Chapter 10
Estimating Person Proficiency and Person Separation

Statistics Review 10: Bernoulli and Binomial random variables.

We continue with the dichotomous Rasch model and with the context of the assessment of proficiency. In this chapter, we use the set of responses of persons to items to estimate their proficiencies, given the estimates of item difficulties. Where the previous chapter was concerned with item calibration, this chapter is concerned with person measurement.

In theory, by conditioning on the total scores of items, we can estimate the person parameters independently of all item parameters. However, that has only recently been made operational and it is not yet practical. Instead, in estimating the person parameters it is assumed that the item parameters are known. This can be assumed if they have been estimated as described in the previous chapter.

Let the probability of a correct response of person n to item i be denoted simply as P_{ni}. Then, according to the dichotomous Rasch model, this probability is given by

$$P_{ni} = \Pr\{x_{ni} = 1\} = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}}. \tag{10.1}$$

Make sure you understand the probability of a Bernoulli random variable in *Statistics Review* 10 that helps in understanding the use and meaning of P_{ni}.

This section could be in a statistics review. However, because we consider that the way the items are formalized, how the probability statements are interpreted, and how the scores on items can be summed, is integral to the interpretation of statistical analyses of assessments, we have included it in the main part of the book.

Solution Equations in the Rasch Model

Recall from previous chapters that the total score is a *sufficient statistic* for its parameter, in this case, the proficiency of the person. Thus, the total person score $r_n = \sum_{i=1}^{I} x_{ni}$ is the sufficient statistic for the estimate of the person proficiency

β_n where I is the number of items responded to by person n. That is, all of the information about β_n resides in the total score r_n.

In the previous chapters, we used the sufficiency of the total score to show how the person parameter can be eliminated to produce equations for estimating the item difficulties without knowledge of the person proficiencies. We now use the total score for a second purpose, to estimate the proficiency β_n of person n, given that we have the estimates of the item difficulties. We can now relate this estimation to *Statistics Review* 10.

We build the equation for the estimation of the person parameter by analogy. We then write out the formal equation and show how it can be derived using the idea of maximum likelihood introduced in the last chapter.

In *Statistics Review* 10, we show an example where the outcomes of Bernoulli variables are summed and where the responses are replications of each other in the sense that the probability of a positive response is the same (i.e. a Binomial experiment). Below we show an example where the outcomes of Bernoulli variables are summed but where the probability of each response is different. Suppose person n responds to 10 items which are analogous to the 10 tosses of a coin in *Statistic Review* 10. The responses and the total score are shown in Table 10.1.

Now, rather than each response being a replication of the same person answering the same item, each item is different and will have a different difficulty from every other item. As a result, the items are not *replications* of each other as in the case of replicated Bernoulli variables (Binomial experiment).

Therefore, we need to imagine the mean score of each item in a different way. We take two steps to build up this imaginary set up. First, imagine that each item is given many times to the person, and consider the estimate of the probability that the person answers each item correctly. This would be the mean number of times of the many replications that the person answers the item correctly. However, we recognize that it is not reasonable to ask the same person to answer the same item many times. Therefore, second, imagine that it is not the identical item that is administered on more than one occasion, but that there are many different items of *exactly the same difficulty* that are administered to the person. In this case, the theoretical mean number of correct responses will be the estimate of the probability that the person will answer correctly any one of these items with the same difficulty. The distinction in the second last sentence above between the *identity* of an item and the *difficulty* of an item will appear throughout this book.

Table 10.1 Responses of person n to 10 items

Random variables	x_{n1}	x_{n2}	x_{n3}	x_{n4}	x_{n5}	x_{n6}	x_{n7}	x_{n8}	x_{n9}	x_{n10}	Total Score r_n
Value	1	1	1	0	1	1	0	0	1	0	6

Table 10.2 Probabilities of responses of a person to 10 items: the proficiency of person n is $\beta_n = 0.5$ and the difficulties of the items are $\delta_1 = -2.5$, $\delta_2 = -2.0$, $\delta_3 = -1.5$, $\delta_4 = -1.0$, $\delta_5 = -0.5$, $\delta_6 = 0.5$, $\delta_7 = 1.0$, $\delta_8 = 1.05$, $\delta_9 = 1.5$, $\delta_{10} = 2.0$

Random variables	x_{n1}	x_{n2}	x_{n3}	x_{n4}	x_{n5}	x_{n6}	x_{n7}	x_{n8}	x_{n9}	x_{n10}	Total score r_n
Observed value	1	1	1	0	1	1	0	0	1	0	6
Average $\bar{X}_{ni}$	0.95	0.92	0.88	0.82	0.73	0.50	0.38	0.37	0.27	0.18	6.0
Probability P_{ni}	0.95	0.92	0.88	0.82	0.73	0.50	0.38	0.37	0.27	0.18	6.0

Table 10.2 shows a set of observed responses and such estimated probabilities of a correct response for each item which satisfy another condition. This condition is as follows:

The sum of the probabilities (theoretical means) of the number of times each item is answered correctly is equal to the number of items that are answered correctly. Thus, the sum of each row in Table 10.2 is 6.

To think about this, it might help if you imagine first that 10 items of exactly the same difficulty were answered by a person. If the probability of being correct on these items is 0.6, then one would expect that the number of times a correct answer would be given is 6.

That is,

$$0.6 + 0.6 + 0.6 + 0.6 + 0.6 + 0.6 + 0.6 + 0.6 + 0.6 + 0.6 = 10(0.6) = 6$$

as in the coin example in *Statistics Review* 10.

