) model.

In particular, this application will help the user to identify differences in the IRS of dichotomous MIRT models; differentiate between compensatory and non-compensatory MIRT models; examine discrepancies between logistic and normal ogive MIRT formulations; observe the influence of item location and discrimination parameters on the IRS; and rotate the IRS to gain a better perspective of the precise shape of the 3-dimensional surface.

Users can run Application 5 via running the R command below. They can also access the same application via https://irtdemo.shinyapps.io/mirt/.

```
# Multidimensional item response theory
irtDemo("mirt")
```

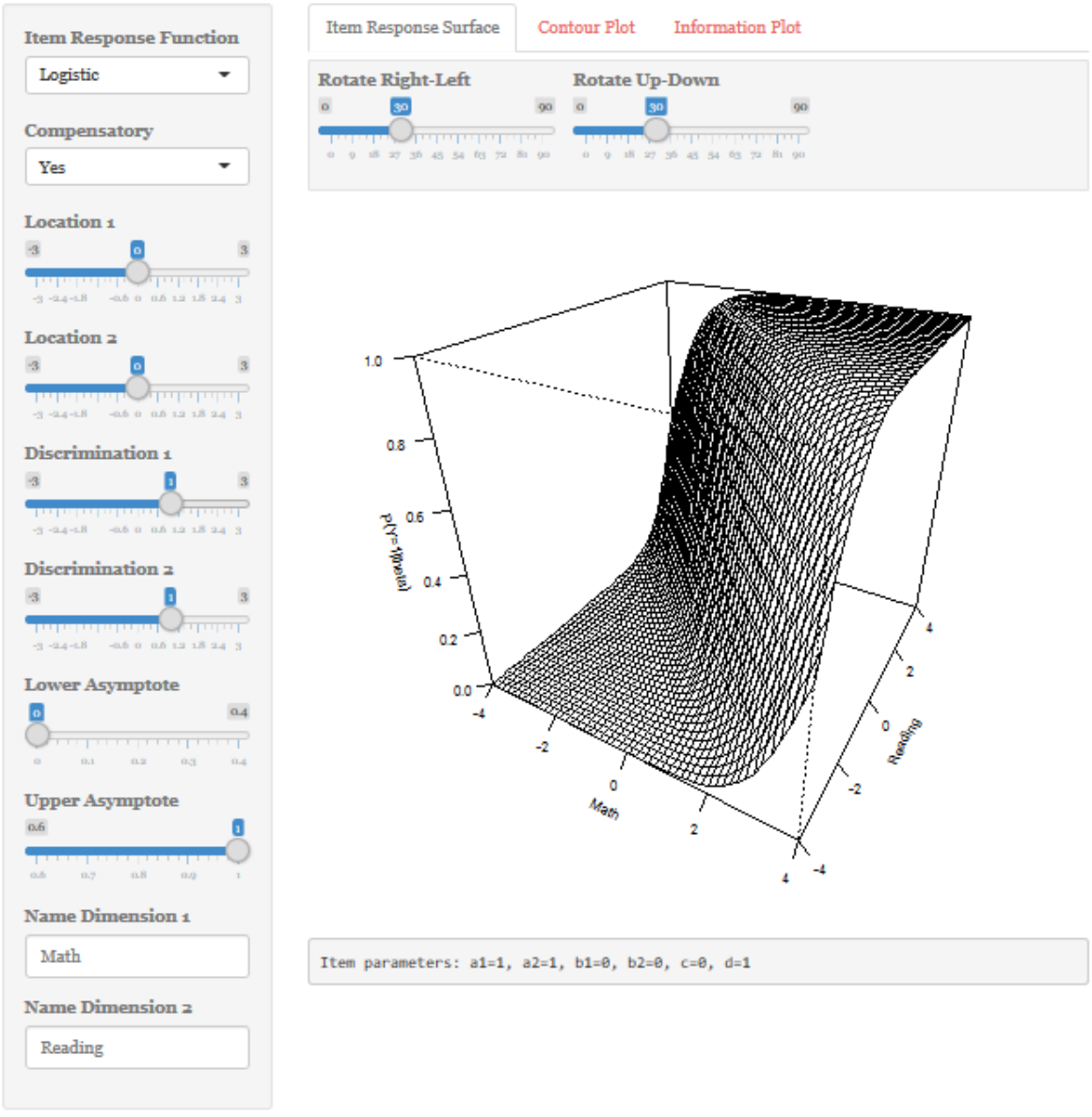

Figure 6: *Item response surface*

There are three panels; the plot in the first panel shows the IRS, the plot in the second panel shows the contour plot based on the IRS, and the plot in the third panel shows the information plot. These

plots change based on the parameters specified with slider controls in the panel on the left side. The first plot in Figure 6 shows probability of endorsing the correct category given both ability levels (e.g. math and reading). Since probability is a function of two dimensions, instead of a traceline, the response plot is a surface. The orientation and peakedness of this surface depends on each ability level, discrimination, lower- and upper-asymptote.

