

# Item Response Theory with Latent Classes

## 潜在クラス項目反応理論

OKAMOTO Yasuharu

岡 本 安 晴

[**Abstract**] To assess the probabilities that a person belongs to latent classes, item response theory (IRT) is modified and the continuous ability parameter is replaced by discrete latent classes. In psychology, whether a theoretical concept is continuous or discrete is a fundamental problem. However, IRT measures only continuous concepts. Thus, as a method of measurement for discrete concepts, item response theory with latent classes (IRT-LC) is proposed, such that latent classes correspond to discrete states of the concept. The probability of a response to an item is given for each latent class. The probability of a person's response to an item is calculated based on the probabilities that a person belongs to each latent class and the probabilities of responses to the item, which are determined for latent classes. To estimate parameters of the model, Metropolis-Hastings (M-H) within Gibbs algorithm was adopted. It should be emphasized that M-H within Gibbs algorithm for IRT-LC works well and IRT-LC can be statistically discriminated from IRT with a continuous ability parameter for data with appropriate information.

## Introduction

In psychology, whether a theoretical concept is continuous or discrete is a fundamental problem. In developmental psychology, Nolen-Hoeksema, Fredrickson, Loftus and Wagenaar (2009) chose as a central question "Is development best understood as a gradual, continuous process of change or as a series of abrupt, qualitatively distinct stages?" (p. 70). Regarding learning theories, Restle and Greeno (1970) presented a chapter entitled "Learning: Gradual or All-or-None." In psychopathology, Meehl (1995), in an edited version of lectures given on receipt of the Joseph Zubin Award at the meeting of the Society for Research in Psychopathology, discussed the empirical question "Is the latent structure of these phenotypic indicator correlations taxonic (categories) or nontaxonic (dimensions, factors)?" Macmillan and Creelman (2005) classified signal detection theories (SDT) into continuous theories, which include the standard SDT of Green and Swets (1966), and discrete theories, i.e., threshold models.

Hence, assessment of a psychological concept needs two types of methods, one corresponding to a continuous concept and the other corresponding to discrete concept. However, item response theory (IRT), the current well-known modern method, was developed to measure continuous concepts. Hence, a method for a discrete concept is needed. Shojima (2007) proposed Neural Test Theory (NTT), which maps people onto discrete latent ranks. NTT is based on Self-Organizing Maps (SOM). SOM compresses information

and maps high-dimensional input data manifolds onto a low-dimensional array (Kohonen, 2001). NTT groups people into ranks, the number of ranks is set not on psychological theory but merely by expedience. A model, in which latent classes are linked to concepts or stages in a psychological theory, is needed. Okamoto (2008, 2009) presented a model, item response theory with latent classes (IRT-LC), which is a discrete version of IRT. However, this model was presented in Japanese, so in this paper the essentials of the model are presented in English.

## Model and Algorithm

In IRT, the probability of a yes response by a person with ability  $\theta$  is denoted as

$$P(\text{yes}|\theta).$$

$\theta$  represents strength of a continuous characteristic of the person relevant to the response. In the case of a discrete version of IRT,  $\theta$  is replaced by a latent class  $R_k$  (the  $k$ th rank), to which the person belongs, and  $P(\text{yes}|\theta)$  is denoted as

$$P(\text{yes}|R_k).$$

Here,  $R_k$ s are assumed to be ordered so that

$$P(\text{yes}|R_k) \leq P(\text{yes}|R_{k+1}).$$

Under this assumption,  $R_k$ s are identifiable.

Consider the case where there are  $Q$  latent classes,  $R_1, \dots, R_Q$ ,  $M$  items,  $item_1, \dots, item_M$ , and  $N$  persons,  $person_1, \dots, person_N$ . Set variable  $X_{ji}$  as follows:

$$X_{ji} = \begin{cases} 1 & \text{response of } person_j \text{ on } item_i \text{ is yes} \\ 0 & \text{response of } person_j \text{ on } item_i \text{ is no} \end{cases}$$

A likelihood function is defined as follows:

$$L = \prod_{j=1}^N \left[ \sum_{k=1}^Q \left\{ P(R_k | person_j) \prod_{i=1}^M P(X_{ji} = 1 | R_k)^{X_{ji}} P(X_{ji} = 0 | R_k)^{1-X_{ji}} \right\} \right], \quad (1)$$

where

$$P(R_k | person_j) = \text{probability that } person_j \text{ belongs to } R_k$$

In equation (1), the parameters to be estimated are

$$P(R_k | person_j), P(X_{ji} = 1 | R_k) \text{ and } P(X_{ji} = 0 | R_k),$$

and the following constraints are set

$$\sum_{k=1}^Q P(R_k | person_j) = 1, \quad (2)$$

$$P(X_{ji} = 1 | R_k) \leq P(X_{ji} = 1 | R_{k+1}), \quad (3)$$

$$P(X_{ji} = 1 | R_k) + P(X_{ji} = 0 | R_{k+1}) = 1, \quad (4)$$

Equation (4) means that only one of the two parameters needs to be estimated. Hence, parameter  $P(X_{ji} = 1|R_k)$  is chosen for estimation.

In equation (3), probabilities are conditional on  $R_k$  to which  $person_j$  belongs, so the index  $j$  of  $X_{ji}$  can be ignored and equation (3) can be rewritten as

$$P(\text{yes on item}_i|R_k) \leq P(\text{yes on item}_i|R_{k+1}) \quad (5)$$

Furthermore, using a cumulative logistic distribution function

$$F(x) = \frac{1}{1 + \exp(-x)} \quad (6)$$

put

$$F(C_{ik}) = P(\text{yes on item}_i|R_k)$$

where

$$C_{ik} \leq C_{i(k+1)}.$$

Put

$$C_{(d)ik} = C_{ik} - C_{i(k-1)}$$

where

$$C_{i0} = 0 \text{ and } C_{(d)ik} \geq 0 \text{ for } k > 1.$$

$C_{(d)ik}$  s are used as parameters to be estimated instead of  $P(\text{yes on item}_i|R_k)$ .

