Chapter 1
The Item Characteristic Curve

1.1 Introduction

In many educational and psychological measurement situations there is an underlying variable of interest. This variable is often something that is intuitively understood, such as "intelligence." People can be described as being bright or average and the listener has some idea as to what the speaker is conveying about the object of the discussion. Similarly, one can talk about scholastic ability and its attributes such as gets good grades, learns new material easily, relates various sources of information, and uses study time effectively. In academic areas, one can use descriptive terms such as reading ability and arithmetic ability. Each of these is what psychometricians refer to as an unobservable or latent trait. While such a variable is easily described and knowledgeable persons can list its attributes, it cannot be measured directly as can height or weight, since the variable is a concept rather than a physical dimension. A primary goal of educational and psychological measurement is the determination of how much of such a latent trait a person possesses. Since most of the research has dealt with variables such as scholastic, reading, mathematical, and arithmetic abilities, the generic term "ability" is used within item response theory to refer to such latent traits.

If one is going to measure how much of a latent trait a person has, it is necessary to have a scale of measurement, that is, a ruler having a given metric. For a number of technical reasons, defining the scale of measurement, the numbers on the scale, and the amount of the trait that the numbers represent is a very difficult task. For the purposes of the first six chapters, this problem shall be solved by simply defining an arbitrary underlying ability scale. It will be assumed that, whatever the ability, it can be measured on a scale having a midpoint of zero, a unit of measurement of one, and a range from negative infinity to positive infinity. Since there is a unit of measurement and an arbitrary zero point, such a scale is referred to as existing at an interval level of measurement. The underlying idea here is that, if one could physically ascertain the ability of a person, this ruler would be used to tell how much

ability a given person has, and the ability of several persons could be compared. While the theoretical range of ability is from negative infinity to positive infinity, practical considerations usually limit the range of values from, say, -3 to $+3$. Consequently, the discussions in the text and the computer sessions will only deal with ability values within this range. However, you should be aware that values beyond this range are possible.

The usual approach taken to measure an ability is to develop a test consisting of a number of items (i.e., questions). Each of these items measures some facet of the particular ability of interest. From a purely technical point of view such items should be free response items where the examinee can write any response that seems appropriate. The person scoring the test then must decide whether the response is correct or not. When the item response is determined to be correct, the examinee receives a score of one, an incorrect answer receives a score of zero, that is, the item is dichotomously scored. Under classical test theory, the examinee's raw test score would be the sum of the scores received on the items in the test. Under item response theory, the primary interest is in whether an examinee got each individual item correct or not rather than in the raw test score. This is because the basic concepts of item response theory rest upon the individual items of a test rather than upon some aggregate of the item responses such as a test score.

From a practical point of view, free response items are difficult to use in a test. In particular, they are difficult to score in a reliable manner. As a result, most tests used under item response theory consist of multiple-choice items. These are scored dichotomously with the correct answer receiving a score of one and each of the distractors yielding a score of zero. Items scored dichotomously are often referred to as binary items.

