Assessing Prior Distributions for the Item Parameters in the Two-Parameter Logistic IRT Model

Kentaro Kato†
Benesse Educational Research and Development Center/Center for Research on Educational Testing, Tokyo, Japan

Introduction

- Item response theory (IRT) models the probability of a correct response to each test item as a function of latent ability θ.
- Practical demands in IRT item paremeter estimation
  - Reasonably high accuracy, while
  - Keeping the sample size minimum to reduce cost
- Can prior information be used to improve estimation?
  1. Estimates of existing, similar test items
  2. Prediction by item properties (e.g., contents, required skills, and format)
  3. Expert (item writer, teacher, etc.) opinion to construct informative prior distributions for item parameters

◊ In this study we consider the 3rd approach

The 2-Parameter Logistic IRT Model (2PLM)

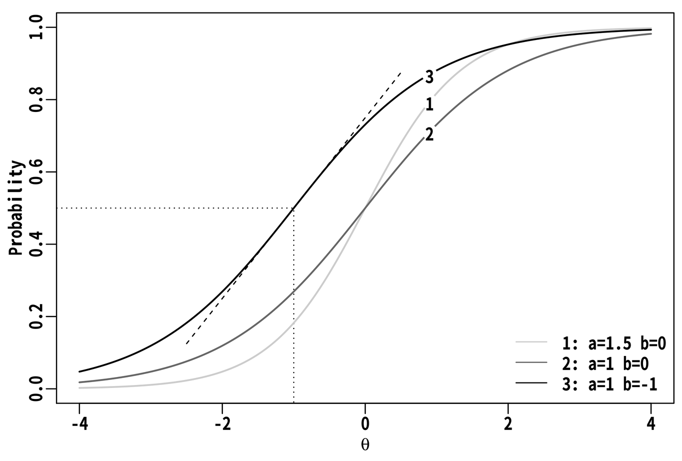

Figure 1: Item response functions

- Item response function (IRF; Fig. 1)

$$\Pr(X_j = 1|\theta, \boldsymbol{\xi}_j) = \frac{1}{1+\exp[-a_j(\theta - b_j)]} \qquad (1)$$

Item $j = 1, \ldots, J$

Item response $x_j \in \{0, 1\}$

$1 =$ correct response, $0 =$ incorrect response

Item parameters $\boldsymbol{\xi}_j = (a_j, b_j)$

$a_j =$ discrimination (slope), $b_j =$ difficulty (location)

Ability parameter $\theta \in (-\infty, \infty)$ (latent variable)

- $\boldsymbol{\xi}_j$, $j = 1, \ldots, J$, are estimated from field-testing data

Probability Assessment of Item Parameters (Tsutakawa & Lin, 1986)

- Hard to work on $\boldsymbol{\xi}$ directly without knowledge on IRT
  - It would be easier to deal with correct-response probabilities
  - Since IRT is a model of correct response probabilities conditional on θ, elicitation is also made about the correct-response probabilities at certain θ points.
  - The elicited prior will then be transformed onto the space of item parameters.
- Pick two points, $\theta_1 < \theta_2$, on the θ scale
  - These θ points should represent locations of well-defined populations of test takers (e.g., the population mean of all 8th graders), which is familiar to the expert
- Under the 2PLM let $\mathbf{p} = (p_1, p_2)$, where

$$p_k = \frac{1}{1+\exp[-a(\theta_k - b)]}, \qquad k = 1, 2 \qquad (2)$$

  - Each p_k represents the correct-response probabilities conditional on θ_k
- Tsutakawa and Lin (1986) showed that $\boldsymbol{\xi}$ and $\mathbf{p}$ are one-to-one and derived a prior distribution of $\boldsymbol{\xi}$ from that of $\mathbf{p}$ under the constraint $0 < p_1 < p_2 < 1$ (or $a > 0$)
  1. The range constraint makes it hard to construct in a precise manner a joint prior of $\mathbf{p}$ from its marginals which are assumed independent (they only provided informal justification)
  2. The resulting $\boldsymbol{\xi}$ prior is almost of the conjugate form but not exactly
  3. Marginal priors for p_ks are determined by their moments, but a more probabilistic elicitation procedure may be desirable to work with experts

Proposed Method

- The proposed method features two possible improvements over Tsutakawa and Lin's (1986) procedure: (a) reparameterization of IRF and (b) the elicitation method