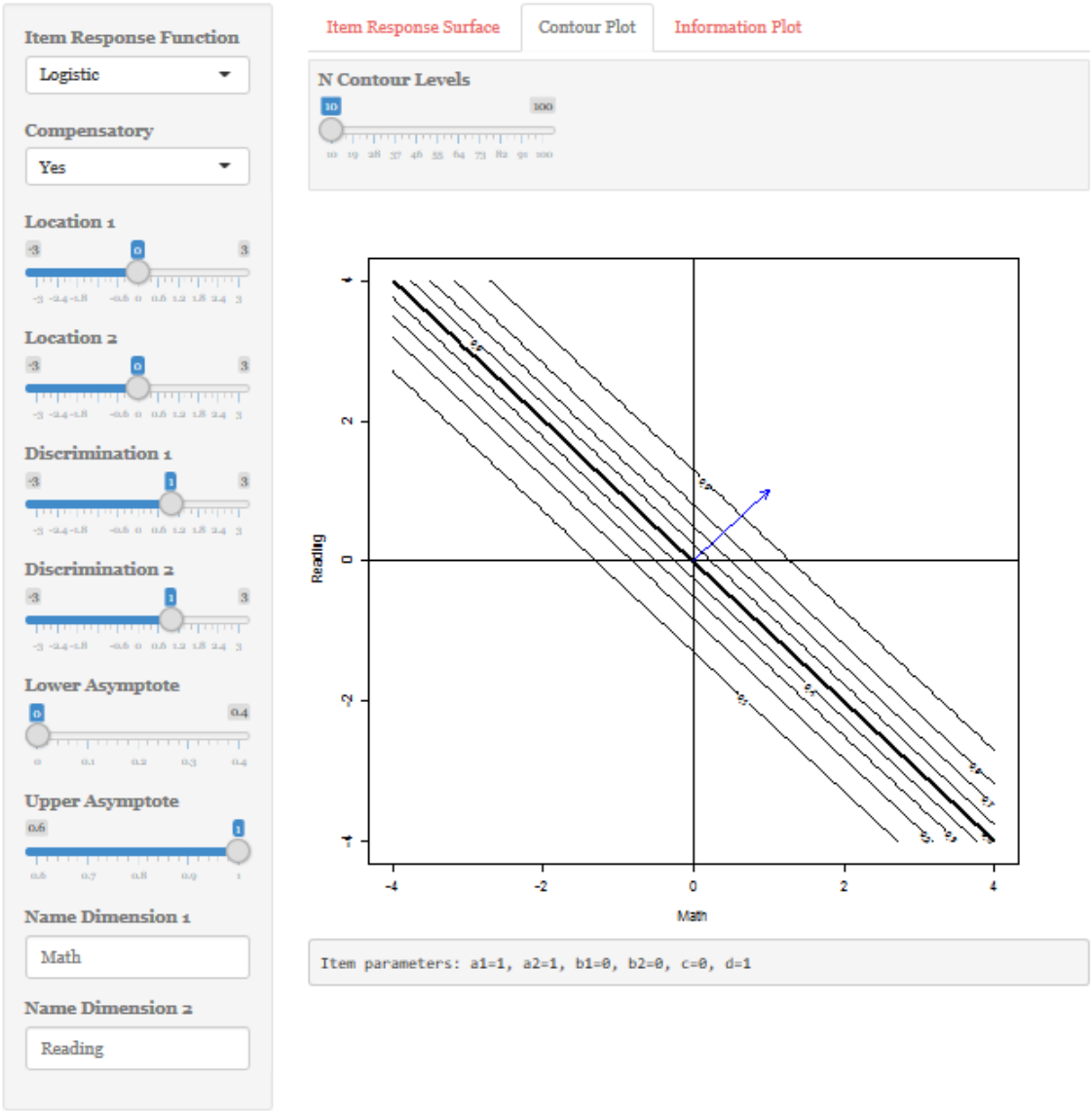

Figure 7: *Contour plot tab*

The second plot in Figure 7 is the contour plot for the IRS in Figure 6. The contour plot is a useful technique to project a three dimensional plot onto a two dimensional plot. The IRS in Figure 6 is sliced through horizontally with equal probability intervals, and the intersection of the slices with the plot is projected onto the two dimensional space. This is useful to demonstrate the vector representing an item as function of parameters on the left panel.