The case in Table 10.2 is analogous, except that every item has a different probability (theoretical mean number) of correct responses. For example, starting from the left, the items are successively more difficult for the person. Nevertheless, the sum of all of these probabilities should equal the number of correct responses, in this case 6.

The Solution Equation for the Estimate of *Person Proficiency*

The above rationale permits setting up equations to estimate the proficiency of each person, given the estimates of the difficulty for each item.

Table 10.2 shows the set up that the sum of the probabilities (means) of each item correct should be equal to the total number of correct responses. In equation form,

$$r_n = \sum_{i=1}^{10} x_{ni} = \sum_{i=1}^{10} P_{ni}. \tag{10.2}$$

However, the probability that a person answers an item correctly can be expressed in terms of the person's proficiency and the item's difficulty, that is, the dichotomous RM equation:

$$P_{ni} = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}} \tag{10.3}$$

Therefore, Eq. (10.2) can be written as

$$r_n = \sum_{i=1}^{10} \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}}$$

$$= \frac{e^{\beta_n - \delta_1}}{1 + e^{\beta_n - \delta_1}} + \frac{e^{\beta_n - \delta_2}}{1 + e^{\beta_n - \delta_2}} + \cdots\cdots \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}} \cdots\cdots + \frac{e^{\beta_n - \delta_{10}}}{1 + e^{\beta_n - \delta_{10}}} \tag{10.4}$$

In words, the sum of the probabilities (or theoretical means) of answering each item correctly, must be equal to the number correct. In general, replacing the 10 items by any number of, say I, items gives

$$r_n = \sum_{i=1}^{I} \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}}. \tag{10.5}$$

Thus, given that the difficulties of the items are known, for example, estimated using the procedures from the last chapter, the one unknown value β_n in Eq. (10.5) can be calculated.

Solving the Equation by Iteration

This equation cannot be solved explicitly and we rely on computers to solve it. The equation is solved iteratively in a systematic way but iteratively. In particular, an initial value of β_n is started with, the probabilities calculated and summed. If this sum is greater than r_n that indicates that our first estimate of β_n is too large and that we should reduce it a little. On the other hand, if it is less than r_n, it indicates that our estimate is too small and that we should increase it a little. That is one *iteration*. The same procedure is continued with the new value, and this is the second iteration. When the sum of the probabilities is close enough to r_n according to some criterion that is set, for example, only 0.001 different from r_n, then the iterations are stopped and it is said that the iterations have *converged* on a solution to the chosen criterion of accuracy. The chosen criterion is called the *convergence criterion*. You do not have to carry out these calculations, but it helps to have an idea of how they are done.

For example, in the above case of 10 items, suppose we knew the items to have the difficulties shown in Table 10.2, and that we know the person's total score was

$r_n = 6$ as above. Our first estimate for the proficiency might be $\beta_n^{(0)} = 0.25$ (based on experience).

Then inserting this value in Eq. (10.5) gives

$$\sum_{i=1}^{10} \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}} = \frac{e^{\beta_n^{(0)} - \delta_1}}{1 + e^{\beta_n^{(0)} - \delta_1}} + \frac{e^{\beta_n^{(0)} - \delta_2}}{1 + e^{\beta_n^{(0)} - \delta_2}} + \cdots\cdots + \frac{e^{\beta_n^{(0)} - \delta_{10}}}{1 + e^{\beta_n^{(0)} - \delta_{10}}}$$

$$= \frac{e^{0.25+2.5}}{1 + e^{0.25+2.5}} + \frac{e^{0.25+2.0}}{1 + e^{0.25+2.0}} + \cdots\cdots + \frac{e^{0.25-2.0}}{1 + e^{0.25-2.0}}$$

$$= 0.94 + 0.90 + 0.85 + 0.78 + 0.68 + 0.44$$

$$+ 0.32 + 0.31 + 0.22 + 0.15 = 5.59.$$

This means that with a proficiency of $\beta_n = 0.25$, the person would be expected to obtain a score of 5.59. However, the person has a score of 6.0, therefore, the proficiency estimate should be a little greater.

We could try $\beta_n^{(1)} = 0.40$. In that case, we would obtain

$$\sum_{i=1}^{10} \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}} = \frac{e^{\beta_n^{(1)} - \delta_1}}{1 + e^{\beta_n^{(1)} - \delta_1}} + \frac{e^{\beta_n^{(1)} - \delta_2}}{1 + e^{\beta_n^{(1)} - \delta_2}} + \cdots\cdots + \frac{e^{\beta_n^{(1)} - \delta_{10}}}{1 + e^{\beta_n^{(1)} - \delta_{10}}}$$

$$= \frac{e^{0.40+2.5}}{1 + e^{0.40+2.5}} + \frac{e^{0.40+2.0}}{1 + e^{0.40+2.0}} + \cdots\cdots + \frac{e^{0.40-2.0}}{1 + e^{0.40-2.0}}$$

$$= 0.95 + 0.92 + 0.87 + 0.80 + 0.71 + 0.48 + 0.35$$

$$+ 0.34 + 0.25 + 0.17 = 5.84.$$

The value of β_n must be a little greater than 0.40, and so we could try 0.45. By this successive process, we would reach $\beta_n = 0.50$ correct to two decimal places.

It sometimes happens that the process does not converge to a solution. However, this is rare in the Rasch model and if it occurs, there are mechanisms to make the algorithm a little more sophisticated and to obtain convergence. Most computer programs have this sophistication built into them. If the Rasch model equation really does not converge, then this is a property of the data and not the model. Again, this is rare, but it is possible. Fischer (1981) describes such a case.