1.2   The Item Characteristic Curve

A reasonable assumption is that each examinee responding to a test item possesses some amount of the underlying ability. Thus, one can consider each examinee to have a numerical value, a score, that places the examinee somewhere on the ability scale. This ability score will be denoted by the Greek letter theta, θ. At each ability level there will be a certain probability that an examinee with that ability will give a correct answer to the item. This probability will be denoted by $P(\theta)$. In the case of a typical test item, this probability will be small for examinees of low ability and large for examinees of high ability. If one plotted $P(\theta)$ as a function of ability, the result would be a smooth S-shaped curve such as shown in Fig. 1.1. The probability of correct response is near zero at the lowest levels of ability and increases until at the highest levels of ability the probability of correct response approaches unity. This S-shaped curve describes the relationship between the probability of correct response to an item and the ability scale. In item response theory, it is known as the item characteristic curve. Each item in a test will have its own item characteristic curve.
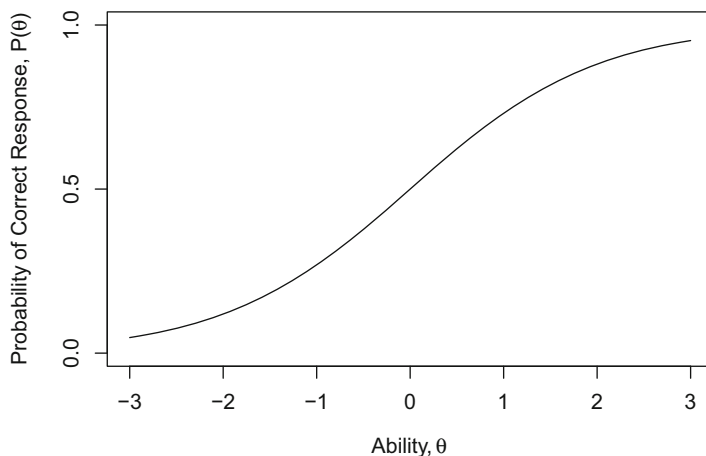
Fig. 1.1 A typical item characteristic curve

The item characteristic curve is the basic building block of item response theory and all the other constructs of the theory depend upon this curve. Because of this, considerable attention will be devoted to this curve and its role within the theory. There are two technical properties of an item characteristic curve that are used to describe it. The first is the difficulty of the item. Under item response theory, the difficulty of an item describes where the item functions along the ability scale. For example, an easy item functions among the low-ability examinees while a hard item would function among the high-ability examinees; thus, item difficulty is a location index. The second technical property is the discrimination of an item, which describes how well an item can differentiate between examinees having abilities below the item location and those having abilities above the item location. This property essentially reflects the steepness of the item characteristic curve in its middle section. The steeper the curve the better the item can discriminate. The flatter the curve the less the item is able to discriminate since the probability of correct response at low ability levels is nearly the same as it is at high ability levels. Using these two descriptors, one can describe the general form of the item characteristic curve. These descriptors are also used to discuss the technical properties of an item. It should be noted that these two properties say nothing about whether the item really measures some facet of the underlying ability or not; that is a question of validity. These two properties simply describe the form of the item characteristic curve.

1.3 Item Difficulty and Item Discrimination

The idea of item difficulty as a location index will be examined first. In Fig. 1.2, three item characteristic curves are presented on the same graph. All have the same level of item discrimination but differ with respect to item difficulty. The left-hand
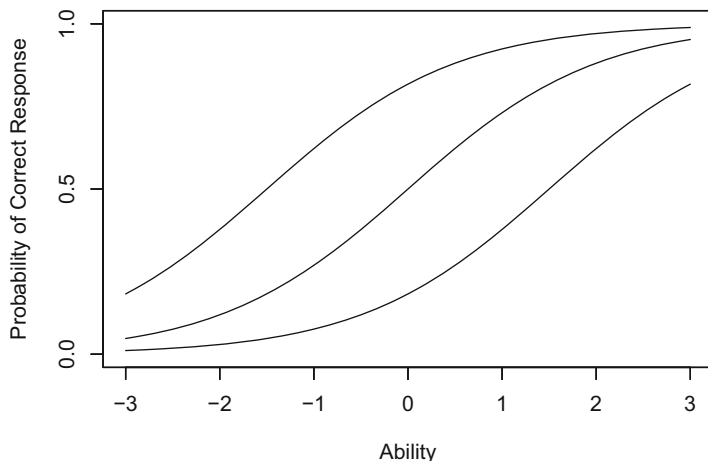
Fig. 1.2 Three item characteristic curves with the same item discrimination but different levels of item difficulty

curve represents an easy item because the probability of correct response is high for low-ability examinees and approaches 1 for high-ability examinees. The center curve represents an item of medium item difficulty because the probability of correct response is low at the lower ability levels, around 0.5 in the middle of the ability scale, and near 1 at the highest ability level. The right-hand curve represents a hard item. The probability of correct response is low for most of the ability scale and increases only when the higher ability levels are reached. Even at the highest ability level shown (i.e., $\theta = 3$) the probability of correct response is only 0.8 for the most difficult item.