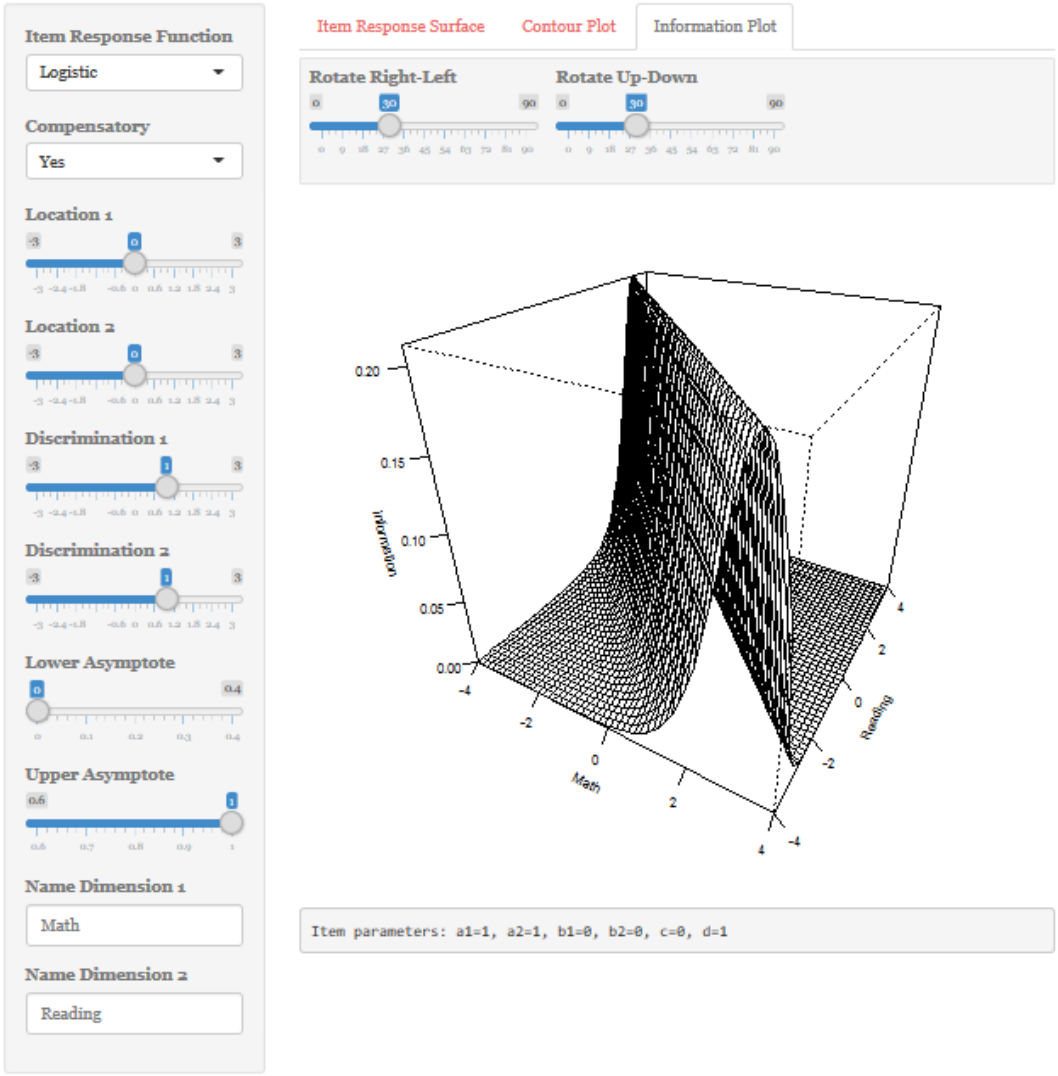

Figure 8: *Information plot tab*

The third plot in Figure 8 shows the information surface. The function to draw this information surface is described in "*Conceptual Underpinnings*" section. It shows the ability levels at which the information for the item is maximized.

4. SUMMARY

Statistics instruction can benefit from interactive R-based applications that supplement the traditional teaching of abstract theorems and equations. The IRT applications described in here will enable students and other users to investigate fundamental estimation and scoring procedures and multidimensional response functions in detail. Not only will this improve comprehension of potentially complicated topics and techniques, it will also prepare students to be insightful applied researchers who are aware of the advantages and disadvantages that exist in various approaches to estimation and scoring.

This study contributes to the body of pedagogical statistical applications targeting audience of advanced and complicated measurement topics. It does not aim to provide an interface for data analysis

of IRT models of any kind. Applications 1 and 4 are limited in the sense that they employ specific estimation algorithms based on examples provided in de Ayala (2009) for Rasch models. Applications 2 and 3 use brute force to find maximum points based on 50 simulated responses, whereas in practice maximum points are found using derivatives akin to Application 1.

REFERENCES

Anderson, E. B. (1972). The numerical solution of a set of conditional estimation equations. *Journal of the Royal Statistical Society*, Series B, 34, 42-54.

