



CHAPTER 5
Estimating an Examinee's
Ability

CHAPTER 5

Estimating an Examinee's Ability

Under item response theory, the primary purpose for administering a test to an examinee is to locate that person on the ability scale. If such an ability measure can be obtained for each person taking the test, two goals can be achieved. First, the examinee can be evaluated in terms of how much underlying ability he or she possesses. Second, comparisons among examinees can be made for purposes of assigning grades, awarding scholarships, etc. Thus, the focus of this chapter is upon the examinees and the procedures for estimating an ability score (parameter) for an examinee.

The test used to measure an unknown latent trait will consist of N items, each of which measures some facet of the trait. In the previous chapters dealing with item parameters and their estimation, it was assumed that the ability parameter of each examinee was known. Conversely, to estimate an examinee's unknown ability parameter, it will be assumed that the numerical values of the parameters of the test items are known. A direct consequence of this assumption is that the metric of the ability scale will be the same as the metric of the known item parameters. When the test is taken, an examinee will respond to each of the N items in the test, and the responses will be dichotomously scored. The result will be a score of either a 1 or a zero for each item in the test. It is common practice to refer to the item score of 1 or 0 as the examinee's item response. Thus, the list of 1's and 0's for the N items is called the examinee's item response vector. The task at hand is to use this item response vector and the known item parameters to estimate the examinee's unknown ability parameter.

Ability Estimation Procedures

Under item response theory, maximum likelihood procedures are used to estimate an examinee's ability. As was the case for item parameter estimation, this procedure is an iterative process. It begins with some *a priori* value for the ability of the examinee and the known values of the item parameters. These are used to compute the probability of correct response to each item for that examinee. Then an adjustment to the ability estimate is obtained that improves the agreement of the computed probabilities with the examinee's item response vector. The process is repeated until the adjustment becomes small

enough that the change in the estimated ability is negligible. The result is an estimate of the examinee's ability parameter. This process is then repeated separately for each examinee taking the test. In Chapter 7, a procedure will be presented through which the ability levels of all examinees are estimated simultaneously. However, this procedure is based upon an approach that treats each examinee separately. Hence, the basic issue is how the ability of a single examinee can be estimated.

The estimation equation used is shown below:

$$\hat{\theta}_{s+1} = \hat{\theta}_s + \frac{\sum_{i=1}^N a_i [u_i - P_i(\hat{\theta}_s)]}{\sum_{i=1}^N a_i^2 P_i(\hat{\theta}_s) Q_i(\hat{\theta}_s)} \quad [5-1]$$

where: $\hat{\theta}_s$ is the estimated ability of the examinee within iteration s

a_i is the discrimination parameter of item i , $i = 1, 2, \dots, N$

u_i is the response made by the examinee to item i :

$u_i = 1$ for a correct response

$u_i = 0$ for an incorrect response

$P_i(\hat{\theta}_s)$ is the probability of correct response to item i , under the given item characteristic curve model, at ability level $\hat{\theta}$ within iteration s .

$Q_i(\hat{\theta}_s) = 1 - P_i(\hat{\theta}_s)$ is the probability of incorrect response to item i , under the given item characteristic curve model, at ability level $\hat{\theta}$ within iteration s .

The equation has a rather simple explanation. Initially, the $\hat{\theta}_s$ on the right side of the equal sign is set to some arbitrary value, such as 1. The probability of correct response to each of the N items in the test is calculated at this ability level using the known item parameters in the given item characteristic curve model. Then the second term to the right of the equal sign is evaluated. This is the adjustment term, denoted by $\Delta\theta$. The value of $\hat{\theta}_{s+1}$ on the left side of the equal sign is obtained by adding $\Delta\theta$ to $\hat{\theta}_s$. This value, $\hat{\theta}_{s+1}$, becomes $\hat{\theta}_s$ in the next iteration. The numerator of the adjustment term contains the essence of

the procedure. Notice that $(u_i - P_i(\hat{\theta}_s))$ is the difference between the examinee's item response and the probability of correct response at an ability level of $\hat{\theta}_s$. Now, as the ability estimate gets closer to the examinee's ability, the sum of the differences between u_i and $P_i(\hat{\theta}_s)$ gets smaller. Thus, the goal is to find the ability estimate yielding values of $P_i(\hat{\theta}_s)$ for all items simultaneously that minimizes this sum. When this happens, the $\Delta\theta$ term becomes as small as possible and the value of $\hat{\theta}_{s+1}$ will not change from iteration to iteration. This final value of $\hat{\theta}_{s+1}$ is then used as the examinee's estimated ability. The ability estimate will be in the same metric as the numerical values of the item parameters. One nice feature of equation 5-1 is that it can be used with all three item characteristic curve models, although the three-parameter model requires a slight modification.

