Lecture 7: Rasch Model & Conditional Maximum Likelihood Estimation

(Baker & Kim (2004), Chapter 5; Embretson & Reise (2000), Chapter 8)

The Rasch model

Rasch used two symbols: η_j for the ability of the examinee $(0 \leq \eta_j \leq \infty)$, and δ_i for the difficulty of the item $(0 \leq \delta_i \leq \infty)$. Because of the meaningful location of a zero point, the scales of these latent variables are regarded as having a ratio-level interpretation. For example, examinee A has twice the ability of examinee B provided $\eta_A = 2\eta_B$ and the difficulty of item a is twice as great as the difficulty of item b if $\delta_a = 2\delta_b$. Moreover, since the first examinee has twice ability of the second, and the first item is twice as difficult as the second, the ratio of the first examinee's ability to the first item should be the same as the ratio of the second examinee's ability to the second item:

$$\frac{\eta_A}{\delta_a} = \frac{2\eta_B}{2\delta_b} = \frac{\eta_B}{\delta_b}$$

This ratio is called the *situational parameter*, and is denoted by ξ_{ij}:

$$\xi_{ij} = \frac{\eta_j}{\delta_i}.$$

This situational parameter is related to the probability of answering the item correctly. But since probability is scaled between 0 and 1, the following transformation of ξ_{ij} is used to produce a probability:

$$f(\xi_{ij}) = \frac{\xi_{ij}}{1+\xi_{ij}}.$$

We can think of f as defining item response function, where

$$P(u_{ij} = 1|\xi_{ij}) = \frac{\xi_{ij}}{1+\xi_{ij}}$$

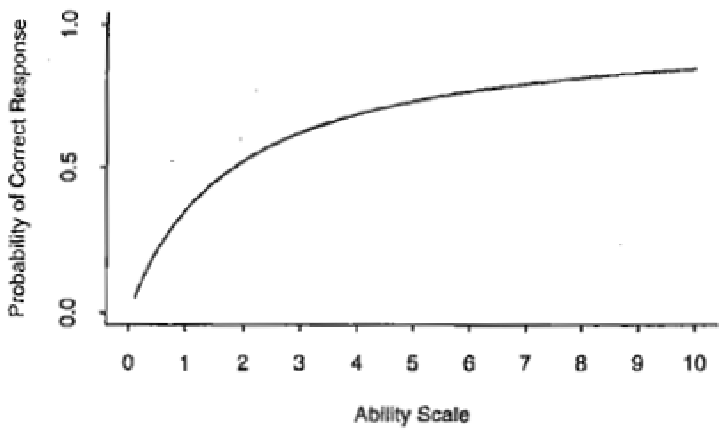
$$P(u_{ij} = 0|\xi_{ij}) = 1 - \frac{\xi_{ij}}{1+\xi_{ij}} = \frac{1}{1+\xi_{ij}}$$

or in terms of the η_j and δ_i parameters:

$$P(u_{ij} = 1|\eta_j, \delta_i) = \frac{\eta_j / \delta_i}{1 + \eta_j / \delta_i}$$

$$P(u_{ij} = 0|\eta_j, \delta_i) = 1 - \frac{\eta_j / \delta_i}{1 + \eta_j / \delta_i} = \frac{1}{1 + \eta_j / \delta_i}$$

When scaling ability according to η_j and δ_i, the ICC is shown below (Figure 5.1, p.113).

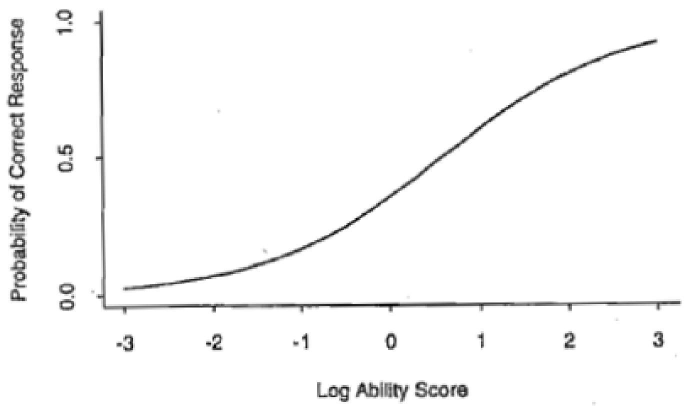


The more common IRT version of this model simply transforms the ability and difficulty scales as:

$$\theta_j = \log \eta_j$$
$$\beta_i = \log \delta_i$$

This transformation results in the Rasch model of the form: $P(u_{ij}|\theta) = \frac{\exp(\theta_j - \beta_i)u_{ij}}{1 + \exp(\theta_j - \beta_i)}$.

See below for Rasch ICC obtained after logarithm transformation (Figure 5.2, p. 115).