Baker, F.B., & Kim, S.H. (2004) *Item Response Theory: Parameter Estimation Techniques* (2nd Ed.), CRC Press, Boca Raton.

Birnbaum, A. (1968). *Some latent trait models and their use in inferring an examinee's ability*. In F. M. Lord and M. R. Novick (Eds.), Statistical theories of mental test scores (pp. 397-472), Reading, MA: Addison-Wesley.

Birnbaum, A. (1969). Statistical theory for logistic mental test models with a prior distribution of ability. *Journal of Mathematical Psychology*, *6*(2), 258-276.

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, *6*(4), 431-444.

Braun, J. W., White, B. J., & Craig, G. (2014). R tricks for kids. *Teaching Statistics*, *36*, 7–12.

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

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

Bulus, M., & Bonifay, W. (2016). irtDemo: Item response theory demo collection. R package version 0.1.2. https://CRAN.R-project.org/package=irtDemo

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

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2017). shiny: Web Application Framework for R. R package version 1.0.3. https://CRAN.R-project.org/package=shiny

De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.

Garfield, J., and Ben-Zvi, D. (2007). How students learn statistics revisited: a current review of research on teaching and learning statistics. *International Statistical Review*, *75*, 372–396.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Reckase, M. D. (2009*).* Multidimensional item response theory models. In *multidimensional item response theory* (pp. 79-112). New York, NY: Springer.