A three-item test will be used to illustrate the ability estimation process. Under a two-parameter model, the known item parameters are:

$$\begin{aligned} b = -1, & \quad a = 1.0 \\ b = 0, & \quad a = 1.2 \\ b = 1, & \quad a = .8 \end{aligned}$$

The examinee's item responses were:

item	response
1	1
2	0
3	1

The *a priori* estimate of the examinee's ability is set to $\hat{\theta}_s = 1.0$

First iteration:

<i>item</i>	<i>u</i>	<i>P</i>	<i>Q</i>	<i>a(u-P)</i>	<i>a*a(PQ)</i>
1	1	.88	.12	.119	.105
2	0	.77	.23	-.922	.255
3	1	.5	.5	.400	.160
			sum	-.403	.520

$$\Delta\hat{\theta}_s = -.403/.520 = -.773, \hat{\theta}_{s+1} = 1.0 - .773 = .227$$

Second iteration:

<i>item</i>	<i>u</i>	<i>P</i>	<i>Q</i>	<i>a(u-P)</i>	<i>a*a(PQ)</i>
1	1	.77	.23	.227	.175
2	0	.57	.43	-.681	.353
3	1	.35	.65	.520	.146
			sum	.066	.674

$$\Delta\hat{\theta}_s = .066/.674 = .097, \hat{\theta}_{s+1} = .227 + .097 = .324$$

Third iteration:

item	u	P	Q	$a(u-P)$	$a^2a(PQ)$
1	1	.79	.21	.2102	.1660
2	0	.60	.40	-.7152	.3467
3	1	.37	.63	.5056	.1488
			sum	.0006	.6615

$$\Delta \hat{\theta}_s = .0006 / .6615 = .0009, \hat{\theta}_{s+1} = .324 + .0009 = .3249$$

At this point, the process is terminated because the value of the adjustment (.0009) is very small. Thus, the examinee's estimated ability is .33. Unfortunately, there is no way to know the examinee's actual ability parameter. The best one can do is estimate it. However, this does not prevent us from conceptualizing such a parameter. Fortunately, one can obtain a standard error of the estimated ability that provides some indication of the precision of the estimate. The underlying principle is that an examinee, hypothetically, could take the same test a large number of times, assuming no recall of how the previous test items were answered. An ability estimate $\hat{\theta}$ would be obtained from each testing. The standard error is a measure of the variability of the values of $\hat{\theta}$ around the examinee's unknown parameter value θ . In the present case, an estimated standard error can be computed using the equation given below:

$$SE(\hat{\theta}) = \frac{1}{\sqrt{\sum_{i=1}^N a_i^2 P(\hat{\theta}) Q(\hat{\theta})}} \quad [5-2]$$

It is of interest to note that the term under the square root sign is exactly the denominator of equation 5-1. As a result, the estimated standard error can be obtained as a side product of estimating the examinee's ability. In the example given above, it was:

$$SE(\hat{\theta}) = 1 / \sqrt{.6615} = 1.23$$

Thus, the examinee's ability was not estimated very precisely because the standard error is very large. This is primarily due to the fact that only three items were used here and one would not expect a very good estimate. As will be shown in the next chapter, the standard error of an examinee's estimated ability plays an important role in item response theory.

There are two cases for which the maximum likelihood estimation procedure fails to yield an ability estimate. First, when an examinee answers none of the items correctly, the corresponding ability estimate is negative infinity. Second, when an examinee answers all the items in the test correctly, the corresponding ability estimate is positive infinity. In both of these cases it is impossible to obtain an ability estimate for the examinee (the computer literally cannot compute a number as big as infinity). Consequently, the computer programs used to estimate ability must protect themselves against these two conditions. When they detect either a test score of zero or a perfect test score, they will eliminate the examinee from further analysis and set the estimated ability to some symbol such as ******* to indicate what has happened.