Initial Estimates

To set initial estimates for each β_n it is common to assume all items have the same difficulty of 0. In that case, Eq. (10.5) reduces to

$$r_n = \sum_{i=1}^{I} \frac{e^{\beta_n}}{1 + e^{\beta_n}} = I \frac{e^{\beta_n}}{1 + e^{\beta_n}}$$

$$\frac{r_n}{I} = \frac{e^{\beta_n}}{1 + e^{\beta_n}} \tag{10.6}$$

$$\text{and} \quad 1 - \frac{r_n}{I} = 1 - \frac{e^{\beta_n}}{1 + e^{\beta_n}} \quad \text{i.e.} \quad \frac{I - r_n}{I} = \frac{1}{1 + e^{\beta_n}}$$

and inverting gives

$$\frac{I}{I - r_n} = 1 + e^{\beta_n} \tag{10.7}$$

Multiplying Eq. (10.6) by (10.7) gives

$$\frac{r_n}{I}\left(\frac{I}{I - r_n}\right) = \frac{e^{\beta_n}}{1 + e^{\beta_n}}(1 + e^{\beta_n}) \quad \text{i.e.} \quad \frac{r_n}{I - r_n} = e^{\beta_n}$$

$$\text{and } \beta_n = \log\left(\frac{r_n}{I - r_n}\right) \tag{10.8}$$

Proficiency Estimates for Each Person

Below is an analysis of the data from Table 5.3 of Chap. 5. Table 10.3 shows the proficiency associated with each total score for the set of items, and three features are noted.

For Responses to the Same Items, the Same Total Score Leads to the Same Person Estimate

Where students respond to the same items, then irrespective of the pattern of responses, the *same total score leads to the same proficiency estimate*. This is evident from Eq. (10.5), in which there is no information about the actual responses—only the total score is used. It is a manifestation of the sufficiency of the total score for the person parameter β.

Estimate for a Score of 0 or Maximum Score

For a person with the *maximum total score* of 18, the *proficiency estimate is infinite* ($+\infty$). This is because the person's proficiency is above the limit of the difficulty of the test, and the probability of a correct response must be 1.00 for all the items. It is as if an adult stood on a weighing machine for babies, and the indicator hit the top of

Table 10.3 Proficiency estimates for the dichotomous items of Table 5.3, persons are ordered by proficiency and items by difficulty

Person	Responses	Total Score r_n	Location $\hat{\beta}$ (MLE)	SE
38	101101001010000000	6	−0.920	0.550
2	101101110100000000	7	−0.625	0.537
40	010111110000101000	8	−0.340	0.531
42	110011111110010000	10	0.227	0.538
41	111101111101000000	10	0.227	0.538
44	111101111111000000	11	0.523	0.551
8	111110110101110000	11	0.523	0.551
35	111111011101100000	11	0.523	0.551
11	101111111110011000	12	0.837	0.572
9	110111111011011000	12	0.837	0.572
46	111011011011011010	12	0.837	0.572
29	111101111011110000	12	0.837	0.572
25	111110101111110000	12	0.837	0.572
27	011101111101100111	13	1.181	0.602
18	110111110111101001	13	1.181	0.602
36	111011110111111000	13	1.181	0.602
37	111101111111101000	13	1.181	0.602
20	111110011111101100	13	1.181	0.602
48	111110101111111000	13	1.181	0.602
13	111111011011110100	13	1.181	0.602
34	111111011100111100	13	1.181	0.602
32	111111101011111000	13	1.181	0.602
22	111111101101101001	13	1.181	0.602
43	111111111101110000	13	1.181	0.602
14	111110101111011110	14	1.570	0.647
12	111111100101011111	14	1.570	0.647
15	111111100110111110	14	1.570	0.647
21	111111111000111110	14	1.570	0.647
5	111111111010011110	14	1.570	0.647
4	111111111110011100	14	1.570	0.647
16	111111111110111000	14	1.570	0.647
45	111111111111000011	14	1.570	0.647
17	111111111111100100	14	1.570	0.647
7	111111110111111100	15	2.030	0.714
50	111111111110011110	15	2.030	0.714
49	111111111110111100	15	2.030	0.714

Table 10.3 (continued)

Person	Responses	Total Score r_n	Location $\hat{\beta}$ (MLE)	SE
23	111111111111110001	15	2.030	0.714
6	111111111111111000	15	2.030	0.714
24	111111111111111000	15	2.030	0.714
33	111110111111111101	16	2.615	0.826
26	111111011111111101	16	2.615	0.826
10	111111111110110111	16	2.615	0.826
31	111111111111101110	16	2.615	0.826
19	111111111111111010	16	2.615	0.826
30	101111111111111111	17	3.481	1.081
28	111011111111111111	17	3.481	1.081
1	111111111111101111	17	3.481	1.081
39	111111111111111101	17	3.481	1.081
47	111111111111111101	17	3.481	1.081
3	111111111111111111	18	$+\infty$ (4.762)	∞ (1.658)

the available scale, in which case the person's weight is unknown. Clearly, the person is beyond the limit measurable by the particular machine. In that case, it would be necessary to use a weighing machine that measures greater weights. In the example with items, the person would take more difficult items to establish a finite estimate of proficiency. Thus, the person is not thought to actually have infinite proficiency, it is just that a finite estimate cannot be obtained from these particular items. Likewise, if a person answered all the items incorrectly, then the person would have a proficiency estimate of $-\infty$. In order to get a finite estimate of proficiency for such a person, easier items should be used.