The concept of item discrimination is illustrated in Fig. 1.3. This figure contains three item characteristic curves having the same item difficulty but differing with respect to item discrimination. The upper curve on the positive ability range has a high level of item discrimination since the curve is quite steep in the middle where the probability of correct response changes very rapidly as ability increases. Just a short distance to the left of the middle of the curve, the probability of correct response is much less than 0.5; and a short distance to the right, the probability is much greater than 0.5. The middle curve represents an item with a moderate level of item discrimination. The slope of this curve is much less than the previous curve and the probability of correct response changes less dramatically than the previous curve as the ability level increases. However, the probability of correct response is near zero for the lowest-ability examinees and near unity for the highest-ability examinees. The third curve represents an item with low item discrimination. The curve has a very small slope and the probability of correct response changes slowly over the full range of abilities shown. Even at low ability levels, the probability of correct response is reasonably large and it increases only slightly when high ability levels are reached. The reader should be warned that although the figures only show a range of ability from -3 to $+3$, the theoretical range of ability is from negative
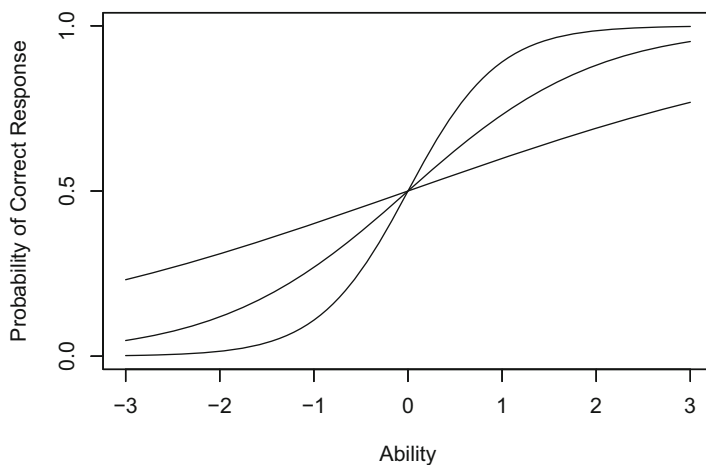
Fig. 1.3 Three item characteristic curves with the same item difficulty but different levels of item discrimination
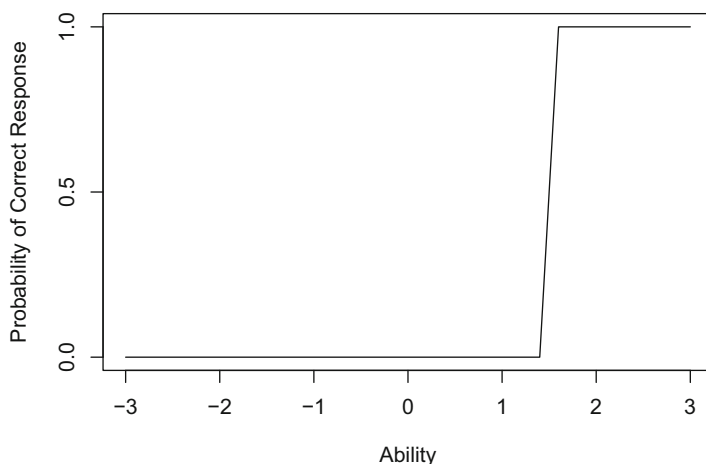

Fig. 1.4 An item that discriminates perfectly at $\theta = 1.5$

infinity to positive infinity. Thus, all item characteristic curves of the type used here actually become asymptotic to a probability of zero at one tail and to unity at the other tail. The restricted range employed in the figures is necessary to fit the curves on the computer screen reasonably and to provide a uniform frame of reference.