Item Invariance of an Examinee's Ability Estimate

Another basic principle of item response theory is that the examinee's ability is invariant with respect to the items used to determine it. This principle rests upon two conditions: first, all the items measure the same underlying latent trait; second, the values of all the item parameters are in a common metric. To illustrate this principle, assume that an examinee has an ability score of zero, which places him at the middle of the ability scale. Now, if a set of ten items having an average difficulty of -2 were administered to this examinee, the item

responses could be used to estimate the examinee's ability, yielding $\hat{\theta}_1$ for this test. Then if a second set of ten items having an average difficulty of +1 were administered to this examinee, these item responses could be used to estimate the examinee's ability, yielding $\hat{\theta}_2$ for this second test. Under the item-invariance principle, $\theta_1 = \theta_2$; i.e., the two sets of items should yield the same ability estimate, within sampling variation, for the examinee. In addition, there is no requirement that the discrimination parameters be the same for the two sets of items. This principle is just a reflection of the fact that the item characteristic curve spans the whole ability scale. Just as any subrange of the ability scale can be used in the estimation of item parameters, the corresponding segments of several item characteristic curves can be used to estimate an examinee's ability. Items with a high average difficulty will have a point on their item characteristic curves that corresponds to the ability of interest. Similarly, items with a low average difficulty will have a point on their item characteristic curves that corresponds to the ability of interest. Consequently, either set of items can be used to estimate the ability of examinees at that point. In each set, a different part of the item characteristic curve is involved, but that is acceptable.

The practical implication of this principle is that a test located anywhere along the ability scale can be used to estimate an examinee's ability. For example, an examinee could take a test that is "easy" or a test that is "hard" and obtain, on the average, the same estimated ability. This is in sharp contrast to classical test theory, where such an examinee would get a high test score on the easy test, a low score on the hard test, and there would be no way of ascertaining the examinee's underlying ability. Under item response theory, the examinee's ability is fixed and invariant with respect to the items used to measure it. A word of warning is in order with respect to the meaning of the word "fixed." An examinee's ability is fixed only in the sense that it has a particular value in a given context. For example, if an examinee took the same test several times and it could be assumed he or she would not remember the items or the responses from testing to testing, the examinee's ability would be fixed. However, if the examinee received remedial instruction between testings or if there were carryover effects, the examinee's underlying ability level would be different for each testing. Thus, the examinee's underlying ability level is not immutable. There are a number of applications of item response theory that depend upon an examinee's ability level changing as a function of changes in the educational context.

The item invariance of an examinee's ability and the group invariance of an item's parameters are two facets of what is referred to, generically, as the invariance principle of item response theory. This principle is the basis for a number of practical applications of the theory.

Computer Session for Chapter 5

This session has two purposes that result in apparently similar outcomes that are actually different in their conceptual basis. The first purpose is to illustrate how an examinee's estimated ability varies when the same test is taken a number of times. A test consisting of a few items will be established, the value of the examinee's ability parameter will be set, and the computer will generate the examinee's item responses. These will be used in equation 5-1 to estimate the examinee's ability. The computer will then generate a new set of item responses to these same items, and another ability estimate will be obtained. After several estimates are obtained, the mean and standard deviation of the estimates will be computed and compared to their theoretical values. The intent is to allow you to develop a sense of how ability estimates for a single examinee are distributed under repeated use of the same test.

The second purpose is to illustrate the item invariance of an examinee's ability. A small test will be established through the values of its item parameters, the examinee's ability will be set, and the computer will generate the examinee's item responses. These will be used in equation 5-1 to obtain an ability estimate for the examinee. Then a new test will be established, item responses for the same examinee generated, and another ability estimate obtained. This process will be repeated for several different tests, resulting in a set of ability estimates. If the invariance principle holds, all the estimates should be clustered around the value of the examinee's ability parameter.

Procedure for Investigating the Sampling Variability of Estimated Ability

This exercise will illustrate the sampling variability of a given examinee's estimated ability when the same test is administered several times.

