ESTIMATION OF LATENT ABILITY USING A RESPONSE PATTERN OF GRADED SCORES

FUMIKO SAMEJIMA

THE UNIVERSITY OF NEW BRUNSWICK
FREDERICTON, CANADA

Printed by The William Byrd Press, Richmond, Va.

TABLE OF CONTENTS

INTRODUCTION

This monograph is a part of a more comprehensive treatment of estimation of latent traits, when the entire response pattern is used.

The fundamental structure of the whole theory comes from the latent trait model, which was initiated by Lazarsfeld as the latent structure analysis [Lazarsfeld, 1959], and also by Lord and others as a theory of mental test scores [Lord, 1952]. Similarities and differences in their mathematical structures and tendencies were discussed by Lazarsfeld [Lazarsfeld, 1960] and the recent book by Lord and Novick with contributions by Birnbaum [Lord & Novick, 1968] provides the dichotomous case of the latent trait model in the context of mental measurement.

The main difference between these two works is that Lazarsfeld tends to study all possible response patterns of dichotomous items, whereas Lord prefers to summarize the information on the examinee's answer sheet by using test scores. That is to say, with n dichotomous items there are 2^n possible response patterns, which are represented by such vectors as $(0, 1, 0, 0, 1, \cdots)$, while these response patterns are classified into $(n + 1)$ score categories, if we use the number of right answers as the test score.

In the present treatment by the author, because of the reasons given in Chapters 2 and 4, not only the response pattern is used as the basis of estimation of latent trait instead of the test score, but also the reponse pattern itself is expanded to cover the situation where item reponses are classified into more than two categories. For the sake of convenience the terminology of mental test theory is used throughout this paper. However, the discussions in this paper are applicable in other fields of psychology, such as measurement of attitude, preference measurement, etc., as well as in the measurement of ability.

The main purpose of the present paper is: (1) to set forth a method of estimation of latent ability which makes use of graded scoring of each of a number of test items; (2) to investigate the uniqueness of such estimators; (3) to compare them with methods which make use of conventional test scores for dichotomous items in terms of the amounts of information, the standard errors of estimators, and the mean-square errors of estimators.

Throughout this paper we shall confine our discussion to the situation where the latent space is unidimensional, and a respondent is allowed to answer each item in a free way, without being forced to choose one of the given alternatives. We shall assume that the principle of local independence is admissible, so that the probabilities of joint occurrence of specified responses to the items should be simple products of the separate probabilities.

We shall also assume that the values of item parameters are known, so that we do not have to estimate them from the data.

Stress will be put upon formulating sufficient conditions for the existence of a unique maximum likehood estimator and a unique Bayes modal estimator when the entire response pattern is used. These conditions will guide us when we try to find operating characteristics for individual item responses which give a unique maximum likehood estimator and a unique Bayes modal estimator. Stress will also be put upon finding a way of increasing the amount of information given by a set of items, and for this purpose operating characteristics of graded item responses are introduced and discussed. The Bayes estimator will be developed, with the loss function taken to be the mean-square error multiplied by the density function of the latent variate, when the distribution is known. In the case where this distribution satisfies certain conditions, the Bayes modal estimator will also be obtained, and its utility as an approximation to the Bayes estimator will be observed and discussed especially in connection with the mean-square errors of estimators. The computational procedures for obtaining the estimates will briefly be discussed. Finally, the relationship between the formula for the item characteristic function and the philosophy of scoring will be observed and discussed.

CHAPTER 1

DEFINITION OF RESPONSE PATTERN AND OPERATING CHARACTERISTIC

Let g, h or j denote an item, and k_g, k_h or k_j be any specified response to item g, h or j. By response pattern, denoted by V_k or V, we mean a vector defined for specified n items by

$$V_k = (k_1, k_2, k_3, \cdots, k_n), \tag{1-1}$$

which is a sequence of specified item responses given by a respondent. In the case of free-response items each item can elicit an almost infinite number of different responses, and hence each test or a set of questionnaires can elicit an almost infinite number of different response patterns. In the process of analyzing data, however, it may be more practical for us to categorize these infinite numbers of possibly different responses to an item in accordance with some criteria. When any response to item g can be evaluated according to its degree of attainment to the solution of the problem in the measurement of ability, or to its degree of intensity of favorableness to the statement in the measurement of attitude, all the possible responses to item g can be classified into a certain limited number of categories arranged in the order of attainment or intensity. We shall call these categories item scores to item g, and denote them by x_g. For convenience, let the possible value of x_g be successive integers, 0 to m_g, assigned in the order of intensity or attainment. Then in such a case the response pattern may be changed into a sequence of integers expressed by

$$(x_1, x_2, x_3, \cdots, x_n), \tag{1-2}$$

and for the sake of simplicity we shall also call it a response pattern and denote it as V.

These specified responses to item g can can come, with possibly different probabilities, from a subject anywhere on the latent continuum. By the operating characteristic of item response we mean the probability of a specified item response, k_g, being expressed as a function of the latent variable θ. Let $P_{k_g}(\theta)$ or P_{k_g} denote the operating characteristic of a specified item response, k_g, so that we have

$$P_{k_g}(\theta) = \Pr\{k_g \mid \theta\}. \tag{1-3}$$

Thus to any response pattern V we could assign n specified operating characteristics of item responses expressed as functions of θ. They can be monotonically increasing or decreasing in θ, unimodal, or multimodal, or of any shape.

We shall define the operating characteristic of response pattern as the probability of a specified response pattern V_k being expressed as a function of θ. Let $P_V(\theta)$ denote the operating characteristic of response pattern V_k, so that we have

$$P_V(\theta) = \Pr\{V_k \mid \theta\}, \tag{1-4}$$

and by the principle of local independence we can write

$$P_V(\theta) = \prod_{k_g \in V} P_{k_g}(\theta). \tag{1-5}$$

In the case where any response to item g is scored in the way explained earlier, we may assign an operating characteristic to any score category x_g, which may be called the operating characteristic of graded response x_g, denoted by $P_{x_g}(\theta)$ or P_{x_g}. Thus in this case we can write

$$\begin{aligned} P_{x_g}(\theta) &= \Pr\{x_g \mid \theta\} \\ &= \Pr\{k_g \in x_g \mid \theta\} \end{aligned} \tag{1-6}$$

and

$$P_V(\theta) = \prod_{x_g \in V} P_{x_g}(\theta). \tag{1-7}$$

When all the items are scored dichotomously, 1 or 0, as is often the case in aptitude or achievement testings, m_g is unity and the resulting response pattern becomes a sequence of 1 and 0. In such a case $P_1(\theta)$ becomes what is usually called the item characteristic function of item g in the terminology of mental test theory [Lord & Novick, 1968]. We shall denote $P_1(\theta)$ as $P_g(\theta)$ and $P_0(\theta)$ as $Q_g(\theta)$ in this particular case, as distinct from the others, where items are scored in more graded ways.

If for some reason all the possible response patterns for specified n items are reclassified into a certain limited number of categories, then $P_T(\theta)$, the operating characteristic of such a category, T, will be given by

$$\begin{aligned} P_T(\theta) &= \sum_{V \in T} P_V(\theta) \\ &= \sum_{V \in T} \prod_{k_g \in V} P_{k_g}(\theta). \end{aligned} \tag{1-8}$$

A simple test score, for instance, is an example of such a category, but (1-8) holds for any other principle of categorization.

CHAPTER 2

MAXIMUM LIKEHOOD ESTIMATOR AND BAYES MODAL ESTIMATOR BASED ON THE RESPONSE PATTERN

When the functional forms for $P_{k_g}(\theta)$ and their parameter values are specified with respect to n particular items, the response pattern given by a respondent can be regarded as a sample of n independent observations from possibly different distributions, with θ as the single unknown parameter. In maximum likehood estimation we take as the estimator of this single unknown parameter that value of θ which makes the joint probability of n independent observations as large as possible. In the present case the likelihood function is the operating characteristic of the response pattern itself so that

$$\text{(2-1)} \qquad L_V(\theta) = \prod_{k_g \in V} P_{k_g}(\theta) = P_V(\theta),$$

and the likelihood equation will be given by

$$\text{(2-2)} \qquad \frac{\partial}{\partial\theta} \log L_V(\theta) = \frac{\partial}{\partial\theta} [\sum_{k_g \in V} \log P_{k_g}(\theta)] = 0.$$

It has been pointed out by Lord [1953] that on the normal ogive model for dichotomous items a sufficient statistic for θ exists only if all the items are equivalent, and observation and discussion are focused on the maximum likelihood and confidence interval estimation of an examinee's ability within the restriction of equivalence of items in the case of free-response items as well as multiple-choice items. It has also been pointed out by Birnbaum [in Lord & Novick, 1968] that on the logistic model a simple sufficient statistic for θ always exists with respect to any response pattern, and this model with a specified scaling factor can be used as a good approximation to the normal ogive model.

In virtue of the fact that high-speed computers are now available and complicated calculations can be make at a very small cost, we can use the entire response pattern on the variety of models in estimating a respondent's latent trait without such a restrictive assumption that the items are equivalent, nor with any approximation to one model by another. Although the sufficiency of a statistic is a desirable property especially in maximum likelihood estimation, it does not hold with respect to many models for the operating characteristic of an individual item response, even if the items are scored dichotomously. In such a case we should use a maximum likelihood estimator, if it exists, based on the entire response pattern rather than change the model itself or add some restrictive assumptions, unless there is a good reason for so doing. For this reason, and aiming towards a broad range of applicability, we shall discuss in the next chapter the sufficient conditions for the existence of the unique maximum likelihood estimator with respect to the model when the entire response pattern is used.

One convenient property of the maximum likelihood estimator with respect to the likelihood function given by (2-1) will be its transformation-free character, which does not pertain to other estimators like the median and Bayesian estimators. Suppose there is another variable τ, which has a one-to-one correspondence with θ. In an ordinary case it may be reasonable to assume that one can be expressed as a monotonically increasing function of the other. From the definition of the operating characteristic of item response we can write

$$\begin{aligned} P_{k_g}(\theta) &= P_r\{k_g \mid \theta\} \\ &= P_r\{k_g \mid \tau\}. \end{aligned} \tag{2-3}$$

We can see from (2-1) and (2-3) that, if a maximum likelihood estimator $\hat{\theta}$ exists for the variable θ, $\tau(\hat{\theta})$ is also the maximum likelihood estimator for the variable τ, which is denoted by $\hat{\tau}$. Thus whenever there is a good reason why θ should be transformed, the new maximum likelihood estmiator can easily be obtained from the old one through the functional relationship between the two variables.

To make this point clearer, suppose θ and τ are continuous and P_{k_g} is differentiable with respect to θ as well as with respect to τ throughout their ranges, and further that τ is a monotonically increasing and twice-differentiable function of θ.

From (2-2) and (2-3) we would write

$$\begin{aligned} \frac{\partial \log L_V}{\partial \theta} &= \sum_{k_g \in V} \frac{(\partial/\partial\theta)P_{k_g}}{P_{k_g}} \\ &= \left[\sum_{k_g \in V} \frac{(\partial/\partial\tau)P_{k_g}}{P_{k_g}} \right] \frac{\partial \tau}{\partial \theta} \\ &= \frac{\partial \log L_V}{\partial \tau} \frac{\partial \tau}{\partial \theta}. \end{aligned} \tag{2-4}$$

Since $\partial\tau/\partial\theta$ is positive, it is clear from (2-4) that $\hat{\tau}$ exists whenever $\hat{\theta}$ exists, and that the value of τ which corresponds to the maximum likelihood estimator $\hat{\theta}$ will also make likelihood function L_V equal to zero regardless of the functional form for $\partial\tau/\partial\theta$.

Generally speaking, we cannot decide which one of the possible monotonic transformations is the most appropriate one, since it is difficult for us to define a reasonable unit and origin of the latent continuum in a strictly objective way. Thus there is a philosophical difficulty in deciding whether a certain difference of the latent ability at one place of the continuum is equivalent to a difference at another place. This problem may be solved, or at least ameliorated, if we succeed in relating the ability measure to some important external variable. Then we might want to choose a transformation which would make that relationship linear. Without such an external criterion we can say, for instance, that examinee A with ability θ_1 is superior to examinee B with ability θ_2, and examinee B is superior to examinee C with ability θ_3, but we cannot say that the superiority of A at B is greater than, equal to or less than, that of B to C. Thus in general cases the origin and unit of the latent continuum are of arbitrary natures, and they may be transformed to another arbitrarily. The transformation-free character, therefore, is a desirable property of an estimator.

When the density function of the latent variable, $f(\theta)$, is unknown, the maximum likelihood estimator $\hat{\theta}$ may be the most reasonable and useful estimator of θ. If the latent density function is known, however, we can add this information to the likelihood function. We shall define another function, $B_V(\theta)$, so that

$$B_V(\theta) = f(\theta)L_V(\theta) = f(\theta) \prod_{k_g \in V} P_{k_g}(\theta), \tag{2-5}$$

and then we have

$$\frac{\partial}{\partial\theta} \log B_V(\theta) = \frac{\partial}{\partial\theta} [\log f(\theta) + \sum_{k_g \in V} \log P_{k_g}(\theta)] \tag{2-6}$$
$$= 0$$

in the same way that the likelihood equation is defined.

The new estimator, which makes $B_V(\theta)$ absolutely maximal, is denoted by $\hat{\hat{\theta}}$, as distinct from $\hat{\theta}$. Hereafter, we shall call this estimator the Bayes modal estimator in distinction from the maximum likelihood estimator and also from the Bayes estimator, which will be introduced and discussed in Chapter 7.

Since in general cases $f(\theta)$ does not have a transformation-free character, the Bayes modal estimator $\hat{\hat{\theta}}$ does not have a transformation-free property except in very limited situations. The density function of τ, denoted by $g(\tau)$, is transformed from the density function of θ by

$$g(\tau) = f(\theta)\cdot\frac{\partial\theta}{\partial\tau}, \tag{2-7}$$

and then from (2-6) and (2-7) we have

$$\begin{aligned}
(2\text{-}8) \qquad \frac{\partial \log B_V(\theta)}{\partial \theta} &= \frac{(\partial/\partial\theta)f(\theta)}{f(\theta)} + \sum_{k_g \in V} \frac{(\partial/\partial\theta)P_{k_g}}{P_{k_g}} \\
&= \frac{(\partial/\partial\theta)\{g(\tau)\cdot\partial\tau/\partial\theta\}}{g(\tau)\cdot\partial\tau/\partial\theta} + \left[\sum_{k_g \in V} \frac{(\partial/\partial\tau)P_{k_g}}{P_{k_g}}\right] \frac{\partial\tau}{\partial\theta} \\
&= \left[\frac{(\partial/\partial\tau)g(\tau)}{g(\tau)} + \sum_{k_g \in V} \frac{(\partial/\partial\tau)P_{k_g}}{P_{k_g}}\right] \frac{\partial\tau}{\partial\theta} + \frac{\partial^2\tau/\partial\theta^2}{\partial\tau/\partial\theta} \\
&= \frac{\partial \log B_V^*(\tau)}{\partial\tau} \frac{\partial\tau}{\partial\theta} + \frac{\partial^2\tau/\partial\theta^2}{\partial\tau/\partial\theta} ,
\end{aligned}$$

where

$$\begin{aligned}
(2\text{-}9) \qquad B_V^*(\tau) &= g(\tau) \prod_{k_g \in V} P_\tau\{k_g \mid \tau\} \\
&= g(\tau) \prod_{k_g \in V} P_{k_g}(\theta).
\end{aligned}$$

Thus it has been made clear that $\hat{\hat{\tau}}$ does not coincide with $\tau(\hat{\hat{\theta}})$ unless $\partial^2\tau/\partial\theta^2 = 0$. This means that the Bayes modal estimator obtained by using $B_V(\theta)$ as the function to be minimized does not have a transformation-free property in the way $\hat{\theta}$ does, unless the transformation is of linear form.

CHAPTER 3

SUFFICIENT CONDITIONS FOR THE EXISTENCE OF THE UNIQUE MAXIMUM LIKELIHOOD ESTIMATOR AND THE UNIQUE BAYES MODAL ESTIMATOR WHEN THE ENTIRE RESPONSE PATTERN IS USED

In using maximum likelihood estimation, it is needless to say that the uniqueness of the maximum likelihood estimator is a desirable character for the likelihood function, although we can choose the largest of all the local maxima, if there are more than one. We shall, therefore, confine our attention to situations where only one local maximum exists for the likelihood function $L_V(\theta)$ and also for the function $B_V(\theta)$.

If we assume that the likelihood function is differentiable with respect to θ throughout its range, a sufficient condition for the existence of an absolute maximum for the likelihood function is that the likelihood equation should hold at a unique value of θ, say θ_0 , and it is monotonically decreasing in θ for $\theta_0 - \epsilon < \theta < \theta_0 + \epsilon$, where ϵ is some positive number. The same is true with respect to the function $B_V(\theta)$ and the equation given by (2-6). Of course, this is not a necessary condition. It is apparent that the likelihood function $L_V(\theta)$ as well as the function $B_V(\theta)$ can be possessed of a unique local maximum even if its first derivative does not exist throughout the range of θ. However, we shall proceed in line with the above assumption and requirement, since we must content ourselves with formulating rules which will give us fruitful results and a relatively wide range of applicability from the practical point of view.

It seems appropriate for us to set up sufficient conditions for the likelihood function $L_V(\theta)$ or the function $B_V(\theta)$ to have a unique maximum with respect to individual items, rather than with respect to an entire test or a set of questionnaires. Conditions which are set up for a particular battery of items will be of little use, since we have to take into consideration the testing situations where we are free to choose items with optimal quality, from the standpoint of testing purposes and respondents' conditions, out of the item library in which a sufficient number of items are stored with reliable quality controls.

For the reasons described above, we shall assume that $P_{k_g}(\theta)$ has a first derivative with respect to θ throughout its range. Let us denote the range of θ by $(\underline{\theta}, \bar{\theta})$ (i.e., $\underline{\theta} < \theta < \bar{\theta}$) or $[\underline{\theta}, \bar{\theta}]$ (i.e., $\underline{\theta} \leq \theta \leq \bar{\theta}$), with $\underline{\theta}$ and $\bar{\theta}$ as the lower and upper bounds of the range, for any possible response to item g. Further, we shall define $A_{k_g}(\theta)$ so that

$$A_{k_g}(\theta) = \frac{(\partial/\partial\theta)P_{k_g}(\theta)}{P_{k_g}(\theta)}, \tag{3-1}$$

and two asymptotes, $C_{k_g,\underline{\theta}}$ and $C_{k_g,\bar{\theta}}$, such that

$$\left.\begin{aligned} C_{k_g,\underline{\theta}} &= \lim_{\theta\to\underline{\theta}} A_{k_g}(\theta) \\ C_{k_g,\bar{\theta}} &= \lim_{\theta\to\bar{\theta}} A_{k_g}(\theta) \end{aligned}\right\}. \tag{3-2}$$

For the likelihood function L_V, defined in the previous chapter, sufficient (though not necessary) conditions for unique maximum with respect to an individual item can be expressed by the following two statements.

(i) $A_{k_g}(\theta)$ should be monotonically decreasing in θ for any response k_g to item g.

(ii) $C_{k_g,\underline{\theta}}$ and $C_{k_g,\bar{\theta}}$ should be positive and negative respectively, for any response k_g to item g. These values can be positive and negative infinities. This second statement is expressed by the following inequalities.

$$\left.\begin{aligned} C_{k_g,\underline{\theta}} &> 0 \\ C_{k_g,\bar{\theta}} &< 0 \end{aligned}\right\}. \tag{3-3}$$

Hereafter, we shall call these statements conditions (i) and (ii) respectively. We may use the above pair of statements as a criterion for a model to produce a unique maximum likelihood estimator, $\hat{\theta}$, with respect to any possible response pattern. It is also apparent that if a test consists of items, each of which has responses satisfying both conditions (i) and (ii), but has a model for the operating characteristic which is different from each other, the resulting likelihood function also supplies a unique local maximum with any response pattern. This fact suggests the possibility of combining several such models whenever it is necessary, provided that the ranges of the latent variables defined are the same.

If we content ourselves with obtaining a terminal maximum at either of the positive and negative extreme values of θ for a very limited number of possible response patterns, (3-3) may be changed into the following.

$$\left.\begin{aligned} C_{k_g,\underline{\theta}} &\geq 0 \\ C_{k_g,\bar{\theta}} &\leq 0 \end{aligned}\right\} \tag{3-4}$$

where one at least is a strict inequality. In this case, when an equality holds for a specified one of them with respect to every element of a given response pattern, a terminal maximum is obtained. Hereafter, we shall call this condition (ii)*, as a substitute for condition (ii). On many models of practical importance condition (ii) does not hold for every individual item response, while condition (ii)* does instead. The normal ogive model and the logistic model for dichotomous items are examples of this case, as we shall see in Chapter 5.

The applicability of these pairs of conditions as criteria for producing a unique maximum likelihood estimator is broadened when we take into consideration the possible transformations of the latent variable, which we denote as τ in the previous chapter. Since maximum likelihood estimator $\hat{\theta}$ has a transformation-free character, as was discussed in the previous chapter, $\hat{\tau}$ also exists whenever $\hat{\theta}$ exists, even if the transformed operating characteristics for individual item responses do not satisfy the above requirements. Thus, even in case a given model does not provide the operating characteristics of individual item responses which satisfy the requirements in the above conditions, we could try to find another variable, which can be expressed as a monotonically increasing function of the original variable, and with which the transformed operating characteristics of item responses satisfy the requirements in the above conditions, in order to see whether the original model supplies the likelihood function which gives a unique maximum likelihood estimator with respect to any possible response pattern.

For illustrative purposes, we shall see an example in which $A_{k_g}^{[\tau]}(\tau)$ defined by

$$A_{k_g}^{[\tau]}(\tau) = \frac{(\partial/\partial\tau)P_{k_g}}{P_{k_g}} \tag{3-5}$$

fulfills the requirement stated in condition (i), while $A_{k_g}(\theta)$ defined by (3-1) does not, and the one variable is a monotonically increasing function of the other.

Example:

We shall consider a variable θ with the range, $(0, \infty)$, and dichotomous item g, all the responses to which are reasonably classified into two categories, 1 and 0. Suppose their operating characteristics are given by

$$P_1 = \frac{\theta}{1 + \theta} \tag{3-6}$$

and

$$P_0 = \frac{1}{1 + \theta}, \tag{3-7}$$

where

$$P_1 + P_0 = 1. \tag{3-8}$$

Differentiating (3-6) with respect to θ, we have

$$\frac{\partial}{\partial\theta}P_1 = \frac{1}{(1 + \theta)^2} \tag{3-9}$$

and then by a further differentiating

$$\frac{\partial^2}{\partial \theta} P_1 = -\frac{2}{(1+\theta)^3}. \tag{3-10}$$

From these two equations and (3-6), we obtain the following inequality

$$P_1\left\{\frac{\partial^2}{\partial \theta^2} P_1\right\} - \left\{\frac{\partial}{\partial \theta} P_1\right\}^2 = -\frac{2\theta + 1}{(1+\theta)^4} < 0\cdot \tag{3-11}$$

Since the left-hand side of (3-11) is the numerator of the derivative of $A_{k_o}(\theta)$, (3-11) suggests that $A_1(\theta)$ satisfies condition (i).

On the other hand, by differentiating (3-7) with respect to θ and proceeding in the same way, we have

$$\frac{\partial}{\partial \theta} P_0 = -\frac{1}{(1+\theta)^2}, \tag{3-12}$$

$$\frac{\partial^2}{\partial \theta^2} P_0 = \frac{2}{(1+\theta)^3} \tag{3-13}$$

and

$$P_0\left\{\frac{\partial^2}{\partial \theta^2} P_0\right\} - \left\{\frac{\partial}{\partial \theta} P_0\right\}^2 = \frac{1}{(1+\theta)^4} > 0, \tag{3-14}$$

the result indicating that $A_0(\theta)$ is a monotonically *increasing* function of θ. Thus it is clear in this case that $A_0(\theta)$ does not fulfill condition (i).

Now we shall consider another variate τ, with the range, $(-\infty, \infty)$, and whose functional relationship with θ is given by

$$\tau = \log \theta. \tag{3-15}$$

Then the preceding operating characteristics as functions of θ will be rewritten as the following.

$$P_1 = \frac{1}{1 + e^{-\tau}}, \tag{3-16}$$

$$P_0 = \frac{1}{1 + e^{\tau}}. \tag{3-17}$$

Differentiating these equations with respect to τ and following the same way as before, we obtain

$$\text{(3-18)} \qquad \frac{\partial}{\partial\tau}P_1 = \frac{1}{\{1+e^{\tau}\}\{1+e^{-\tau}\}}$$
$$= P_1P_0,$$

$$\text{(3-19)} \qquad \frac{\partial}{\partial\tau}P_0 = -\frac{1}{\{1+e^{\tau}\}\{1+e^{-\tau}\}}$$
$$= -P_1P_0,$$

$$\text{(3-20)} \qquad \frac{\partial^2}{\partial\tau^2}P_1 = P_1P_0\{P_0 - P_1\}$$

and

$$\text{(3-21)} \qquad \frac{\partial^2}{\partial\tau^2}P_0 = P_1P_0\{P_1 - P_0\},$$

and then finally,

$$\text{(3-22)} \qquad P_1\left\{\frac{\partial^2}{\partial\tau^2}P_1\right\} - \left\{\frac{\partial}{\partial\tau}P_1\right\}^2 = -P_1^3P_0$$
$$< 0$$

and

$$\text{(3-23)} \qquad P_0\left\{\frac{\partial^2}{\partial\tau^2}P_0\right\} - \left\{\frac{\partial}{\partial\tau}P_0\right\}^2 = -P_1P_0$$

The results indicate that both $A_1^{[\tau]}(\tau)$ and $A_0^{[\tau]}(\tau)$ satisfy the requirement stated in condition (i).

Thus we have seen in this example that condition (i) does not hold when we take θ as the variable and (3-6) and (3-7) as the operating characteristics of item responses, while it does when we take τ, the logarithmic transformation of θ, as the variable and the transformed equations, (3-16) and (3-17), as the operating characteristics.

We could easily see that both $A_1^{[\tau]}(\tau)$ and $A_0^{[\tau]}(\tau)$ satisfy the requirement stated in condition (ii)*, since we have from (3-5), (3-18), and (3-19)

$$\text{(3-24)} \qquad A_1^{[\tau]}(\tau) = \frac{(\partial/\partial\tau)P_1}{P_1}$$
$$= P_0$$

and

$$\text{(3-25)} \qquad A_0^{[\tau]}(\tau) = \frac{(\partial/\partial\tau)P_0}{P_0}$$
$$= -P_1,$$

and then

$$\left.\begin{aligned} C_{1,\underline{\tau}} &= \lim_{\tau\to-\infty} P_0 \\ &= 1 \\ C_{1,\bar{\tau}} &= \lim_{\tau\to\infty} P_0 \\ &= 0 \end{aligned}\right\} \tag{3-26}$$

and

$$\left.\begin{aligned} C_{0,\underline{\tau}} &= \lim_{\tau\to-\infty} [-P_1] \\ &= 0 \\ C_{0,\bar{\tau}} &= \lim_{\tau\to\infty} [-P_1] \\ &= -1 \end{aligned}\right\}. \tag{3-27}$$

The above results suggest that both $\hat{\theta}$ and $\hat{\tau}$ exist with respect to any possible response pattern except for two extreme cases for a test consisting of such items, and one is transformed to the other through the functional form given by (3-15). In both cases terminal maxima are obtained instead with respect to the two particular response patterns where all the elements are 1, and where they are 0, since neither $A_1^{[\tau]}(\tau)$ nor $A_0^{[\tau]}(\tau)$ satisfies the requirement stated in condition (ii), that is, strict inequalities.