One special case is of interest; namely, that of an item with perfect discrimination. The item characteristic curve of such an item is a vertical line at some point along the ability scale. Figure 1.4 shows such an item. To the left of the vertical line at $\theta = 1.5$, the probability of correct response is zero and to the right of the line the probability of correct response is unity. Thus, the item discriminates perfectly

between examinees whose abilities are above and below an ability score of 1.5. Such items would be ideal for distinguishing between examinees with abilities just above and below 1.5. However, such an item makes no distinction among those examinees with abilities above 1.5 nor among those examinees with abilities below 1.5.

1.4   Verbal Terms of Item Difficulty and Item Discrimination

At the present point in the presentation of item response theory, the goal is to allow you to develop an intuitive understanding of the item characteristic curve and its properties. In keeping with this goal, item difficulty and item discrimination will be defined in verbal terms. Item difficulty will have the following levels:

very easy
easy
medium
hard
very hard

Item discrimination will have the following levels:

none
low
moderate
high
perfect

These terms will be used in the computer session to specify item characteristic curves.

1.5   Computer Session

The purpose of this session is to enable you to develop a sense of how the shape of the item characteristic curve is related to item difficulty and item discrimination. To accomplish this, you will be able to select verbal terms describing the item difficulty and item discrimination of an item. The computer program R will then calculate and display the corresponding item characteristic curve on the screen. You should do examples in this section and exercises in the next section, then try various combinations of levels of item difficulty and item discrimination and relate these to the resulting curves. After a bit of such exploratory practice, you should be able to predict what the item characteristic curve will look like for a given combination of item difficulty and item discrimination.

1.5.1 Procedures for an Example Case

When R is ready for input through the R console window, it prints out its prompt character with an invisible, horizontal space after it:

```
>
```

A command line in R will be executed by pressing the enter key:

Enter
⟵⟶

Such a special character that indicates the end of a command line, usually entered by pressing the enter or return key will be treated as an invisible character here. When an incomplete command line (e.g., the end of the expression cannot have occurred yet) gets the enter key, R prints out the continuation prompt character with an invisible space after it:

```
+
```

To improve readability and because a rather long command line can be typed in without pressing the enter key in the middle, a long command line will not be separated by the continuation prompt character but will be continued to the next line with indentation.

The followings are the simplest command lines, each with the R prompt in front, to display an item characteristic curve for an item with medium item difficulty and moderate item discrimination:

```
> theta <- seq(-3, 3, .1)
> bmedium <- 0
> amoderate <- 1
> P <- 1 / (1 + exp(-amoderate * (theta - bmedium)))
> plot(theta, P, type="l")
```

By pressing the enter key in the end of each line, five times as a total, the computer will display an item characteristic curve shown in Fig. 1.5 in the graphics window.

By pressing the enter key in the end of the first line, a sequence of numbers (i.e., a vector) will be created with -3 as a starting number and 3 as an ending number with an increment of 0.1. The length or the total number of elements of the sequence is 61. The name of such a sequence is given as theta, and it is to be done by using the assign function <- for which the less than character < and the hyphen character - cannot be separated with a space. You can read the first line as "theta gets a sequence of numbers . . ." or "a sequence of numbers . . . is saved under the name theta." The first line is equivalent to:

```
> assign("theta", seq(-3,3,.1))
```

