Use of Crude Prior Information for Item Parameter Estimation in the Item Response Theory

Kentaro Kato†
Benesse Educational Research and Development Institute, Tokyo, Japan

Introduction

- Use of appropriate prior information (expert opinions) on item parameters can improve estimation.
- Methods based on “probability assessment” (PA) on item parameters are available (Kato, 2012; Tsutakawa & Lin, 1986), but detailed PA can be difficult and time-cosuming.
- Crude prior information: An expert gives each item his/her “difficulty rating”such as Easy, Medium, and Difficult.
- Does this type of information improve estimation?

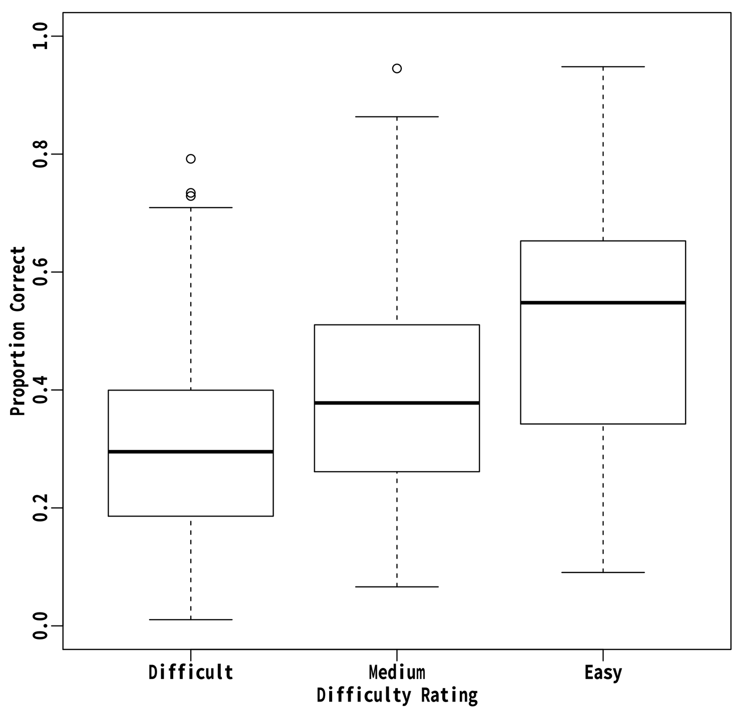

Figure 1: Distribution of proportion correct by difficulty rating from an expert (254 items)

Bayesian Hierarchical Modeling of Item Parameters

- The 2-parameter logistic model (2PLM): The probability that respondent i ($i = 1, \ldots, N$) answers item j ($j = 1, \ldots, J$) correctly is given by

$$P(u_{ij} = 1|\theta_i, \alpha_j, \beta_j) = \frac{1}{1 + \exp(-\alpha_j(\theta_i - \beta_j))} \quad (1)$$

- Prior distributions
  - Ability parameter: $\theta_i \overset{i.i.d.}{\sim} N(0, 1)$
  - **Model 1** (Fox, 2010, p. 72)

$$\begin{bmatrix} \ln \alpha_j \\ \beta_j \end{bmatrix} \overset{i.i.d.}{\sim} MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (2)$$
$$\boldsymbol{\Sigma} \sim IW(\boldsymbol{\Sigma}_0, m) \quad (3)$$
$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim MVN(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}/n) \quad (4)$$

  - **Model 2** assumes different prior means for difficulty parameters based on the difficulty level $k(j) = 1, \ldots, K$ assined a priori to each item:

$$\begin{bmatrix} \ln \alpha_j \\ \beta_j \end{bmatrix} \overset{i.i.d.}{\sim} MVN(\boldsymbol{\mu}_{k(j)}, \boldsymbol{\Sigma}) \quad (5)$$
$$\boldsymbol{\Sigma} \sim IW(\boldsymbol{\Sigma}_0, m) \quad (6)$$
$$\boldsymbol{\mu}_{k(j)} \overset{i.i.d.}{\sim} MVN(\boldsymbol{\mu}_0, \boldsymbol{\Psi}) \quad (7)$$