Separation of Parameters

Following Rasch's notation let $\varepsilon_i = \dfrac{1}{\delta_i}$ be an *easiness parameter* that is an inverse of the difficulty parameter. Then the probability of correct response for an item i by an examinee j can be written as:

$$P(u_{ij}|\eta_j) = \frac{(\eta_j \varepsilon_i)^{u_{ij}}}{(1+\eta_j \varepsilon_i)}$$

So the likelihood function can be written as:

$$P(u_{ij}|\eta_j) = \frac{\prod_{i=1}^{n} \eta_j^{u_{ij}} \varepsilon_i^{u_{ij}}}{\prod_{i=1}^{n} (1+\eta_j \varepsilon_i)} = \frac{\eta_j^{s_j} \prod_{i=1}^{n} \varepsilon_i^{u_{ij}}}{\prod_{i=1}^{n} (1+\eta_j \varepsilon_i)}$$

where $s_j = \sum_{i=1}^{n} u_{ij}$, or the total test score for person j.

This formula can be used to consider the probability that an examinee's pattern is one of many that all result in the same total score $s_j = r$. For example when $s_j = 1$, there are a total of n different possible response patterns, and the conditional probability of each is given by:

$$P(u_{1j} = 1|\eta_j) = \frac{\eta_j \varepsilon_1}{\prod_{i=1}^{n} (1+\eta_j \varepsilon_i)}$$

$$P(u_{2j} = 1 | \eta_j) = \frac{\eta_j \varepsilon_2}{\prod_{i=1}^{n} (1 + \eta_j \varepsilon_i)}$$

$$\vdots$$

$$P(u_{nj} = 1 | \eta_j) = \frac{\eta_j \varepsilon_n}{\prod_{i=1}^{n} (1 + \eta_j \varepsilon_i)}$$

To simplify the notation let $d(\eta_j) = \prod_{i=1}^{n} (1 + \eta_j \varepsilon_i)$. Then the probability of $s_j = 1$ is computed by simply summing the probabilities of each of the patterns above:

$$P(s_j = 1 | \eta_j) = \frac{\sum_{i=1}^{n} \eta_j \varepsilon_i}{d(\eta_j)} = \frac{\eta_j \sum_{i=1}^{n} \varepsilon_i}{d(\eta_j)}$$

For a test score of $s_j = 2$, there are $\binom{n}{2}$ different possible response patterns. Their probabilities for an examinee of ability η_j are as follows:

$$P(u_{1j} = 1, u_{2j} = 1 | \eta_j) = \frac{\eta_j^2 \varepsilon_1 \varepsilon_2}{d(\eta_j)}$$

$$P(u_{1j} = 1, u_{3j} = 1 | \eta_j) = \frac{\eta_j^2 \varepsilon_1 \varepsilon_3}{d(\eta_j)}$$

$$\vdots$$

$$P(u_{n-1,j} = 1, u_{nj} = 1 | \eta_j) = \frac{\eta_j^2 \varepsilon_{n-1} \varepsilon_n}{d(\eta_j)}$$

And then the sum of these probabilities, which is equivalent to the probability of obtaining a total test score of 2, is:

$$P(s_j = 2 | \eta_j) = \frac{\eta_j^2 (\varepsilon_1 \varepsilon_2 + \varepsilon_1 \varepsilon_3 + ... + \varepsilon_{n-1} \varepsilon_n)}{d(\eta_j)}$$

The general pattern for evaluating the probability of obtaining a total test score of $s_j = r$ is:

$$P(s_j = r|\eta_j) = \frac{\eta_j^r}{d(\eta_j)} \sum_{(u_{ij})}^{r} \left(\prod_{i=1}^{n} \varepsilon_i^{u_{ij}} \right)$$

The element $\sum_{(u_{ij})}^{r} \left(\prod_{i=1}^{n} \varepsilon_i^{u_{ij}} \right)$ is called an *elementary symmetric function* and is denoted as

$$\gamma_r = \sum_{(u_{ij})}^{r} \left(\prod_{i=1}^{n} \varepsilon_i^{u_{ij}} \right).$$

where $\sum_{(u_{ij})}^{r}$: the number of terms in the sum as a function of $\binom{n}{r}$ for a given value of r, and (u_{ij}) denotes an item response vector for person j that yields a test score of r.