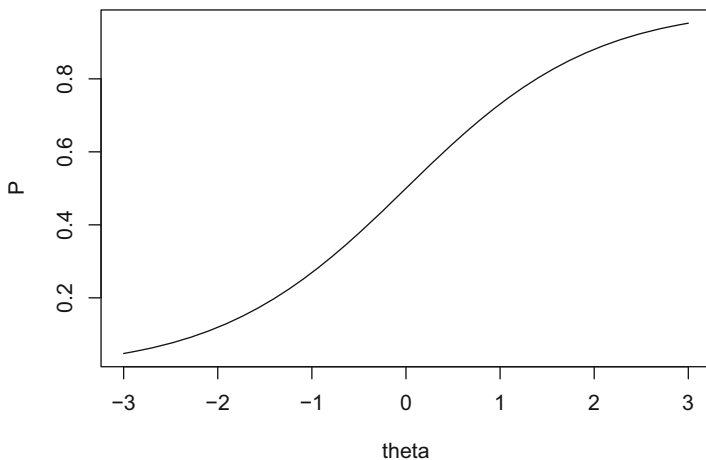or

```
> seq(-3, 3, .1) -> theta
```

Fig. 1.5 An example item characteristic curve

Use of spaces is optional in the R command line as shown in `seq(-3,3,.1)`. Especially, spaces before and after a comma, a left or right parenthesis, an elementary arithmetic operator (e.g., `+`, `-`, `*`, `/`), an exponentiation operator (i.e., `∧`, `**`), a common mathematical function (e.g., `log`, `exp`, `sin`, `cos`, `tan`, `sqrt`), a logical operator (e.g., `<`, `<=`, `>`, `>=`, `==`, `!=`, `&`, `|`), and an assign function (e.g., `->`, `<-`) are all optional. To improve readability of the code, at least one space can be put before and after some operators or symbols.

Multiple command lines can be typed in a single line by separating them with a semicolon. Real numbers in the integer form can be entered with a decimal place (i.e., a period) in the end of the number, and the number 0 can be added after the decimal place. The number 0 can be added before the decimal place if the number is less than unity that is to be expressed as a decimal. The second and third lines of the above command lines can be combined as:

```
> bmedium <- 0; amoderate <- 1
```

Equivelently, we may use the numbers with a decimal place as:

```
> bmedium <- 0.0; amoderate <- 1.0
```

The item characteristic curve is a function of ability and item characteristics, that is, item difficulty and item discrimination. Because ability is a latent variable, it was denoted by θ and the variable name `theta` was used in the command line. The item difficulty and item discrimination are denoted as numerical values of b and a, respectively, and the verbal terms contain the letters `b` and `a` as mnemonic clues. The names of objects in the R command can be constructed with the upper- and lower-case roman letters as an initial character and the digits and the period as any non-initial characters. Names of built-in, intrinsic functions and system variables (e.g., `seq`, `exp`, `plot`, `type` in the above lines, and `c`, `q`, `s`, `t`, `C`, `D`, `F`, `I`, `T`,

diff, length, mean, sd, pi, range, rank, time, tree, var, etc.) should not be used as the names of variables or functions of your own. Note that R is case sensitive, so the variable P will be different from the variable p unless both are defined to be equivalent.

The command line

```
P <- 1 / (1 + exp(-amoderate * (theta - bmedium)))))
```

will create a vector of 61 values of the probabilities of correct response based on the respective 61 ability points, item difficulty, and item discrimination. We will explore the exact meaning of this function in the subsequent chapter.

The plot based on the 61 sets of points from the ability variable as the horizontal axis, abscissa, and the probability of correct response as a vertical axis, ordinate, can be constructed via:

```
> plot(theta, P)
```

With the enter key pressed, the above line can open an R graphics window that contains the plot of the two variables. The default setting of the function plot will create a plot with a symbol ○ as each point. Because we want to have a plot with connected lines that ultimately yield a curve, the optional argument type="l" was added in the earlier command line. Whenever R opens up its graphics window, it treats the graphics window as a current window. If you want to continue to use command lines, you may click the R console window (especially the caption on the top of the R console window or any inside portion of the R console window) to make it current before you type in a new command line.