To show this, the likelihood equation is given by

$$\begin{aligned} \frac{\partial \log L_V(\theta)}{\partial \theta} &= n_1 A_1(\theta) + n_0 A_0(\theta) \\ &= \frac{n_1}{\theta(1+\theta)} - \frac{n_0}{(1+\theta)} \\ &= 0, \end{aligned} \tag{3-28}$$

where n_1 and n_0 are numbers of response 1 and 0 respectively in a given response pattern, satisfying

$$n = n_1 + n_0 \,. \tag{3-29}$$

From (3-28) we obtain

$$\hat{\theta} = \frac{n_1}{n_0} \tag{3-30}$$

which provides terminal maxima, 0 and positive infinity, when $n_1 = 0$ and $n_0 = 0$ respectively, and reasonable maximum likelihood estimates otherwise.

On the other hand, if the likelihood equation is defined with respect to τ, we have

$$\text{(3-31)} \qquad \frac{\partial \log L_V}{\partial \tau} = n_1 A_1^{[\tau]}(\tau) + n_0 A_0^{[\tau]}(\tau)$$

$$= \frac{n_1}{\{1+e^{\tau}\}} - \frac{n_0}{\{1+e^{-\tau}\}}$$

$$= \frac{n_1}{\{1+e^{\tau}\}} - \frac{n_0 e^{\tau}}{\{1+e^{\tau}\}}$$

$$= 0.$$

Rearranging (3-31) we could write

$$\text{(3-32)} \qquad \hat{\tau} = \log n_1 - \log n_0$$

$$= \log (\hat{\theta}),$$

which provides terminal maxima, positive and negative infinities, when $n_0 = 0$ and $n_1 = 0$ respectively, and reasonable maximum likelihood estimates otherwise, the results directly obtainable from $\hat{\theta}$ through (3-15).

In a specific case where one of the variables is a linear transformation of the other, $A_{k_g}(\theta)$ is monotonically decreasing in θ if $A_{k_g}^{[\tau]}(\tau)$ is monotonically decreasing in τ, and vice versa. To prove this, we shall assume that P_{k_g} is twice-differentiable with respect to τ as well as with respect to θ. We could write

$$\text{(3-33)} \qquad \frac{\partial}{\partial \theta} P_{k_g} = \left\{\frac{\partial}{\partial \tau} P_{k_g}\right\} \frac{\partial \tau}{\partial \theta},$$

$$\text{(3-34)} \qquad \frac{\partial^2}{\partial \theta^2} P_{k_g} = \left\{\frac{\partial^2}{\partial \tau^2} P_{k_g}\right\}\left\{\frac{\partial \tau}{\partial \theta}\right\}^2 + \left\{\frac{\partial}{\partial \tau} P_{k_g}\right\} \frac{\partial^2 \tau}{\partial \theta^2}$$

and then

$$\text{(3-35)} \qquad P_{k_g}\left\{\frac{\partial^2}{\partial \theta^2} P_{k_g}\right\} - \left\{\frac{\partial}{\partial \theta} P_{k_g}\right\}^2$$

$$= P_{k_g}\left\{\frac{\partial^2}{\partial \tau^2} P_{k_g}\right\}\left\{\frac{\partial \tau}{\partial \theta}\right\}^2 + P_{k_g}\left\{\frac{\partial}{\partial \tau} P_{k_g}\right\} \frac{\partial^2 \tau}{\partial \theta^2} - \left\{\frac{\partial}{\partial \tau} P_{k_g}\right\}^2 \left\{\frac{\partial \tau}{\partial \theta}\right\}^2$$

$$= \left\{\frac{\partial \tau}{\partial \theta}\right\}^2 \left[P_{k_g}\left\{\frac{\partial^2}{\partial \tau^2} P_{k_g}\right\} - \left\{\frac{\partial}{\partial \tau} P_{k_g}\right\}^2\right] + P_{k_g}\left\{\frac{\partial}{\partial \tau} P_{k_g}\right\} \frac{\partial^2 \tau}{\partial \theta^2}.$$

We can see that the first term of the right-hand side of (3-35) will be negative only if $A_{k_g}^{[\tau]}(\tau)$ is monotonically decreasing in τ, and that

$$\text{(3-36)} \qquad \frac{\partial^2 \tau}{\partial \theta^2} = 0$$

only if τ is a linear transformation of θ. Thus in this case the left-hand side of (3-35) is negative, so it has been proved that $A_{k_v}(\theta)$ as well as $A_{k_v}^{[\tau]}(\tau)$ is a monotonically decreasing function of θ. The other half of the proof can be attained simply by exchanging θ to τ and τ to θ in (3-33) through (3-36).

So far we have discussed sufficient conditions for the likelihood function L_V to provide a maximum likelihood estimator. Now we shall proceed to discuss those for the function B_v to supply a unique Bayes modal estimator with respect to any possible response pattern.

For the function B_v, sufficient (though not necessary) conditions for the existence of an absolute maximum will be expressed by the following two statements concerning the density function of θ, in addition to the previous two conditions, (i) and (ii)*.

Throughout the rest of this chapter we assume that $f(\theta)$ is differentiable with respect to θ. Here we shall define $G(\theta)$, and the two asymptotes, $C_{f,\underline{\theta}}$ and $C_{f,\bar{\theta}}$, so that

$$G(\theta) = \frac{(\partial/\partial\theta)f(\theta)}{f(\theta)} \tag{3-37}$$

and

$$\left.\begin{aligned} C_{f,\underline{\theta}} &= \lim_{\theta\to\underline{\theta}} G(\theta) \\ C_{f,\bar{\theta}} &= \lim_{\theta\to\bar{\theta}} G(\theta) \end{aligned}\right\}. \tag{3-38}$$

These additional conditions are the following:

(iii) $G(\theta)$ defined by (3-37) should be monotonically decreasing in θ.

(iv) $C_{f,\underline{\theta}}$ and $C_{f,\bar{\theta}}$ should be positive and negative respectively, and these values can be positive and negative infinities. This statement will be expressed by the following set of inequalities.

$$\left.\begin{aligned} C_{f,\underline{\theta}} &> 0 \\ C_{f,\bar{\theta}} &< 0 \end{aligned}\right\}. \tag{3-39}$$

Here it should be noted that, in order for a model to provide a unique local maximum with respect to any possible response pattern without exception, the model does not have to produce the operating characteristics of individual item responses which satisfy the requirement stated in condition (ii), that is, strict inequalities, but what is required is only the fulfillment of condition (ii)* instead, which holds more frequently in practical situations, as we shall see in Chapter 5. In such cases we can always get the Bayes modal estimator $\hat{\theta}$ with respect to every possible response pattern whenever an appropriate latent density function is given, although we can only obtain terminal maxima with respect to the extreme response patterns if we try to get the maximum likelihood estimator $\hat{\theta}$. This fact supports the utility of the Bayes modal estimator $\hat{\hat{\theta}}$ as a good computational compromise for the Bayes estimator as we shall see in Chapter 8.

As examples of the latent density function satisfying the requirements stated in conditions (iii) and (iv), the normal and logistic density function with μ and σ as the parameters are taken here. In these cases the range of variate θ is $(-\infty, \infty)$, and each density function is expressed as follows:

$$f(\theta) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(\theta-\mu)^2/2\sigma^2} \tag{3-40}$$

$$f(\theta) = \frac{d}{\sigma}\{1 + e^{-d(\theta-\mu)/\sigma}\}^{-1}\{1 + e^{d(\theta-\mu)/\sigma}\}^{-1} \tag{3-41}$$

where d is a scaling factor.

For the normal density function, we have by differentiating (3–40) with respect to θ,

$$\begin{aligned} \frac{\partial}{\partial\theta} f(\theta) &= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(\theta-\mu)^2/2\sigma^2}\left\{-\frac{(\theta-\mu)}{\sigma^2}\right\} \\ &= f(\theta)\left\{-\frac{(\theta-\mu)}{\sigma^2}\right\}, \end{aligned} \tag{3-42}$$

and then inserting this result into (3-37), we obtain

$$\underline{G(\theta)} = -\frac{(\theta-\mu)}{\sigma^2}. \tag{3-43}$$

Obviously this result shows that the normal density function satisfies the requirement stated in condition (iii), and will give the two asymptotes such that

$$\left.\begin{aligned} C_{f,\underline{\theta}} &= \infty \\ C_{f,\bar{\theta}} &= -\infty \end{aligned}\right\} \tag{3-44}$$

which fulfill the requirement stated in condition (iv).

As for the logistic density function, the first derivative of (3-41) is given by

$$\begin{aligned} \frac{\partial}{\partial\theta} f(\theta) &= \frac{d^2}{\sigma^2}\{1+e^{-d(\theta-\mu)/\sigma}\}^{-1}\{1+e^{d(\theta-\mu)/\sigma}\}^{-1}\{1-2(1+e^{-d(\theta-\mu)/\sigma})^{-1}\} \\ &= f(\theta)\cdot\frac{d}{\sigma}\{1-2(1+e^{-d(\theta-\mu)/\sigma})^{-1}\}. \end{aligned} \tag{3-45}$$

Inserting this result into (3-37), we obtain

$$G(\theta) = \frac{d}{\sigma}\{1-2(1+e^{-d(\theta-\mu)/\sigma})^{-1}\} \tag{3-46}$$

which is again monotonically decreasing in θ and will give the two asymptotes such that

$$\left.\begin{aligned} C_{f,\underline{\theta}} &= \frac{d}{\sigma} \\ C_{f,\bar{\theta}} &= -\frac{d}{\sigma} \end{aligned}\right\}. \tag{3-47}$$

Thus conditions (iii) and (iv) are also met with respect to the logistic density function.

CHAPTER 4

THE OPERATING CHARACTERISTIC OF GRADED RESPONSE WHEN THE THINKING PROCESS IS HOMOGENEOUS

In many cases of estimating a respondent's latent trait, or ability in the terminology of mental test theory, the situation is limited to one in which a sample of respondents has answered a test or a set of questionnaires consisting of a certain number of dichotomous items. In other words, any response to an item should be either positive or negative, that is, "correct" or "incorrect" in the terminology of mental testing and "favorable" or "unfavorable" in the terminology of attitude measurement. Thus the resulting response pattern is a sequence of positive and negative responses, and, on the basis of such a response pattern or of some kind of score which is nothing but an aggregation of more than one response pattern, some kind of average position is assigned to the respondent.

If, however, we try to measure an ability which is essential to profound thinking, for instance, we may have to prepare items requiring complex reasoning processes, and consequently it may require a considerable amount of time for an examinee to solve even one of them. Since the entire time for testing is more or less limited and we cannot expect an examinee to solve many such items, it may be that the more profound the items, the less information we get about an examinee's ability, so far as the answers are evaluated dichotomously, *i.e.*, success or failure. In this instance we shall be able to get more information if we modify the items so that we may evaluate their responses in a more graded way, without changing the qualities of the items.

If an examinee succeeds in solving a given problem perfectly, his reasoning should be regarded as complete. If he fails in solving it, his reasoning process should be considered as incomplete, possibly to various extents. We should score the response to an item in accordance with the extent of the respondent's attainment toward the goal manifested in his protocol. In other words, how close his reasoning process has attained toward the goal should be evaluated. Sometimes the reasoning required in solving the problem may be fairly homogeneous throughout the whole thinking process, while sometimes it may be a heterogeneous one consisting of somewhat different subprocesses like a chain. In the former case the operating characteristics of graded responses may be expressed in a much simpler way, since it may reasonably be assumed that the discriminating power should be almost constant throughout the whole thinking process required in solving the problem. In the latter case, however, they may be more complex, since each subprocess may supply a different value of discriminating power. Here we shall deal only with the former case where the thinking process used in solving a given item is assumed to be homogeneous, leaving discussion on the latter case to some other opportunity.

As we have defined in Chapter 1, when any response to item g is classified into $(m_g + 1)$ categories, 0 through m_g, the resulting response patterns are sequences of integers. We shall call such a scored response a "graded response," and an item all of whose responses are scored in this way a "graded item." A dichotomous item is a special case of a graded item, in which $m_g = 1$, as was mentioned in Chapter 1.

The response pattern for n graded items is given by the following form,

$$V = (x_1, x_2, x_3, \cdots, x_n), \tag{4-1}$$

where x_g is a nonnegative integer.

It is easily understood that any graded item can be reduced to a dichotomous item, if only we rescore a given graded item in such a way that any item score less than x_g is 0 and that more than or equal to x_g is 1. Since there are m_g category bounds for a graded item, we can obtain m_g sets of $P_g(\theta)$ and $Q_g(\theta)$, the operating characteristics defined for a dichotomous item in chapter 1. Let $P^*_{x_g}(\theta)$ denote $P_g(\theta)$ obtained in this way for category bound x_g, which varies 1 through m_g. We must note that these $P^*_{x_g}(\theta)$ may be regarded as a set of item characteristic functions *with the same value of discrimination index*, but with different values of difficulty index. Further, we shall define $P^*_0(\theta)$ and $P^*_{(m_g+1)}(\theta)$ so that

$$P^*_0(\theta) = 1 \tag{4-2}$$

and

$$P^*_{(m_g+1)}(\theta) = 0. \tag{4-3}$$

We can write for a specified graded response x_g that

$$P_{x_g}(\theta) = P^*_{x_g}(\theta) - P^*_{(x_g+1)}(\theta) > 0, \tag{4-4}$$

which is the basic formula for the operating characteristic of graded response, when the thinking process is homogeneous, for any specification of the model with difficulty and discrimination index as the parameters for the item characteristic function. When these item characteristic functions are monotonically increasing in θ, the operating characteristic of graded response is neither monotonically increasing nor decreasing in θ, unless $x_g = m_g$ or $x_g = 0$.

In the measurement of a speed factor, for instance, we could use this model of graded dichotomy, although it may be better to set up a new model for that purpose by taking into account various elements peculiar to it. In this example, items themselves are usually easy enough for an examinee of the target population to solve, if only ample time is allowed. The point is not whether an examinee can solve an individual item, but how many items he can solve within a limited time. For this reason we should rather treat as a unit an aggregation of homogeneous items with a strict time limit, and specify operating characteristics with respect to this aggregation in a graded way, for the various degrees of attainment, just like we do for an individual item of a power test. If, for instance, we try to measure the examinee's speed factor in calculation, we could prepare a certain number of different units with possibly different discriminating powers, one of which consists of homogeneous items of addition, another of which consists of homogeneous items of subtraction, and so forth. We may evaluate the results for each unit in a graded way, with, perhaps, more than two grades, and specify an operating characteristic for each graded response category of the unit.

In the measurement of attitude, this model may reasonably correspond to the situation in which the respondent's intensity of positivity toward a given statement is manifested. In this instance the intensity of positivity may vary continuously, and yet it is usually expressed in a discrete manner.

There may be more instances and situations available for this general model of graded dichotomy. In any case it may be a fruitful device for us to obtain as much information as possible from testing results.

CHAPTER 5

NORMAL OGIVE AND LOGISTIC MODELS FOR THE OPERATING CHARACTERISTIC OF GRADED RESPONSE†

As examples of models which provide operating characteristics of graded item responses, the normal ogive model and the logistic model expanded for graded items in the manner discussed in the previous chapter are introduced here. It is needless to say that whenever we specify a model it is preferable that the resulting operating characteristics supply the maximum likelihood estimator $\hat{\theta}$ at least with respect to most of the possible response patterns. The normal ogive and the logistic models expanded for graded item responses are such models, which satisfy the requirements stated in conditions (i) and (ii)*, as we shall see in the rest of this chapter.

On the normal ogive model, the formula for $P_{x_g}^*(\theta)$ is given by

$$P_{x_g}^*(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_g(\theta - b_{x_g})} e^{-t^2/2}\, dt, \tag{5-1}$$

where a_g is an item parameter indicating discriminating power and b_{x_g} is a parameter indicating difficulty specified for each response category bound x_g. We could also write

$$P_{(x_g+1)}^*(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_g[\theta - b_{(x_g+1)}]} e^{-t^2/2}\, dt, \tag{5-2}$$

where $b_{(x_g+1)}$ is an item response parameter indicating difficulty specified for category bound $(x_g + 1)$, satisfying

$$b_{(x_g+1)} > b_{x_g} . \tag{5-3}$$

From equations (5-1), (5-2) and the basic formula for the operating characteristic of graded response given as equation (4-4) in the previous chapter, we have

$$P_{x_g}(\theta) = \frac{1}{\sqrt{2\pi}} \int_{a_g[\theta - b_{(x_g+1)}]}^{a_g(\theta - b_{x_g})} e^{-t^2/2}\, dt, \tag{5-4}$$

and this formula gives the operating characteristic of graded item response on the normal ogive model.

Differentiating equation (5-4) with respect to θ, we obtain

$$\frac{\partial}{\partial \theta} P_{x_g}(\theta) = \frac{a_g}{\sqrt{2\pi}} [e^{-a_g^2[\theta - b_{x_g}]^2/2} - e^{-a_g^2[\theta - b_{(x_g+1)}]^2/2}], \tag{5-5}$$

† This topic is discussed from more general standpoints. See Samejima, [1967 and 1968b].

which equals zero only when

$$\theta = \frac{b_{x_g} + b_{(x_g+1)}}{2}. \tag{5-6}$$

Since the second derivative of $P_{x_g}(\theta)$ is given by

$$\frac{\partial^2}{\partial\theta^2} P_{x_g}(\theta) = \frac{-a_g^3}{\sqrt{2\pi}} [[\theta - b_{x_g}]e^{-a_g^2[\theta - b_{x_g}]^2/2} - [\theta - b_{(x_g+1)}]e^{-a_g^2[\theta - b_{(x_g+1)}]^2/2}] \tag{5-7}$$

and this takes a negative value when (5-6) holds, the operating characteristic has an absolute maximum at this point of θ. Since by definition of $P_0^*(\theta)$ and $P_{(m_g+1)}^*(\theta)$ we have

$$b_0 = -\infty \tag{5-8}$$

and

$$b_{(m_g+1)} = \infty, \tag{5-9}$$

the operating characteristic has the negative and positive terminal maxima when $x_g = 0$ and $x_g = m_g$ respectively. In any other case it is a unimodal curve, and it is symmetric, as we have from (5-4)

$$\begin{aligned} P_{x_g}(\theta) &= \frac{1}{\sqrt{2\pi}} \int_{-a_g(\theta - b_{x_g})}^{-a_g[\theta - b_{(x_g+1)}]} e^{-t^2/2}\, dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{a_g\{-\theta + b_{x_g} + b_{(x_g+1)} - b_{(x_g+1)}\}}^{a_g\{-\theta + b_{x_g} + b_{(x_g+1)} - b_{x_g}\}} e^{-t^2/2}\, dt \\ &= P_{x_g}[-\theta + b_{x_g} + b_{(x_g+1)}]. \end{aligned} \tag{5-10}$$

Taking $(-\infty, \infty)$ as the range of θ, we shall define $A_{x_g}(\theta)$ (or A_{x_g}) and the two asymptotes, $C_{x_g,\underline{\theta}}$ and $C_{x_g,\bar{\theta}}$, after (3-1) and (3-2) in Chapter 3, so that

$$A_{x_g}(\theta) = \frac{(\partial/\partial\theta)P_{x_g}(\theta)}{P_{x_g}(\theta)}, \tag{5-11}$$

and

$$\left.\begin{aligned} C_{x_g,\underline{\theta}} &= \lim_{\theta\to-\infty} A_{x_g}(\theta) \\ C_{x_g,\bar{\theta}} &= \lim_{\theta\to\infty} A_{x_g}(\theta) \end{aligned}\right\}. \tag{5-12}$$

Hereafter, we shall call $A_{x_g(\theta)}$ or A_{x_g} the *basic function* of a given graded item response x_g.

Now we shall prove that this basic function satisfies the requirements stated in conditions (i) and (ii) presented in Chapter 3 in any case where $x_g \neq 0$ or m_g, and conditions (i) and (ii)* hold otherwise. That is to say, we shall prove that the normal ogive model for graded responses supplies an absolute maximum for likelihood function L_V for any possible response pattern except for two extreme cases where all the elements are 0 or m_g. In these two cases negative and positive terminal maxima are given instead.

Proof:

We can rewrite (5-5) on this specific model in the following form:

$$\frac{\partial}{\partial\theta} P_{x_g}(\theta) = -\frac{a_g}{\sqrt{2\pi}} \int_{a_g[\theta - b_{(x_g+1)}]}^{a_g(\theta - b_{x_g})} te^{-t^2/2}\, dt. \tag{5-13}$$

Defining u, c and $\mu(\theta)$ such that

$$u = a_g(\theta - b_{x_g}) \tag{5-14}$$

$$c = a_g[b_{(x_g+1)} - b_{x_g}] \tag{5-15}$$

and

$$\mu(\theta) = \frac{1/\sqrt{2\pi}\int_{u-c}^{u} te^{-t^2/2}\, dt}{1/\sqrt{2\pi}\int_{u-c}^{u} e^{-t^2/2}\, dt} \tag{5-16}$$

we have from (5-4), (5-11) and (5-13)

$$A_{x_g}(\theta) = -a_g\mu(\theta). \tag{5-17}$$

We note that $\mu(\theta)$ is the expectation of t with the distribution function $N(0, 1)$ for the limited range given by

$$u - c < t < u. \tag{5-18}$$

When x_g takes any value, 1 through $(m_g - 1)$, both b_{x_g} and $b_{(x_g+1)}$ are finite and the above range is a function of u with a constant distance between the upper and lower limits.

If x_g is 0, the above range becomes

$$v < t < \infty \tag{5-19}$$

where

$$v = a_g[\theta - b_{(x_g+1)}], \tag{5-20}$$

since b_{x_g} is negative infinity in this case. Then (5-16) can be rewritten as

$$\mu(\theta) = \frac{1/\sqrt{2\pi}\int_{v}^{\infty} te^{-t^2/2}\, dt}{1/\sqrt{2\pi}\int_{v}^{\infty} e^{-t^2/2}\, dt}. \tag{5-21}$$

Similarly, if x_g is m_g, the range of t is given by

$$-\infty < t < u, \tag{5-22}$$

since $b_{(x_g+1)}$ is positive infinity. The equation (5-16) can be rewritten as

$$\mu(\theta) = \frac{1/\sqrt{2\pi}\int_{-\infty}^{u} te^{-t^2/2}\,dt}{1/\sqrt{2\pi}\int_{-\infty}^{u} e^{-t^2/2}\,dt} \tag{5-23}$$

in this case.

Thus it is easily seen that $\mu(\theta)$ is a monotonically increasing function of u for any value of x_g, 1 through m_g, and is a monotonically increasing function of v when $x_g = 0$, and hence it is a monotonically increasing function of θ in any case throughout the range of θ given by

$$-\infty < \theta < \infty, \tag{5-24}$$

as is obvious from the definition of u and v.

Since a_g is a positive constant specified for item g, from (5-17) we can conclude that the basic function of the graded item response on the normal ogive model is monotonically decreasing in θ throughout the whole range given by (5-24), and, therefore, the requirement stated in condition (i) is satisfied.

The numerator of the right-hand side of (5-16) tends to positive and negative infinities as θ tends to positive and negative infinities respectively, while the denominator tends to zero in either case, with respect to any value of x_g, 1 through $(m_g - 1)$. This fact implies that the upper and lower asymptotes of the basic function are positive and negative infinities respectively, for the value of x_g, 1 through $(m_g - 1)$, as is obvious from (5-17).

If $x_g = 0$, the numerator of the right-hand side of (5-21) tends to positive infinity and zero as θ tends to positive and negative infinities respectively, while the denominator tends to zero in the former case and to unity in the latter case. Thus the upper and lower asymptotes of the basic function are zero and negative infinity.

If $x_g = m_g$, the numerator of the right-hand side of (5-23) tends to zero and negative infinity as θ tends to positive and negative infinities respectively, while the denominator tends to unity in the former case and to zero in the latter case. Thus the upper and lower asymptotes of the basic function are positive infinity and zero.

We can conclude, therefore, that

$$C_{x_g,\underline{\theta}} = \infty \tag{5-25}$$

for any x_g, 1 through m_g, and

$$C_{0,\underline{\theta}} = 0, \tag{5-26}$$

and further,

$$C_{x_g,\bar{\theta}} = -\infty \tag{5-27}$$

for any x_g, 0 through $(m_g - 1)$, and

$$C_{m_g,\bar{g}} = 0. \tag{5-28}$$

Condition (ii) is satisfied for any value of x_g except for 0 and m_g, and condition (ii)* holds when $x_g = 0$ and $x_g = m_g$.

It is easily seen from this result that on the normal ogive model, if the items are second dichotomously, condition (ii) does not hold for any individual response, but condition (ii)* is met instead, since in this case x_g is either 0 or $m_g(=1)$.

Thus we have proved that the normal ogive model for graded responses supplies a unique local maximum for likelihood function L_V, defined in Chapter 2, for any possible response pattern except for two extreme cases where either all the elements are 0 or m_g, in which the negative and positive terminal maxima are given respectively. It will give a unique local maximum for the function B_V, defined in Chapter 3, for every response pattern without exception, only if a distribution of θ satisfying the requirements stated in conditions (iii) and (iv) is given.

It is easy for us to deduce this model for graded responses from the assumption of bivariate normal distribution for the latent ability and an item variable [Samejima, 1962a].

For the purpose of illustration, items C_2, $D_{2.2}$ and $A_{2.2}$ of the LIS Measurement Scale for Non-verbal Reasoning Ability [Indow & Samejima, 1962, 1966] are taken here. The examinees were 883 junior high school students in the suburbs of Tokyo in April through July 1960, boys and girls of approximately 13 through 15 years of age.

Item C_2 requires an examinee to find out the principle of classification of simple figures, as shown in Appendix. If an examinee succeeds in discovering the principle, it will be easy for him to respond to 18 test figures correctly, and 392 out of 883 sample examinees succeeded in so doing. There were 67 examinees among the remaining 491 students who responded to 17 figures correctly but failed in one figure, and also 20 who succeeded in 16 figures but failed in two figures. In these cases it may reasonably be assumed that these 87 examinees had also discovered the principle of classification somehow or other, though in a more uncertain way than the 392 successful students, since the probabilities of chance occurrences of these phenomena are negligibly small. Assuming the number of test figures responded to correctly reflects the intensity of cognition of the principle, all the examinees were classified into four groups, that is, 392 of category 3, 67 of category 2, 20 of category 1, and the remaining 404 of category 0.

On the other hand, both items $D_{2.2}$ and $A_{2.2}$ require an examinee to discover the correspondence between alphabetical letters and numbers by following the arithmetical reasoning processes of multiplication and addition, which is also presented in Appendix. In these cases we can follow the examinee's attainment to some extent by tracing his protocol in the following way.

In item $D_{2.2}$, the first clue is to find out that $I = 0$ and then $U = 9$. There were 187 examinees who had discovered these two correspondences, but failed in proceeding further, while 202 students solved the entire problem. Thus all the examinees were classified into three categories in accordance with the degree of attainment, that is, 202 of category 2, 187 of category 1, and the remaining 494 of category 0. In solving item $A_{2.2}$, on the other hand, the first clue is to find out that $I = 0$ and U is an even number. In this case, however, it is more likely that an examinee will start by trial and error, since it is easily seen that I is either 0 or 5, and yet it requires a certain amount of insight to discover directly that it should be 0. Anyway 140 students responded to I and U in the way mentioned above but failed in proceeding further, while 107 examinees solved the problem perfectly. Thus the total 883 students were divided into three categories in this case: 107 of category 2, 140 of category 1, and the remaining 636 of category 0.