  - $k(j)$ represents a prior difficulty level assigned to item j (e.g., if item 1 is given level 3, then $k(1) = 3$)
  - Let

$$\boldsymbol{\mu}_{k(j)} = \left[\mu_{\alpha k(j)} \ \mu_{\beta k(j)}\right]^T \quad (8)$$
$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{bmatrix} \quad (9)$$

  - $\mu_{\beta k(j)}$ represents the mean difficulty in class $k(j)$.
  - If the prior difficulty level $k(j)$ reflects the reality...
    - $\mu_{\beta k(j)}$s will be well separated from each other and in the predicted order.
    - Since $k(j)$ accounts for the variation of item difficulty, the “within-level” variance σ_β^2 will get smaller than in the case of Model 1.
  - **Model 3** is the same as Model 2 but imposes inequality constraints on $\mu_{bk(j)}$s according to the prior difficulty ordering (i.e., more explicit formulation of a prior “hypothesis” on item difficulty):

$$\mu_{b1} > \mu_{b2} > \cdots > \mu_{bK} \quad (10)$$

Method

- Data
  - $J = 39$ Japanese vocabulary items (multiple choice with 5 response options)
  - $N = 484$ respondents (college students and adults)
- Difficulty ratings
  - An expert rated each item at $K = 3$ difficulty levels:
    - Difficult ($k(j) = 1$), 10 items
    - Medium ($k(j) = 2$), 13 items
    - Easy ($k(j) = 3$), 16 items
- Parameter estimation
  - Models 1 through 3
  - Specification of hyperparameters

$$\boldsymbol{\mu}_0 = \left[0 \ 0\right]^T \quad (11)$$
$$m = 4 \quad (12)$$
$$n = 2 \quad (13)$$
$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} \quad (14)$$
$$\boldsymbol{\Psi} = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} \quad (15)$$

  - MCMC computation was perfomed by OpenBUGS and R (`R2OpenBUGS`)
  - 6000 draws from the posterior distribution (3 chains, 4000 iterations for each chain, and the first 2000 discarded as burn-in)

Results

- DIC was comparable for all three models (DIC $= 20910.0$)
- Covariance of item parameters ($\hat{\boldsymbol{\Sigma}}$)

	Model 1		Model 2		Model 3
$\hat{\sigma}_\alpha^2$	0.10	>	0.08	>	0.08
$\hat{\sigma}_\beta^2$	2.21	>	1.89	>	1.85
$\hat{\sigma}_{\alpha\beta}$	−0.20	<	−0.10	<	−0.10

  - Variance of item difficulty ($\hat{\sigma}_\beta^2$): 14% (Model 2) and 16% (Model 3) reduction from Model 1 by incorporating prior information
- Means of item parameters ($\hat{\boldsymbol{\mu}}$ or $\hat{\boldsymbol{\mu}}_{k(j)}$)

	Model 1	Model 2	Model 3
$\hat{\mu}_\alpha$	−0.29		
$\hat{\mu}_{\alpha 1}$ (Difficult)		−0.57	−0.56
$\hat{\mu}_{\alpha 2}$ (Medium)		−0.27	−0.27
$\hat{\mu}_{\alpha 3}$ (Easy)		−0.15	−0.14
$\hat{\mu}_\beta$	−0.19		
$\hat{\mu}_{\beta 1}$ (Difficult)		1.26	1.27
$\hat{\mu}_{\beta 2}$ (Medium)		0.03	0.15
$\hat{\mu}_{\beta 3}$ (Easy)		−0.34	−0.42

  - In Model 2, $\hat{\mu}_{\beta k(j)}$s are well separated from each other and follow the predicted order ($\hat{\mu}_{\beta 1} > \hat{\mu}_{\beta 2} > \hat{\mu}_{\beta 3}$).
  - Model 3 imposed inequality constraints on the Model 2 means, but the estimates were almost the same as those in Model 2.
  - $\hat{\mu}_{\alpha k(j)}$ tends to get smaller as the difficulty level goes up (Models 2 and 3; more difficult, less discriminative).
- Item discrimination parameter estimates ($\hat{\alpha}_j$)