You may notice that the number of ticks used in the default setting of the function plot may not be an appropriate one you want to use in your own figure. The numbers of ticks in the horizontal and vertical axes can be modified with the use of the graphical parameters function par and its labels argument lab, for example:

```
> par(lab=c(7,3,3))
```

The labels argument was defined by the three parameters in the above line. The set of the three parameters (i.e., the three numbers separated with two commas in the combine function c) specifies the number of ticks on the horizontal axes to be 7, that on the vertical axes to be 3, and the length of characters in the labels to be 3 (but the character length will be most likely ignored in R). If you want to explore or read the full description of the R function, for example par, you can obtain it by typing:

```
> ?par
```

or

```
> help(par)
```

Assuming that your computer is connected to Internet, such a command line will open up a file in html, the HyperText Markup Language, that explains the function in a default browser you are using.

It can be noticed in Fig. 1.5 that the variable names are appeared as the respective labeling texts along the horizontal and vertical axes. You can change them with the arguments of, for example, `xlab="Ability"` and `ylab="Probability of Correct Response"` in the function `plot`. The ranges of the horizontal and vertical variables can be precisely controlled with the use of the arguments and parameters of `xlim=c(-3,3)` and `ylim=c(0,1)`, by specifying the lower limit number and the upper limit number in the combine or concatenate function `c`.

The following command lines can be used to obtain Fig. 1.1:

```
> par(lab=c(7,3,3))
> theta <- seq(-3, 3, .1)
> b <- 0
> a <- 1
> P <- 1 / (1 + exp(-a * (theta - b)))
> plot(theta, P, type="l", xlim=c(-3,3), ylim=c(0,1),
    xlab=expression(paste("Ability, ",theta)),
    ylab=expression(paste(
    "Probability of Correct Response, P(",theta,")")))
```

As shown in Fig. 1.1, you may use the function `expression` to add characters in Greek. We may simply use `xlab="Ability"` and `ylab="Probability of Correct Response"`. You can add the main heading on the top of the plot and the subheading to the bottom of the plot with the arguments in the function `plot`. These will be helpful when you are making figures for a presentation purpose. For example, you may construct a figure (n.b., a figure is not displayed here) that is more elaborate but nevertheless similar to Fig. 1.5 using the following command lines:

```
> par(lab=c(7,3,3))
> theta <- seq(-3, 3, .1)
> bmedium <- 0
> amoderate <- 1
> P <- 1 / (1 + exp(-amoderate * (theta - bmedium)))
> plot(theta, P, type="l", xlim=c(-3,3), ylim=c(0,1),
    xlab="Ability", ylab="Probability of Correct Response",
    main="Figure 1. An Item Characteristic Curve with
    Medium Item Difficulty and Moderate Item Discrimination",
    sub="See Baker and Kim (2016).")
```

1.5.2 An R Function for Item Characteristic Curves

It is possible to create your own R functions. Each time the plot of the item characteristic curve is created, you may notice that a nearly identical set of R command lines are executed. To avoid the repetition of typing in the same command

lines, an R function for plotting an item characteristic curve can be constructed. Consider the following function named `iccplot`:

```
> iccplot <- function(b, a) {
    par(lab=c(7,3,3))
    theta <- seq(-3, 3, .1)
    P <- 1 / (1 + exp(-a * (theta - b)))
    plot(theta, P, type="l", xlim=c(-3,3), ylim=c(0,1),
      xlab="Ability", ylab="Probability of Correct Response")
  }
```

After defining the function, by specifying the numerical values of item difficulty and item discrimination, a plot of the item characteristic curve can be constructed in the R graphics window. For example, you may use the following line to obtain a plot of item characteristic curve with medium item difficulty and moderate item discrimination:

```
> iccplot(0, 1)
```

The two arguments, `b` and `a`, are the named actual arguments in R. When the numerical values are specified without the arguments, R recognizes the first number to be the value of item difficulty and the second number to be the value of item discrimination. The arguments in the function can be specified in arbitrary order by exactly defining them with names. You can obtain the same plot by typing:

```
> iccplot(a=1, b=0)
```