Sometimes groups of people need to be compared say for relative improvement or for baseline data. For example, we may have responses from boys and girls and we might have assessed them on some proficiency before some program of teaching is in place. If different numbers of boys and girls obtain a 0 score or maximum score, it would bias the comparison of the boys and girls if they were left out. Although they are left out in the item calibration, they cannot be left out of the group comparisons. Therefore, there is the need to provide an estimate of a person with a maximum (or minimum score of 0). Different methods have been devised for this purpose, and they all involve some extra assumption or reasoning that goes beyond the model itself. Some methods make it explicit that the person with a maximum or minimum score belongs to the population of persons. In RUMM2030 (Andrich, Sheridan, & Luo, 2018), a value is extrapolated by observing that the relative differences in successive values at the extremes increase. Thus, in this example, the successive difference between the scores of 15 and 14, 16 and 15 and 17 and 16 are 0.46, 0.59 and 0.87, showing successive increases. The procedure in RUMM2030 specifically uses the

geometric mean of the differences of the three scores before the maximum score. As a result, the extrapolated value for a score of 18 is 4.762. The same principle is used to extrapolate the value for a score of 0, shown later for this example.

The Standard Error of Measurement of a Person

There is an extra column in Table 10.3 giving the standard error of measurement for each person. We do not derive this equation in this book, but it also arises directly from maximum likelihood theory.

The equation for estimating this standard error is relatively simple, it is given by

$$\sigma_{\hat{\beta}} = \frac{1}{\sqrt{\sum_{i=1}^{I} P_{ni}(1 - P_{ni})}} \tag{10.9}$$

Unlike CTT, the *standard error is not the same for all persons*. Compare Eq. (10.9) with Eq. (3.5) of Chap. 3 where the CTT standard error is a function of test reliability and variance. In the dichotomous RM the standard error of measurement *depends on the total score*; if a person has answered few or very many items correctly, then the standard error is greater than if the person has answered a moderate number of items correctly. If a person has answered all the items correctly, then the standard error is infinitely large. This is consistent with not having a finite estimate of the person's proficiency. Again, RUMM2030 provides a value using Eq. (10.9) for the extrapolated value. As expected, it is large and larger than the standard error for the score one less than the maximum.

Proficiency Estimate for Each Total Score When All Persons Respond to the Same Items

In the case that all persons have responded to all items, there is another way of displaying the information in Table 10.3. It is displayed by the total score, the proficiency estimate associated with the total score, and the standard error. Table 10.4 is such a table.

Two special features of Table 10.4 are noted. First, total scores with zero frequency (e.g. a score of 9) have proficiency estimates, and second transformation from a total score to an estimate is non-linear.

Table 10.4 Total scores, frequencies, proficiency estimates and standard errors

Raw score	Frequency	Location (MLE)	Std Error
0	0	$-\infty$ (−4.543)	∞ (1.679)
1	0	−3.316	1.050
2	0	−2.515	0.784
3	0	−1.993	0.672
4	0	−1.584	0.611
5	0	−1.235	0.573
6	1	−0.920	0.550
7	1	−0.625	0.537
8	1	−0.340	0.531
9	0	−0.059	0.531
10	2	0.227	0.538
11	3	0.523	0.551
12	5	0.837	0.572
13	11	1.181	0.602
14	9	1.570	0.647
15	6	2.030	0.714
16	5	2.615	0.826
17	5	3.481	1.081
18	1	$+\infty$ (4.762)	∞ (1.658)

Estimates for Every Total Score

There are some scores in Table 10.4 which no one has achieved. For example, there is no one with a score of 1, 2, 3, 4, 5 and 9. Nevertheless, there is a proficiency estimate associated with these scores. This is because, given the difficulty of the items, the proficiency for each total score can be estimated from Eq. (10.5). Likewise, the standard error for these scores can be estimated from Eq. (10.9).

For example, for a score of 9, Eq. (10.5) becomes

$$9 = \sum_{i=1}^{18} \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}}$$

and every person who has responded to these items obtains the same estimate.

Non-linear Transformation from Raw Score to Person Estimate

Although the distance between all successive total scores is 1, the distances between the proficiency estimates for successive total scores are different. For example, the proficiency difference between scores of 4 and 5 is -1.584 to $(-1.235) = -0.349$ while the difference between scores of 10 and 11 is $0.227-(0.523) = -0.296$. These differences reflect the non-linear transformation of the raw scores to the estimates. This non-linear transformation is an attempt to undo the implicit effects of constrained, finite, minimum and maximum scores. When there are many maximum scores because there are not enough items of greater difficulty, and when there are many scores of 0 because there are not enough easier items, it is said that there is a *ceiling* and *floor* effect, respectively.