Symmetric Functions

r	γ_r	Terms	Number of terms
0	γ_0	1	1
1	γ_1	$\varepsilon_1+\varepsilon_2+\varepsilon_3+\ldots+\varepsilon_n$	n
2	γ_2	$\varepsilon_1\varepsilon_2+\varepsilon_1\varepsilon_3+\ldots+\varepsilon_{n-1}\varepsilon_n$	$\binom{n}{2}$
3	γ_3	$\varepsilon_1\varepsilon_2\varepsilon_3+\varepsilon_1\varepsilon_2\varepsilon_4+\ldots+\varepsilon_{n-2}\varepsilon_{n-1}\varepsilon_n$	$\binom{n}{3}$
:	:	:	:
:	:	:	:
r	γ_r	$(\varepsilon_1\varepsilon_2\varepsilon_3\varepsilon_4\ldots\varepsilon_r)+(\varepsilon_2\varepsilon_3\varepsilon_4\ldots\varepsilon_{r-1}\varepsilon_{r+1})+(\varepsilon_{n-r}\varepsilon_{n-r+1}\ldots\varepsilon_n)$	$\binom{n}{r}$
:	:	:	:
:	:	:	:
n	γ_n	$\prod_{i=1}^{n}\varepsilon_i$	1

Since the sum of all the item response pattern probabilities across all possible test scores is 1,

$$\sum_{r=0}^{n}\frac{\eta_j^r\gamma_r}{d(\eta_j)}=1 .$$

Because $d(\eta_j)$ does not depend on r,

$$\sum_{r=0}^{n}\eta_j^r\gamma_r=d(\eta_j) ,$$

and since $d(\eta_j)=\prod_{i=1}^{n}(1+\eta_j\varepsilon_i)$,

$$\sum_{r=0}^{n}\eta_j^r\gamma_r=\prod_{i=1}^{n}(1+\eta_j\varepsilon_i) .$$

To model the probability of a single response pattern for an examinee j, we can then write

$$P(u_{ij}|s_j = r) = \frac{P(u_{ij}|\eta_j)}{P(s_j = r|\eta_j)} = \frac{\eta_j^r \prod_{i=1}^{n} \varepsilon_i^{u_{ij}}}{d(\eta_j)} \Bigg/ \frac{\eta_j^r \gamma_r}{d(\eta_j)} = \frac{\prod_{i=1}^{n} \varepsilon_i^{u_{ij}}}{\sum_{(u_{ij})}^{r} \left\{ \prod_{i=1}^{n} \varepsilon_i^{u_{ij}} \right\}}.$$

The important result here is that the probability of getting a particular item correct is no longer a function of η_j and it is conditional upon total test score. In other words, the total test score is a sufficient statistic for η_j .

Example: Suppose three items with $\beta_1 = -1.0$, $\beta_2 = 0.0$, $\beta_3 = 1.0$. Suppose a student answers one item correctly. Based on the above result, we can compute a probability that each item is answered correctly without knowing the student's η_j .

The same approach can be followed with respect to the items. Taking the total correct score for the item, we can write the likelihood of a response pattern across examinees for the item as a function of only the η_j s for the examinees and without any reference to the difficulty parameter (or easiness parameter) of the items. In other words, knowing the total number of examinees that answered the item correctly is a sufficient statistic for the difficulty parameter for the item.

These two results allow for a separation of the two groups of parameters (the η s and the δ s or ε s) so that one set can be estimated without direct reference to the other.

- Once number correct is known for an examinee, nothing else about the examinee contributes to helping estimate the item parameters.
- Once number correct is known for an item, nothing else about the item contributes to helping estimate the ability parameter for an examinee.

Conditional Maximum Likelihood Procedure

The basic idea in conditional maximum likelihood (CML) estimation is to use the known total test scores for each examinee, and find the set of item parameters β_1 , β_2 , …, β_n that make the likelihood function conditional upon the test scores and unknown β parameters as large as possible.

Since the parameterization of the model we are most accustomed to involves β s instead of δ s or ε s, we also redefine in this notation the likelihood of each examinee response pattern. The conditional probability of a response pattern given test score is:

$$P(u_j | s_j = r) = \frac{\exp\left(-\sum_{i=1}^{n} u_{ij}\beta_i\right)}{\gamma(r,\beta)}$$

where

$$\gamma(r,\beta) = \sum_{(u_{ij})}^{r_j} \exp\left(-\sum_{i=1}^{n} u_{ij}\beta_i\right).$$

The likelihood of data across all examinees given test scores and unknown parameters is:

$$L = \frac{\exp\left(-\sum_{i=1}^{n} s_i\beta_i\right)(-\sum_{i=1}^{n} s_i\beta_i)}{\prod_{r=0}^{n}[\gamma(r,\beta)]^{f_r}}$$

where s_i is the number of examinees that answer correctly, and f_r is the number of examinees that had test scores of r .

We can compute the first and second derivatives of the log of this expression and then apply the Newton-Raphson procedure to get the maximum-likelihood item parameter estimates. The computational work of the procedure is mainly due to having to evaluate the elementary symmetric functions. Fortunately the derivatives of the elementary symmetric functions are expressed easily themselves as other elementary symmetric functions. There have been several procedures proposed for computing the elementary symmetric functions.

- CML estimation is implemented among other places in the program WINMIRA. There is a student version of this program that can be downloaded for free at http://www.von-davier.com/

- Rasch Measurement Analysis Software Directory is available at http://www.rasch.org/software.htm