Instead of using the numerical values to define item difficulty and item discrimination, the verbal terms described earlier can be used to plot item characteristic curves. The numerical definitions of the verbal terms for item difficulty are as follows:

```
> bveryeasy <- -2.625
> beasy <- -1.5
> bmedium <- 0
> bhard <- 1.5
> bveryhard <- 2.625
```

The numerical definitions of the verbal terms for item discrimination are as follows:

```
> anone <- 0
> alow <- 0.4
> amoderate <- 1
> ahigh <- 2.1
> aperfect <- 999
```

The following command lines now can display two item characteristic curves in the R graphics window (see Fig. 1.6). Note that you should click the R console window after pressing the enter key in the end of the first line, that is, after creating the R graphics window.
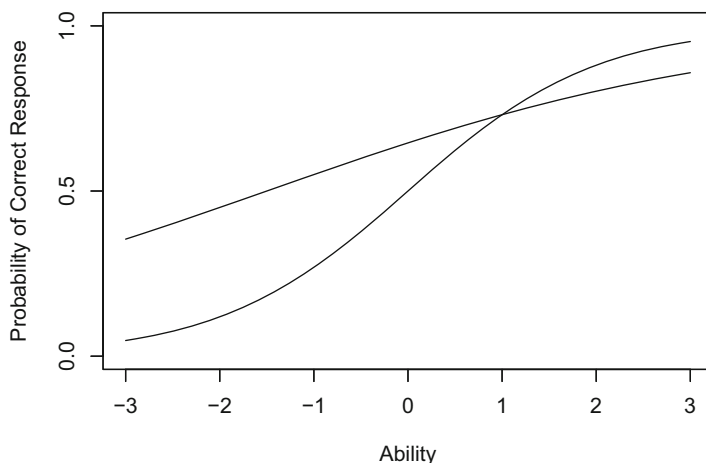
Fig. 1.6 Two item characteristic curves

```
> iccplot(bmedium, amoderate)
> par(new=T)
> iccplot(beasy, alow)
```

The first item characteristic curve is for an item with medium item difficulty and moderate item discrimination. The line

```
> par(new=T)
```

is an overlying figures parameter and equivalent to:

```
> par(new=TRUE)
```

The line resets a graphics parameter new to be TRUE, meaning that the R graphics window is treated now as a new graphics device, so it is assumed that there are currently no plots on it. In such a case, a call to a high-level plotting function could not erase the canvas of the graphics window before putting up a plot. After entering the third line, you may notice that the second item characteristic curve was overlaid on the same graph as the previous curve and you can compare the two. The new curve is rather flat and has higher probabilities of correct response in the lower range of abilities than did the previous item characteristic curve. This is because it was an easier item and low-ability examinees should do well on it. The low item discrimination shows up in the curve having only a slight bend over the range of ability scores employed. At the high ability levels the probability of correct response was somewhat lower than that of the previous item. This is reflection of the lower discrimination on the new item. When the function is specified without par(new=T), a new plot is created in the R graphics window.

Note that you can obtain a figure that contains several item characteristic curves. For example, to obtain Fig. 1.2 you can type the following command lines with

clicking the console window after the first line and the third line, respectively:

```
> iccplot(beasy, amoderate)
> par(new=T)
> iccplot(bmedium, amoderate)
> par(new=T)
> iccplot(bhard, amoderate)
```

1.6   Exercises

For the exercises, it is assumed that you have defined the function `iccplot` and the numerical values of the verbal terms of item difficulty and item discrimination by typing in the respective command lines.

1. An item with easy item difficulty and high item discrimination is to be plotted.

   (a) Use the function `iccplot` to plot an item characteristic curve of an item with easy item difficulty and high item discrimination.
   (b) From the graph it can be seen that the probability of correct response will be rather high over most of the ability scale. The item characteristic curve will be steep in the lower part of the ability scale.
   (c) After you have studied the curve, make sure to click the R console window to make it current and to type in a new command line.
   (d) The next graph will be plotted in a new graphics window.