Considering constraint (2), put

$$P(R_k|person_j) = \begin{cases} F(r_1^{(j)}) & \text{for } k=1 \\ F(r_k^{(j)}) - F(r_{k-1}^{(j)}) & \text{for } 1 < k < Q \\ 1 - F(r_{Q-1}^{(j)}) & \text{for } k=Q \end{cases}$$

where  $F(x)$  is given by equation (6) and

$$r_k^{(j)} \leq r_{k+1}^{(j)}.$$

Put

$$r_{(d)k}^j = r_k^{(j)} - r_{k-1}^{(j)},$$

where

$$r_0^j = 0 \text{ and } r_{(d)k}^j \geq 0 \text{ for } k > 1$$

$r_{(d)k}^j$  s are used as parameters to be estimated instead of  $P(R_k|person_j)$ .

The parameters  $C_{(d)k}^j$  s and  $r_{(d)k}^j$  s are estimated by the following M-H within Gibbs algorithm.

The posterior distribution is given by

$$P(\dots C_{(d)ik} \dots r_{(d)k}^j \dots) \propto P_0(\dots C_{(d)ik} \dots r_{(d)k}^j \dots) \times \prod_{j=1}^N \left[ \sum_{k=1}^Q \left\{ P(R_k|person_j) \prod_{i=1}^M \left( P(X_{ji} = 1|R_k)^{X_{ji}} P(X_{ji} = 0|R_k)^{1-X_{ji}} \right) \right\} \right]$$

where  $P(R_k|person_j)$  and  $P(X_{ji} = 1|R_k)$  are functions of  $r_{(d)k}^j$  and  $C_{(d)ik}$ , respectively, and the prior distribution is given as follows:

$$P_0(\dots C_{(d)ik} \dots r_{(d)k}^j \dots) = \begin{cases} \text{const or} & k=1 \\ & k>1 \text{ and } C_{(d)ik} \geq 0, r_{(d)k}^j \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Prior distribution is considered to be nonzero only within a sufficiently large region, so  $P_0$  is a proper probability distribution.  $P_0$  is zero if  $C_{(d)ik}$  or  $r_{(d)k}^j$  is negative for  $k>1$ , i.e., constraints (2) and (3) do not hold, so sampling in M–H within Gibbs algorithm are constrained to satisfy restrictions (2) and (3).

Samplings from the posterior distribution are repeated  $NC$  cycles as follows:

Step 0.

Set initial values  $C_{(d)ik}(0)$ s and  $r_{(d)k}^j(0)$ s, where  $C_{(d)ik}(t)$ s and  $r_{(d)k}^j(t)$ s represent values of  $C_{(d)ik}$ s and  $r_{(d)k}^j$ s in iteration  $t$ .

Set  $t \leftarrow 0$ .

Step 1.

Update  $C_{(d)ik}(t+1)$ s one by one based on the values  $C_{(d)i_1k_1}(t)$ , or  $C_{(d)i_1k_1}(t+1)$  if it is already updated, and  $r_{(d)k_1}^j(t)$ , according to M–H within Gibbs algorithm.

Step 2.

Update  $r_{(d)k}^j(t+1)$ s one by one based on the values  $C_{(d)i_1k_1}(t+1)$ , and  $r_{(d)k_1}^j(t)$ , or  $r_{(d)k_1}^j(t+1)$  if it is already updated.

Step 3.

When  $t=NC$ , finish sampling.

Otherwise set  $t \leftarrow t+1$  and jump to Step1.

## Discussion

To assess a person with respect to a discrete concept requires a discrete type of IRT. In this paper, the model presented is a direct extension of IRT to a model for a discrete concept. Okamoto (2008) showed by simulation that IRT–LC model works well. Of course, for a discrete concept, assessment should be made by IRT–LC instead of IRT. Okamoto (2009) compared IRT–LC with IRT and suggested that IRT–LC can be statistically discriminated from IRT if the structure of the data contains information about differences between IRT–LC and IRT. Hence, we can expect that IRT–LC would provide a useful method to assess whether a concept in psychology is discrete or continuous.

## References

- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Los Altos: Peninsula Publishing.
- Kohonen, T. (2001). *Self-Organizing Maps, 3<sup>rd</sup> edition*. Berlin: springer.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide, 2<sup>nd</sup> edition*. Mahwah: Lawrence Erlbaum Associates, Publishers.
- Nolen-Hoeksema, S., Fredrickson, B. L., Loftus, G. R., & Wagenaar, W. A. (2009). *Atkinson & Hilgard's introduction to psychology, 15<sup>th</sup> edition*. Andover: Cengage Learning EMEA.
- Okamoto, Y. (2008). Bayesian analysis for IRT based on ability categories. *Proceedings of the 6th Annual Conference of The Japan Association for Research on Testing*, 98-101. (in Japanese).
- Okamoto, Y. (2009). Comparison between discrete and continuous IRTs. *Proceedings of the 7th Annual Conference of The Japan Association for Research on Testing*, 142-145. (in Japanese).
- Restle, F., & Greeno, J. G. (1970). *Introduction to mathematical psychology*. Reading: Addison-Wesley Publishing Company.
- Shojima, K. (2007). Neural test theory. *Proceedings of the 5th Annual Conference of The Japan Association for Research on Testing*, 174-177. (in Japanese).