The values of b_{x_g} were computed on these numbers in view of the assumption of the homogeneous thinking process and the bivariate normal distribution for the latent ability and each item variable [Samejima, 1962b]. In this case b_{x_g} is the normal deviate corresponding to the proportion of examinees who are scored below x_g, divided by the factor loading of item g on the principal factor which was obtained by the principal factor solution of factor analysis on the tetrachoric correlation matrix of the entire 30 LIS items being scored dichotomously. The values of a_g were also computed from these factor loadings [Indow & Samejima, 1962, 1966]. The resulting values are:

$a_g = 1/0.901$, $b_1 = -0.148$, $b_2 = -0.067$, and $b_3 = 0.188$ for item C_2,

$a_g = 1/0.944$, $b_1 = 0.206$, and $b_2 = 1.018$ for item $D_{2.2}$, and

$a_g = 1/0.863$, $b_1 = 0.766$, and $b_2 = 1.546$ for item $A_{2.2}$, [Samejima, 1962b].

In order to test the goodness of fit of the model, 12 other items were selected from LIS, which had been dichotomously scored and whose test scores were to be used for giving the categorization for the examinee. The parameter values of these 12 items are presented in Table 5–1. The operating characteristic of test score T for these 12 items was computed, which is denoted by $P_T(\theta)$, as was discussed in Chapter 1. The theoretical frequency distribution, $G_{x_g}(T)$, for each graded category was computed by the formula

$$G_{x_g}(T) = N \int_{-\infty}^{\infty} f(\theta) P_{x_g}(\theta) P_T(\theta)\, d\theta, \tag{5-29}$$

where $N = 883$, assuming the normality, $N(0, 1)$, for the latent distribution. Table 5-2 presents these results in comparison with the actual frequencies, together with χ^2 -values computed without joining together any categories with small frequencies. It may be observed that the fit is fairly good for items C_2 and $D_{2.2}$, especially if we compare them with that of total examinees,

TABLE 5-1

Item Parameters of Twelve Dichotomous Items
Selected from LIS Measurement Scale
for Non-verbal Reasoning Ability

Item	$1/a_g$	b_g
$C_{1.1}$	0.77	-1.43
$A_{3.1}$	0.97	-0.86
$D_{1.1}$	0.62	-0.55
$D_{2.1}$	0.59	-0.52
$C_{3.1}$	0.71	-0.33
$A_{3.2}$	0.75	-0.31
$B_{1.1}$	0.67	-0.28
$A_{3.3}$	0.97	0.08
$A_{2.1}$	0.72	0.47
$C_{1.3}$	1.36	1.57
$D_{1.4}$	1.11	2.10
$B_{3.3}$	1.13	2.36

while it is by no means good for item $A_{2.2}$. This may suggest that the reasoning process in solving item $A_{2.2}$ is more heterogeneous than the others. These three items are examples of good and poor fits among the eight items being taken up, and the results obtained for the other five items are presented elsewhere (cf. Samejima, 1968a, Appendix 2).

For the purpose of illustration, the operating characteristics of graded responses for items C_2 and $D_{2.2}$ are shown in Figure 5-1.

We have already observed that the basic function of the graded response on the normal ogive model is a monotonically decreasing function of θ throughout its range with positive and negative infinities as the limit values, except for two extreme categories, 0 and m_g. From the definition of A_{x_g} and (5-4) and (5-17), we can easily see that in an asymptotic case where b_{x_g} and $b_{(x_g+1)}$ tend to negative and positive infinities, we have

(5-30) $$\lim_{\substack{b_{x_g}\to -\infty \\ b_{(x_g+1)}\to\infty}} A_{x_g} = 0.$$

TABLE 5-2

Observed (O(T)) and Theoretical (G(T)) Frequency Distributions of Category T for Each Graded Response Group to Items C_2 , $D_{2.2}$ and $A_{2.2}$

Item C_2

T	$x_g = 3$			$x_g = 2$			$x_g = 1$			$x_g = 0$		
	O(T)	G(T)	$\frac{(O(T)-G(T))^2}{G(T)}$	O(T)	G(T)	$\frac{(O(T)-G(T))^2}{G(T)}$	O(T)	G(T)	$\frac{(O(T)-G(T))^2}{G(T)}$	O(T)	G(T)	$\frac{(O(T)-G(T))^2}{G(T)}$
12	9	7.42	0.33	0	0.11	0.11	0	0.02	0.02	2	0.12	29.87
11	21	28.00	1.75	1	0.91	0.01	0	0.22	0.22	2	1.24	0.46
10	58	63.03	0.40	3	3.78	0.16	1	0.98	0.00	6	6.39	0.02
9	87	84.75	0.06	7	8.20	0.18	4	2.25	1.37	25	17.24	3.50
8	66	66.75	0.01	5	9.98	2.48	4	2.90	0.42	27	26.88	0.00
7	59	47.67	2.69	13	9.76	1.07	3	2.98	0.00	33	33.48	0.01
6	36	33.15	0.25	7	8.67	0.32	1	2.77	1.13	34	37.72	0.37
5	25	23.12	0.15	8	7.39	0.05	1	2.46	0.87	33	40.92	1.53
4	14	16.05	0.26	7	6.11	0.13	2	2.12	0.01	43	43.88	0.02
3	9	10.87	0.32	5	4.88	0.00	2	1.77	0.03	54	47.16	0.99
2	4	6.88	1.20	7	3.64	3.11	0	1.38	1.38	53	51.19	0.06
1	3	3.63	0.11	3	2.28	0.23	2	0.91	1.32	42	53.12	2.33
0	1	1.14	0.02	1	0.86	0.02	0	0.36	0.36	50	43.52	0.97

$\chi^2 = 7.5581$ $\chi^2 = 7.8745$ $\chi^2 = 7.1426$ $\chi^2 = 40.1298$

TABLE 5-2 (Continued)

Item $D_{2.2}$

T	$x_g = 2$			$x_g = 1$			$x_g = 0$		
	O(T)	G(T)	$\frac{(O(T)-G(T))^2}{G(T)}$	O(T)	G(T)	$\frac{(O(T)-G(T))^2}{G(T)}$	O(T)	G(T)	$\frac{(O(T)-G(T))^2}{G(T)}$
12	10	6.51	1.88	1	0.87	0.02	0	0.30	0.30
11	18	22.05	0.74	4	5.65	0.48	2	2.68	0.17
10	40	43.47	0.28	12	18.60	2.34	16	12.11	1.25
9	53	50.25	0.15	33	32.98	0.00	37	29.22	2.07
8	27	32.75	1.01	34	32.77	0.05	41	40.98	0.00
7	26	19.73	1.99	33	27.25	1.21	49	46.92	0.09
6	13	11.75	0.13	18	21.14	0.47	47	49.41	0.12
5	5	7.12	0.63	20	16.05	0.97	42	50.73	1.50
4	6	4.32	0.65	13	11.97	0.09	47	51.87	0.46
3	4	2.56	0.81	9	8.66	0.01	57	53.45	0.24
2	0	1.40	1.40	8	5.85	0.79	56	55.83	0.00
1	0	0.63	0.63	0	3.31	3.31	50	55.99	0.64
0	0	0.17	0.17	2	1.13	0.68	50	44.58	0.66

$\chi^2 = 10.4813$ $\chi^2 = 10.4212$ $\chi^2 = 7.4966$

On the other hand, if we define $\Delta\theta$ so that

$$\Delta\theta = b_{(x_g+1)} - b_{x_g} , \tag{5-31}$$

and could write

$$A_{x_g} = \frac{\{(\partial/\partial t)P^*_{(x_g+1)}(t) \mid_{t=\theta+\Delta\theta}\} - \{(\partial/\partial t)P^*_{(x_g+1)}(t) \mid_{t=\theta}\}}{P^*_{(x_g+1)}(\theta + \Delta\theta) - P^*_{(x_g+1)}(\theta)} \tag{5-32}$$

$$= \frac{1/\Delta\theta[\{(\partial/\partial t)P^*_{(x_g+1)}(t) \mid_{t=\theta+\Delta\theta}\} - \{(\partial/\partial t)P^*_{(x_g+1)}(t) \mid_{t=\theta}\}]}{1/\Delta\theta[P^*_{(x_g+1)}(\theta + \Delta\theta) - P^*_{(x_g+1)}(\theta)]}$$

TABLE 5-2 (Continued)

Item $A_{2.2}$ Total 883 Examinees

T	$x_g = 2$			$x_g = 1$			$x_g = 0$					
	O(T)	G(T)	$\frac{(O(T)-G(T))^2}{G(T)}$	O(T)	G(T)	$\frac{(O(T)-G(T))^2}{G(T)}$	O(T)	G(T)	$\frac{(O(T)-G(T))^2}{G(T)}$	O(T)	G(T)	$\frac{(O(T)-G(T))^2}{G(T)}$
12	9	5.54	2.17	1	1.44	0.14	1	0.70	0.13	11	7.68	1.44
11	15	16.67	0.17	4	8.03	2.02	5	5.68	0.08	24	30.38	1.34
10	24	28.44	0.69	9	22.46	8.07	35	23.27	5.91	68	74.18	0.51
9	36	27.79	2.42	18	33.55	7.20	69	51.11	6.27	123	112.44	0.99
8	14	14.48	0.02	16	27.16	4.58	72	64.86	0.79	102	106.50	0.19
7	5	7.07	0.61	17	18.64	0.14	86	68.18	4.66	108	93.90	2.12
6	0	3.46	3.46	18	12.07	2.91	60	66.78	0.69	78	82.31	0.23
5	2	1.74	0.04	13	7.71	3.63	52	64.45	2.40	67	73.90	0.64
4	1	0.88	0.02	7	4.84	0.96	58	62.45	0.32	66	68.17	0.07
3	0	0.43	0.43	15	2.92	49.97	55	61.32	0.65	70	64.67	0.44
2	1	0.19	3.43	17	1.61	146.81	46	61.28	3.81	64	63.08	0.01
1	0	0.07	0.07	3	0.72	7.20	47	59.14	2.49	50	59.93	1.65
0	0	0.01	0.01	2	0.18	17.92	50	45.68	0.41	52	45.88	0.82
	$\chi^2 = 13.5272$			$\chi^2 = 251.5619$			$\chi^2 = 28.6014$			$\chi^2 = 10.4495$		

and then we obtain the following equation for another asymptotic case where $b_{(x_g+1)}$ tends to b_{x_g} ,

$$\text{(5-33)} \qquad \lim_{b_{(x_g+1)} \to b_{x_g}} A_{x_g}$$

$$= \lim_{\Delta\theta \to 0} A_{x_g}$$

$$= \frac{\lim_{\Delta\theta \to 0} [\{(\partial/\partial t) P^*_{(x_g+1)}(t) \mid_{t=\theta+\Delta\theta}\} - \{(\partial/\partial t) P^*_{(x_g+1)}(t) \mid_{t=\theta}\}]/\Delta\theta}{\lim_{\Delta\theta \to 0} [P^*_{(x_g+1)}(\theta + \Delta\theta) - P^*_{(x_g+1)}(\theta)]/\Delta\theta}$$

$$= \frac{P^{*\prime\prime}_{(x_g+1)}}{P^{*\prime}_{(x_g+1)}}$$

$$= \frac{P^{*\prime\prime}_{x_g}}{P^{*\prime}_{x_g}} ,$$

which gives the basic function in the imaginary situation, where the degree of attainment toward the goal of a given power test item, or intensity of positivity toward a given attitude survey statement, is observed continuously.

On the normal ogive model, we have

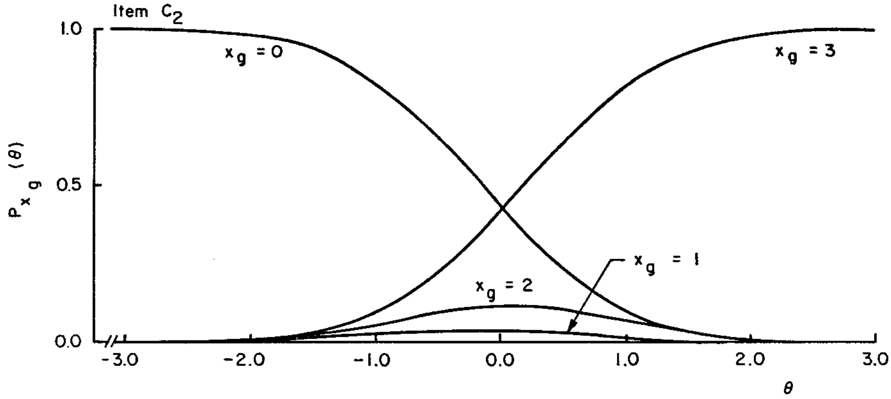
$$\text{(5-34)} \qquad P^{*\prime\prime}_{x_g} = -a_g^2(\theta - b_{x_g}) P^{*\prime}_{x_g} ,$$

and inserting this result into (5-33), we obtain

$$\text{(5-35)} \qquad \lim_{b_{(x_g+1)} \to b_{x_g}} A_{x_g} = -a_g^2(\theta - b_{x_g}),$$

We can easily see that the result is a straight line with $(-a_g^2)$ as the slope and $\theta = b_{x_g}$ as the cutting point of the abscissa.

For the purpose of illustration, Figure 5-2 presents the basic functions on the normal ogive model for five cases in which $a_g = 1.0$, and b_{x_g} and $b_{(x_g+1)}$



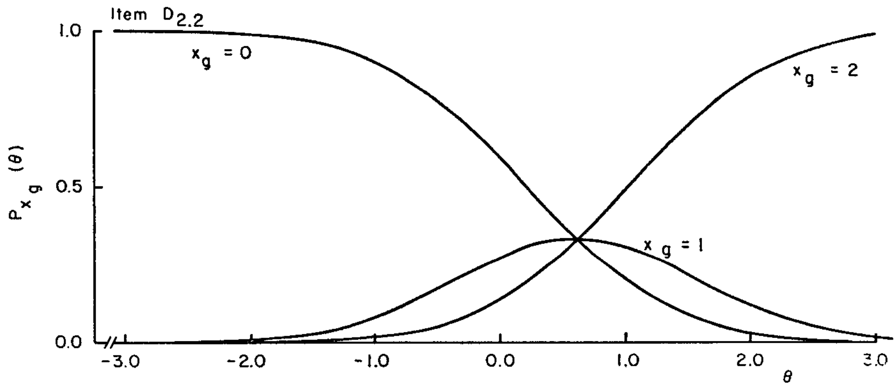


FIGURE 5-1

Operating characteristics of graded responses for items C_2 and $D_{2.2}$ of LIS measurement scale for Non-verbal Reasoning Ability on the normal ogive model.

are: -1.0, 1.0; -2.0, 2.0; -3.0, 3.0; -4.0, 4.0; and -5.0, 5.0 respectively. The basic function of the imaginary case given by equation (5-35) when $a_g = 1$ and $b_{x_g} = 0$ is drawn by a dotted line in the same figure. We can see, for instance, that the basic function of case (5) is practically equal to zero for the range $(-3.0, 3.0)$. These results are connected with the amount of information given by an individual item response, which will be discussed in the following chapter.

On the logistic model, the formula for $P^*_{x_g}$ is given by

$$P^*_{x_g} = \{1 + e^{-Da_g(\theta - b_{x_g})}\}^{-1}, \tag{5-36}$$

where D is a scaling factor satisfying

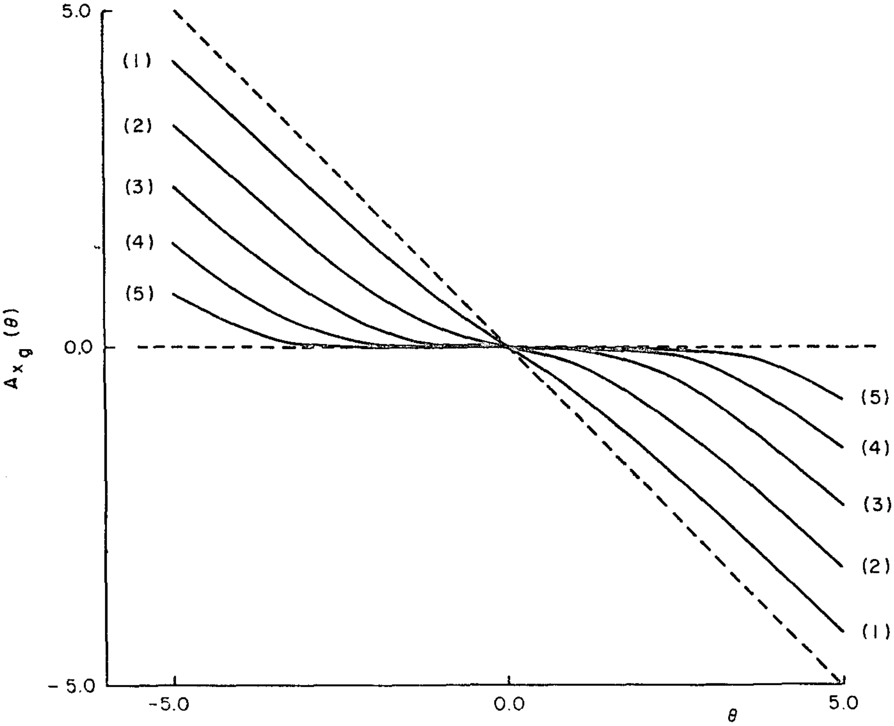


FIGURE 5-2

$A_{x_g}(\theta)$, basic function for response category x_g, on the normal ogive model, for five hypothetical cases in which $a_g = 1.0$; and b_{x_g} and $b_{(x_g+1)}$ are as the following.

Case (1) : $b_{x_g} = -1.0, \quad b_{(x_g+1)} = 1.0$
Case (2) : $b_{x_g} = -2.0, \quad b_{(x_g+1)} = 2.0$
Case (3) : $b_{x_g} = -3.0, \quad b_{(x_g+1)} = 3.0$
Case (4) : $b_{x_g} = -4.0, \quad b_{(x_g+1)} = 4.0$
Case (5) : $b_{x_g} = -5.0, \quad b_{(x_g+1)} = 5.0$

(5-37) $$D > 0,$$

and, just as on the normal ogive model, a_g is an item parameter indicating discriminating power, and b_{x_g} is an item response parameter indicating difficulty specified for category bound x_g. We could also write

(5-38) $$P^*_{(x_g+1)} = \{1 + e^{-Da_g[\theta - b_{(x_g+1)}]}\}^{-1},$$

where

(5-39) $$b_{(x_g+1)} > b_{x_g}.$$

The basic formula for the operating characteristic of graded response on the logistic model is given and developed as the following.

$$(5\text{-}40)\qquad \begin{aligned} P_{x_g}(\theta) &= \{1 + e^{-Da_g(\theta - b_{x_g})}\}^{-1} - \{1 + e^{-Da_g[\theta - b(x_g+1)]}\}^{-1} \\ &= \{1 - e^{-Da_g[b(x_g+1) - b_{x_g}]}\} \\ &\qquad \cdot \{1 + e^{-Da_g(\theta - b_{x_g})}\}^{-1}\{1 + e^{Da_g[\theta - b(x_g+1)]}\}^{-1} \\ &= \{1 - e^{-Da_g[b(x_g+1) - b_{x_g}]}\} \\ &\quad \cdot \{1 + e^{Da_g[-\theta + b_{x_g} + b(x_g+1) - b(x_g+1)]}\}^{-1} \\ &\quad \cdot \{1 + e^{-Da_g[-\theta + b_{x_g} + b(x_g+1) - b_{x_g}]}\}^{-1} \\ &= P_{x_g}[-\theta + b_{x_g} + b_{(x_g+1)}]. \end{aligned}$$

It is apparent from the above result that the operating characteristic of a graded response is symmetric except for two extreme cases, where $x_g = 0$ and $x_g = m_g$. Differentiating this equation with respect to θ, we have

$$(5\text{-}41)\quad P'_{x_g} = Da_g\{1 - e^{-Da_g[b(x_g+1) - b_{x_g}]}\}P^*_{x_g}\{1 - P^*_{(x_g+1)}\}\{1 - P^*_{x_g} - P^*_{(x_g+1)}\},$$

which equals zero only when

$$(5\text{-}42)\qquad \theta = \frac{b_{x_g} + b_{(x_g+1)}}{2}.$$

From these results we can easily see that the operating characteristic of a graded response in the logistic model is again unimodal and symmetric with the modal point given by (5-42) for any x_g except for 0 and m_g, and in these two cases it gives negative and positive terminal maxima respectively, exactly the same result as in the normal ogive model so far.

From (5-40) and (5-41) the basic function in the logistic model will be given by

$$(5\text{-}43)\qquad A_{x_g} = Da_g\{1 - P^*_{x_g} - P^*_{(x_g+1)}\}.$$

Differentiating this equation with respect to θ, we have

$$(5\text{-}44)\qquad \begin{aligned} \frac{\partial}{\partial\theta} A_{x_g} &= -D^2a_g^2[P^*_{x_g}\{1 - P^*_{x_g}\} + P^*_{(x_g+1)}\{1 - P^*_{(x_g+1)}\}] \\ &< 0 \end{aligned}$$

which apparently satisfies condition (i) for any x_g.

From (5-43) we obtain

$$(5\text{-}45)\qquad C_{x_g,\underline{\theta}} = Da_g$$

for any x_g satisfying $x_g \neq 0$, and

$$(5\text{-}46)\qquad C_{x_g,\bar{\theta}} = -Da_g$$

for any x_g satisfying $x_g \neq m_g$, and we have

(5-47) $$C_{0,\underline{\theta}} = 0$$

and

(5-48) $$C_{m_g,\bar{\theta}} = 0$$

for these exceptional cases. Thus we have seen that condition (ii) holds for any graded response other than 0 and m_g , and condition (ii)* holds in these two cases. The values of asymptotes of the basic functions are different from those on the normal ogive model given earlier in this chapter, except for $C_{0,\underline{\theta}}$ and $C_{m_g,\bar{\theta}}$ which are zero in both cases.

It is easily seen from this result that in the logistic model, if the items are scored dichotomously, condition (ii) does not hold for any individual response, but condition (ii)* is satisfied instead, just as was the case on the normal ogive model.

The asymptotic formulas for the basic function on the logistic model will be given by

(5-49) $$\lim_{b_{(x_g+1)} \to b_{x_g}} A_{x_g} = Da_g\{1 - 2P^*_{x_g}\}$$

and

(5-50) $$\lim_{\substack{b_{x_g} \to -\infty \\ b_{(x_g+1)} \to \infty}} A_{x_g} = 0,$$

again the former being quite different from that on the normal ogive model. Although the cutting point of the abscissa is again b_{x_g} , this is not a straight line, having finite values of asymptotes, Da_g and $-Da_g$.

Five examples of the basic functions with the same parameter values as on the normal ogive model are shown elsewhere (cf. Samejima, 1968a, Appendix 3) together with those on the normal ogive model.

CHAPTER 6

AMOUNT OF INFORMATION

Throughout this chapter we shall assume that the operating characteristic of graded response is twice-differentiable with respect to θ for the range $(\underline{\theta}, \bar{\theta})$.

Let $I(\theta)$ denote the amount of information given by a set of n items for the fixed value of θ, or the information function of the test consisting of n items. This function is defined as the expectation of the square of the first derivative of $\log L_V(\theta)$ such that

$$(6\text{-}1) \qquad I(\theta) = E\left\{\frac{\partial \log L_V}{\partial \theta}\right\}^2.$$

In view of the fact that

$$(6\text{-}2) \qquad E\left\{\frac{\partial \log L}{\partial \theta}\right\}^2 = -E\left\{\frac{\partial^2 \log L}{\partial \theta^2}\right\}$$

holds for any likelihood function L when it is the joint frequency function of a sample of independent observations, (6-1) can be rewritten and rearranged in the following way,

$$(6\text{-}3) \qquad \begin{aligned} I(\theta) &= -E\left\{\frac{\partial^2 \log L_V}{\partial \theta^2}\right\} \\ &= -E\left\{\sum_{g=1}^{n} \frac{\partial^2 \log P_{k_g}(\theta)}{\partial \theta^2}\right\} \\ &= \sum_{g=1}^{n}\left[-E\left\{\frac{\partial^2 \log P_{k_g}(\theta)}{\partial \theta^2}\right\}\right] \\ &= \sum_{g=1}^{n} I_g(\theta), \end{aligned}$$

where $I_g(\theta)$ denotes the amount of information given by an individual item, g, for the fixed value of θ, or the information function of item g.

Let $I_{x_g}(\theta)$ be the information function given by a specific graded response to item g for the fixed value of θ, when the item is scored in a graded way. From equation (6-3) we have

$$(6\text{-}4) \qquad \begin{aligned} I_g(\theta) &= -E\left\{\frac{\partial^2 \log P_{x_g}(\theta)}{\partial \theta^2}\right\} \\ &= \sum_{x_g=0}^{m_g}\left\{-\frac{\partial^2 \log P_{x_g}(\theta)}{\partial \theta^2}\right\} P_{x_g}(\theta) \end{aligned}$$

$$= \sum_{x_g=0}^{m_g} I_{x_g}(\theta)P_{x_g}(\theta)$$

$$= E\{I_{x_g}(\theta)\},$$

and finally

$$I_{x_g}(\theta)P_{x_g}(\theta) = \left\{-\frac{\partial}{\partial\theta} A_{x_g}(\theta)\right\}P_{x_g}(\theta) \tag{6-5}$$

$$= \frac{\{P'_{x_g}\}^2}{P_{x_g}} - P''_{x_g}$$

$$= \frac{\{P^{*\prime}_{x_g} - P^{*\prime}_{(x_g+1)}\}^2}{\{P^*_{x_g} - P^*_{(x_g+1)}\}} - \{P^{*\prime\prime}_{x_g} - P^{*\prime\prime}_{(x_g+1)}\}.$$

Equation (6-5) supplies the amount of information shared by an individual graded response for any specified model for $P^*_{x_g}$, provided that it is twice differentiable with respect to θ. This amount of information share is the information function of an individual graded response, which is the negative of the first derivative of the basic function, multiplied by the operating characteristic as shown in equation (6-5). It is easily understood from this fact that both on the normal ogive and the logistic models the information share of the individual item response, x_g, should be symmetrical with $\theta = b_{x_g} + b_{(x_g+1)}/2$ as the axis of symmetry, except for the case where $x_g = 0$ or $x_g = m_g$, since A_{x_g} equals zero at that value of θ and is symmetrical with that value as the center of symmetry, and P_{x_g} is symmetrical with that value as the axis of symmetry.

For the purpose of illustration, Figure 6-1 presents the amount of information shares of the individual item responses on the normal ogive model,

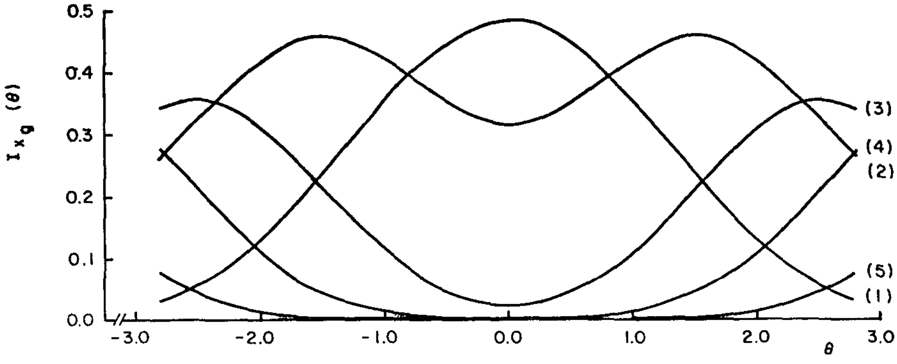


FIGURE 6–1

Amount of information shared by each of the same hypothetical item responses as used in Figure 5–2, on the normal ogive model.

using the same hypothetical data whose basic functions were presented in the previous chapter as Figure 5-2. In these instances the axis of symmetry is $\theta = 0$. This figure gives us an idea of optimal distance between b_{x_g} and $b_{(x_g+1)}$, for increasing the amount of information share for a specified range of θ. If, for instance, we intend to get a large amount of information share for a fairly narrow range of θ, the distance between b_{x_g} and $b_{(x_g+1)}$ should not be so large, while it should be much larger in case we want to get a moderate share of information for a wider range of θ, as is suggested by cases (1) and (2) in Figure 6-1. Cases (3), (4), and (5) may be taken as examples where the distances are too large to give enough information share around $\theta = b_{x_g} + b_{(x_g+1)}/2$.