- n. Follow the start-up procedures described in the Introduction.
- o. Use the mouse to highlight the ABILITY ESTIMATION session and click on [SELECT].
- p. Read the explanatory screen and click on [CONTINUE]. The SET EXERCISE SPECIFICATIONS screen will appear.
- q. After NEW EXAMINEE? click on YES.
- r. Click on [NUMBER OF ABILITY ESTIMATES TO MAKE] and set the number of estimates (K) to 6.
- s. After ADMINISTER THE SAME TEST SEVERAL TIMES? click on YES.
- t. Respond to the question SPECIFICATIONS OK? by clicking on the YES button. The TEST SPECIFICATION screen will appear.
- u. Click on [NUMBER OF ITEMS] and set the number of items to 5.
- v. Under SELECT ITEM PARAMETER MODEL, click on TWO PARAMETER.
- w. Under SELECT ITEM PARAMETER CREATION METHOD, click on USER INPUT OF ITEM PARAMETER VALUES.
- x. Respond to the question SETTINGS OK? by clicking on the YES button. The ESTABLISH ITEM PARAMETER VALUES screen will appear.
- y. Click on [ENTER ITEM PARAMETERS] and then set the following values:

$b = -.5,$	$a = 1.0$
$b = -.25,$	$a = 1.5$
$b = 0,$	$a = .7$
$b = .25,$	$a = .6$
$b = .50,$	$a = 1.8$

- z. Study the table of item parameters for a moment. If you need to change a value, click on the value and the data input box will appear, allowing you to enter a new value.
- aa. When you are satisfied with the parameter values, respond to the question PARAMETER VALUES OK by clicking on the YES button.
- bb. The computer will generate and display the examinee's item responses (1 = correct, 0 = incorrect) in the table of item parameters.
- cc. Click on [CONTINUE]. The ABILITY ESTIMATION RESULTS screen will appear.
- dd. The computer will display the following information: the item response vector, the raw score, the estimated ability $\hat{\theta}$ and its standard error, S.E. ($\hat{\theta}$), and the sequence number of the ability estimate.
- ee. Study the results, then click on [CONTINUE]. The ABILITY ESTIMATION RESULTS screen for the next estimate will appear. Repeat K times.
- ff. After the K 'th ability estimate, a summary table similar to the following will appear on the screen:

TRUE ABILITY OF EXAMINEE = 1.25

ABILITY ESTIMATES

1.33 .28 .28 .69 1.25 1.86

MEAN OF ABILITY ESTIMATES = .95

STANDARD ERROR

OBSERVED = .64 THEORETICAL = .98

The mean of the ability estimates (.95 in the table above) should be reasonably close to the ability parameter value (1.25 in the table above) set by the computer for the examinee. The observed standard error of the ability estimates should approximate the theoretical value. However, with such a small number of items and replications, the results will probably deviate somewhat from their theoretical values.

- a. Respond to DO ANOTHER EXERCISE? by clicking on the YES button. The EXERCISE SPECIFICATION screen will appear.

Exercises

Exercise 1

- a. Respond to NEW EXAMINEE? by clicking on YES.
- b. Click on [NUMBER OF ABILITY ESTIMATES TO MAKE] and set the number of estimates (K) to 10.
- c. Respond to ADMINISTER THE SAME TEST SEVERAL TIMES? by selecting YES.
- d. Respond to SPECIFICATIONS OK? by selecting YES. The TEST SPECIFICATION screen will appear.
- e. Click on [NUMBER OF ITEMS] and set the number of items to 5.
- f. Under SELECT ITEM PARAMETER MODEL, click on a model of your choice.
- g. Under SELECT ITEM PARAMETER CREATION METHOD, click on COMPUTER GENERATED ITEM PARAMETER VALUES.

- h. Respond to the question SETTINGS OK? by clicking on the YES button. The ESTABLISH ITEM PARAMETER VALUES screen will appear.
- i. Click on [GENERATE ITEM PARAMETERS]. A table of item parameter values will appear. Respond to PARAMETER VALUES OK? by clicking on YES. Then click on [CONTINUE].
- j. When the summary table is shown on the screen, write the value shown for the TRUE ABILITY OF EXAMINEE on a piece of paper so it can be used in the next exercise.
- k. Respond to DO ANOTHER EXERCISE? by selecting YES.

Exercise 2

The intent of this exercise is to see if you can improve upon the previous estimates of the examinee's ability parameter by proper selection of the test's item parameters.