2. An item with hard item difficulty and low item discrimination is to be plotted.

   (a) Use the function `iccplot` to plot an item characteristic curve of an item with hard item difficulty and low item discrimination.
   (b) From the graph it can be seen that the probability of correct response will have a low general level over most of the ability scale. The item characteristic curve will not be very steep.
   (c) After you have studied the curve, make sure to click the R console window to make it current and to type in a new command line.
   (d) The next graph will be plotted in a new graphics window.

3. An item with medium item difficulty and low item discrimination is to be plotted.

   (a) Use the function `iccplot` to plot an item characteristic curve of an item with medium item difficulty and low item discrimination.
   (b) From the graph it can be seen that the probability of correct response will be between 0.2 and 0.8 over the range of ability shown. The item characteristic curve will be nearly linear over the range of ability employed.
   (c) After you have studied the curve, make sure to click the R console window to make it current and to type in a new command line.
   (d) The next graph will be plotted in a new graphics window.

4. In this exercise, all the items will have the same level of item difficulty but different levels of item discrimination. The intent is to relate the steepness of the curve to the level of item discrimination.

   (a) Use the function `iccplot` to plot an item characteristic curve of an item with medium item difficulty and moderate item discrimination.
   (b) From the graph it can be seen that the probability of correct response will be small at low ability levels and large at high ability levels. The item characteristic curve will be moderately steep in the middle part of the ability scale.
   (c) After you have studied the curve, make sure to click the R console window to make it current and to type in a new command line.
   (d) The next graph will be plotted on the same graph in the graphics window. Type in:

   ```
   > par(new=T)
   ```

   (e) Now repeat steps a through d several times using medium item difficulty for each item and item discrimination levels of your choice (e.g., none, low, high, perfect).
   (f) The next graph will be plotted in a new graphics window.

5. In this exercise, all the items will have the same level of item discrimination but different levels of item difficulty. The intent is to relate the location of the item on the ability scale to its level of item difficulty.

   (a) Use the function `iccplot` to plot an item characteristic curve of an item with very easy item difficulty and moderate item discrimination.
   (b) From the graph it can be seen that the probability of correct response will be reasonably large over most of the ability scale. The item characteristic curve will be moderately steep in the lower part of the ability scale.
   (c) After you have studied the curve, make sure to click the R console window to make it current and to type in a new command line.
   (d) The next graph will be plotted on the same graph in the graphics window. Type in:

   ```
   > par(new=T)
   ```

   (e) Now repeat steps a through d several times using items with moderate item discrimination and item difficulty levels of your choice (e.g., easy, medium, hard, very hard).
   (f) The next graph will be plotted in a new graphics window.

6. Experiment with various combinations of item difficulty and item discrimination of your own choice until you are confident that you can predict the shape of the item characteristic curve corresponding to the levels chosen. You may find it useful to make a rough sketch of what you think the curve will look like before you have the computer display it on the screen.

1.7   **Things to Notice**

1. When item discrimination is less than moderate, the item characteristic curve is nearly linear and appears rather flat.
2. When item discrimination is greater than moderate, the item characteristic curve is S-shaped and rather steep in its middle section.
3. When item difficulty is less than medium, most of the item characteristic curve has a probability of correct response that is greater than 0.5.
4. When item difficulty is greater than medium, most of the item characteristic curve has a probability of correct response less than 0.5.
5. Regardless of the level of item discrimination, item difficulty locates the item along the ability scale. Therefore item difficulty and item discrimination are independent of each other.
6. When an item has no item discrimination, all choices of item difficulty yield the same horizontal line at a value of $P(\theta) = 0.5$. This is because the value of item difficulty for an item with no item discrimination is undefined.
7. If you have been very observant, you may have noticed the point at which $P(\theta) = 0.5$ corresponds to item difficulty. When an item is easy, this value occurs at a low ability level. When an item is hard, this value corresponds to a high ability level.