Inserting (6-5) into (6-4) and rearranging, we obtain

$$
\text{(6-6)} \qquad
\begin{aligned}
I_g(\theta) &= \sum_{x_g=0}^{m_g} \frac{\{P'_{x_g}\}^2}{P_{x_g}} - \sum_{x_g=0}^{m_g} P''_{x_g} \\
&= \sum_{x_g=0}^{m_g} \frac{\{P'_{x_g}\}^2}{P_{x_g}} \\
&= \sum_{x_g=0}^{m_g} \frac{\{P^{*\prime}_{x_g} - P^{*\prime}_{(x_g+1)}\}^2}{\{P^*_{x_g} - P^*_{(x_g+1)}\}}
\end{aligned}
$$

as the information function of the graded item, since we have

$$
\text{(6-7)} \qquad
\begin{aligned}
\sum_{x_g=0}^{m_g} P''_{x_g} &= \sum_{x_g=0}^{m_g} \{P^{*\prime\prime}_{x_g} - P^{*\prime\prime}_{(x_g+1)}\} \\
&= \sum_{x_g=0}^{m_g} P^{*\prime\prime}_{x_g} - \sum_{x_g=1}^{m_g+1} P^{*\prime\prime}_{x_g} \\
&= 0.
\end{aligned}
$$

From (6-3) and (6-6) we have for the information function of a test consisting of n graded items

$$
\text{(6-8)} \qquad
\begin{aligned}
I(\theta) &= \sum_{g=1}^{n} \sum_{x_g=0}^{m_g} \frac{\{P'_{x_g}\}^2}{P_{x_g}} \\
&= \sum_{g=1}^{n} \sum_{x_g=0}^{m_g} \frac{\{P^{*\prime}_{x_g} - P^{*\prime}_{(x_g+1)}\}^2}{\{P^*_{x_g} - P^*_{(x_g+1)}\}}.
\end{aligned}
$$

In the case where items are dichotomously scored, equation (6-6) is rewritten as

$$
\text{(6-9)} \qquad
\begin{aligned}
I_g(\theta) &= \frac{\{P'_g(\theta)\}^2}{P_g(\theta)} + \frac{\{-P'_g(\theta)\}^2}{Q_g(\theta)} \\
&= \frac{\{P'_g(\theta)\}^2}{P_g(\theta)Q_g(\theta)}.
\end{aligned}
$$

This is the information function of a dichotomous item, which was named the item information function by Birnbaum. He has made an important use of this item information function and the information function of the test defined in the same connection as (6-3) with respect to the estimation of ability, devoting many sections to this purpose (in Lord & Novick, 1968).

Figure 6-2 illustrates the amounts of information given by items C_2 and $D_{2.2}$ of the LIS Measurement Scale for Non-verbal Reasoning Ability when they are scored in the graded ways described in the previous chapter, as well as dichotomously, on the normal ogive model. In the same figure the shares of information given by the individual response categories are indicated by dotted lines.

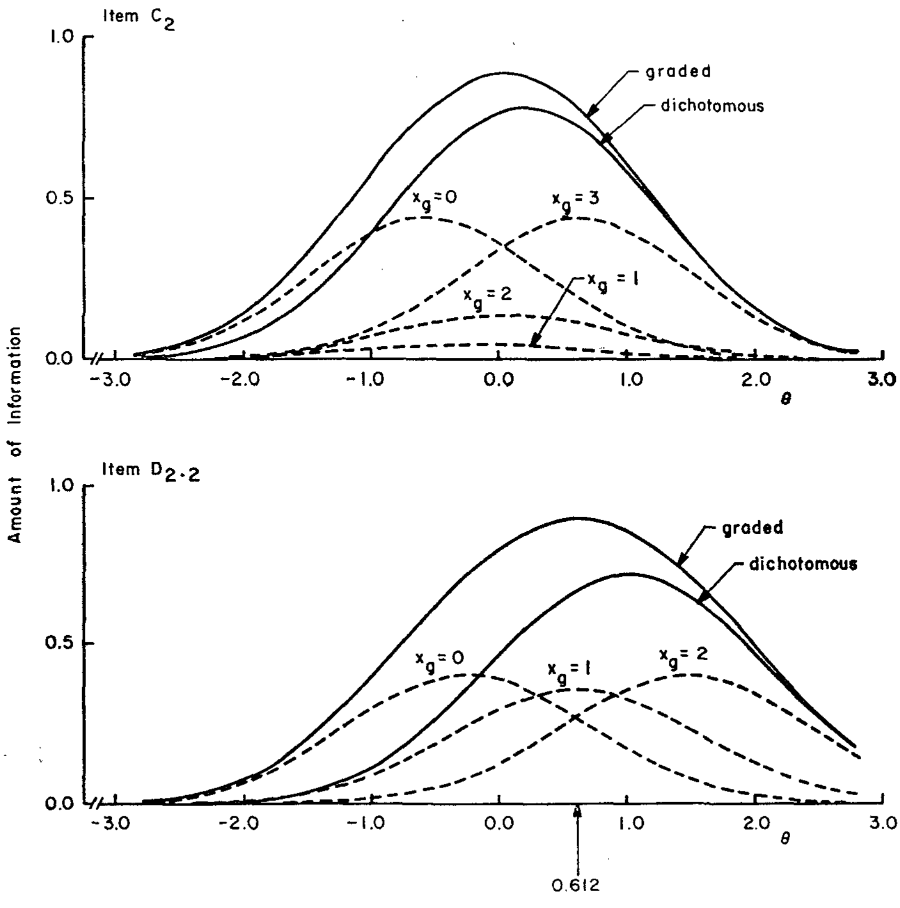


FIGURE 6-2

Amount of information given by each of items C_2 and $D_{2.2}$, when they are scored in graded ways as well as dichotomously (solid line), and share of information given by each individual response category (dotted line).

We can see that the increment in the amount of information by being scored in a more graded way is greater in the case of item $D_{2.2}$, although the response categories are more in the case of C_2. This is again connected with the distance between difficulty indices, as we have seen earlier in this chapter. In fact the distance between b_1 and b_3 in item C_2 is only 0.336, while that between b_1 and b_2 in item $D_{2.2}$ is 0.812.

The result concerning item C_2 might give an impression that it is of no use for us to set up many categories unless we can assign a considerable amount of distance between adjacent difficulty indices. Although the amount of information added may be small, however, this is not true. We shall prove, therefore, that more information will be given by an item if a response category is added between two adjacent categories.

Suppose that a given item, g, has $(m_g + 1)$ response categories, 0 through m_g. Let r denote a response category, x_g, which is less than m_g, and let $(r + 2)$ be its adjacent response, $(x_g + 1)$. Equation (6-6) indicates that the amount of contribution of response category r, which is to be denoted as J_r, to the total information given by item g is

$$J_r = \frac{\{P_r^{*\prime} - P_{(r+2)}^{*\prime}\}^2}{\{P_r^* - P_{(r+2)}^*\}}. \tag{6-10}$$

Suppose, further, that we divide this category into two and set up one more category, $(r + 1)$. The operating characteristic of response r should necessarily be changed by the establishment of this new response category and consequently the amount of contribution of response r toward the entire item information will be reduced, while that of any other original response category remains unchanged. Let J_r^* be the new amount of contribution given by r, and $J_{(r+1)}^*$ denote the amount of contribution given by the additional response category, $(r + 1)$. These amounts of contribution will be given by

$$J_r^* = \frac{\{P_r^{*\prime} - P_{(r+1)}^{*\prime}\}^2}{\{P_r^* - P_{(r+1)}^*\}} \tag{6-11}$$

and

$$J_{(r+1)}^* = \frac{\{P_{(r+1)}^{*\prime} - P_{(r+2)}^{*\prime}\}^2}{\{P_{(r+1)}^* - P_{(r+2)}^*\}}, \tag{6-12}$$

where $P_{(r+1)}^*$ is a function of θ defined in Chapter 4 and specified for response category $(r + 1)$, and $P_{(r+1)}^{*\prime}$ is its first derivative.

Here we shall define S_i, a positive real number, so that

$$S_i > 0, \tag{6-13}$$

and also s_i, any real number, for

$$i = 1, 2, 3, \cdots, u. \tag{6-14}$$

By Schwarz's inequality, we can write

$$\left\{\sum_{i=1}^{u} S_i\right\}\left[\sum_{i=1}^{u} \frac{s_i^2}{S_i}\right] \geq \left\{\sum_{i=1}^{u} s_i\right\}^2, \tag{6-15}$$

and, dividing both sides of inequality (6-15) by $\{\sum_{i=1}^{u} S_i\}$, we have

$$\sum_{i=1}^{u} \frac{s_i^2}{S_i} \geq \frac{\left\{\sum_{i=1}^{u} s_i\right\}^2}{\sum_{i=1}^{u} S_i}. \tag{6-16}$$

Since we could rewrite (6-10) in the following way

$$J_r = \frac{[\{P_r^{*\prime} - P_{(r+1)}^{*\prime}\} + \{P_{(r+1)}^{*\prime} - P_{(r+2)}^{*\prime}\}]^2}{[\{P_r^* - P_{(r+1)}^*\} + \{P_{(r+1)}^* - P_{(r+2)}^*\}]}, \tag{6-17}$$

by setting

$$s_1 = P_r^{*\prime} - P_{(r+1)}^{*\prime}, \tag{6-18}$$

$$s_2 = P_{(r+1)}^{*\prime} - P_{(r+2)}^{*\prime}, \tag{6-19}$$

$$\begin{aligned} S_1 &= P_r^* - P_{(r+1)}^* \\ &> 0, \end{aligned} \tag{6-20}$$

$$\begin{aligned} S_2 &= P_{(r+1)}^* - P_{(r+2)}^* \\ &> 0 \end{aligned} \tag{6-21}$$

and

$$u = 2, \tag{6-22}$$

and inserting these equations into (6-11), (6-12), and (6-17), we obtain from (6-16)

$$\begin{aligned} J_r^* + J_{(r+1)}^* &= \frac{s_1^2}{S_1} + \frac{s_2^2}{S_2} \\ &\geq \frac{\{s_1 + s_2\}^2}{S_1 + S_2} \\ &= J_r . \end{aligned} \tag{6-23}$$

Thus it has been proved that the amount of contribution resulting by adding a new response category $(r + 1)$ is no less than the original one. Since in general cases s_1/S_1 does not coincide with s_2/S_2 for all the values of θ and an equality in (6-23) does not always hold, the amount of information increases in the total by setting up a new response category. More information will be obtained, therefore, if more response categories are set up.

Figure 6-3 presents examples of information functions obtained on the normal ogive model, of a test consisting of six items with the same discriminating powers, in each of the three cases where $1/a_g$ is 0.484, 1.020, and 1.732, respectively. In each hypothetical test $b_g = -1.500$ for a pair of items, $b_g = 0.750$ for another pair of items, and $b_g = 3.000$ for the remaining two items, when they are scored dichotomously. The resulting information function is the lowest curve in each figure of Figure 6-3. The curve in the middle of each figure represents the information function obtained when one more grade is added, by using the above values of b_g as b_2 for each pair of items, and setting $b_1 = -2.625$, $b_1 = -0.375$, and $b_1 = 1.875$, respectively. The curve in the highest vertical position in each figure is the information function obtained when each item is scored with four response categories, by using the above values of b_g as b_3 and setting $b_1 = -3.000$ and $b_2 = -2.250$ for the first pair of items, $b_1 = -0.750$ and $b_2 = 0.000$ for the second pair of items, and $b_1 = 1.500$ and $b_2 = 2.250$ for the last pair of items. We can see that in each of the three cases the increment in the amount of information is especially great when one more response category is added to each of the six dichotomous items. Especially in the case of items with high discriminating powers the effect of the so-called attenuation paradox disappears as we can see in the upper figure of Figure 6-3. That is to say, the amount of information given by the six hypothetical items is very small around the values of θ near $(-1.500 + 0.750)/2$ and $(0.750 + 3.000)/2$ in a relative sense when they are scored dichotomously, while such a conspicuous tendency does not appear in either of the other two cases in which the items are scored in the more graded ways.

Now we shall proceed to prove that if some aggregation of response patterns is used instead of the response pattern itself, the amount of information given by the test may be reduced. The simple test score and some kind of weighted test score are typical examples of such aggregations. Let T denote an aggregation of response patterns, and $P_T(\theta)$ its operating characteristic, as we have done in Chapter 1. We could write

$$P_T(\theta) = \sum_{V \in T} P_V(\theta), \tag{6-24}$$

where $P_V(\theta)$ is the operating characteristic of response pattern V. Since we have

$$L_V(\theta) = P_V(\theta), \tag{6-25}$$

we obtain from the definition of the information function of a test given by (6-1),

$$I(\theta) = E\left\{\frac{\partial \log P_V(\theta)}{\partial \theta}\right\}^2 \tag{6-26}$$

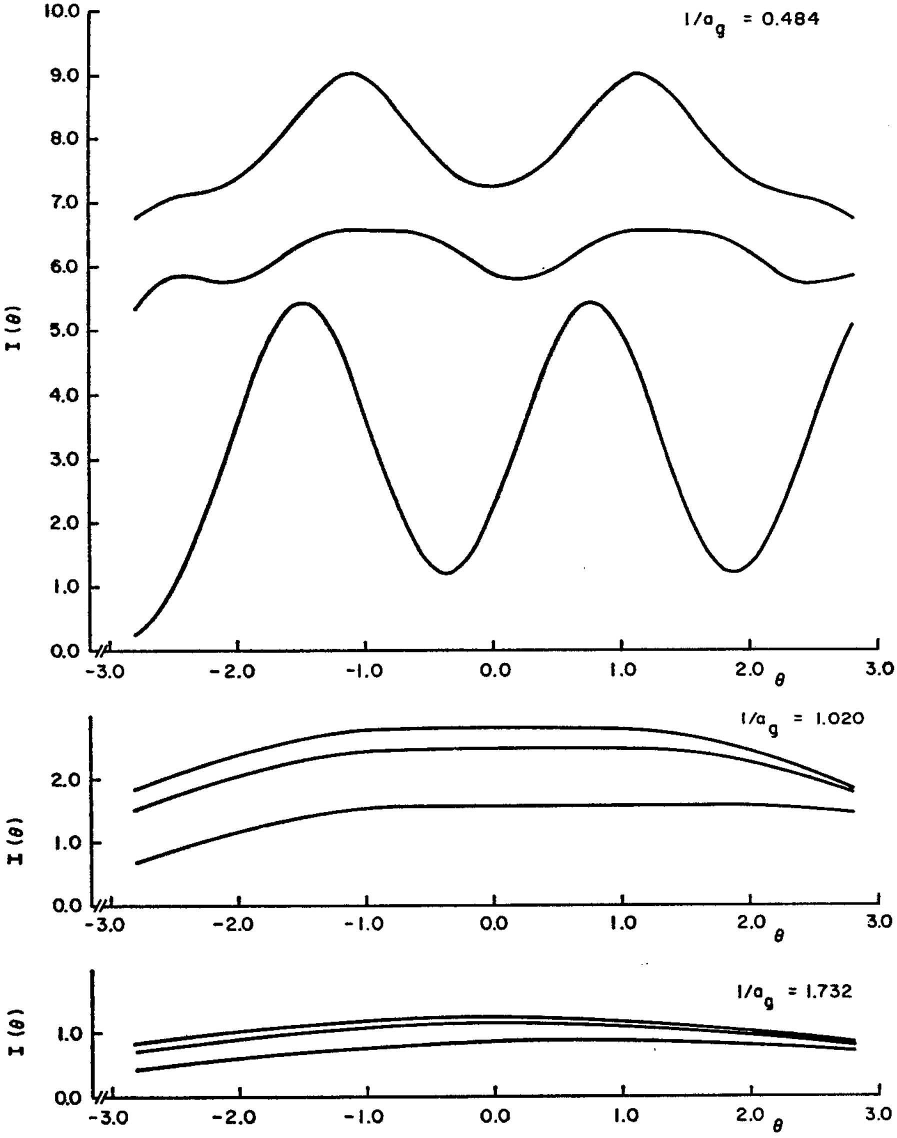


FIGURE 6-3

Information functions given by six hypothetical items, when they are scored dichotomously, with three graded response categories, and with four graded response categories.

$$= \sum_{V} \frac{\{(\partial/\partial\theta)P_V(\theta)\}^2}{P_V(\theta)}$$

$$= \sum_{T} \sum_{V \in T} \frac{\{(\partial/\partial\theta)P_V(\theta)\}^2}{P_V(\theta)}.$$

If we define another likelihood function, $L_T(\theta)$, based on aggregation T instead of V, so that

(6-27) $$L_T(\theta) = P_T(\theta),$$

and denote the resulting information function as $I^*(\theta)$, we obtain from equation (6-24) and the definition of the information function

(6-28) $$I^*(\theta) = E\left\{\frac{\partial \log P_T(\theta)}{\partial\theta}\right\}^2$$

$$= \sum_{T} \frac{\{(\partial/\partial\theta)P_T(\theta)\}^2}{P_T(\theta)}$$

$$= \sum_{T} \frac{\{\sum_{V \in T} (\partial/\partial\theta)P_V(\theta)\}^2}{\sum_{V \in T} P_V(\theta)}.$$

Setting

(6-29) $$S_i = P_V(\theta) > 0,$$

and

(6-30) $$s_i = \frac{\partial}{\partial\theta} P_V(\theta),$$

and defining u as the number of response patterns contained by T, we have from (6-16) the following inequality

(6-31) $$\sum_{V \in T} \frac{\{(\partial/\partial\theta)P_V(\theta)\}^2}{P_V(\theta)} \geq \frac{\{\sum_{V \in T} (\partial/\partial\theta)P_V(\theta)\}^2}{\sum_{V \in T} P_V(\theta)}.$$

Then from this inequality and (6-26) and (6-28), we obtain

(6-32) $$I(\theta) \geq I^*(\theta).$$

Thus it has been proved that the usage of aggregation T instead of response pattern V may reduce the amount of information given by the total test. Since the response pattern whose elements are less graded item responses are nothing but an aggregation of response patterns whose elements are more graded, this proof includes the former one.

When all the items are dichotomously scored and equivalent with one another and the simple test score is used as aggregation T, we obtain from (6-28)

$$
\begin{aligned}
(6\text{-}33) \qquad I^*(\theta) &= \sum_T \frac{\{n_T(\partial/\partial\theta)P_{V_T}(\theta)\}^2}{n_T P_{V_T}(\theta)} \\
&= \sum_T n_T \frac{\{(\partial/\partial\theta)P_{V_T}(\theta)\}^2}{P_{V_T}(\theta)} \\
&= \sum_V \frac{\{(\partial/\partial\theta)P_V(\theta)\}^2}{P_V(\theta)} \\
&= I(\theta),
\end{aligned}
$$

where n_T is the number of V's contained by T and $P_{V_T}(\theta)$ is the common operating characteristic of the response pattern contained by T. Thus in this case there is no reduction in the amount of information given by the test, even if we use test score T instead of response pattern V.

CHAPTER 7

BAYES ESTIMATOR BASED ON THE RESPONSE PATTERN

When the latent density function, denoted by $f(\theta)$,is known or reasonably assumed, it is possible for us to obtain an estimator of θ with which the expectation of its mean-square error about the true value of θ is minimized. If we temporarily denote this estimator defined on the response pattern, V, as θ_V^* , its mean-square error is given by

$$(7\text{-}1) \qquad E\{\theta_V^* - \theta\}^2 = \sum_v \{\theta_V^* - \theta\}^2 P_V(\theta),$$

where $P_V(\theta)$ is the operating characteristic of response pattern V. Then the expectation of this mean-square error over the whole range of θ can be expressed as

$$(7\text{-}2) \qquad \int_{-\infty}^{\infty} E\{\theta_V^* - \theta\}^2 f(\theta)\, d\theta = \int_{-\infty}^{\infty} \sum_V \{\theta_V^* - \theta\}^2 P_V(\theta) f(\theta)\, d\theta$$

$$= \sum_V \int_{-\infty}^{\infty} \{\theta_V^* - \theta\}^2 \psi(V, \theta)\, d\theta,$$

where $\psi(V, \theta)$ is the bivariate density function of the response pattern V and latent variate θ.

Let V_0 denote a specified response pattern and $\theta_{V_0}^*$ be such an estimator on this response pattern. Differentiating (7-2) with respect to $\theta_{V_0}^*$, we have

$$(7\text{-}3) \qquad \frac{\partial}{\partial \theta_{V_0}^*} \int_{-\infty}^{\infty} E\{\theta_V^* - \theta\}^2 f(\theta)\, d\theta$$

$$= 2 \int_{-\infty}^{\infty} \{\theta_{V_0}^* - \theta\} \psi(V_0 , \theta)\, d\theta$$

$$= 2\left[\theta_{V_0}^* \int_{-\infty}^{\infty} \psi(V_0 , \theta)\, d\theta - \int_{-\infty}^{\infty} \theta \psi(V_0 , \theta)\, d\theta \right].$$

We obtain by setting the above equation equal to zero and rearranging

$$(7\text{-}4) \qquad \theta_{V_0}^* = \frac{\int_{-\infty}^{\infty} \theta \psi(V_0 , \theta)\, d\theta}{P(V_0)} ,$$

where $P(V)$ is the probability of response pattern V given by

$$(7\text{-}5) \qquad P(V) = \int_{-\infty}^{\infty} \psi(V, \theta)\, d\theta.$$

Equation (7-4) is nothing but the expectation of θ defined on the specified response pattern V_0, or the first-order moment of θ about the origin for the latent density function specified for response pattern V_0.

Thus we have seen that the expectation or first-order moment about the origin of θ is the estimator with which the expectation of the mean-square error over the whole range of θ is minimized. Since this is a Bayes estimator obtained by using the mean-square error multiplied by the latent density function of θ as the loss function, hereafter we shall call this estimator simply the Bayes estimator and denote it as μ'_{1V}, which is given by the formula

$$\mu'_{1V} = \frac{\int_{-\infty}^{\infty} \theta\psi(V, \theta)\, d\theta}{P(V)}. \tag{7-6}$$

In view of the fact that the minimization of the mean-square error about the true value of θ is the most desirable property for an estimator when $f(\theta)$ is known or reasonably assumed, it will be preferable for us to take μ'_{1V} as our estimator whenever the necessary computation is available.

Another big advantage of using μ'_{1V} as our estimator is that it is unnecessary for us to restrict the functional form of the operating characteristic of item response and also that of the latent density function as we have to do in order to obtain the maximum likelihood estimator $\hat{\theta}$ or the Bayes modal estimator $\hat{\hat{\theta}}$, since neither $P_V(\theta)$ nor $f(\theta)P_V(\theta)$ need necessarily be unimodal in this case. The operating characteristic of item response can be monotonically increasing or decreasing, unimodal or multimodal, or of any shape, while the latent density function can be rectangular, multimodal, or anything else.†

Let μ_{2V} denote the variance of θ defined on the density function for a given response pattern, V, or the second-order moment of θ about μ'_{1V}. The square root of μ_{2V} may be taken as the standard error of measurement of μ'_{1V}. We can write

$$\mu_{2V} = \frac{\int_{-\infty}^{\infty} \{\theta - \mu'_{1V}\}^2 \psi(V, \theta)\, d\theta}{P(V)} \tag{7-7}$$

$$= \mu'_{2V} - \{\mu'_{1V}\}^2,$$

where μ'_{2V} is the second-order moment about the origin, defined by

$$\mu'_{2V} = \frac{\int_{-\infty}^{\infty} \theta^2 \psi(V, \theta)\, d\theta}{P(V)}. \tag{7-8}$$

We shall prove that in general cases the expectation of the standard errors of measurement is smaller when items are scored in a more graded way, and also that the expectation of the standard errors of measurement obtained

† As we have discussed in Chapter 3, conditions (i) through (iv) may hold when θ is transformed to τ, even if they do not meet without transformation.

on the density function for response pattern is no more than the standard error of measurement obtained on the density function for some kind of aggregation of response patterns.

Let μ'_{1T} denote the estimator obtained on the density function for T, some kind of aggregation of response patterns, as distinct from μ'_{1V}. In proving the above two propositions, it will be enough for us to show that the expectation of the standard errors of measurement with respect to μ'_{1V} over all the response patterns contained by T is no more than the standard error of measurement with respect to μ'_{1T}, since a response pattern defined on less graded items is nothing but an aggregation of response patterns possibly obtained when the same items are scored in a more graded way, as we have already observed in the previous chapter.

Let $\mu'_{rV}(\alpha)$ denote the r-th moment of θ about an arbitrary value α defined on the density function for V, and $\mu'_{rT}(\alpha)$ be that on the density function for T. We could write

$$\text{(7-9)} \qquad \mu'_{rT}(\alpha) = \frac{\int_{-\infty}^{\infty} \{\theta - \alpha\}^r P_T(\theta) f(\theta)\, d\theta}{\int_{-\infty}^{\infty} P_T(\theta) f(\theta)\, d\theta}$$

$$= \frac{\sum_{V \in T} \int_{-\infty}^{\infty} \{\theta - \alpha\}^r \psi(V, \theta)\, d\theta}{\sum_{V \in T} \int_{-\infty}^{\infty} \psi(V, \theta)\, d\theta}$$

$$= \frac{\sum_{V \in T} \mu'_{rV}(\alpha) P(V)}{\sum_{V \in T} P(V)},$$

since we have

$$\text{(7-10)} \qquad P_T(\theta) = \sum_{V \in T} P_V(\theta).$$

Thus $\mu'_{rT}(\alpha)$, the moment of order r about an arbitrary value α on T, is rewritten as the expected value of $\mu'_{rV}(\alpha)$ over the response patterns contained by T, no matter the value of r. From (7-9) we have

$$\text{(7-11)} \qquad \mu'_{1T} = \frac{\sum_{V \in T} \mu'_{1V} P(V)}{\sum_{V \in T} P(V)}$$

and

$$\text{(7-12)} \qquad \mu'_{2T} = \frac{\sum_{V \in T} \mu'_{2V} P(V)}{\sum_{V \in T} P(V)}.$$

These results make it possible to deduce the following inequality by the aid of Schwarz's inequality.

$$
\begin{aligned}
\text{(7-13)} \qquad \mu_{2T} &= \mu'_{2T} - \{\mu'_{1T}\}^2 \\
&= \frac{\sum\limits_{V \in T} \mu'_{2V} P(V)}{\sum\limits_{V \in T} P(V)} - \left[\frac{\sum\limits_{V \in T} \mu'_{1V} P(V)}{\sum\limits_{V \in T} P(V)} \right]^2 \\
&\geq \frac{\sum\limits_{V \in T} \mu'_{2V} P(V)}{\sum\limits_{V \in T} P(V)} - \frac{\sum\limits_{V \in T} \{\mu'_{1V}\}^2 P(V)}{\sum\limits_{V \in T} P(V)} \\
&= \frac{\sum\limits_{V \in T} \mu_{2V} P(V)}{\sum\limits_{V \in T} P(V)} \\
&\geq \left[\frac{\sum\limits_{V \in T} \{\mu_{2V}\}^{1/2} P(V)}{\sum\limits_{V \in T} P(V)} \right]^2 .
\end{aligned}
$$

Also from (7-13) we have

$$
\text{(7-14)} \qquad \{\mu_{2T}\}^{1/2} \geq \frac{\sum\limits_{V \in T} \{\mu_{2V}\}^{1/2} P(V)}{\sum\limits_{V \in T} P(V)} .
$$

Thus it has been proved that the variance of θ on the density function for T is no less than the expectation of the variances of θ on the density functions for the response patterns contained by T, and also the same is true with the standard error of measurement.