Reparameterization and Transformation from p to $\boldsymbol{\xi}$

- Reparameterize the IRF as

$$p_k = \frac{1}{1+\exp[-(a\theta_k + d)]}, \qquad k = 1, 2 \qquad (3)$$

where $d = -ab$ (redefine $\boldsymbol{\xi} = (a, d)$). Then

$$a = \frac{L_2 - L_1}{\theta_2 - \theta_1}, \qquad d = \frac{\theta_2 L_1 - \theta_1 L_2}{\theta_2 - \theta_1}, \qquad (4)$$

where $L_k = \ln[p_k/(1-p_k)]$, $k = 1, 2$

- There is a one-to-one correspondence between

$$\mathcal{P} = \{(p_1, p_2) | 0 < p_k < 1, k = 1, 2\} \text{ and} \qquad (5)$$
$$\Omega = \{(a, d) | -\infty < a < \infty, -\infty < d < \infty\} \qquad (6)$$

- Jacobian of transformation $\mathbf{p} \to \boldsymbol{\xi}$ is

$$\mathbf{J} = \left|\frac{\partial \mathbf{p}}{\partial \boldsymbol{\xi}}\right| = (\theta_1 - \theta_2)p_1(1-p_1)p_2(1-p_2) \neq 0 \qquad (7)$$

with p_ks replaced by (3)

- If we assume independent beta priors for p_1 and p_2,

$$f(\mathbf{p}) = \prod_{k=1}^{2} p_k^{r_k - 1}(1-p_k)^{s_k - 1}, \qquad \mathbf{p} \in \mathcal{P}, \qquad (8)$$

and then the prior for $\boldsymbol{\xi}$ is

$$\begin{aligned} g(\boldsymbol{\xi}) &= f(\mathbf{p})|\mathbf{J}| && (9) \\ &\propto \prod_{k=1}^{2} p_k^{r_k}(1-p_k)^{s_k} && (10) \\ &\propto \prod_{k=1}^{2} \frac{\exp[-s_k(a\theta_k + d)]}{[1+\exp[-(a\theta_k + d)]]^{r_k + s_k}}, \quad \boldsymbol{\xi} \in \Omega && (11) \end{aligned}$$

(Fig. 2). This leads to the conjugate form to the likelihood function based on the reparameterized IRF (3).

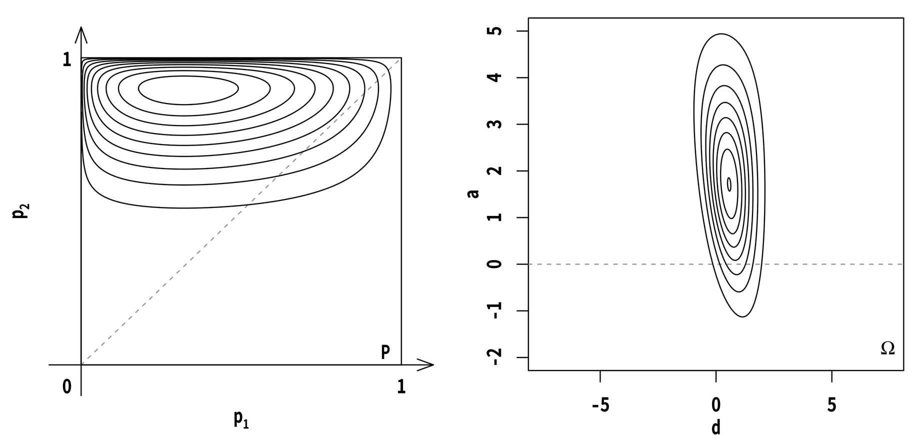

Figure 2: Transformation from $\mathcal{P}$ to Ω

Elicitation of Beta Priors

- If we adopt the independent beta marginal priors (8), elicitation can be made for p_1 and p_2 independently to determine the values of (r_k, s_k), $k = 1, 2$, in (11)
- We use a simple two-point fractile method and fit a beta distribution independently for each p_k
  1. Specify two quantiles $0 < x_{q_1} < x_{q_2} < 1$ on the p_k scale ($x_{q_1} = 0.5$ and $x_{q_2} = 0.8$ are used in the current study; different values may be used for different k)
  2. Ask the expert to provide his/her estimate of the probability $q_{km} = P(p_k \leq x_{q_m})$, $m = 1, 2$, $k = 1, 2$; the question to be asked may be like:

     "What is your estimate of the probability that the correct-response rate for people with θ =[θ_k] is no greater than [x_{q_m}]?"
  3. Step 2 yields two-point quantile-probability "data" (x_{q_m}, q_{km}), $m = 1, 2$, $k = 1, 2$, to which a beta distribution is fit for each k
  4. Use a bisecting search algorithm by van Dorp and Mazzuchi (2004) to estimate the beta parameters (r_k, s_k) (the program BETA-CALCULATOR is available for this computation) (Fig. 3)

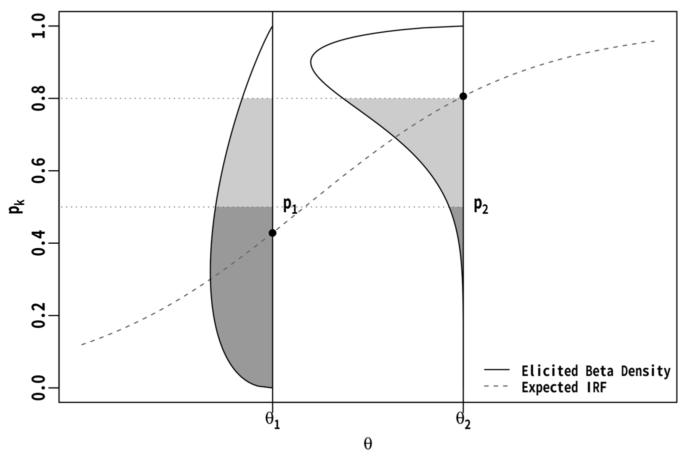

Figure 3: Elicited marginal beta priors of p_1 and p_2 at θ_1 and θ_2, respectively

A Real Experiment

- An item editor was asked to elicit his prior distribution for a reading comprehension item
- Specifications

$\theta_1 = -0.50$ (mean of a particular high school)

$\theta_2 = 0.50$ (mean of a particular college)

$x_{q_1} = 0.50$

$x_{q_2} = 0.80$

- His estimated probabilities (q_{km}) were

Quantile	$\theta_1 = -0.50$	$\theta_2 = 0.50$
$x_{q_1} = 0.50$	50%	5%
$x_{q_2} = 0.80$	70%	40%

- His beta parameters were estimated as $(r_1, s_1) = (0.48, 0.48)$ and $(r_2, s_2) = (4.54, 1.10)$, and the joint beta prior was transformed onto Ω (Fig. 4)

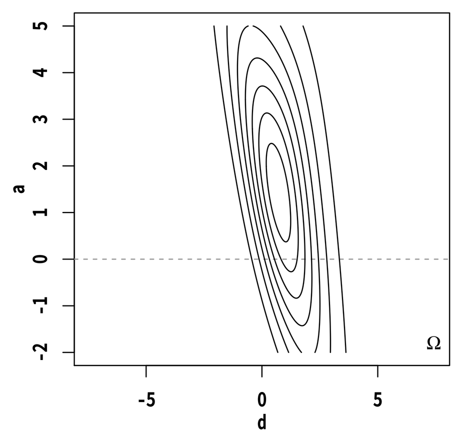

Figure 4: Elicited joint prior for (a, d)

Further Considerations

- A program which implements parameter estimation with the proposed prior distribution needs to be developed
- Elicitation at two different points of θ can be independent or should be dependent?
- The use of the two-point fractile method could easily lead to an overfitting problem
  - If more than two q_{km}s are elicited for each k, the result could be very different
  - What kind of fitting method is available?

References

Tsutakawa, R. K., & Lin, H.-Y. (1986). Bayesian estimation of item resopnse curves. *Psychometrika*, *51*, 251–267.

van Dorp, J. R., & Mazzuchi, T. A. (2004). Parameter specification of the beta distribution and its Dirichlet extensions utilizing quantiles. In A. K. Gupta & S. Nadarajah (Eds.), *Handbook of beta distribution and its applications* (pp. 283–318). NY: Marcel Dekker.

† 1-34 Ochiai, Tama, Tokyo, Japan 206-8686. mailto: nekotak@mail.benesse.co.jp