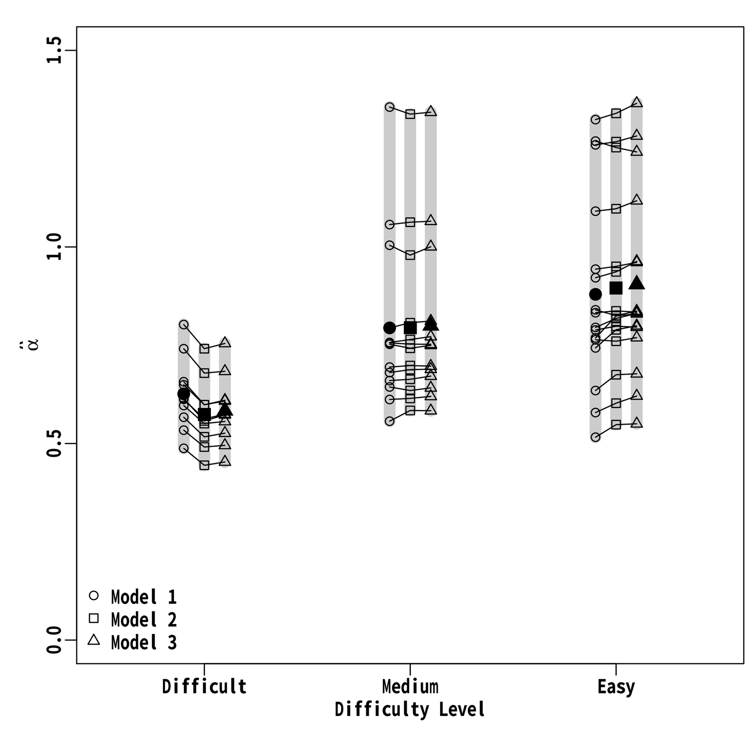

Figure 2: Estimates of item discrimination parameters

- Item difficulty parameter estimates ($\hat{\beta}_j$)