Inequality (7-13) suggests that the variance with respect to the density function for T coincides with the expectation of variances with respect to the density function for V if, and only if, the values of μ'_{1V} are constant for all the V's pertaining to T. This holds in the case where all the items are dichotomously scored and their item characteristic functions are equivalent, and the simple test score is used as T, while it does not hold in general cases. Especially when T is taken as a response pattern and V is also a response pattern obtained by rescoring items in a more graded way, it never happens that all the values of μ'_{1V} for V's contained by T are constant, if more than one V is contained. Thus we can see that in the more graded way items are scored, the less will become the expectation of the variances, and hence the same is true for the expectation of the standard errors of measurement. Inequality (7-13) also suggests that the standard error of measurement on the density function for T agrees with the expectation of standard errors of measurement with respect to the density functions for V if, and only if, the values of μ'_{2V} are constant for all the V's contained by T, in addition to the requirement stated above.

TABLE 7-1

The Bayes Estimate, μ'_{1V} , and Its Standard Error of Measurement, $[\mu_{2V}]^{\frac{1}{2}}$, with Respect to Each of the 128 Possible Response Patterns on LIS-U

Response Pattern	μ'_{1V}	$[\mu_{2V}]^{\frac{1}{2}}$	P(V)	T
RESP(1) = 1 1 1 1 1 1 1	1.47968	0.58765	0.11683	7
RESP(2) = 1 1 1 1 1 1 0	0.84917	0.45724	0.04817	6
RESP(3) = 1 1 1 1 1 0 1	0.82493	0.44003	0.02420	6
RESP(4) = 1 1 1 1 0 1 1	0.91608	0.45310	0.01738	6
RESP(5) = 1 1 1 0 1 1 1	1.05266	0.48513	0.02023	6
RESP(6) = 1 1 0 1 1 1 1	0.91975	0.45043	0.00912	6
RESP(7) = 1 0 1 1 1 1 1	0.52068	0.36164	0.00074	6
RESP(8) = 0 1 1 1 1 1 1	0.83466	0.43031	0.00224	6
RESP(9) = 1 1 1 1 1 0 0	0.40369	0.42151	0.03394	5
RESP(10) = 1 1 1 1 0 1 0	0.48682	0.41905	0.02009	5
RESP(11) = 1 1 1 0 1 1 0	0.58226	0.42854	0.01809	5
RESP(12) = 1 1 0 1 1 1 0	0.49797	0.41459	0.01037	5
RESP(13) = 1 0 1 1 1 1 0	0.21481	0.36393	0.00181	5
RESP(14) = 0 1 1 1 1 1 0	0.44375	0.40195	0.00297	5
RESP(15) = 1 1 1 1 0 0 1	0.48685	0.40761	0.01032	5
RESP(16) = 1 1 1 0 1 0 1	0.57651	0.41575	0.00927	5
RESP(17) = 1 1 0 1 1 0 1	0.49736	0.40342	0.00533	5
RESP(18) = 1 0 1 1 1 0 1	0.22654	0.35682	0.00092	5
RESP(19) = 0 1 1 1 1 0 1	0.44595	0.39195	0.00153	5
RESP(20) = 1 1 1 0 0 1 1	0.65479	0.41958	0.00594	5
RESP(21) = 1 1 0 1 0 1 1	0.57135	0.40438	0.00329	5
RESP(22) = 1 0 1 1 0 1 1	0.28947	0.35129	0.00050	5
RESP(23) = 0 1 1 1 0 1 1	0.51632	0.39181	0.00092	5
RESP(24) = 1 1 0 0 1 1 1	0.66200	0.41611	0.00310	5
RESP(25) = 1 0 1 0 1 1 1	0.35283	0.35299	0.00040	5
RESP(26) = 0 1 1 0 1 1 1	0.59970	0.40142	0.00084	5
RESP(27) = 1 0 0 1 1 1 1	0.30094	0.34781	0.00026	5
RESP(28) = 0 1 0 1 1 1 1	0.52517	0.38837	0.00048	5
RESP(29) = 0 0 1 1 1 1 1	0.26909	0.34103	0.00008	5
RESP(30) = 1 1 1 1 0 0 0	0.08684	0.41351	0.03074	4
RESP(31) = 1 1 1 0 1 0 0	0.17651	0.41254	0.02241	4
RESP(32) = 1 1 0 1 1 0 0	0.10945	0.40683	0.01526	4
RESP(33) = 1 0 1 1 1 0 0	-0.11502	0.38108	0.00478	4
RESP(34) = 0 1 1 1 1 0 0	0.07554	0.39669	0.00484	4
RESP(35) = 1 1 1 0 0 1 0	0.26374	0.40527	0.01187	4
RESP(36) = 1 1 0 1 0 1 0	0.19729	0.39788	0.00780	4
RESP(37) = 1 0 1 1 0 1 0	-0.02653	0.36644	0.00215	4
RESP(38) = 0 1 1 1 0 1 0	0.16059	0.38776	0.00243	4
RESP(39) = 1 1 0 0 1 1 0	0.27982	0.40027	0.00602	4
RESP(40) = 1 0 1 0 1 1 0	0.04409	0.36282	0.00147	4
RESP(41) = 0 1 1 0 1 1 0	0.23883	0.38943	0.00184	4
RESP(42) = 1 0 0 1 1 1 0	-0.00567	0.36086	0.00105	4
RESP(43) = 0 1 0 1 1 1 0	0.17808	0.38286	0.00122	4
RESP(44) = 0 0 1 1 1 1 0	-0.02802	0.35415	0.00034	4
RESP(45) = 1 1 1 0 0 0 1	0.27503	0.39575	0.00606	4
RESP(46) = 1 1 0 1 0 0 1	0.21146	0.38902	0.00396	4
RESP(47) = 1 0 1 1 0 0 1	-0.00527	0.35938	0.00107	4
RESP(48) = 0 1 1 1 0 0 1	0.17576	0.37965	0.00123	4
RESP(49) = 1 1 0 0 1 0 1	0.29006	0.39104	0.00308	4
RESP(50) = 1 0 1 0 1 0 1	0.06230	0.35601	0.00074	4
RESP(51) = 0 1 1 0 1 0 1	0.25047	0.38104	0.00094	4
RESP(52) = 1 0 0 1 1 0 1	0.01421	0.35414	0.00052	4
RESP(53) = 0 1 0 1 1 0 1	0.19212	0.37502	0.00062	4
RESP(54) = 0 0 1 1 1 0 1	-0.00806	0.34783	0.00017	4
RESP(55) = 1 1 0 0 0 1 1	0.36403	0.38775	0.00172	4
RESP(56) = 1 0 1 0 0 1 1	0.13111	0.34780	0.00037	4
RESP(57) = 0 1 1 0 0 1 1	0.32180	0.37728	0.00052	4
RESP(58) = 1 0 0 1 0 1 1	0.08441	0.34510	0.00025	4
RESP(59) = 0 1 0 1 0 1 1	0.26316	0.37001	0.00033	4
RESP(60) = 0 0 1 1 0 1 1	0.06064	0.33898	0.00008	4
RESP(61) = 1 0 0 0 1 1 1	0.14587	0.34396	0.00018	4
RESP(62) = 0 1 0 0 1 1 1	0.33415	0.37355	0.00026	4
RESP(63) = 0 0 1 0 1 1 1	0.12004	0.33768	0.00006	4
RESP(64) = 0 0 0 1 1 1 1	0.07624	0.33506	0.00004	4

TABLE 7-1 (Continued)

Response Pattern		μ'_{1V}	$[\mu_{2V}]^{\frac{1}{2}}$	P(V)	T
RESP(65) =	1 1 1 0 0 0 0	-0.13404	0.41280	0.03048	3
RESP(66) =	1 1 0 1 0 0 0	-0.19694	0.41136	0.02335	3
RESP(67) =	1 0 1 1 0 0 0	-0.40396	0.40475	0.01085	3
RESP(68) =	0 1 1 1 0 0 0	-0.21633	0.40141	0.00777	3
RESP(69) =	1 1 0 0 1 0 0	-0.10364	0.40491	0.01462	3
RESP(70) =	1 0 1 0 1 0 0	-0.30789	0.39060	0.00604	3
RESP(71) =	0 1 1 0 1 0 0	-0.12709	0.39516	0.00481	3
RESP(72) =	1 0 0 1 1 0 0	-0.36363	0.39348	0.00492	3
RESP(73) =	0 1 0 1 1 0 0	-0.18454	0.39352	0.00367	3
RESP(74) =	0 0 1 1 1 0 0	-0.37323	0.38500	0.00167	3
RESP(75) =	1 1 0 0 0 1 0	-0.00454	0.39185	0.00663	3
RESP(76) =	1 0 1 0 0 1 0	-0.20201	0.37085	0.00240	3
RESP(77) =	0 1 1 0 0 1 0	-0.03119	0.38256	0.00215	3
RESP(78) =	1 0 0 1 0 1 0	-0.25210	0.37154	0.00188	3
RESP(79) =	0 1 0 1 0 1 0	-0.08570	0.37948	0.00158	3
RESP(80) =	0 0 1 1 0 1 0	-0.26553	0.36438	0.00063	3
RESP(81) =	1 0 0 0 1 1 0	-0.17523	0.36411	0.00114	3
RESP(82) =	0 1 0 0 1 1 0	-0.00859	0.37706	0.00105	3
RESP(83) =	0 0 1 0 1 1 0	-0.19129	0.35729	0.00038	3
RESP(84) =	0 0 0 1 1 1 0	-0.23765	0.35755	0.00030	3
RESP(85) =	1 1 0 0 0 0 1	0.01812	0.38396	0.00329	3
RESP(86) =	1 0 1 0 0 0 1	-0.17392	0.36369	0.00115	3
RESP(87) =	0 1 1 0 0 0 1	-0.00843	0.37527	0.00106	3
RESP(88) =	1 0 0 1 0 0 1	-0.22216	0.36427	0.00089	3
RESP(89) =	0 1 0 1 0 0 1	-0.06107	0.37240	0.00078	3
RESP(90) =	0 0 1 1 0 0 1	-0.23623	0.35756	0.00030	3
RESP(91) =	1 0 0 0 1 0 1	-0.14901	0.35738	0.00055	3
RESP(92) =	0 1 0 0 1 0 1	0.01267	0.37006	0.00052	3
RESP(93) =	0 0 1 0 1 0 1	-0.16547	0.35096	0.00018	3
RESP(94) =	0 0 0 1 1 0 1	-0.21028	0.35115	0.00014	3
RESP(95) =	1 0 0 0 0 1 1	-0.06924	0.34543	0.00024	3
RESP(96) =	0 1 0 0 0 1 1	0.08945	0.36235	0.00025	3
RESP(97) =	0 0 1 0 0 1 1	-0.08762	0.33944	0.00008	3
RESP(98) =	0 0 0 1 0 1 1	-0.12985	0.33860	0.00006	3
RESP(99) =	0 0 0 0 1 1 1	-0.06833	0.33494	0.00004	3
RESP(100) =	1 1 0 0 0 0 0	-0.42081	0.41766	0.03352	2
RESP(101) =	1 0 1 0 0 0 0	-0.63285	0.42749	0.02047	2
RESP(102) =	0 1 1 0 0 0 0	-0.42937	0.40727	0.01136	2
RESP(103) =	1 0 0 1 0 0 0	-0.70363	0.43795	0.01887	2
RESP(104) =	0 1 0 1 0 0 0	-0.49017	0.40960	0.00969	2
RESP(105) =	0 0 1 1 0 0 0	-0.69591	0.42498	0.00640	2
RESP(106) =	1 0 0 0 1 0 0	-0.57687	0.41172	0.00871	2
RESP(107) =	0 1 0 0 1 0 0	-0.38801	0.39771	0.00512	2
RESP(108) =	0 0 1 0 1 0 0	-0.57666	0.40149	0.00297	2
RESP(109) =	0 0 0 1 1 0 0	-0.63744	0.40850	0.00267	2
RESP(110) =	1 0 0 0 0 1 0	-0.43679	0.38158	0.00282	2
RESP(111) =	0 1 0 0 0 1 0	-0.27200	0.37948	0.00192	2
RESP(112) =	0 0 1 0 0 1 0	-0.44293	0.37366	0.00096	2
RESP(113) =	0 0 0 1 0 1 0	-0.49391	0.37704	0.00082	2
RESP(114) =	0 0 0 0 1 1 0	-0.40738	0.36513	0.00043	2
RESP(115) =	1 0 0 0 0 0 1	-0.39918	0.37387	0.00128	2
RESP(116) =	0 1 0 0 0 0 1	-0.24051	0.37284	0.00091	2
RESP(117) =	0 0 1 0 0 0 1	-0.40660	0.36647	0.00044	2
RESP(118) =	0 0 0 1 0 0 1	-0.45543	0.36953	0.00037	2
RESP(119) =	0 0 0 0 1 0 1	-0.37363	0.35848	0.00020	2
RESP(120) =	0 0 0 0 0 1 1	-0.27964	0.34238	0.00008	2
RESP(121) =	1 0 0 0 0 0 0	-0.99611	0.48452	0.05555	1
RESP(122) =	0 1 0 0 0 0 0	-0.72121	0.42328	0.02054	1
RESP(123) =	0 0 1 0 0 0 0	-0.96726	0.46497	0.01837	1
RESP(124) =	0 0 0 1 0 0 0	-1.06392	0.48843	0.01946	1
RESP(125) =	0 0 0 0 1 0 0	-0.88125	0.43948	0.00696	1
RESP(126) =	0 0 0 0 0 1 0	-0.69139	0.39409	0.00171	1
RESP(127) =	0 0 0 0 0 0 1	-0.64400	0.38568	0.00072	1
RESP(128) =	0 0 0 0 0 0 0	-1.49598	0.59447	0.10501	0
			TOTAL	1.00000	

Again this holds when all the items are dichotomously scored and equivalent and the simple test score is used as T, but is not true in general.

For the purpose of illustration, μ'_{1V} and its standard error of measurement with respect to every possible response pattern are computed for LIS-U [Indow & Samejima, 1962, 1966], with $N(0, 1)$ as the c.d.f. of the latent ability. Table 7-1 presents these results together with $P(V)$ and T, which is simple test score in this case. LIS-U consists of seven non-verbal reasoning items including item C_2 illustrated in Appendix. All the items are scored dichotomously in this case, and the item characteristic functions have been obtained from the tetrachoric correlation coefficients for the pairs of item variables on the assumption of bivariate normal distribution concerning the latent ability and each of all the 30 item variables of LIS scale. Item parameters thus obtained for these seven items on the normal ogive model for dichotomous items are shown in Table 7-2, where a_g is the discrimination index and b_g is the difficulty index.

We can see from Table 7-1 that there is a considerable variety among the values of μ'_{1V} for the response patterns belonging to the same test score category, as well as among the values of $\{\mu_{2V}\}^{1/2}$, the standard error of measurement. For example, the highest of all the μ'_{1V}'s in the test category 6 is approximately 1.053, while the lowest is 0.521, and the difference is about half of the standard deviation of the latent distribution. To make this point more obvious, Table 7-3 presents the bivariate frequency distribution of the 128 possible response patterns with respect to the value of μ'_{1V} and test score T. Except for the two extreme categories where only one response pattern belongs to each, response patterns contained by one test score category distribute for more than or equal to five categories of μ'_{1V} .

The values of the estimate μ'_{1T} and its standard error of measurement are also computed with respect to every test score category and are shown in Table 7-4. In the same table are the values of expectation of the standard

TABLE 7-2

Item Parameters for the Seven Items of LIS-U and for Those of the Hypothetical Tests 1 and 2

	g	1	2	3	4	5	6	7
LIS-U	$1/a_g$	0.970	0.590	0.980	1.250	0.900	0.720	0.730
	b_g	-0.860	-0.520	-0.220	-0.030	0.190	0.470	0.760
Hypothetical Test 1	$1/a_g$	1.333	0.484	1.732	0.750	3.180	1.020	2.293
	b_g	-1.500	-1.000	-0.500	0.000	0.500	1.000	1.500
Hypothetical Test 2	$1/a_g$	3.180	3.180	3.180	3.180	3.180	3.180	3.180
	b_g	-1.500	-1.000	-0.500	0.000	0.500	1.000	1.500

TABLE 7-3

Bivariate Frequency Distribution of the 128 Possible Response Patterns Concerning Estimate μ'_{1V} and Test Score T of LIS-U

μ'_{1V} (Greater or Equal) — (Less) \ T	0	1	2	3	4	5	6	7	TOTAL
-2.0 -- -1.9	0	0	0	0	0	0	0	0	0
-1.9 -- -1.8	0	0	0	0	0	0	0	0	0
-1.8 -- -1.7	0	0	0	0	0	0	0	0	0
-1.7 -- -1.6	0	0	0	0	0	0	0	0	0
-1.6 -- -1.5	0	0	0	0	0	0	0	0	0
-1.5 -- -1.4	1	0	0	0	0	0	0	0	1
-1.4 -- -1.3	0	0	0	0	0	0	0	0	0
-1.3 -- -1.2	0	0	0	0	0	0	0	0	0
-1.2 -- -1.1	0	0	0	0	0	0	0	0	0
-1.1 -- -1.0	0	1	0	0	0	0	0	0	1
-1.0 -- -0.9	0	2	0	0	0	0	0	0	2
-0.9 -- -0.8	0	1	0	0	0	0	0	0	1
-0.8 -- -0.7	0	1	1	0	0	0	0	0	2
-0.7 -- -0.6	0	2	3	0	0	0	0	0	5
-0.6 -- -0.5	0	0	2	0	0	0	0	0	2
-0.5 -- -0.4	0	0	9	1	0	0	0	0	10
-0.4 -- -0.3	0	0	3	3	0	0	0	0	6
-0.3 -- -0.2	0	0	3	8	0	0	0	0	11
-0.2 -- -0.1	0	0	0	11	1	0	0	0	12
-0.1 -- -0.0	0	0	0	9	5	0	0	0	14
-0.0 -- 0.1	0	0	0	3	8	0	0	0	11
0.1 -- 0.2	0	0	0	0	10	0	0	0	10
0.2 -- 0.3	0	0	0	0	8	4	0	0	12
0.3 -- 0.4	0	0	0	0	3	2	0	0	5
0.4 -- 0.5	0	0	0	0	0	7	0	0	7
0.5 -- 0.6	0	0	0	0	0	6	1	0	7
0.6 -- 0.7	0	0	0	0	0	2	0	0	2
0.7 -- 0.8	0	0	0	0	0	0	0	0	0
0.8 -- 0.9	0	0	0	0	0	0	3	0	3
0.9 -- 1.0	0	0	0	0	0	0	2	0	2
1.0 -- 1.1	0	0	0	0	0	0	1	0	1
1.1 -- 1.2	0	0	0	0	0	0	0	0	0
1.2 -- 1.3	0	0	0	0	0	0	0	0	0
1.3 -- 1.4	0	0	0	0	0	0	0	0	0
1.4 -- 1.5	0	0	0	0	0	0	0	1	1
1.5 -- 1.6	0	0	0	0	0	0	0	0	0
1.6 -- 1.7	0	0	0	0	0	0	0	0	0
1.7 -- 1.8	0	0	0	0	0	0	0	0	0
1.8 -- 1.9	0	0	0	0	0	0	0	0	0
1.9 -- 2.0	0	0	0	0	0	0	0	0	0
TOTAL	1	7	21	35	35	21	7	1	128

TABLE 7-4

Estimate μ'_{1T}, Its Standard Error of Measurement $\{\mu_{2T}\}^{\frac{1}{2}}$ and Expectation of the Standard Errors of Measurement of μ'_{1V} over All the Response Patterns Contained by Test Score Category T on LIS-U

T	μ'_{1T}	$\{\mu_{2T}\}^{\frac{1}{2}}$	$E[\{\mu_{2V}\}^{\frac{1}{2}}]$	Amount of Reduction	Number of Response Patterns Contained
7	1.47968	0.58765	0.58765	0.00000	1
6	0.89062	0.46409	0.45628	0.00781	7
5	0.49239	0.42493	0.41548	0.00945	21
4	0.15098	0.41259	0.40088	0.01172	35
3	-0.17773	0.41447	0.39992	0.01454	35
2	-0.52904	0.43379	0.41597	0.01781	21
1	-0.94395	0.48255	0.46765	0.01490	7
0	-1.49598	0.59447	0.59447	0.00000	1

errors of measurement of μ'_{1V} over all the response patterns contained by the test score category, computed in accordance with the formula given by the right-hand side of (7-14). The amounts of reduction in the standard error of measurement are approximately 0.008 through 0.018 in this instance, except for the two extreme score categories each of which contains only one response pattern. These values of reduction are not so large, as is expected from the fact that all the discriminating powers of the items in LIS-*U* are uniformly high, as we can see in Table 7-2.

For the hypothetical test 1, which also consists of seven items, both the variety in the values of μ'_{1V} within a test score category and the reduction in the standard error of measurement are much larger than those of LIS-*U*, as we can see in Tables 7-5 and 7-6. The discriminating powers of these seven hypothetical items have been set up to be considerably different from one another and have been arranged randomly, as shown in Table 7-2. They were computed on the assumption that the correlation coefficients between item variables and latent ability are 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 respectively, which is not unrealistic. The normal ogive model for dichotomous items was applied and $N(0, 1)$ was used as the latent distribution, just as in the case of LIS-*U*.

With this hypothetical set of data the amounts of reduction in the standard error of measurement are 0.034 through 0.095 and approximately

TABLE 7-5

Bivariate Frequency Distribution of the 128 Possible Response Patterns Concerning Estimate μ'_{1V} and Test Score T of Hypothetical Test 1

μ'_{1V} (Greater or Equal) (Less) \ T	0	1	2	3	4	5	6	7	TOTAL
-2.0 -- -1.9	0	0	0	0	0	0	0	0	0
-1.9 -- -1.8	0	0	0	0	0	0	0	0	0
-1.8 -- -1.7	1	0	0	0	0	0	0	0	1
-1.7 -- -1.6	0	0	0	0	0	0	0	0	0
-1.6 -- -1.5	0	1	0	0	0	0	0	0	1
-1.5 -- -1.4	0	2	0	0	0	0	0	0	2
-1.4 -- -1.3	0	1	1	0	0	0	0	0	2
-1.3 -- -1.2	0	0	2	0	0	0	0	0	2
-1.2 -- -1.1	0	0	3	0	0	0	0	0	3
-1.1 -- -1.0	0	1	0	2	0	0	0	0	3
-1.0 -- -0.9	0	1	2	2	0	0	0	0	5
-0.9 -- -0.8	0	0	5	1	1	0	0	0	7
-0.8 -- -0.7	0	1	1	6	0	0	0	0	8
-0.7 -- -0.6	0	0	3	5	3	0	0	0	11
-0.6 -- -0.5	0	0	1	4	5	0	0	0	10
-0.5 -- -0.4	0	0	1	3	3	2	0	0	9
-0.4 -- -0.3	0	0	1	2	3	1	0	0	7
-0.3 -- -0.2	0	0	1	2	1	3	0	0	7
-0.2 -- -0.1	0	0	0	3	2	0	1	0	6
-0.1 -- -0.0	0	0	0	3	3	0	0	0	6
-0.0 -- 0.1	0	0	0	1	4	1	0	0	6
0.1 -- 0.2	0	0	0	1	2	1	0	0	4
0.2 -- 0.3	0	0	0	0	4	2	0	0	6
0.3 -- 0.4	0	0	0	0	2	2	0	0	4
0.4 -- 0.5	0	0	0	0	1	2	0	0	3
0.5 -- 0.6	0	0	0	0	1	2	1	0	4
0.6 -- 0.7	0	0	0	0	0	2	0	0	2
0.7 -- 0.8	0	0	0	0	0	1	2	0	3
0.8 -- 0.9	0	0	0	0	0	1	0	0	1
0.9 -- 1.0	0	0	0	0	0	1	0	0	1
1.0 -- 1.1	0	0	0	0	0	0	1	0	1
1.1 -- 1.2	0	0	0	0	0	0	1	0	1
1.2 -- 1.3	0	0	0	0	0	0	1	0	1
1.3 -- 1.4	0	0	0	0	0	0	0	0	0
1.4 -- 1.5	0	0	0	0	0	0	0	1	1
1.5 -- 1.6	0	0	0	0	0	0	0	0	0
1.6 -- 1.7	0	0	0	0	0	0	0	0	0
1.7 -- 1.8	0	0	0	0	0	0	0	0	0
1.8 -- 1.9	0	0	0	0	0	0	0	0	0
1.9 -- 2.0	0	0	0	0	0	0	0	0	0
TOTAL	1	7	21	35	35	21	7	1	128

TABLE 7-6

Estimate μ'_{1T}, Its Standard Error of Measurement $\{\mu_{2T}\}^{\frac{1}{2}}$ and Expectation of the Standard Errors of Measurement of μ'_{1V} over All the Response Patterns Contained by Test Score Category T on Hypothetical Test 1

T	μ'_{1T}	$\{\mu_{2T}\}^{\frac{1}{2}}$	$E[\{\mu_{2V}\}^{\frac{1}{2}}]$	Amount of Reduction	Number of Response Patterns Contained
7	1.48517	0.69298	0.69298	0.00000	1
6	1.04789	0.66750	0.63392	0.03359	7
5	0.59583	0.63138	0.58068	0.05069	21
4	0.15052	0.59543	0.53435	0.06108	35
3	-0.29143	0.56817	0.49626	0.07191	35
2	-0.75610	0.57412	0.47933	0.09479	21
1	-1.26858	0.57442	0.51220	0.06223	7
0	-1.72305	0.59884	0.59884	0.00000	1

5% through 17% of the values of $\{\mu_{2T}\}^{1/2}$ for test score category T. The values of the Bayes estimate μ'_{1V} and its standard error of measurement $\{\mu_{2V}\}^{1/2}$ with respect to every possible response pattern are given elsewhere (cf. Samejima, 1968a, Appendix 4).

On the other hand, for the hypothetical test 2, also consisting of seven items, the variety in the values of μ'_{1V} within a test score category is so small that the reduction in the standard errors of measurement is almost nothing, as is obvious from Tables 7-7 and 7-8. In this hypothetical set of data the discriminating powers of the items are of the lowest value of those in hypothetical test 1, as shown in Table 7-2, and all the other conditions are the same as those in the previous instance.