Figure 10.1 shows the non-linear transformation graphically for the responses in Table 10.4. The estimates are not symmetrical around 0 because the items are not uniformly spaced. Figure 10.2 shows how the standard errors of the estimates are greater at the extremes than in the middle of the score range.
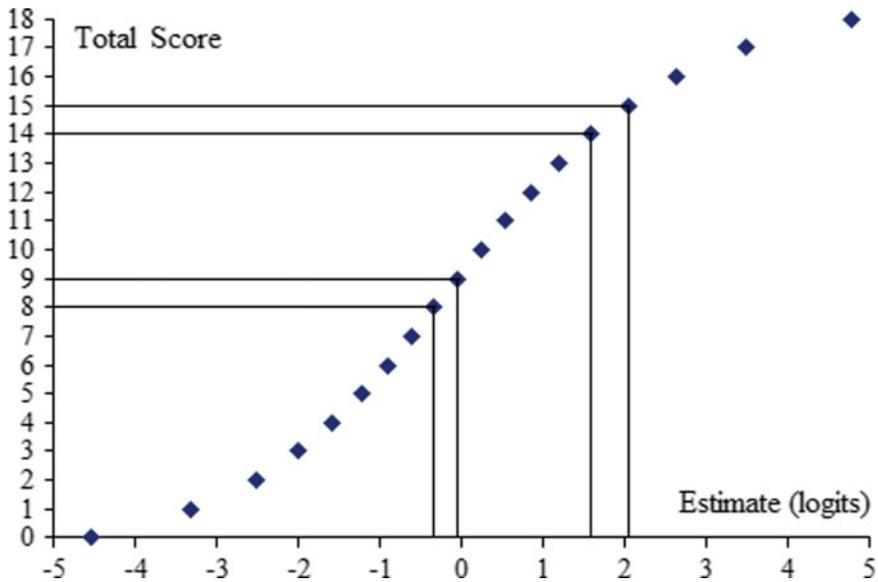

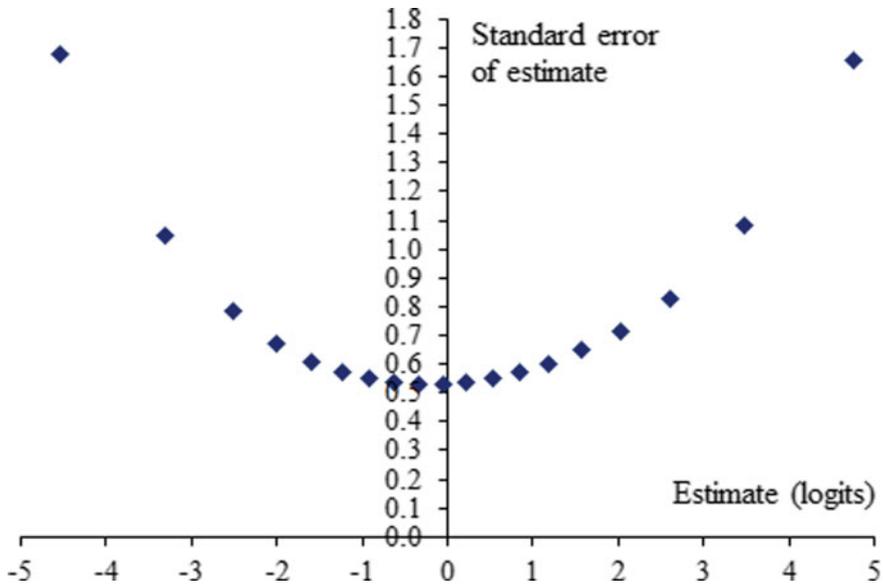**Fig. 10.1** Non-linear transformation of the total score to an estimate

Fig. 10.2 Standard errors as a function of the estimates

Displaying Person and Item Estimates on the Same Continuum

Figure 10.3 shows a graphical display of the person location estimates of Table 10.4 in this chapter and the item location estimates from Table 9.4 in the previous chapter. It shows the information from these two tables as histograms on the same scale, one above the horizontal axis showing the person distribution, and one below the horizontal axis showing the item distribution. This graph makes the interpretation of the person values more tangible in terms of the locations of the items. It is clear from this Figure that the persons overall found this test relatively easy. Thus, with a person mean of the order of 1.58, and the mean of the item difficulties defined to be 0.0, the probability that the average student will answer correctly a question with difficulty 0, given by Eq. (10.1), is 0.829. In tests of proficiency, we might expect success rates to be of more than 50%. However, there are individual students whose success is very low. In principle, it is possible to have different students attempt different questions which are adapted to their proficiencies, so students do not attempt items that are either too difficult or too easy for them. We consider this possibility, and the facilities of the Rasch model to cater to it, in a subsequent chapter.
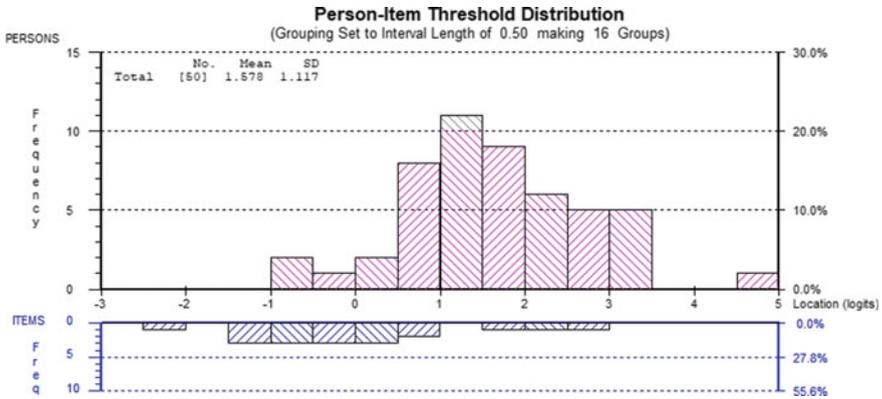
Fig. 10.3 Item and person estimates on the latent continuum

CTT Reliability Calculated from Rasch Person Parameter Estimates

The calculation of a reliability index has not been very common in modern test theory. However, it is possible to construct an index of reliability which is analogous in calculation and interpretation, and generally in value, using Rasch measurement theory. We demonstrate its construction first and then comment on its interpretation.