- a. After NEW EXAMINEE? click on NO.
- b. Click on [NUMBER OF ABILITY ESTIMATES TO MAKE] and set the number of estimates (K) to 10.
- c. After ADMINISTER THE SAME TEST SEVERAL TIMES? click on YES.
- d. Respond to SPECIFICATIONS OK? by clicking on the YES button. The TEST SPECIFICATION screen will appear.
- e. Click on [NUMBER OF ITEMS] and set the number of items to 5.
- f. Under SELECT ITEM PARAMETER MODEL, click on the same model as used in Exercise 1.
- g. Under SELECT ITEM PARAMETER CREATION METHOD, click on USER INPUT OF ITEM PARAMETER VALUES.

- h. Since you know the value of the examinee's ability parameter, choose values of the item difficulty parameters that are close to this value and use large values of the discrimination parameters.
- i. Follow Steps n through s of the example case.
- j. The summary table will be shown on the screen. If you chose the parameter values wisely, the mean of the ability estimates should have been close to the value of the examinee's ability parameter. The observed standard error should have also approximated the theoretical value. If such was not the case, think about some reasons for the lack of a match. You need to keep in mind that the obtained results are subject to considerable sampling variability due to the small numbers of items being used (increasing N to 10 will help) and the small number of replications used.

Exercise 3

Experiment with different types of models and item parameter values to see if you can determine what influences the distribution of the estimated abilities.

Procedures for Investigating the Item Invariance of an Examinee's Ability

In this example, a given examinee will be administered a number of different tests. The intent is to illustrate that the estimated abilities should cluster about the value of the examinee's ability parameter.

- a. Follow the start-up procedures described in the Introduction.
- b. Use the mouse to highlight on the ABILITY ESTIMATION session and click on [SELECT].
- c. Read the explanatory screen and click on [CONTINUE]. The SET EXERCISE SPECIFICATIONS screen will appear.
- d. After NEW EXAMINEE?, click on YES.
- e. Click on [NUMBER OF ABILITY ESTIMATES TO MAKE] and set the number of estimates (K) to 10.
- f. After ADMINISTER SAME TEST SEVERAL TIMES? click on NO.
- g. Respond to question SPECIFICATIONS OK? by clicking on the YES button. The TEST SPECIFICATION screen will appear.
- h. Click on [NUMBER OF ITEMS] and set the number of items to 5.
- i. Under SELECT ITEM PARAMETER MODEL, click on RASCH.
- j. Under SELECT ITEM PARAMETER CREATION METHOD, click on COMPUTER GENERATION OF ITEM PARAMETER VALUES.

- k. Respond to the question SETTINGS OK? by clicking on the YES button. The ESTABLISH ITEM PARAMETER VALUES screen will appear.
- l. Click on [ENTER ITEM PARAMETERS]. The computer will generate the random item difficulty values for the 5 items.
- m. Study the table of item parameters for a moment and take note of their average value and the distribution of the values.
- n. When you are satisfied with the parameter values, respond to the question PARAMETERS OK? by clicking on the YES button. Then click on [CONTINUE].
- o. The computer will generate and display the examinee's item responses in the table of item parameters, using 1 for a correct response and 0 for an incorrect one.
- p. Click on [CONTINUE]. The ABILITY ESTIMATION RESULTS screen will appear.
- q. Study the results, then click on [CONTINUE]. Repeat Steps l through q K times.
- r. The computer will display the following information: the item response vector, the raw score, the estimated ability θ and its standard error, S.E. (θ) , and the sequence number of the ability estimate.
- s. After the K 'th ability estimate, a summary table similar to the following will appear on the screen:

ABILITY OF EXAMINEE = .50

ABILITY ESTIMATES

.11 .22 .33 .44 .55 .66 .77 .88 .99 .00

MEAN OF ABILITY ESTIMATES = .50

STANDARD ERROR

OBSERVED = 1.23 THEORETICAL = 1.35

The average value of the estimates should be close to the value of the examinee's ability parameter. There should not be a large amount of scatter in the estimates. Again, due to the small number of items and estimates used, the item invariance of the ability estimate may not be readily apparent.

- t. Respond to DO ANOTHER EXERCISE? by selecting YES. The EXERCISE SPECIFICATION screen will appear.