All the above results are naturally expected from the fact that simple test scores are sufficient statistics on the normal ogive model for dichotomous items if all the items are equivalent [Lord, 1953] and on the logistic model if

TABLE 7-7

Bivariate Frequency Distribution of the 128 Possible Response Patterns Concerning Estimate μ'_{1V} and Test Score T of Hypothetical Test 2

μ'_{1V} (Greater or Equal) (Less) \ T	0	1	2	3	4	5	6	7	TOTAL
-2.0 -- -1.9	0	0	0	0	0	0	0	0	0
-1.9 -- -1.8	0	0	0	0	0	0	0	0	0
-1.8 -- -1.7	0	0	0	0	0	0	0	0	0
-1.7 -- -1.6	0	0	0	0	0	0	0	0	0
-1.6 -- -1.5	0	0	0	0	0	0	0	0	0
-1.5 -- -1.4	0	0	0	0	0	0	0	0	0
-1.4 -- -1.3	0	0	0	0	0	0	0	0	0
-1.3 -- -1.2	1	0	0	0	0	0	0	0	1
-1.2 -- -1.1	0	0	0	0	0	0	0	0	0
-1.1 -- -1.0	0	0	0	0	0	0	0	0	0
-1.0 -- -0.9	0	5	0	0	0	0	0	0	5
-0.9 -- -0.8	0	2	0	0	0	0	0	0	2
-0.8 -- -0.7	0	0	0	0	0	0	0	0	0
-0.7 -- -0.6	0	0	0	0	0	0	0	0	0
-0.6 -- -0.5	0	0	21	0	0	0	0	0	21
-0.5 -- -0.4	0	0	0	0	0	0	0	0	0
-0.4 -- -0.3	0	0	0	0	0	0	0	0	0
-0.3 -- -0.2	0	0	0	0	0	0	0	0	0
-0.2 -- -0.1	0	0	0	35	0	0	0	0	35
-0.1 -- -0.0	0	0	0	0	0	0	0	0	0
-0.0 -- 0.1	0	0	0	0	0	0	0	0	0
0.1 -- 0.2	0	0	0	0	35	0	0	0	35
0.2 -- 0.3	0	0	0	0	0	0	0	0	0
0.3 -- 0.4	0	0	0	0	0	0	0	0	0
0.4 -- 0.5	0	0	0	0	0	0	0	0	0
0.5 -- 0.6	0	0	0	0	0	21	0	0	21
0.6 -- 0.7	0	0	0	0	0	0	0	0	0
0.7 -- 0.8	0	0	0	0	0	0	0	0	0
0.8 -- 0.9	0	0	0	0	0	0	2	0	2
0.9 -- 1.0	0	0	0	0	0	0	5	0	5
1.0 -- 1.1	0	0	0	0	0	0	0	0	0
1.1 -- 1.2	0	0	0	0	0	0	0	0	0
1.2 -- 1.3	0	0	0	0	0	0	0	1	1
1.3 -- 1.4	0	0	0	0	0	0	0	0	0
1.4 -- 1.5	0	0	0	0	0	0	0	0	0
1.5 -- 1.6	0	0	0	0	0	0	0	0	0
1.6 -- 1.7	0	0	0	0	0	0	0	0	0
1.7 -- 1.8	0	0	0	0	0	0	0	0	0
1.8 -- 1.9	0	0	0	0	0	0	0	0	0
1.9 -- 2.0	0	0	0	0	0	0	0	0	0
TOTAL	1	7	21	35	35	21	7	1	128

TABLE 7-8

Estimate μ'_{1T}, Its Standard Error of Measurement $\{\mu_{2T}\}^{\frac{1}{2}}$ and Expectation of the Standard Errors of Measurement of μ'_{1V} over All the Response Patterns Contained by Test Score Category T on Hypothetical Test 2

T	μ'_{1T}	$\{\mu_{2T}\}^{\frac{1}{2}}$	$E[\{\mu_{2V}\}^{\frac{1}{2}}]$	Amount of Reduction	Number of Response Patterns Contained
7	1.28186	0.85506	0.85506	0.00000	1
6	0.90908	0.84708	0.84704	0.00004	7
5	0.54291	0.84189	0.84184	0.00004	21
4	0.18055	0.83933	0.83930	0.00004	35
3	-0.18055	0.83933	0.83930	0.00004	35
2	-0.54291	0.84189	0.84184	0.00004	21
1	-0.90908	0.84708	0.84704	0.00004	7
0	-1.28186	0.85506	0.85506	0.00000	1

all the discriminating powers of the items are of the same value [Birnbaum, in Lord & Novick, 1968], and that among the above three sets of data hypothetical test 2 is nearest to the situation of equivalent items and hypothetical test 1 is the farthest from it. The values of individual μ'_{1V} for the hypothetical test 2 are given elsewhere (cf. Samejima, 1968a, Appendix 4).

CHAPTER 8

THE MEAN-SQUARE ERRORS OF ESTIMATORS

As was observed in the previous chapter, the expectation or the first-order moment of θ specified for the density function for response pattern V is the estimator for which the expectation of the mean-square error about the true value of θ is minimized.

Again it is easily seen that the expectation of the mean-square error of μ'_{1T}, defined on the density function for aggregation T, is no less than that of μ'_{1V}, defined on the density function for response pattern V, because of the fact that

$$
\begin{aligned}
\text{(8-1)}\qquad \int_{-\infty}^{\infty} E\{\mu'_{1V} - \theta\}^2 f(\theta)\, d\theta &= \int_{-\infty}^{\infty} \sum_V \{\mu'_{1V} - \theta\}^2 P_V(\theta) f(\theta)\, d\theta \\
&= \sum_V P(V) \cdot \frac{\int_{-\infty}^{\infty} \{\theta - \mu'_{1V}\}^2 \psi(V, \theta)\, d\theta}{P(V)} \\
&= \sum_V \mu_{2V} P(V) \\
&= \sum_T \sum_{V \in T} \mu_{2V} P(V) \\
&= \sum_T \frac{\sum_{V \in T} \mu_{2V} P(V)}{\sum_{V \in T} P(V)} P(T),
\end{aligned}
$$

and

$$
\begin{aligned}
\text{(8-2)}\qquad \int_{-\infty}^{\infty} E\{\mu'_{1T} - \theta\}^2 f(\theta)\, d\theta &= \int_{-\infty}^{\infty} \sum_T \{\mu'_{1T} - \theta\}^2 P_T(\theta) f(\theta)\, d\theta \\
&= \sum_T P(T) \cdot \frac{\int_{-\infty}^{\infty} \{\theta - \mu'_{1T}\}^2 P_T(\theta) f(\theta)\, d\theta}{P(T)} \\
&= \sum_T \mu_{2T} P(T),
\end{aligned}
$$

where $P(T)$ is the probability of T specified in connection with the latent density function $f(\theta)$, and that the following inequality has been obtained in the previous chapter.

$$
\text{(8-3)}\qquad \mu_{2T} \geq \frac{\sum_{V \in T} \mu_{2V} P(V)}{\sum_{V \in T} P(V)}.
$$

Thus we have

$$\int_{-\infty}^{\infty} E\{\mu'_{1T} - \theta\}^2 f(\theta)\, d\theta \geq \int_{-\infty}^{\infty} E\{\mu'_{1V} - \theta\}^2 f(\theta)\, d\theta. \tag{8-4}$$

An equality holds in (8-4) when all the items are scored dichotomously and their item characteristic functions are equivalent, if we take the simple test score as T and the response pattern as V, while a strict inequality always holds if we take the response pattern of less graded item responses as T and that of more graded responses as V, as exactly in the same way as discussed in the previous chapter.

Figure 8-1 presents the mean-square error computed by the formula

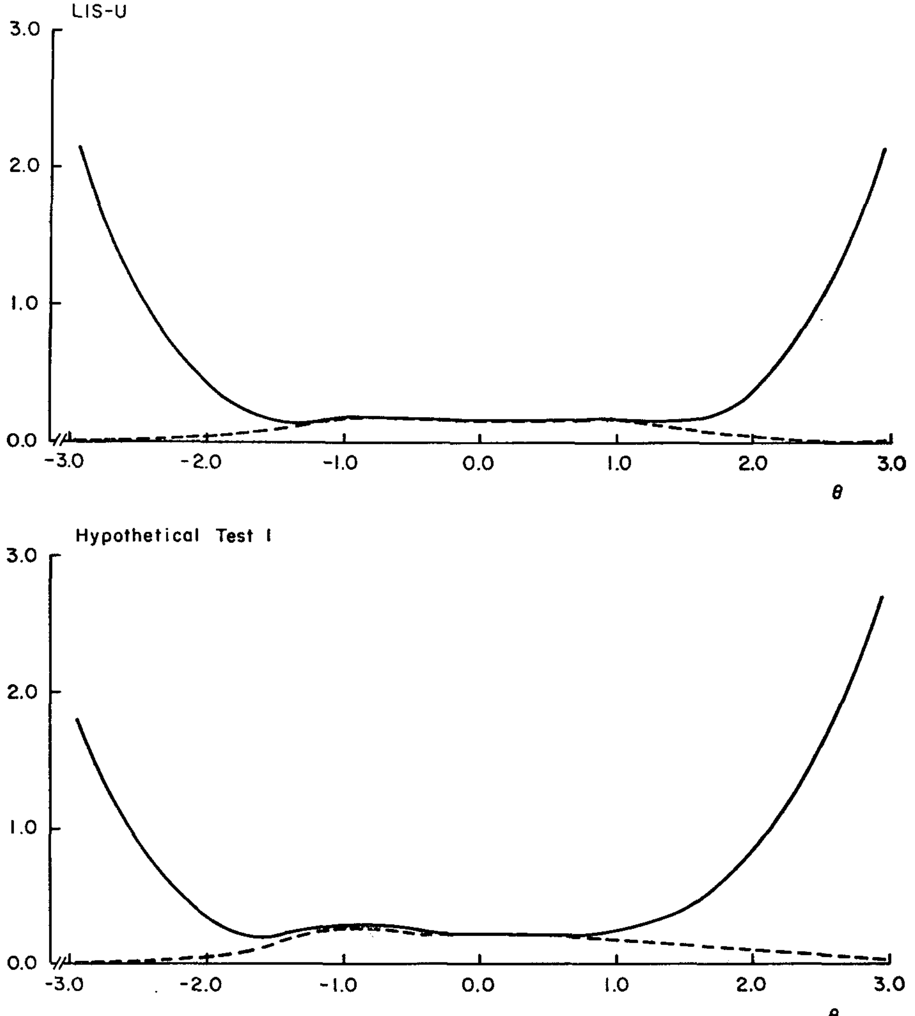


FIGURE 8-1

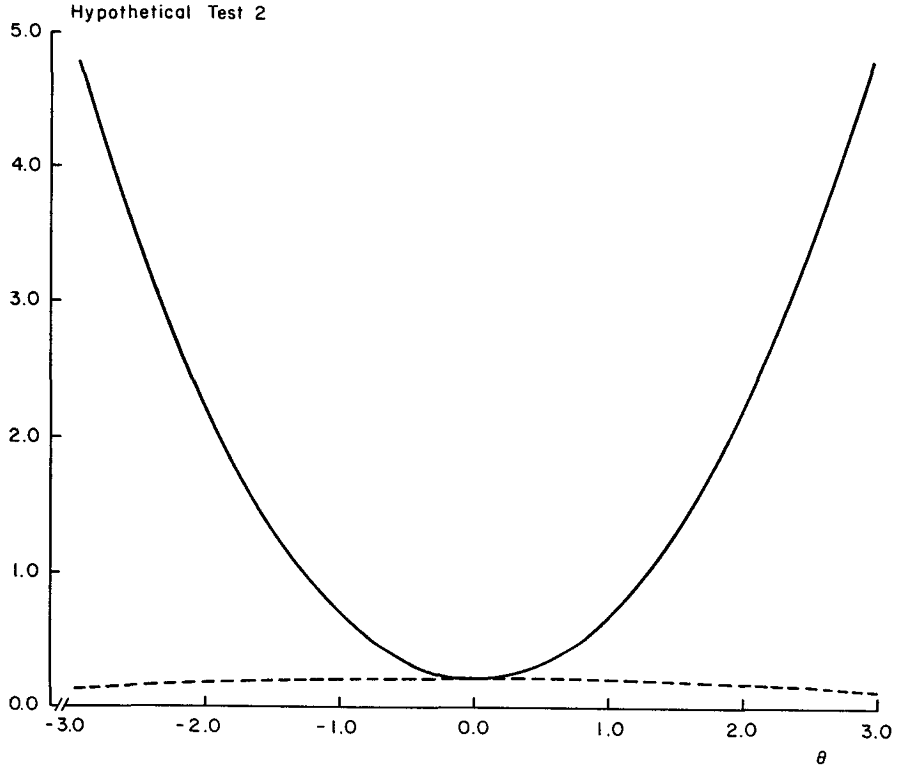


FIGURE 8–1 (*Continued*).

Mean-square error (solid line) as a function of θ with μ'_{1V} as the estimator, and also the variance (dotted line) of μ'_{1V} for the fixed value of θ, for LIS-U, Hypothetical Tests 1 and 2.

$$E\{\mu'_{1V} - \theta\}^2 \underline{\underline{\quad}} \sum_V \{\mu'_{1V} - \theta\}^2 P_V(\theta), \tag{8-5}$$

for the fixed value of θ, and expressed as a function of θ, drawn by solid line, for each of LIS-U, hypothetical set of data 1 and 2, which were introduced and described in the previous chapter. In the same figure is plotted the variance of the Bayes estimate μ'_{1V} for the fixed value of θ (dotted line). Since we have for any estimator θ^* the variance for the fixed value of θ expressed as

$$E\{\theta^* - E\{\theta^*\}\}^2 \tag{8-6}$$

and the following equation

$$E\{\theta^* - \theta\}^2 = E\{\theta^* - E\{\theta^*\}\}^2 + E\{E\{\theta^*\} - \theta\}^2 \tag{8-7}$$

holds, the discrepancy between two curves in each figure of Figure 8-1 is the expectation of the square of the bias of μ'_{1V} in estimating the true value θ, specified for each of these three sets of data.

We can see from these results that in the case of LIS-U the mean-square error is small for a considerably wide range of θ, approximately -1.75 through 1.75, and almost the same is true for hypothetical test 1, while it is by no means small in the case of hypothetical test 2, except for a very narrow range of θ around 0.0. On the other hand, the variances are uniformly small for the entire range of θ in all the three instances, and the amount of bias is the greatest in hypothetical test 2, as is expected from the fact that the discriminating powers of the items in this test are so poor that very little information is given by these items.

The expectation of Bayes estimator for the fixed value of θ is given by the formula

$$E\{\mu'_{1V}\} = \sum_{V} \mu'_{1V} P_V(\theta), \tag{8-8}$$

and it has been computed and plotted against θ with solid lines in Figure 8-2 for these three sets of data. It is observed that in each case the expectation of the estimates for the fixed value of θ is larger than the true ability value at the lower part of the continuum, while it is smaller than the true ability value at the higher part of the continuum. This tendency of regression results from the specification of $f(\theta)$ as the standard normal density, and is the most conspicuous in the case of hypothetical test 2, while it is much milder in the others, especially in LIS-U. In this instance the regression is practically nil for the range of θ -1.0 through 1.0.

These values can also be obtained with μ'_{1T} as the estimator when the simple test score is used as T and compared with the above results obtained with respect to μ'_{1V} . All the dicsrepancies between these two corresponding sets of values concerning hypothetical test 2 proved to be practically nil, as is expected from the results obtained about the expectations and their standard errors of measurement in the previous chapter. The discrepancies are larger in the case of LIS-U and are much larger in hypothetical test 1. The upper figure of Figure 8-3 illustrates the mean-square error obtained on μ'_{1T} , which is drawn by dotted line, in comparison with that on μ'_{1V} , drawn by solid line, in the case of hypothetical test 1. The discrepancy between these two values is small for the range of θ -1.0 through 1.0, but it is larger outside of this range. In the same figure are also plotted the values of mean-square error multiplied by $f(\theta)$ and 10^2, again with dotted and solid lines respectively. They are the lower two curves in this figure. It is apparent that the area under the dotted curve is considerably larger than the one under the solid curve.

These areas computed for the range of θ, -3.0 through 3.0, concerning these three sets of data by

$$(8\text{-}9)\qquad \int_{-3.0}^{3.0} E\{\mu'_{1V} - \theta\}^2 f(\theta)\,d\theta = \sum_V \int_{-3.0}^{3.0} \{\mu'_{1V} - \theta\}^2 P_V(\theta) f(\theta)\,d\theta$$

and

$$(8\text{-}10)\qquad \int_{-3.0}^{3.0} E\{\mu'_{1T} - \theta\}^2 f(\theta)\,d\theta = \sum_T \int_{-3.0}^{3.0} \{\mu'_{1T} - \theta\}^2 P_T(\theta) f(\theta)\,d\theta$$

are shown in Table 8-1. The discrepancy of the areas in hypothetical test 1 is almost eight times as much as that in LIS-U.

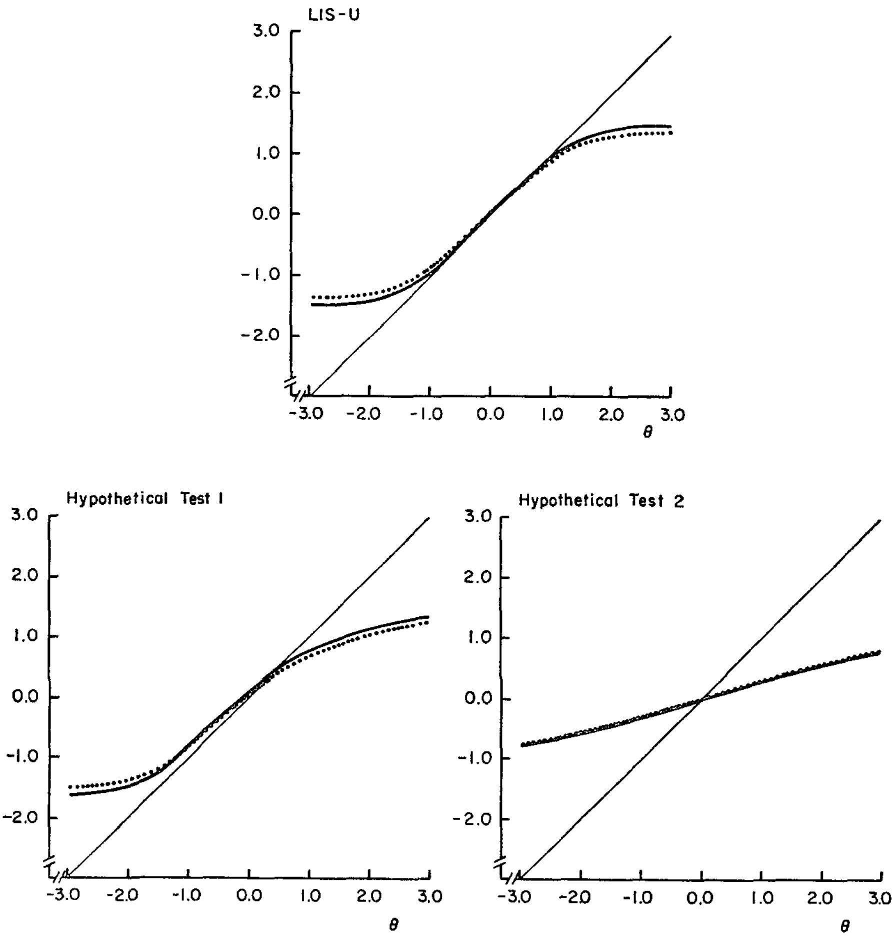
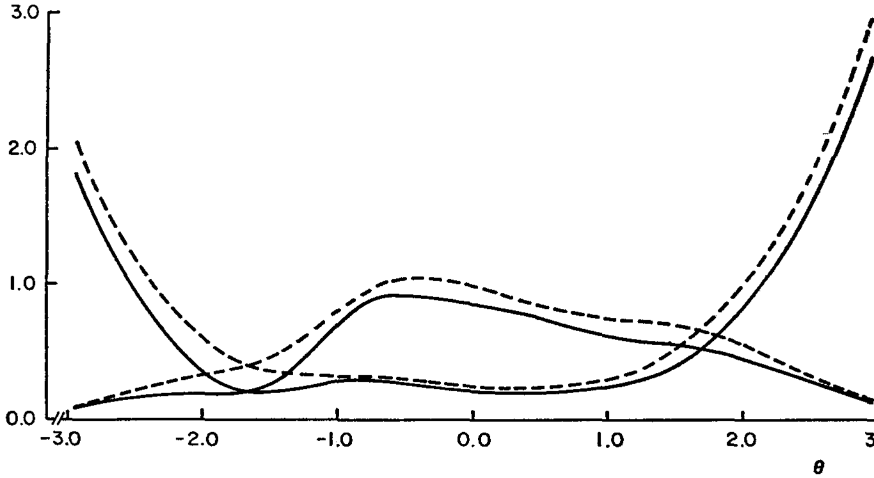


FIGURE 8-2

Expectations of μ'_{1V} (solid line) and $\hat{\hat{\theta}}_V$ (dotted line) plotted against θ, for LIS-U, Hypothetical Tests 1 and 2.



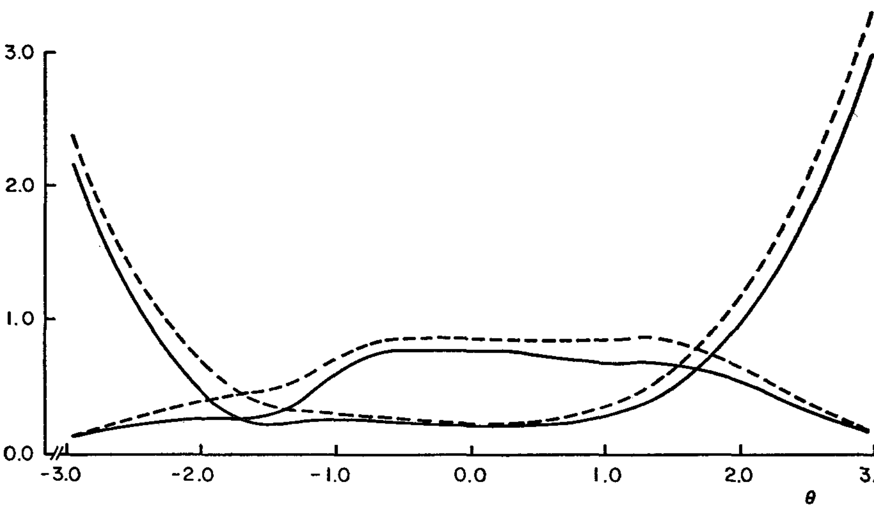


FIGURE 8–3

Mean-square error computed with μ'_{1T} as the estimator (dotted U-shape curve, in the upper figure) and with μ'_{1V} as the estimator (solid U-shape curve, in the upper figure), and also mean-square error with $\hat{\hat{\theta}}_T$ as the estimator (dotted U-shape curve, in the lower figure) and with $\hat{\hat{\theta}}_V$ as the estimator (solid U-shape curve, in the lower figure) for Hypothetical Test 1. Lower curves in each figure are these values multiplied by $f(\theta)$ and 10^2 drawn in the corresponding manner.

TABLE 8-1

Areas under the Mean-Square Error Multiplied by $f(\theta) \times 10^2$, and Also Those under the Variance for the Fixed Value of θ Multiplied by $f(\theta) \times 10^2$, for the Range of θ -3.0 through 3.0, with Each of μ'_{1V}, μ'_{1T}, $\hat{\theta}_V$ and $\hat{\theta}_T$ as the Estimator, for LIS-U and Hypothetical Tests 1 through 5

		Mean-Square Error			Variance (Conditional)		
		r.p.	t.s.	dif.	r.p.	t.s.	dif.
LIS-U	μ'_1	0.2098	0.2181	-0.0083	0.1532	0.1589	-0.0057
	$\hat{\theta}$	0.2133	0.2217	-0.0084	0.1293	0.1339	-0.0046
Hp. 1	μ'_1	0.2949	0.3580	-0.0631	0.1965	0.2226	-0.0261
	$\hat{\theta}$	0.2984	0.3617	-0.0633	0.1740	0.1948	-0.0208
Hp. 2	μ'_1	0.6924	0.6925	-0.0001	0.2060	0.2060	0.0000
	$\hat{\theta}$	0.6923	0.6924	-0.0001	0.2010	0.2028	-0.0018
Hp. 3	μ'_1	0.3159	0.3753	-0.0594	0.1994	0.2226	-0.0232
	$\hat{\theta}$	0.3194	0.3798	-0.0604	0.1809	0.2016	-0.0207
Hp. 4	μ'_1	0.2674	0.2681	-0.0007	0.1940	0.1943	-0.0003
	$\hat{\theta}$	0.2680	0.2688	-0.0008	0.1777	0.1774	0.0003
Hp. 5	μ'_1	0.2405	0.2405	0.0000	0.1707	0.1707	0.0000
	$\hat{\theta}$	0.2432	0.2432	0.0000	0.1468	0.1468	0.0000

Illustrated in the same table are the values obtained by

$$(8\text{-}11) \qquad \int_{-3.0}^{3.0} E\{\mu'_{1V} - E\{\mu'_{1V}\}\}^2 f(\theta)\, d\theta = \sum_V \int_{-3.0}^{3.0} \{\mu'_{1V} - E\{\mu'_{1V}\}\}^2 P_V(\theta) f(\theta)\, d\theta$$

and

$$(8\text{-}12) \qquad \int_{-3.0}^{3.0} E\{\mu'_{1T} - E\{\mu'_{1T}\}\}^2 f(\theta)\, d\theta = \sum_T \int_{-3.0}^{3.0} \{\mu'_{1T} - E\{\mu'_{1T}\}\}^2 P_T(\theta) f(\theta)\, d\theta,$$

and also the values obtained by substituting $\hat{\theta}_V$ for μ'_{1V}, and $\hat{\theta}_T$ for μ'_{1T}, in

(8-9) through (8-12), where $\hat{\hat{\theta}}_V$ is the Bayes modal estimate obtained by using the entire response pattern and $\hat{\hat{\theta}}_T$ is that obtained on test score T. These values of the Bayes modal estimate $\hat{\hat{\theta}}_V$ and $\hat{\hat{\theta}}_T$ are computed and are presented elsewhere (cf. Samejima, 1968a, Appendix 5) together with the values of maximum likelihood estimates $\hat{\theta}_V$ and $\hat{\theta}_T$ obtained with L_V and L_T as the likelihood functions respectively, for these three sets of data. Also presented in Table 8-1 are the corresponding values for three other hypothetical tests consisting of seven dichotomous items. In hypothetical test 3 the difficulty parameters are -3.1, -1.2, -0.9, -0.5, 0.0, 0.8, and 1.1, and all the other conditions are the same as those in hypothetical test 1. In hypothetical test 4 all the discrimination parameters are 1.0, and otherwise the conditions are the same as in hypothetical data 1 and 2. In hypothetical test 5 all the items are equivalent, with 0.0 as the difficulty parameters and 1.0 as the discrimination parameters, and the other conditions are the same as in the other tests.