Derivation of r_β

Given the estimates of proficiency and the standard error of these estimates, it is possible to calculate a reliability index in a simple way.

The key point is to apply the CTT formula for reliability, Eq. (3.3) of Chap. 3:

$$r_{yy} = \frac{s_t^2}{s_y^2} = \frac{s_y^2 - s_e^2}{s_y^2} \tag{10.10}$$

However, instead of using the raw scores in this equation, we use the proficiency estimates. We use the same process in applying Eq. (10.10) except that we do this with the proficiencies. Thus, we consider that the proficiency estimate $\hat{\beta}_n$ for each person n can be expressed as the sum of the true latent proficiency and the error, that is

$$\hat{\beta}_n = \beta_n + \varepsilon_n \tag{10.11}$$

Thus instead of s_x^2 we use $\hat{\sigma}_{\hat{\beta}}^2$ which is the estimate of the variance of the estimates of proficiencies. This is simply given by

$$\hat{\sigma}_{\hat{\beta}}^2 = \frac{\sum_{n=1}^{N} \left(\hat{\beta}_n - \bar{\hat{\beta}} \right)^2}{N - 1} \tag{10.12}$$

where $\bar{\hat{\beta}}$ is the mean of the estimates of the persons.

This variance, being of the estimates, includes the variance of the errors $\hat{\sigma}_{\hat{\varepsilon}}^2$. To account for this variance of errors, the best we can do, even though the errors are a function of the locations of the persons, is take the average of the estimates of the variance of errors for each person. This is given simply by taking the average of the squares of the standard errors of measurement for each person, that is

$$\hat{\sigma}_{\hat{\varepsilon}}^2 = \frac{\sum_{n=1}^{N} \hat{\sigma}_n^2}{N} \tag{10.13}$$

The key feature of reliability in CTT is that it indicates the degree to which there is systematic variance among the persons relative to the error variance—it is the ratio of the estimated true variance relative to the true variance plus the error variance. In CTT, the reliability index can give the impression that it is a property of the test, when it is a property of the persons as identified by the test. The same test administered to people of a similar class of persons, but with a smaller true variance would be shown to have a lower reliability. Thus, the index needs to be interpreted with the distribution of the persons in mind.

Therefore, to focus on this qualification to its interpretation, we refer to this index (which is of the same kind as the traditional reliability index) as the person separation index (PSI) denoted r_β. Finally, therefore, we have

$$r_\beta = \frac{\hat{\sigma}_{\hat{\beta}}^2 - \hat{\sigma}_{\hat{\varepsilon}}^2}{\hat{\sigma}_{\hat{\beta}}^2} \tag{10.14}$$

where the components are given by Eqs. (10.12) and (10.13).

In the case with person/item distributions that are standard and to which the CTT reliability is correctly applied, the values of the coefficient α and those obtained from Eq. (10.14) are very similar (Andrich, 1982). However, in cases where coefficient α should not really be interpreted, the values might vary. The situation can occur when there is an artificially skewed distribution of scores in which there are floor or ceiling effects in the responses. Then the assumption that the sum of the item scores is effectively unbounded is grossly violated, and the coefficient α becomes inflated. It is inflated effectively because the scores of each person on the items may be more similar than they would be if there were no floor or ceiling effect. On the other hand, the error in the Rasch model is larger at the low and high scores and therefore r_β will be larger than α. In cases where every person responds to every item, both can be

calculated and compared. If they are very different, then the person/item distribution should be reexamined before either is interpreted. In any case, the interpretation of this index requires an examination of the person/item distribution and if there are floor and ceiling effects these should be noted. There are of course other factors that can affect the interpretation of this index and this is just one of them. Floor and ceiling effects will generate different values of r_β compared to α.

Example 1 For the data that are analyzed in Table 10.3,

$\hat{\sigma}_{\hat{\beta}}^2 = 1.25$ and $\hat{\sigma}_{\hat{\varepsilon}}^2 = 0.54$

Therefore, $\hat{r}_\beta = \frac{\hat{\sigma}_{\hat{\beta}}^2 - \hat{\sigma}_{\hat{\varepsilon}}^2}{\hat{\sigma}_{\hat{\beta}}^2} = \frac{1.25 - 0.54}{1.25} = \frac{0.71}{1.25} = 0.57$.

Example 2 For the data including graded responses that are analyzed in *Part III* of this book,

$\hat{\sigma}_{\hat{\beta}}^2 = 0.79$ and $\hat{\sigma}_{\hat{\varepsilon}}^2 = 0.43$

Therefore, $\hat{r}_\beta = \frac{\hat{\sigma}_{\hat{\beta}}^2 - \hat{\sigma}_{\hat{\varepsilon}}^2}{\hat{\sigma}_{\hat{\beta}}^2} = \frac{0.79 - 0.43}{0.79} = \frac{0.36}{0.79} = 0.46$.

These are very moderate values and are explained by the fact that the test was a little easy and persons were grouped at the top end of the range, and that the test is short. That this index is smaller when the data are grouped indicates that the responses within the items that are combined have some dependencies and that the dichotomous data gave an artificially high reliability.

In addition to providing the same kind of information as the index α, this index is readily calculated if there are missing data without any extra assumptions needing to be made. Missing data can occur either with some people missing some items at random or when there is some structural missing data, for example, different groups of persons are not all given the same items. This case is considered in the chapter where linking tests with common items are discussed.

In addition, as is considered in *Part II* of this book, the index is relevant in the power to detect misfit of the responses to the Rasch model.