Exercises

Exercise 1

The intent of this exercise is to enable you to experiment with the item sets used to illustrate the item invariance of the ability estimates. Rather than letting the computer set the values of the item parameters, you can choose your own values.

- a. After NEW EXAMINEE?, click on YES.
- b. Click on [NUMBER OF ABILITY ESTIMATES TO MAKE] and set the number of estimates (K) to 4.
- c. After ADMINISTER SAME TEST SEVERAL TIMES?, click on NO.
- d. Respond to SPECIFICATIONS OK? by clicking on YES. The TEST SPECIFICATION screen will appear.
- e. Click on [NUMBER OF ITEMS] and set the number of items to 5.
- f. Under SELECT ITEM PARAMETER MODEL, click on the model of your choice.

- g. Under SELECT ITEM PARAMETER CREATION METHOD, click on USER INPUT OF ITEM PARAMETER VALUES.
- h. Respond to the question SETTINGS OK? by clicking on the YES button. The ESTABLISH ITEM PARAMETER VALUES screen will appear.
- i. Click on [ENTER ITEM PARAMETERS] and enter item parameter values for the 5 items.
- j. When you are satisfied with the parameter values, respond to the question PARAMETER VALUES OK? by clicking on the YES button.
- k. Click on [CONTINUE]. The ABILITY ESTIMATION RESULTS screen will appear.
- l. After each estimate has been shown, enter a new set of values for the item parameters. When doing so, experiment with their average values and range of values.
- m. After the K th estimate has been shown, the summary table will appear and you will be able to see how well you have done.

Exercise 2

To make things easy for you, let the computer generate the sets of item parameter values. Repeat the procedures for exercise 1, but at Step h, respond to DO YOU WANT TO SET THE VALUES OF THE ITEM PARAMETERS? by selecting NO. Now the computer will do the tedious job of setting the parameters.

Things To Notice

1. Distribution of estimated ability.
 - a. The average value of the estimates is reasonably close to the value of the ability parameter for the examinee set by the computer program.
 - b. When the item difficulties are at or near the examinee's ability parameter value, the mean of the estimated abilities will be close to that ability value.
 - c. The standard error of the estimates can be quite large when the items are not located near the ability of the examinee. However, the theoretical values of the standard errors are also quite large, and the obtained standard errors approximate these values.
 - d. When the values of the item discrimination indices are large, the standard error of the ability estimates is small. When the item discrimination indices are small, the standard error of the ability estimates is large.
 - e. The optimum set of items for estimating an examinee's ability would have all its item difficulties equal to the examinee's ability parameter and have items with large values for the item discrimination indices.
2. Item invariance of the examinee's ability.
 - a. The different sets of items yielded values of estimated ability that were near the examinee's actual ability level.
 - b. The mean value of these estimates generally was a close approximation of the examinee's ability parameter. If one used many tests, each having a large number of items, the mean estimated ability would equal the examinee's ability parameter. In addition, these estimates would be very tightly clustered around the parameter value. In such a situation, it would be very clear that the item invariance principle holds.

3. Overall observation.

This session has dealt with two facets of estimating an examinee's ability that are conceptually distinct but look similar in certain respects. The first set of examples focused upon the variability of the ability estimates about the value of the examinee's ability parameter. This will serve as the basis for the next chapter, which deals with how well a test estimates ability over the whole ability scale. The second set of exercises focused upon the item invariance of an examinee's estimated ability. This will serve as part of the basis for Chapter 7 dealing with test calibration.

The reader should keep in mind that an ability estimate is just another type of test score, but it is interpreted within the context of item response theory. Consequently, such ability estimates can be used to compute summary statistics for groups of examinees and other indices of interest.

4. A final comment.

In Chapter 1, the concept of a latent trait was introduced. An integral part of item response theory is that an examinee can be positioned on the scale representing this latent trait. Thus, in theory, each examinee has an ability score (parameter value) that locates that person on the scale. However, in the real world we cannot obtain the value of the examinee's ability parameter. The best we can do is obtain an estimate of it. In the computer session for this chapter it was assumed that we could generate the value of an examinee's ability parameter. This assumption enabled the program to generate the item response vectors used to obtain the ability estimates and hence illustrate the theory.