The comparison of the results obtained on estimator μ_1' and $\hat{\hat{\theta}}$ suggests the remarkable resemblance between two sets of estimates, μ_1' and $\hat{\hat{\theta}}$, in every

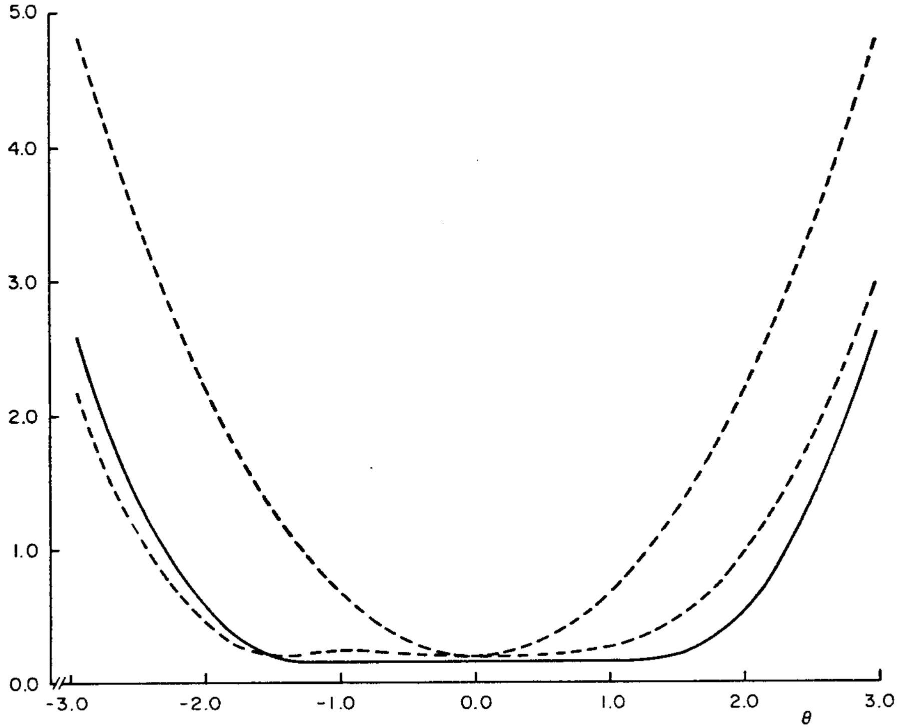
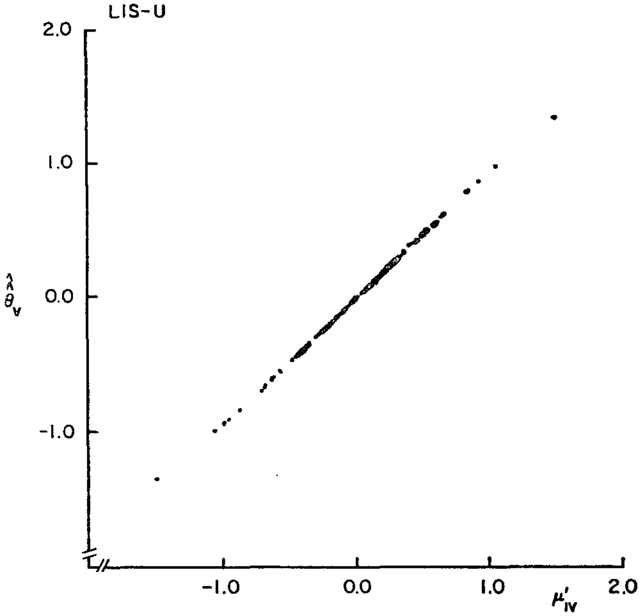


FIGURE 8-4

Mean-square error computed with $\hat{\hat{\theta}}_V$ as the estimator for LIS-U (solid line), for Hypothetical Test 1 (small dotted line) and for Hypothetical Test 2 (large dotted line).



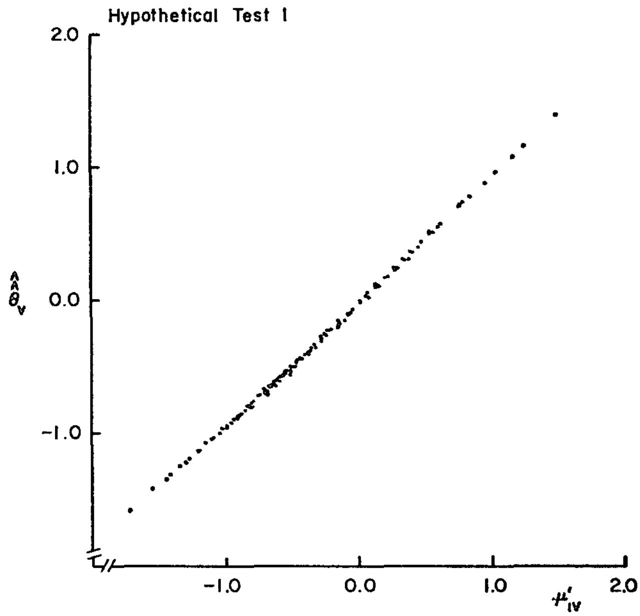


FIGURE 8–5

$\hat{\hat{\theta}}_V$ plotted against μ'_{1V} , with respect to 128 response patterns obtained on LIS-U, Hypothetical Tests 1 and 2.

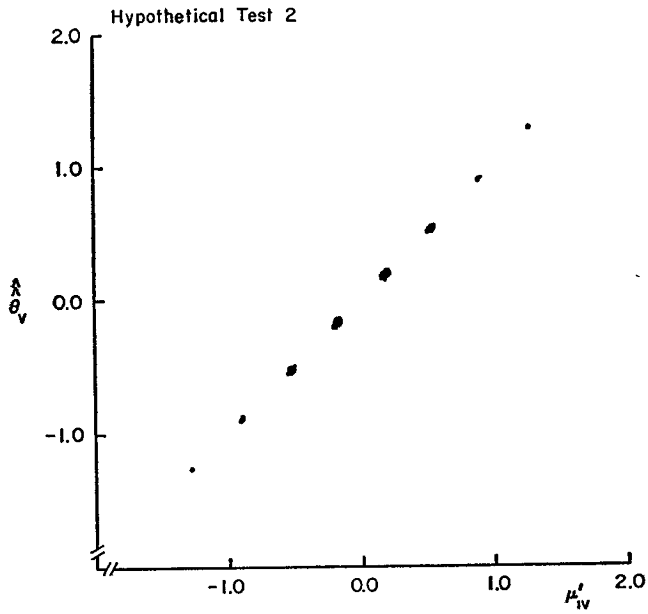


FIGURE 8-5 (*Continued*).

case. To make this point clearer, the mean-square error for estimator $\hat{\theta}_V$ was computed for the first three sets of data, and the results are illustrated as Figure 8-4. Comparing this with Figure 8-1, we can see that the results are very similar. Although the expectation of mean-square error is the smallest when we take μ'_{1V} as the estimator, as was already discussed in the previous chapter, that of the mean-square error with $\hat{\theta}_V$ as the estimator is also small. In fact in hypothetical test 2 the difference between these two kinds of mean-square errors is practically nil.

The expectation of estimator $\hat{\theta}_V$ for the fixed value of θ is plotted in Figure 8-2 with dotted lines, for the three sets of data. In LIS-*U* and hypothetical test 1, regression is a little more conspicuous than in the case of estimator μ'_{1V}, but on the whole these two expectations are similar.

The comparison of the mean-square errors in hypothetical test 1 between those obtained on $\hat{\theta}_V$ and those on $\hat{\theta}_T$ is also done and is illustrated in the lower figure of Figure 8-3 in the same way as those on μ'_{1V} and μ'_{1T} were done. Again the discrepancy, that is, the reduction in the mean-square error by using the entire response pattern instead of the test score, is small for the range of θ — 1.0 through 1.0 but is larger outside of this range, and the area under the lower dotted curve is considerably larger than that under the lower solid curve, in a manner similar to the one in the previous case.

Figure 8-5 presents the values of $\hat{\theta}_V$ plotted against the values of μ'_{1V} with respect to all the 128 possible response patterns, obtained on LIS-*U*, hypothetical test 1 and 2. Although the manners of divergence of these plots are remarkably different from test to test, the correlations between two sets of estimates proved to be uniformly high.

The correlation coefficient, $R_{\mu'_1 V}\hat{\theta}_V$, between the two sets of estimates given by

$$R_{\mu'_1 V}\hat{\theta}_V = \frac{\mathrm{Cov}\,\{\mu'_{1V}\,,\,\hat{\theta}_V\}}{[\mathrm{Var}\,\{\mu'_{1V}\}\,\mathrm{Var}\,\{\hat{\theta}_V\}]^{1/2}}\,, \tag{8-13}$$

where

$$\mathrm{Var}\,\{\mu'_{1V}\} = \sum_V \int_{-\infty}^{\infty} \left\{\mu'_{1V} - \sum_U \int_{-\infty}^{\infty} \mu'_{1U} P_U(\theta) f(\theta)\, d\theta\right\}^2 P_V(\theta)\, d\theta, \tag{8-14}$$

$$\mathrm{Var}\,\{\hat{\theta}_V\} = \sum_V \int_{-\infty}^{\infty} \left\{\hat{\theta}_V - \sum_U \int_{-\infty}^{\infty} \hat{\theta}_U P_U(\theta) f(\theta)\, d\theta\right\}^2 P_V(\theta)\, d\theta \tag{8-15}$$

and

$$\mathrm{Cov}\,\{\mu'_{1V}\,,\,\hat{\theta}_V\} = \sum_V \int_{-\infty}^{\infty} \left\{\mu'_{1V} - \sum_U \int_{-\infty}^{\infty} \mu'_{1U} P_U(\theta) f(\theta)\, d\theta\right\} \cdot \left\{\hat{\theta}_V - \sum_U \int_{-\infty}^{\infty} \hat{\theta}_U P_U(\theta) f(\theta)\, d\theta\right\} P_V(\theta)\, d\theta, \tag{8-16}$$

was computed for the six sets of data shown in Table 8-1. These values are presented in Table 8-2, together with the values of the variances and covariances for the entire range of θ. As is expected from the results so far, the

TABLE 8-2

Expectation of μ'_{1V}, Expectation of $\hat{\theta}_V$, Covariance of μ'_{1V} and $\hat{\theta}_V$, Variance of μ'_{1V}, Variance of $\hat{\theta}_V$, and Correlation between μ'_{1V} and $\hat{\theta}_V$, for LIS-U and Hypothetical Tests 1 through 5

	Expectation of μ'_{1V}	Expectation of $\hat{\theta}_V$	Cov $\{\mu'_{1V}, \hat{\theta}_V\}$	Var $\{\mu'_{1V}\}$	Var $\{\hat{\theta}_V\}$	$R_{\mu'_{1V}\hat{\theta}_V}$
LIS-U	0.000	0.001	0.721	0.781	0.666	0.9998
Hp. 1	0.000	-0.029	0.651	0.696	0.609	0.9993
Hp. 2	0.000	0.000	0.288	0.291	0.284	1.0000
Hp. 3	0.000	-0.042	0.638	0.675	0.603	0.9994
Hp. 4	0.000	0.000	0.695	0.725	0.666	1.0000
Hp. 5	0.000	0.000	0.698	0.750	0.649	0.9999

correlation is very high in every instance. The value of variance for $\hat{\theta}_V$ is smaller than that for μ'_{1V} in every case, which supports a slightly stronger tendency of regression about estimator $\hat{\theta}_V$.

From the above observations of mean-square error, etc., it has been suggested that the Bayes modal estimator $\hat{\theta}_V$ can be used as a good approximation to the Bayes estimator μ'_{1V} , since in computational procedure it is much easier to get $\hat{\theta}_V$ rather than μ'_{1V} , when conditions (i) and (ii) or (ii)*, described in Chapter 3, hold for the model for the operating characteristics of individual item responses. The procedure of obtaining the Bayes modal estimate will be discussed in the following chapter.

CHAPTER 9

COMPUTATIONAL PROCEDURES FOR OBTAINING THE ESTIMATES

As was stated earlier, availability of high-speed computers facilitates the computation of the estimates, especially in maximum likelihood estimation and Bayes modal estimation.

We can obtain the Bayes estimate for a given response pattern directly from (7-6) of Chapter 7,

$$\mu'_{1V} = \frac{\int_{-\infty}^{\infty} \theta\psi(V, \theta)\, d\theta}{P(V)}, \tag{7-6}$$

where $P(V)$ is given by equation (7-5)

$$P(V) = \int_{-\infty}^{\infty} \psi(V, \theta)\, d\theta \tag{7-5}$$

and

$$\psi(V, \theta) = P_V(\theta) f(\theta), \tag{9-1}$$

when the distribution of the latent variate is known or reasonably assumed.

In contrast to this, we cannot obtain the maximum likelihood estimate or the Bayes modal estimate directly by using a simple formula like (7-6). This is true even for the simplified case of the logistic model of dichotomous items, where a sufficient statistic exists. In this instance, $t(V)$, defined by

$$t(V) = \sum_{g=1}^{n} a_g x_g, \tag{9-2}$$

gives all the information about θ available in the response pattern V [Birnbaum in Lord & Novick, 1968], where x_g takes either 1 or 0, and the maximum likelihood estimate is the value of θ which satisfies

$$\begin{aligned} t(V) &= E(t(V) \mid \theta) \\ &= \sum_{g=1}^{n} a_g P_g(\theta). \end{aligned} \tag{9-3}$$

Although we cannot let the computer calculate the maximum likelihood estimate directly from (9-3), we can have it compute the right hand side of the equation for a sequence of θ values, as well as the left hand side for a given response pattern V, and search for the value of θ at which the equality holds.

In the general case where a sufficient statistic does not exist, the maximum likelihood estimate, $\hat{\theta}_V$, can be obtained by the following procedure.

(1) For each element of a given response pattern, k_g , let the computer calculate the values of $A_{k_g}(\theta)$ with respect to a sequence of θ values, which have a sufficiently wide range and a small interval. $A_{k_g}(\theta)$ has been defined by (3-1) in Chapter 3, such that

$$(3\text{-}1) \qquad A_{k_g}(\theta) = \frac{(\partial/\partial\theta)P_{k_g}(\theta)}{P_{k_g}(\theta)}.$$

(2) Let the computer add these $A_{k_g}(\theta)$'s computed with respect to each value of θ, over all the responses in the response pattern.

(3) If the result of the summation equals zero at a specified value of θ, this value is taken as the maximum likelihood estimate, $\hat{\theta}_V$. If not, let the computer specify the two successive values of θ between which the sum total of $A_{k_g}(\theta)$'s transits from the negative value to the positive value, and search for the value of θ between these two values at which the sum total equals zero. A convenient way will be to have the computer specify additional nine values of θ between the two, by making the interval one tenth of the original one, and to let it repeat the steps, (1), (2) and (3), for these eleven values of θ. In this way, we can make the value of maximum likelihood estimate as precise as we wish.

(4) If the transformation of the variable is necessary, let the computer transform $\hat{\theta}_V$ thus obtained into $\hat{\tau}_V$ by a specified formula

$$(9\text{-}4) \qquad \hat{\tau}_V = \tau(\hat{\theta}_V).$$

If the computer has failed in specifying the two successive values of step (3), it must be because of the shortage of the range of θ, provided that $A_{k_g}(\theta)$ defined by (3-1) satisfies condition (i) and (ii) for unique maximum, which have been discussed in Chapter 3, for every k_g contained by the response pattern. Then we have to try to improve the computer program so that enough range of θ will be covered. If all the k_g's in the response pattern satisfy condition (i) and (ii)* for unique maximum, instead of (i) and (ii), the failure in specification may be due to the existence of a terminal maximum.

If $f(\theta)$ is known or reasonably assumed, and it satisfies condition (iii) and (iv) for unique Bayesian modality, we can obtain the Bayes modal estimate in the same way as we get the maximum likelihood estimate. The only additional procedure is that $G(\theta)$, defined by (3-37) in Chapter 3, such that

$$(3\text{-}37) \qquad G(\theta) = \frac{(\partial/\partial\theta)f(\theta)}{f(\theta)}$$

should be computed with respect to the same sequences of θ, and should be used in addition to the $A_{k_g}(\theta)$'s. In this case, however, there exists no terminal maximum, which we have already seen in Chapter 3, provided that conditions (i), (ii)*, (iii) and (iv) are met. Step (4) should be excluded, moreover, since the Bayes modal estimator does not have a transformation-free character, as we have seen in Chapter 2.

This method of obtaining the maximum likelihood estimate or Bayes modal estimate for a given response pattern will particularly be useful in the machine-assisted testing situation, where the examinee is given test items sequentially. We can have the computer obtain the maximum likelihood estimate, $\hat{\theta}_V$, or the Bayes modal estimate, $\hat{\hat{\theta}}_V$, based on the response pattern of an initial few test items given to the examinee, and let it select an item having a maximum amount of information around that value of θ out of the whole item library, as the next item presented to the examinee. In this way we can design an efficient individual test, which makes the convergence of the estimate very rapid. This may be worth examining in practical situations.

CHAPTER 10

SOME OBSERVATIONS CONCERNING THE RELATIONSHIP BETWEEN FORMULAS FOR THE ITEM CHARACTERISTIC FUNCTION AND THE PHILOSOPHY OF SCORING

In this chapter, we shall deal only with dichotomous items. Notations $P_g(\theta)$ and $Q_g(\theta)$, defined in Chapter 1, are to be used, therefore, throughout this chapter instead of $P_{k_g}(\theta)$, the operating characteristic of item response, and $P_g(\theta)$ is called the item characteristic function.

It happens frequently that we can explain a set of testing results not only on one specified model but also on some other model or models. In other words, the fit of more than one model can be simultaneously good to the same empirical data. It is interesting to note, on the other hand, that the philosophies of scoring implied in these models are often different from one another even if the item characteristic functions are very much alike. It seems to us, therefore, that the philosophy of scoring may be used as one of the criteria for determining which model is preferable to the others, in case the empirical data can be explained by more than one model simultaneously. This criterion should be used, however, only when there is no other objective reason why one model should be taken in preference to the others, since it largely depends on a researcher's subjective judgment to decide which philosophy of scoring is better than the others.

In this chapter we shall use $\hat{\theta}$ as our estimator, i.e., the maximum likelihood estimator of θ on the response pattern with $L_V(\theta)$ as the likelihood function, since it can be obtained without specifying the latent distribution. In so doing we have the advantage of generalizing the results obtained with respect to one variable to its transformed variables and, consequently, to the transformed models, because of the estimator's transformation-free character discussed in Chapter 2. As far as we have no external criteria with which the equality of intervals of the latent variable is justified, we have to content ourselves to get only order statistics for the respondents' abilities in the strict sense of the word, and, therefore, the transformation-free property of an estimator is important in preserving the order of respondents' estimated abilities. Thus we must keep in mind that all the results obtained with respect to $\hat{\theta}$ can be transferred to $\hat{\tau}$, if necessary, provided that variate τ has a functional relationship with θ which is expressed as being monotonically increasing in θ.

We shall deal in this chapter solely with cases in which conditions (i) and (ii)*, introduced in Chapter 3, are satisfied with respect to θ.

Let $A_g(\theta)$ or A_g denote the basic function defined in Chapter 3, for the positive response to item g, and $B_g(\theta)$ or B_g be that for the negative response.

We can write

$$A_g(\theta) = \frac{(\partial/\partial\theta)P_g(\theta)}{P_g(\theta)} \tag{10-1}$$

and

$$B_g(\theta) = \frac{(\partial/\partial\theta)Q_g(\theta)}{Q_g(\theta)}, \tag{10-2}$$

by following (3-1) of Chapter 3.

Equation (10-1) can be developed and rewritten as

$$\begin{aligned} A_g(\theta) &= \frac{\{(\partial/\partial\theta)P_g(\theta)\}\{1 - P_g(\theta)\}}{P_g(\theta)Q_g(\theta)} \\ &= B_g(\theta) + S_g(\theta), \end{aligned} \tag{10-3}$$

by defining $S_g(\theta)$ so that

$$S_g(\theta) = \frac{(\partial/\partial\theta)P_g(\theta)}{P_g(\theta)Q_g(\theta)}. \tag{10-4}$$

We can also write from (10-3)

$$B_g(\theta) = A_g(\theta) - S_g(\theta). \tag{10-5}$$

Here we shall define $\alpha_n(\theta)$ and $\beta_n(\theta)$ so that

$$\alpha_n(\theta) = \sum_{g=1}^{n} A_g(\theta) \tag{10-6}$$

and

$$\beta_n(\theta) = \sum_{g=1}^{n} B_g(\theta). \tag{10-7}$$

In order to obtain the value of maximum likelihood estimate $\hat{\theta}$ from a given response pattern on some specified model, the term to be set equal to zero should be given by

$$\begin{aligned} \sum_{g \in G} A_g(\theta) + \sum_{h \in \bar{G}} B_h(\theta) &= \alpha_n(\theta) - \sum_{h \in \bar{G}} S_h(\theta) \\ &= \beta_n(\theta) + \sum_{g \in G} S_g(\theta), \end{aligned} \tag{10-8}$$

where G means the set of items to which the responses of a given response pattern are positive, and $\bar{G}$ is that of items to which the answers are negative. Since both $\alpha_n(\theta)$ and $\beta_n(\theta)$ have fixed values for a given set of n items and a given value of θ, the term which determines the value of estimate for a given response pattern will be either

$$-\sum_{h \in \bar{G}} S_h(\theta) \tag{10-9}$$

or

$$\sum_{g \in G} S_g(\theta), \tag{10-10}$$

and, therefore, the functional form for $S_g(\theta)$ takes an important role in determining the order of the estimates.

From the definition of $S_g(\theta)$ we can easily see that, in the case where the item characteristic function is symmetric with $(b_g\ ,\ 0.5)$ as the center of symmetry and the relationship

$$P_g(\theta) = Q_g(2b_g - \theta) \tag{10-11}$$

holds, where b_g is the value of θ at which

$$P_g(b_g) = Q_g(b_g) = 0.5, \tag{10-12}$$

$S_g(\theta)$ becomes a symmetric curve with $\theta = b_g$ as its axis of symmetry. This means that in such a case $S_g(\theta)$ can never be monotonically increasing nor decreasing in θ, but can be constant, symmetrically convex, etc.

In the case of the logistic model for dichotomous items, which evidently satisfies conditions (i) and (ii)*, and where the item characteristic function is given by

$$P_g(\theta) = \frac{1}{1 + e^{-Da_g(\theta - b_g)}}\ , \tag{10-13}$$

we obtain a formula for $S_g(\theta)$ so that

$$S_g(\theta) = Da_g\ , \tag{10-14}$$

since we have by differentiating (10-13)

$$\frac{\partial}{\partial \theta} P_g(\theta) = Da_g P_g(\theta) Q_g(\theta). \tag{10-15}$$

This result indicates that $S_g(\theta)$ on this model is constant throughout the whole range of θ, and its magnitude is determined solely by discriminating parameter a_g , without being affected by the value of difficulty parameter b_g . Since $\alpha_n(\theta)$ is a monotonically decreasing function of θ and the term to be subtracted from $\alpha_n(\theta)$ is given by

$$\sum_{h \in \bar{G}} S_h(\theta) = D \sum_{h \in \bar{G}} a_h\ , \tag{10-16}$$

it is obvious that the less the value of (10-16) is, the higher is the value of the estimate, regardless of the difficulties of given items. This conclusion is, of course, congruent with the fact that a sufficient statistic t exists for θ on the logistic model for dichotomous items which is given by

$$t(V) = \sum_{g \in G} a_g \tag{10-17}$$

(Birnbaum in Lord & Novick, 1968).

In contrast with the above result obtained on the logistic model, some interesting facts are observed on the normal ogive model, which provides the item characteristic function given by

$$P_g = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_g(\theta - b_g)} e^{-t^2/2}\, dt, \tag{10-18}$$

where t is a dummy variable in this case. On this model $S_g(\theta)$ is not constant for all the values of θ, but is a symmetrically convex curve with

$$\frac{2\sqrt{2}}{\sqrt{\pi}}\, a_g \tag{10-19}$$

as its local minimum attained at the point

$$\theta = b_g \tag{10-20}$$

and with ∞ as its two upper asymptotes. That is to say, the minimal value of $S_g(\theta)$ and the steepness of its curve are determined by the magnitude of a_g, the discrimination index, and the horizontal position of the curve is determined by b_g, the difficulty index of item g.

Thus it has been made clear that on the normal ogive model not only the discriminating powers of items but also the difficulties take part in determining the value of the estimate for a specified response pattern.

We could classify all the possible relationships between $S_g(\theta)$ and $S_h(\theta)$ on the normal ogive model for dichotomous items into the following four categories.

(1) The two curves never converge throughout the whole range of θ.

In a special case where the following relationships

$$\left.\begin{aligned} b_g &= b_h \\ a_g &> a_h \end{aligned}\right\} \tag{10-21}$$

hold, we obtain a strict inequality between $S_g(\theta)$ and $S_h(\theta)$ so that

$$S_g(\theta) > S_h(\theta) \tag{10-22}$$

throughout the range of θ. Then we could easily understand that, whenever the rest of the elements of two response patterns are the same, the response pattern in which the response to g is positive and that to h is negative is given a higher value of estimate than the other, in which the response to h is positive and that to g is negative. For convenience, hereafter, we shall call the former response pattern $\bar{h}$ and the latter $\bar{g}$. This is a typical example of category (1). Inequality (10-22) holds, however, in many other instances where equality between b_g and b_h does not hold, as illustrated in Figure 10-1. In this figure the values of a_g and b_g are given in the parentheses attached to each curve. We can see that curves B and D are in the relationship expressed by (10-21)

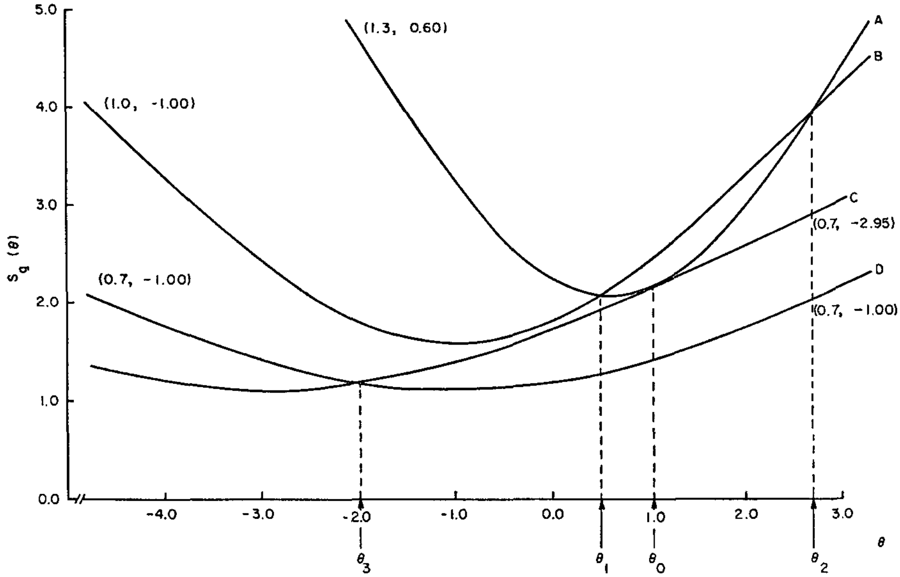


FIGURE 10–1
Examples of the relative positions of curves given by $S_g(\theta)$.

and never converge, and also curves B and C never converge in spite of the fact that the values of the b_g's are considerably different from each other, i.e., -1.00 and -2.95.

(2) The two curves are tangent at one point.

In this case the relationship between $S_g(\theta)$ and $S_h(\theta)$ is expressed by the following inequality

(10-23) $$S_g(\theta) \geq S_h(\theta),$$

where equality holds only for one value of θ. The relative magnitudes of item parameters are given by

(10-24) $$\left.\begin{aligned} a_g &> a_h \\ b_g &\neq b_h \end{aligned}\right\},$$

just like many instances involved in category (1). In Figure 10-1 curves A and C illustrate this relationship, and the value of θ at which these two curves are tangent is denoted by θ_0 , although this figure is more or less schematic and the value of parameter b_g attached to curve C is an approximated one. Except for a special case where the curve expressed by

(10-25) $$\alpha_n(\theta) - \sum_{\substack{i \in G \\ i \neq g, h}} S_i(\theta)$$

intersects curves A and C at the point of their convergence, the same principle for the relative magnitudes of estimates assigned to response patterns $\bar{h}$ and $\bar{g}$ described in the previous category it also valid. In that special case where the three curves cross one another at one point, the same estimate value, θ_0 , is assigned to both response patterns $\bar{h}$ and $\bar{g}$.

(3) The two curves intersect each other at two points.

In this case the relationship between $S_g(\theta)$ and $S_h(\theta)$ is given by

$$\left.\begin{array}{lll} S_g(\theta) > S_h(\theta) & \text{for} \quad \theta < \theta_1 \quad \text{and} \quad \theta > \theta_2 \\ S_g(\theta) = S_h(\theta) & \text{for} \quad \theta = \theta_1 \quad \text{and} \quad \theta = \theta_2 \\ S_g(\theta) < S_h(\theta) & \text{for} \quad \theta_1 < \theta < \theta_2 \end{array}\right\}, \tag{10-26}$$

where

$$\theta_1 < \theta_2 . \tag{10-27}$$

The relationships between the item parameters of g and h are the same as those given by (10-24). In Figure 10-1 the above relationship between $S_g(\theta)$ and $S_h(\theta)$ is illustrated by curves A and B. It is obvious that the relative magnitudes of estimates assigned to response patterns $\bar{h}$ and $\bar{g}$ are reversed whenever the curve expressed by (10-25) intersects curves A and B within the range $\theta_1 < \theta < \theta_2$, and otherwise the same relationship as the one between curves A and C also holds.