Principle of Maximum Likelihood

We now take the opportunity to show more explicitly the idea of maximum likelihood, which is central to estimation in the Rasch model and statistics in general. Again, some of this material could be a statistics review, but because it is central to the Rasch model we have retained it in the main section of the book.

We have in the dichotomous RM that

$$\Pr\{x_{ni} = 1\} = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}}; \quad \Pr\{x_{ni} = 0\} = \frac{1}{1 + e^{\beta_n - \delta_i}} \tag{10.15}$$

We have seen that these two sub equations can be written as one equation:

$$\Pr\{x_{ni}\} = \frac{e^{x_{ni}(\beta_n - \delta_i)}}{1 + e^{\beta_n - \delta_i}} \tag{10.16}$$

Now consider the probability, according to this equation, that the first person (Person 38) in Table 10.3 has those responses. To do this, we apply the principle of statistical independence we broached in Chap. 7. Thus, a person's actual response to one question does not affect the response to any other question, other than through the person's proficiency parameter, which governs responses to all items. Accordingly, the probability of the person's responses is given by the product of the probabilities of responses to individual items. This probability is

$$\Pr\{(x_{1i})\} = \prod_{i=1}^{18} \frac{e^{x_{1i}(\beta_1 - \delta_i)}}{1 + e^{\beta_1 - \delta_i}}$$
$$= \frac{e^{1(\beta_1 - \delta_1)}}{1 + e^{\beta_1 - \delta_1}} \frac{e^{0(\beta_1 - \delta_2)}}{1 + e^{\beta_1 - \delta_2}} \frac{e^{1(\beta_1 - \delta_3)}}{1 + e^{\beta_1 - \delta_3}} \cdots \frac{e^{0(\beta_1 - \delta_{18})}}{1 + e^{\beta_1 - \delta_{18}}}. \tag{10.17}$$

This equation can be simplified by simply summing the exponents of the numerators, and multiplying the denominator term which is exactly the same in every term. This gives

$$\Pr\{(x_{1i})\} = \prod_{i=1}^{18} \frac{e^{x_{1i}(\beta_1 - \delta_i)}}{1 + e^{\beta_1 - \delta_i}}$$
$$= \frac{e^{6\beta_1 - \sum_{i=1}^{k} x_{1i}\delta_i}}{\prod_{i=1}^{18}(1 + e^{\beta_1 - \delta_i})}. \tag{10.18}$$

Notice that the coefficient of the person proficiency β_1 in the numerator is the person's total score—the sufficient statistic. The other term in the numerator is simply the sum of parameters of the items that the person has answered correctly. We will see that this term plays no role in the final equation.

Now consider the same equation for every other person. Equation (10.18) is simply repeated for each person. We now assume statistical independence of responses between persons. For example, we consider that the students have not worked together to provide the same response to any item. Then to obtain the probability of the matrix of responses, we simply multiply Eq. (10.18) across all persons. This is written in general, with $N = 50$ and $I = 18$, as

$$L = \Pr\{(x_{ni})\} = \prod_{n=1}^{N} \prod_{i=1}^{I} \frac{e^{x_{ni}(\beta_n - \delta_i)}}{1 + e^{\beta n - \delta_i}}$$
$$= \frac{e^{1(\beta_1 - \delta_1)}}{1 + e^{\beta_1 - \delta_1}} \frac{e^{0(\beta_1 - \delta_2)}}{1 + e^{\beta_1 - \delta_2}} \frac{e^{1(\beta_1 - \delta_3)}}{1 + e^{\beta_1 - \delta_3}} \cdots \frac{e^{0(\beta_1 - \delta_{18})}}{1 + e^{\beta_1 - \delta_{18}}} \cdots \cdots$$
$$\frac{e^{1(\beta_{49} - \delta_1)}}{1 + e^{\beta_{49} - \delta_1}} \frac{e^{1(\beta_{49} - \delta_2)}}{1 + e^{\beta_{49} - \delta_2}} \frac{e^{1(\beta_{49} - \delta_3)}}{1 + e^{\beta_{49} - \delta_3}} \cdots \frac{e^{1(\beta_{49} - \delta_{18})}}{1 + e^{\beta_{49} - \delta_{18}}}$$

$$= e^{6\beta_1 + \cdots\cdots + 17\beta_{49} - \sum_{n=1}^{N} \sum_{i=1}^{I} x_{ni}\delta_i} \prod_{n=1}^{N} \prod_{i=1}^{I} \frac{1}{1 + e^{\beta n - \delta_i}} \tag{10.19}$$

The last person (Person 3) is not included in Eq. (10.19) because that person has the maximum score of 18 and the person's theoretical estimate is $+\infty$. This person's estimate is extrapolated, not estimated.

L in front of this equation stands for the *Likelihood* of the responses, which is the joint probability of the matrix of all responses across persons and items.

Now take the logarithm of Eq. (10.19) which gives the *log-likelihood*:

$$\ln L = 6\beta_1 + \cdots\cdots + 17\beta_{49} - \sum_{n=1}^{N} \sum_{i=1}^{I} x_{ni}\delta_i - \ln \sum_{n=1}^{N} \sum_{i=1}^{I} 1 + e^{\beta n - \delta_i} \tag{10.20}$$

Now the task is to find the β_n value for each person that gives the maximum value for Eq. (10.20), which is the same value that maximizes the likelihood of Eq. (10.19). For example, we could try different values as we did above in obtaining a person's estimate. To obtain the equation for the maximum value requires calculus. It involves differentiating Eq. (10.20) successively with respect to each person's parameter β_n. It turns out that equation is exactly Eq. (10.5) we used above.