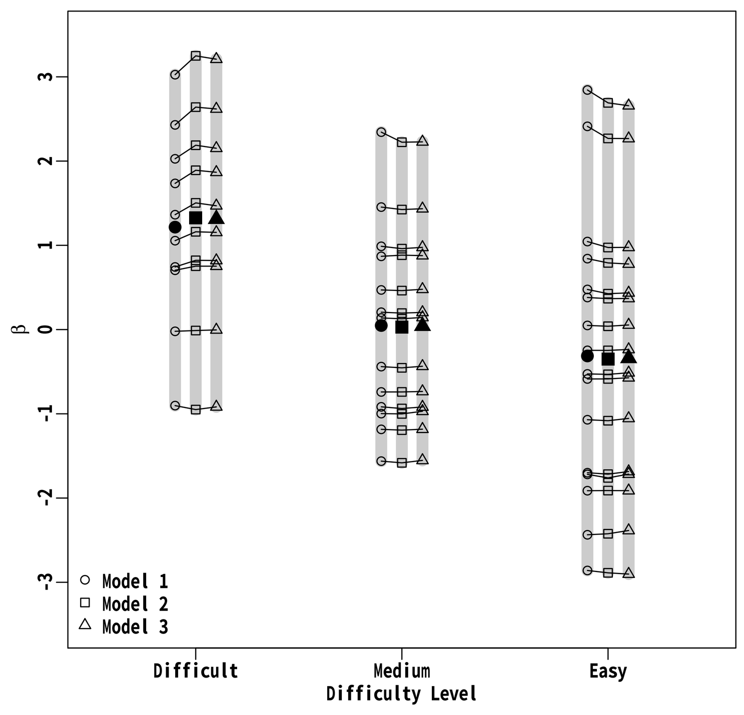

Figure 3: Estimates of item difficulty parameters

- Posterior standard deviations of item discrimination parameters

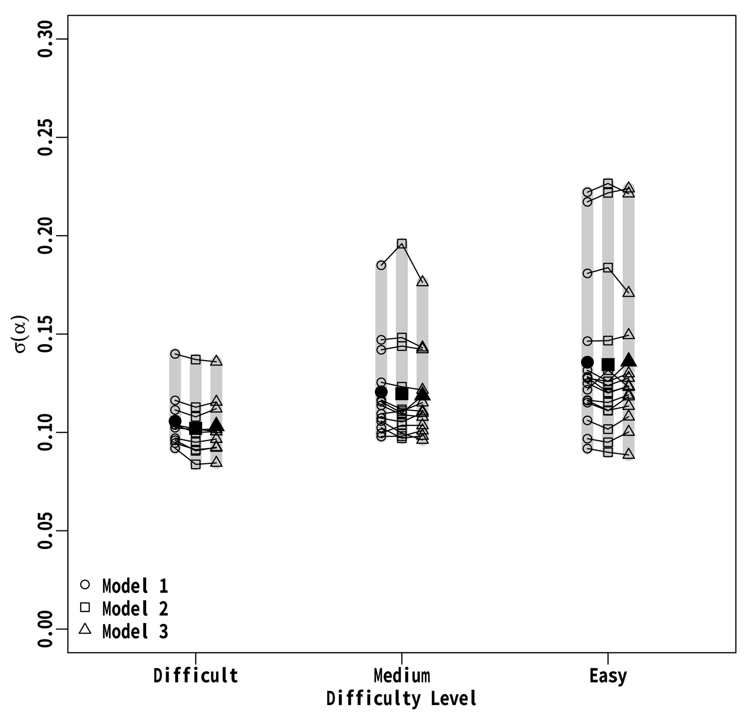

Figure 4: Posterior SDs of item discrimination parameters

- Posterior standard deviations of item difficulty parameters

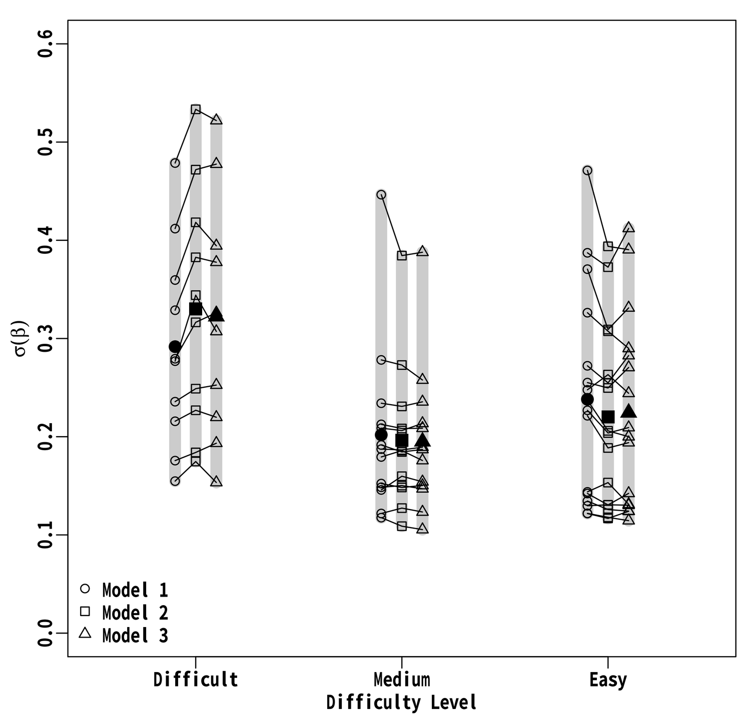

Figure 5: Posterior SDs of item difficulty parameters

  - Much improvement was not found for estimation accuracy.

Conclusions and Further Considerations

- Use of “crude” prior information on item difficulty levels
  - Means of item difficulty ($\boldsymbol{\mu}_{k(j)}$) well reflected the prior difficulty ratings.
  - However, “within-level” variance of item difficulty was not reduced enough for shrinkage to the level mean (and thus improvement of estimation accuracy) to occur.
  - If the prior rating is valid, inequality constraints on the level means would probably be trivial.
- Other elements to consider
  - Number of difficulty levels
  - Effect of feedback (training)
  - Combining ratings from multiple experts

References

Fox, J.-P. (2010). *Bayesian item response modeling*. NY: Springer.

Kato, K. (2012, June). *Assessing prior distributions for the item parameters in the two-parameter logistic IRT model.* Poster presented at the World Meeting of the International Society for Bayesian Analysis, Kyoto, Japan.

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

† Benesse Educational Research and Development Institute, 1-34 Ochiai, Tama, Tokyo, 206-8686 Japan. mailto: nekotak@mail.benesse.co.jp