Thus,

$$r_n = \sum_{i=1}^{I} \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}} \tag{10.21}$$

gives the maximum likelihood estimate of the person parameter in the dichotomous RM.

As indicated in Chap. 9, maximum likelihood estimation is analogous but a different principle from that which is used in regression described in the statistics reviews. In regression, the criterion for the estimates is that the parameter estimates of the model are such that the residuals between the model and the data are minimized. In maximum likelihood, the parameter estimates of the model are such that the likelihood of the data is a maximum. Often, but not always, minimizing residuals and maximizing the likelihood give the same estimates.

Bias in the Estimate

The estimates of person parameters in the Rasch model are biased in the sense that with a fixed number of items, the person parameters at extremes are a little more extreme than they should be. The probabilities that are estimated are not biased, but the non-linear relationship between the person parameters and the probabilities creates a bias. This bias tends to 0 as the number of items is increased, and tends to 0

more quickly if the person and item distributions are well aligned. Various software packages for the person estimation have modifications to the maximum likelihood estimates which shrink the extreme values; RUMM2030 is one of these.

Exercises

1. In this chapter, an example of a person answering 10 items was used to illustrate the estimate of a person's proficiency. The item difficulties were $\delta_1 = -2.5$, $\delta_2 = -2.0$, $\delta_3 = -1.5$, $\delta_4 = -1.0$, $\delta_5 = -0.5$, $\delta_6 = 0.5$, $\delta_7 = 1.0$, $\delta_8 = 1.05$, $\delta_9 = 1.5$, $\delta_{10} = 2.0$.

   The person answered six of the items correctly, that is $r_n = 6 = 6$. An initial value of $\beta_n^{(0)} = 0.25$ was used as an estimate of the person's proficiency, and inserted into Eq. (10.5) to give

   $$\sum_{i=1}^{10} \frac{e^{\beta_n-\delta_i}}{1+e^{\beta_n-\delta_i}} = \frac{e^{\beta_n^{(0)}-\delta_1}}{1+e^{\beta_n^{(0)}-\delta_1}} + \frac{e^{\beta_n^{(0)}-\delta_2}}{1+e^{\beta_n^{(0)}-\delta_2}} + \cdots\cdots + \frac{e^{\beta_n^{(0)}-\delta_{10}}}{1+e^{\beta_n^{(0)}-\delta_{10}}}$$

   $$= \frac{e^{0.25+2.5}}{1+e^{0.25+2.5}} + \frac{e^{0.25+2.0}}{1+e^{0.25+2.0}} + \cdots\cdots + \frac{e^{0.25-2.0}}{1+e^{0.25-2.0}}$$

   $$= 0.94 + 0.90 + 0.85 + 0.78 + 0.68$$
   $$+ 0.44 + 0.32 + 0.31 + 0.22 + 0.15 = 5.59.$$

   Because 5.59 is less than 6, a proficiency value a little greater than 0.25 was tried, in particular, $\beta_n^{(1)} = 0.40$ to give

   $$\sum_{i=1}^{10} \frac{e^{\beta_n-\delta_i}}{1+e^{\beta_n-\delta_i}} = \frac{e^{\beta_n^{(1)}-\delta_1}}{1+e^{\beta_n^{(1)}-\delta_1}} + \frac{e^{\beta_n^{(1)}-\delta_2}}{1+e^{\beta_n^{(1)}-\delta_2}} + \cdots\cdots + \frac{e^{\beta_n^{(1)}-\delta_{10}}}{1+e^{\beta_n^{(1)}-\delta_{10}}}$$

   $$= \frac{e^{0.40+2.5}}{1+e^{0.40+2.5}} + \frac{e^{0.40+2.0}}{1+e^{0.40+2.0}} + \cdots\cdots + \frac{e^{0.40-2.0}}{1+e^{0.40-2.0}}$$

   $$= 0.95 + 0.92 + 0.87 + 0.80 + 0.71 + 0.48$$
   $$+ 0.35 + 0.34 + 0.25 + 0.17 = 5.84.$$

   This sum of 5.84 is again less than 6. Therefore, the proficiency estimate must be greater than 0.40.

   (a) Try the value $\beta_n^{(2)} = 0.45$ in Eq. (10.5) as above. Is the required proficiency estimate greater than 0.45 or less than 0.45?

   (b) Try the value $\beta_n^{(3)} = 0.55$ in Eq. (10.5) as above. Is the required proficiency estimate greater than 0.55 or less than 0.55?

   (c) Try the value $\beta_n^{(4)} = 0.50$ in Eq. (10.5) as above. Which of these values that have been tried is the best estimate?

2. Calculate the PSI index of reliability for data where $\hat{\sigma}_{\hat{\beta}}^2 = 1.51$ and $\hat{\sigma}_{\hat{\varepsilon}}^2 = 0.32$.

References

Andrich, D. (1982). An index of person separation in latent trait theory, the traditional KR.20 index, and the Guttman scale response pattern. *Education Research and Perspectives, 9,* 95–104.

Andrich, D., Sheridan, B. E., & Luo, G. (2018). *RUMM2030: Rasch unidimensional models for measurement. Interpreting RUMM2030 Part I dichotomous data*. Perth, Western Australia: RUMM Laboratory.

Fischer, G. H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika, 46*(1), 59–77.

Further Reading

Andrich, D. (1988). *Rasch models for measurement* (pp. 49–52). Newbury Park, CA: Sage.