(4) The two curves intersect each other at one point.

The relationship between $S_g(\theta)$ and $S_h(\theta)$ is expressed as follows.

$$\left.\begin{array}{lll} S_g(\theta) > S_h(\theta) & \text{for} & \theta < \theta_3 \\ S_g(\theta) = S_h(\theta) & \text{for} & \theta = \theta_3 \\ S_g(\theta) < S_h(\theta) & \text{for} & \theta > \theta_3 \end{array}\right\} \tag{10-28}$$

In this category situations are limited to the ones in which

$$\left.\begin{array}{l} a_g = a_h \\ b_g > b_h \end{array}\right\}, \tag{10-29}$$

provided that the range of θ is $(-\infty, \infty)$. In Figure 10-1 curves D and C illustrate this relationship.

The relative magnitudes of the estimates assigned to response patterns $\bar{h}$ and $\bar{g}$ are reversed depending upon whether the curve expressed as (10-25) intersects these two curves within the range of θ less than θ_3 or greater than θ_3 . If it crosses the two curves exactly at their intersection, the same value, θ_3 , is assigned to both response patterns as their estimates.

So far we have seen how complex the factors for determining the relative magnitudes of the estimates assigned to specified response patterns on the normal ogive model are, even if we take a simple situation where only response patterns $\bar{h}$ and $\bar{g}$ are taken into consideration.

Among those complex relationships, category (4) supplies a relatively simple principle, and for this reason we shall observe this particular case for a while. In this case we can easily see that

$$\theta_3 = \frac{b_g + b_h}{2}. \tag{10-30}$$

Suppose that a test consists of only two dichotomous items, g and h, whose relationship is classified into category (4). We shall express a response pattern supplied by this test as

$$V = (x_h \, , x_g), \tag{10-31}$$

which leads to the result that response patterns $\bar{h}$ and $\bar{g}$ are represented by (0, 1) and (1, 0) respectively.

We can write from equations (10-1), (10-2), and (10-8) that

$$\begin{aligned} \alpha_2(\theta) - S_h(\theta) &= \frac{(\partial/\partial\theta)P_g(\theta)}{P_g(\theta)} + \frac{(\partial/\partial\theta)Q_h(\theta)}{Q_h(\theta)} \\ &= \frac{(\partial/\partial\theta)P_g(\theta)}{P_g(\theta)} - \frac{(\partial/\partial\theta)P_h(\theta)}{1 - P_h(\theta)} \end{aligned} \tag{10-32}$$

and

$$\begin{aligned} \alpha_2(\theta) - S_g(\theta) &= \frac{(\partial/\partial\theta)P_h(\theta)}{P_h(\theta)} + \frac{(\partial/\partial\theta)Q_g(\theta)}{Q_g(\theta)} \\ &= \frac{(\partial/\partial\theta)P_h(\theta)}{P_h(\theta)} - \frac{(\partial/\partial\theta)P_g(\theta)}{1 - P_g(\theta)}, \end{aligned} \tag{10-33}$$

both of which equal zero simultaneously when $\theta = b_g + b_h/2$. In other words the curve given by $\alpha_2(\theta)$ for this test, which coincides with the term expressed by (10-25) in this particular case, intersects each of $S_g(\theta)$ and $S_h(\theta)$ at exactly the same point where these two curves cross each other. Thus the same value, $b_g + b_h/2$, is assigned to both response patterns, (0, 1) and (1, 0), as the estimates.

In the upper figure of Figure 10-2, these three curves are drawn by solid lines. In this example, $b_g = -0.5$, $b_h = -1.5$, and $a_g = a_h = 1.0$.

This fact suggests that if a test consists of more than two dichotomous items including items g and h satisfying the condition given by (10-29), the value of estimate assigned to a special case of response pattern $\bar{h}$, in which all the item responses to the other $(n - 2)$ items exclusive of items g and h are 0, is greater than the one assigned to the corresponding case of response pattern $\bar{g}$, while the value of estimate assigned to another special case of response pattern $\bar{h}$, in which all the item responses to the other $(n - 2)$ items are 1,

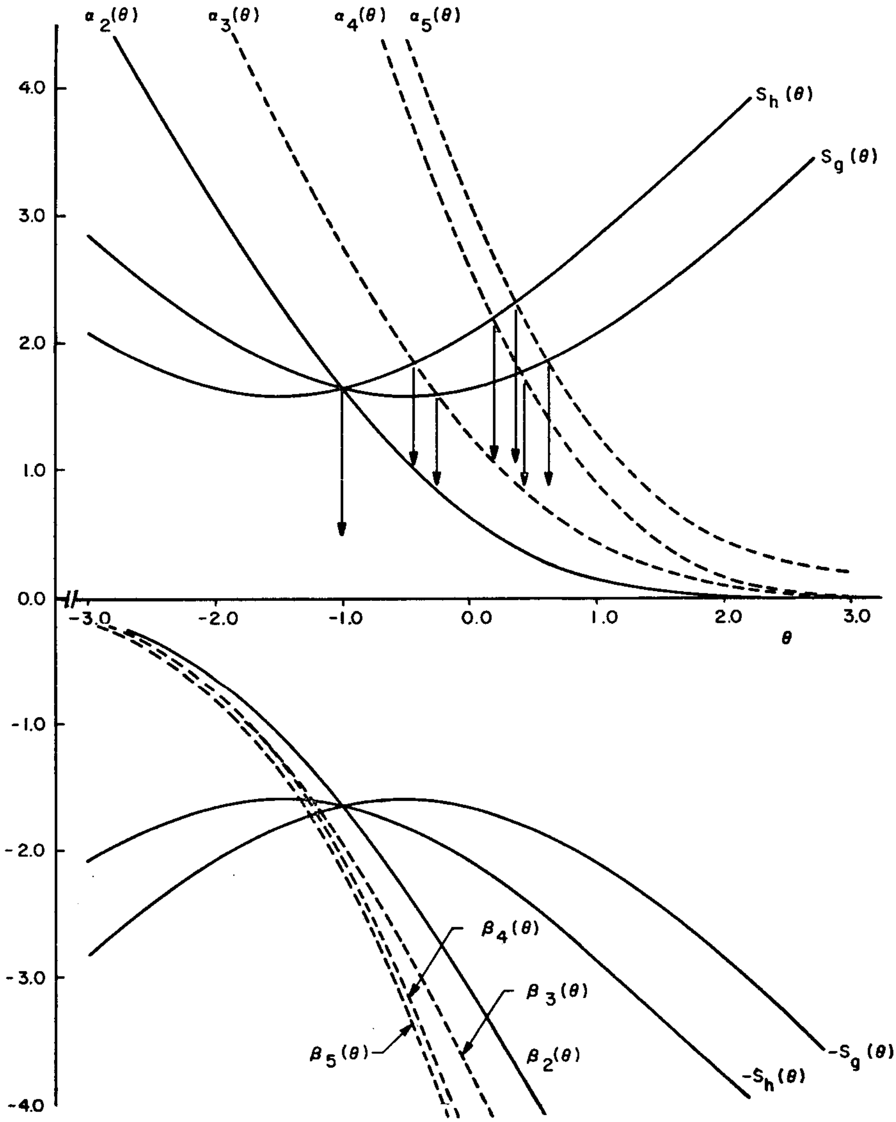


FIGURE 10–2
Relative positions of $S_g(\theta)$, $S_h(\theta)$ and $\alpha_n(\theta)$; $-S_g(\theta)$, $-S_h(\theta)$ and $\beta_n(\theta)$, when $n = 2, 3, 4$ and 5.

is less than the one assigned to the corresponding case of response pattern $\bar{g}$. In the upper figure of Figure 10-2 the situation held in the latter case is illustrated for three cases where $n = 3$, $n = 4$ and $n = 5$, and the parameter values of the items additional to g and h are $a_i = 0.8$ and $b_i = 0.0$ in the case where $n = 3$, $a_i = 1.2$ and $b_i = 0.4$, together with the previous one in the case where $n = 4$, $a_i = 0.4$ and $b_i = 1.5$ in addition to the previous two in the case where $n = 5$. That is to say, the curve drawn by the dotted line in this figure represents $\alpha_n(\theta)$ in each case, which coincides with the term expressed as (10-25). It is obvious from the definition of $\alpha_n(\theta)$ that we always have

$$\alpha_n(\theta) > \alpha_{n-1}(\theta), \tag{10-34}$$

for the entire range of θ, if n is greater than one.

In the lower figure of Figure 10-2 the relationship held in the former case is illustrated for the same three instances. The curve drawn by the dotted line in this figure represents $\beta_n(\theta)$ in each case, which coincides with the term

$$\beta_n(\theta) + \sum_{\substack{i \in G \\ i \neq g, h}} S_i(\theta) \tag{10-35}$$

in this specific case, and the curves drawn by solid lines are $-S_g(\theta)$ and $-S_h(\theta)$ respectively. It is also apparent from the definition of $\beta_n(\theta)$ that

$$\beta_n(\theta) < \beta_{n-1}(\theta) \tag{10.36}$$

always holds for the entire range of θ, provided that n is greater than one.

Now let us consider a situation in which examinees are required to solve n dichotomous items, all of whose discriminating powers are the same, but whose difficulty levels are different from one another. For simplicity, let us suppose that $n = 5$ and the five items are denoted by 1, 2, 3, 4 and 5, in the order of easiness. If there are five examinees who have tried to solve all the five items but succeeded in only one, and each item solved is different from each other, to which of the five examinees should be assigned the highest value of estimate? The answer to this question may largely depend upon subjective judgments or preferences. On the normal ogive model, however, the answer is definite, as we can easily observe by following the preceding reasoning. We can arrange the five response patterns in the order of high evaluation as the following.

$$\begin{array}{c}
(0,\ 0,\ 0,\ 0,\ 1) \\
(0,\ 0,\ 0,\ 1,\ 0) \\
(0,\ 0,\ 1,\ 0,\ 0) \\
(0,\ 1,\ 0,\ 0,\ 0) \\
(1,\ 0,\ 0,\ 0,\ 0)
\end{array}$$

This fact appears to suggest that the philosophy of scoring underlying the normal ogive model is such that an examinee is evaluated as low in the ability tested because he can only solve an easy item, while an examinee who has solved a difficult item is evaluated as high in the ability tested because of the difficulty of the item solved.

Now let us consider another instance. If there are another five examinees who have also tried to solve the same five items and who have succeeded in four of them, but failed in only one, and each item failed is different from each other, to which of these five examinees should be assigned the highest value of estimate? On the normal ogive model, again we can easily see that these five response patterns are arranged in the order of the magnitudes of estimates as follows.

$$
\begin{array}{l}
(1,\ 1,\ 1,\ 1,\ 0)\\
(1,\ 1,\ 1,\ 0,\ 1)\\
(1,\ 1,\ 0,\ 1,\ 1)\\
(1,\ 0,\ 1,\ 1,\ 1)\\
(0,\ 1,\ 1,\ 1,\ 1)
\end{array}
$$

In this case the philosophy of scoring seems to be that an examinee is evaluated as low in the ability tested because of the fact that he cannot solve even an easy item, while an examinee who has failed in a difficult item is evaluated as high in the ability tested because it is no proof of his inferiority that he has failed in a difficult item.

The above two philosophies are, in one sense, contradictory with each other, since the principle is completely reversed. That is to say, the difficulty of an item in the former instance is treated just as the easiness of an item in the latter instance, and vice versa. If we apply the principle in the former case to the latter, the order of evaluation of the five response patterns should be reversed, since, for instance, an examinee with response pattern (0, 1, 1, 1, 1) has succeeded in the most difficult four items, while an examinee with response pattern (1, 1, 1, 1, 0) has succeeded in the easiest four items.

This fact is not necessarily considered as contradictory. As was mentioned earlier in this chapter, it may largely depend upon subjective judgments or preferences. It can be integrated by the statement that stress is put upon the difficulty of the solved item if the number of the solved items is small, and it is put upon the easiness of the unsolved item if the number of the unsolved items is small. This is the due result obtained from the symmetric formula for the item characteristic function expressed by (10-11) and (10-12), with which both success and failure are treated exactly in the same way.

If the number of succeeded items is more than one and less than four, the order of the magnitudes of estimates assigned to the response patterns on the normal ogive model is much less predictable, as we can foresee from the earlier observations. We cannot predict so easily, for example, the result obtained for a rather simple case in which all the distances between two adjacent difficulty parameters of the items are the same. For the purpose of illustration, the result obtained for the response patterns in which three responses are positive, by setting $b_1 = -2.0$, $b_2 = -1.0$, $b_3 = 0.0$, $b_4 = 1.0$ and $b_5 = 2.0$ and all the values of discriminating parameters are 1.0, is shown below. The range of resulting estimates is approximately 0.07 through 0.67.

$$
\begin{array}{ccccc}
(1, & 1, & 0, & 0, & 1) \\
(1, & 1, & 1, & 0, & 0) \\
(1, & 1, & 0, & 1, & 0) \\
(1, & 0, & 1, & 0, & 1) \\
(1, & 0, & 0, & 1, & 1) \\
(1, & 0, & 1, & 1, & 0) \\
(0, & 1, & 0, & 1, & 1) \\
(0, & 1, & 1, & 0, & 1) \\
(0, & 0, & 1, & 1, & 1) \\
(0, & 1, & 1, & 1, & 0)
\end{array}
$$

Since on the logistic model, which is used as an approximation to the normal ogive model, all the values of estimates assigned to the above response patterns are the same (approximately 0.52), this diversity of the values of estimates obtained on the normal ogive model is by no means negligible. And yet it is hard for us to induce a simple philosophy of scoring from the above result, although we can understand why it is so complicated from the earlier observations about $S_g(\theta)$, $\alpha_n(\theta)$, and others.

In the author's experience, even the integrated explanation about the philosophy of scoring on the normal ogive model when we take only the first two observations into our consideration is the subject of opposition for many people's subjective and intuitive judgments. They approve of the principle suggested in the first instance where the response patterns with only one element of success are treated, but oppose the other which is suggested in the second instance where the response patterns with four elements of success are dealt with.

Thus we may have to find some other model on which the philosophy of scoring is such that, among the response patterns with the same number of positive responses, higher values of estimates are assigned to response patterns in which responses to more difficult items are positive, while lower values of estimates are assigned to response patterns in which less difficult items are responded to correctly, in order to satisfy their intuitions. The item characteristic function on such a model should provide a monotonically decreasing function for $S_g(\theta)$, whose horizontal position should be determined by the difficulty of the item. In that case, the item characteristic function cannot be symmetric, as is clear from the earlier observations.

For illustrative purposes, we shall take one specified model here, although we can formulate many other models satisfying the above requirement. On this new model the item characteristic function is given by the following formula, (10-37) $$P_g(\theta) = \{1 + e^{-Da_g(\theta - b_g)}\}^{-2}.$$

We can see that if an item characteristic function, $P_g(\theta)$, can be expressed by

(10-38) $$P_g(\theta) = \{P_g^{**}(\theta)\}^2,$$

where $P_g^{**}(\theta)$ is also the item characteristic function defined on another model, which provides the basic functions satisfying conditions (i) and (ii)* described in Chapter 3, we shall obtain

(10-39) $$A_1(\theta) = \frac{(\partial/\partial\theta)P_g(\theta)}{P_g(\theta)} = \frac{2(\partial/\partial\theta)P_g^{**}(\theta)}{P_g^{**}(\theta)}$$

and

(10-40) $$A_0(\theta) = \frac{(\partial/\partial\theta)Q_g(\theta)}{Q_g(\theta)} = \frac{2P_g^{**}(\theta)}{\{1 + P_g^{**}(\theta)\}} \frac{(\partial/\partial\theta)Q_g^{**}(\theta)}{Q_g^{**}(\theta)}.$$

From these two equations it is concluded that the new model also provides the basic functions satisfying conditions (i) and (ii)*. Since the item characteristic function given by equation (10-37) is the square of the one on the logistic model, the new model also provides basic functions satisfying conditions (i) and (ii)*. Thus it has been proved that the unique maximum likelihood estimator exists with respect to any possible response pattern on this new model, except for two extreme cases where all the elements of the response pattern are uniformly 0 or 1.

From (10-38) and (10-39) we have

(10-41) $$S_g(\theta) = \frac{(\partial/\partial\theta)P_g(\theta)}{P_g(\theta)Q_g(\theta)} = \frac{2}{1 + P_g^{**}(\theta)} \frac{(\partial/\partial\theta)P_g^{**}(\theta)}{P_g^{**}(\theta)Q_g^{**}(\theta)}$$

and on this specific model we can write

(10-42) $$S_g(\theta) = \frac{2Da_g}{1 + P_g^{**}(\theta)}.$$

Since by differentiating (10-42) we have

(10-43) $$\frac{\partial}{\partial\theta} S_g(\theta) = -\frac{2D^2a_g^2P_g^{**}(\theta)Q_g^{**}(\theta)}{\{1 + P_g^{**}(\theta)\}^2} < 0,$$

it is evident that $S_g(\theta)$ is monotonically decreasing in θ.

The upper and lower asymptotes for $S_g(\theta)$ are given by

(10-44) $$\lim_{\theta \to -\infty} S_g(\theta) = 2Da_g$$

and

(10-45) $$\lim_{\theta \to \infty} S_g(\theta) = Da_g \ ,$$

and further we have

(10-46) $$S_g(\theta) = \frac{2Da_g}{1.5} \ , \qquad \text{for} \quad \theta = b_g \ .$$

The upper figure of Figure 10-3 illustrates the four examples of the item characteristic functions where $D = 1.702$ and whose parameter values, a_g and b_g , are shown in parentheses. In the lower figure of Figure 10-3 are shown the corresponding item characteristic functions obtained by transforming θ into τ by

(10-47) $$\tau = e^{\theta},$$

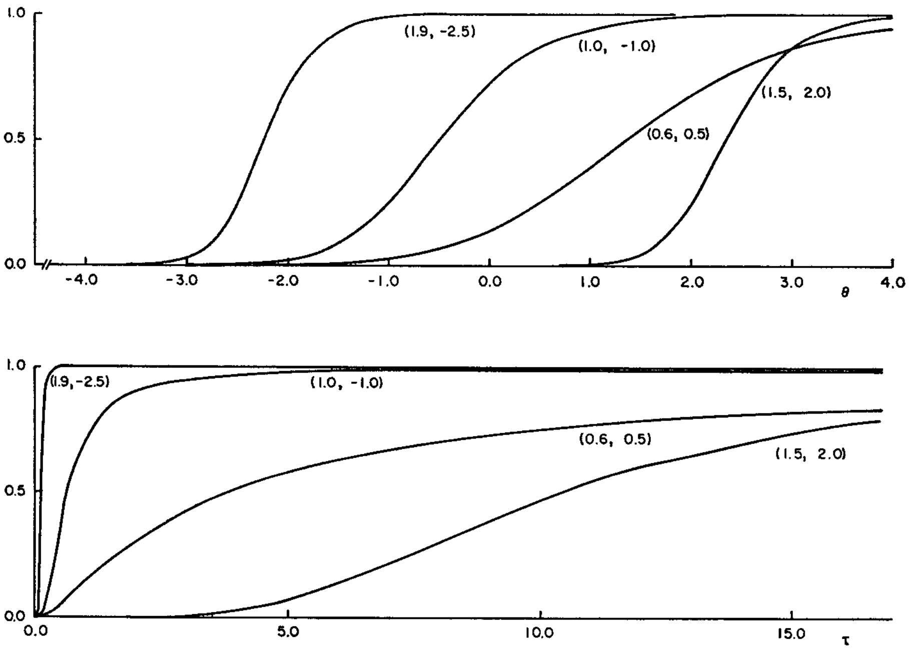


FIGURE 10–3
Item characteristic functions for four hypothetical items on the model defined by equation (10-37) and their exponential transformations.

which also provide the maximum likelihood estimator $\hat{\tau}$, as we have seen earlier.

Figure 10-4 presents eight examples of $S_g(\theta)$, in which $D = 1.702$ and the values of a_g are 1.9, 1.7, 1.5, 1.3, 1.1, 0.9, 0.7 and 0.5, respectively, and all the values of b_g are 0.0.

If we apply this model to the earlier instances of five items instead of the normal ogive model, the order of arrangement of five response patterns with only one positive response is the same as before, but that of the other five response patterns with four positive responses is completely reversed. The order of arrangement of 10 response patterns with three positive responses is quite different from the previous one, as shown below.

$$
\begin{array}{l}
(0,\ 0,\ 1,\ 1,\ 1)\\
(0,\ 1,\ 0,\ 1,\ 1)\\
(1,\ 0,\ 0,\ 1,\ 1)\\
(0,\ 1,\ 1,\ 0,\ 1)\\
(1,\ 0,\ 1,\ 0,\ 1)\\
(1,\ 1,\ 0,\ 0,\ 1)\\
(0,\ 1,\ 1,\ 1,\ 0)\\
(1,\ 0,\ 1,\ 1,\ 0)\\
(1,\ 1,\ 0,\ 1,\ 0)\\
(1,\ 1,\ 1,\ 0,\ 0)
\end{array}
$$

The result is quite predictable, as we can easily see. The basic functions for these response patterns are given elsewhere (cf. Samejima, 1968a, Appendix 6), together with those on the normal ogive model.

This is only an example of the possibilities of using the philosophy of

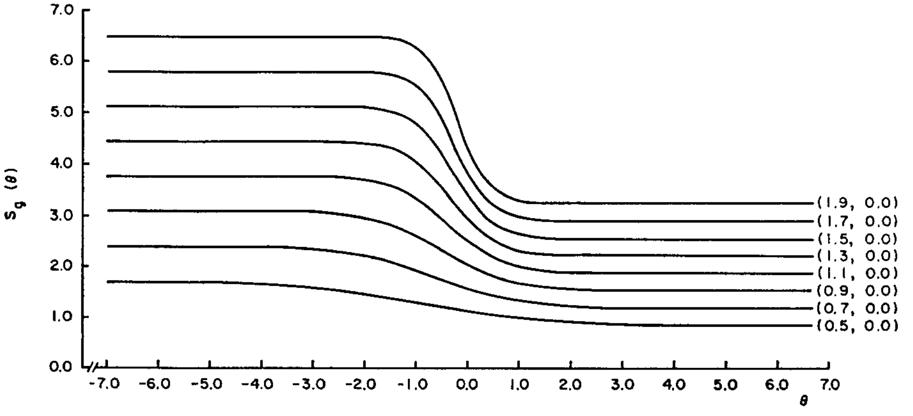


FIGURE 10-4

$S_g(\theta)$ for eight hypothetical items on the model defined by equation (10-37).

scoring in formulating a model. Since in practical situations it is not likely that we solve a very difficult item simply because of the good conditions brought about by chance factors, while it is much more likely that we fail in a very easy item simply because of chance factors, it may be advisable that the treatment of success and failure in a symmetric way should be modified at least in some cases. Asymmetric functional forms for the item characteristic function, therefore, should be investigated from various aspects for that purpose.

CHAPTER 11

DISCUSSION AND PRACTICAL IMPLICATIONS

Sufficient conditions for the existence of a unique maximum likelihood estimator and that of a unique Bayes modal estimator were formulated; and in line with these conditions operating characteristics of graded responses when the thinking process is homogeneous were introduced and discussed especially in connection with the normal ogive model and the logistic model. It has been made clear that the amount of information given by an individual item or by a test will substantially increase if we score an item in a graded way by using data on the normal ogive model for graded item responses as an example. It has also been made clear that the estimator specified on the entire response pattern has a substantial advantage to the one defined on the simple test score on the normal ogive model, when the values of item discriminating parameters are considerably different from one another, in the sense that it provides us with substantially different values of estimates for individual response patterns, reduces the standard errors of measurement when the estimator is the expected value, and decreases the mean-square errors. The Bayes modal estimator $\hat{\theta}$ proved to be a good approximation to the Bayes estimator, the latter being the most accurate estimator in the sense that the expectation of its mean-square errors is the smallest. The relationship between the formula for the item characteristic function and the philosophy of scoring was observed and the utility of asymmetric functional form for the item characteristic function was suggested.

Throughout this paper the principle of local independence is assumed, and testing situations in which a relatively small number of very meaningful items are practiced have been focused on, rather than the situations where a large number of less meaningful items are presented to examinees just as in ordinary paper-and-pencil testings. The method of obtaining the maximum likelihood estimator or the Bayes modal estimator on the response pattern is useful in such situations, especially when it is combined with the graded way of scoring. In sequential testing situations, another advantage is that we could make use of this method in making the computer decide the difficulty of the item which should be presented next by depending solely upon the previous response given by the examinee, since the values of these estimates can easily be obtained by simple addition of the basic functions of given individual responses, provided that they satisfy the requirements stated in conditions (i) and (ii), or, in the case of the Bayes modal estimator $\hat{\theta}_V$, the requirements stated in conditions (i) and (ii)*, and that the latent density function satisfies the requirements stated in conditions (iii) and (iv). This is also possible when we use the Bayes estimator μ'_{1V} on the response pattern as our estimator, although the time required for the computer to decide the difficulty of the next item may be longer. In order to realize the rapid convergence of the estimated values to the true value of examinee's ability, the latent density function should have been obtained for a specified concentrated group of people to which the examinee belongs. Further, it may be possible for us to assume that an examinee's ability itself has a distribution, and to try to specify the functional form for it. If we have succeeded in specifying this distribution, we could use it in the estimation. Especially in the case of sequential testing, this distribution may have to be specified not only on spatial axes, but on temporal axes, since there may be warming-up effects, fatigue and other factors which effect the examinee's ability.

In this paper we avoided dealing with multiple-choice items. One reason why multiple-choice items were not treated here is that the conventional formula for the multiple-choice item characteristic function does not seem appropriate at least to several important cases. Another reason is that, in the conventional usage of multiple-choice alternatives, alternatives themselves may destroy the nature of the item since they give an examinee suggestions for the correct response, and, consequently, the quality of what should be measured may be altered, especially in the measurement of the profound thinking process. This defect will be diminished, however, by such a device as setting a pair of alternatives to one item [Noryoku-Kaihatsu-Kenkyujo, 1966], and with this multiple-choice items can be treated just as free-response items, so that there is no need to modify the model particularly for multiple-choice items. The discussion about the multiple-choice situation will be made separately.

In the measurement of any psychological trait, we can make use of any specified response other than the positive one as a source of information, if only it has high discriminating power for the trait measured. For example, we could make use of some incorrect but plausible response to an item in the measurement of ability, which is often used as a distracter in multiple-choice alternatives, since it requires a certain level of ability for an examinee to discover its plausibility.

In order to specify operating characteristics for such responses, we must find a way to approach operating characteristics without formulating any particular model, and by reducing as many assumptions as possible.

REFERENCES

Birnbaum, A. Efficient design and use of tests of a mental ability for various decision making problems. *Series Report No. 58-16*, Project No. 7755-23, USAF School of Aviation Medicine, Randolph Air Force Base, Texas, January 1957.

Birnbaum, A. On the estimation of mental ability. *Series Report No. 15*, Project No. 7755-23, USAF School of Aviation Medicine, Randolph Air Force Base, Texas, 1958. (a)

Birnbaum, A. Further considerations of efficiency in tests of a mental ability. *Technical Report No. 17*, Project No. 7755-23, USAF School of Aviation Medicine, Randolph Air Force Base, Texas, 1958. (b)

Indow, T., & Samejima, F. *LIS measurement scale for non-verbal reasoning ability*. Tokyo: Nippon Bunka Kagakusha, 1962. (In Japanese)

Indow, T., & Samejima, F. On the results obtained by the absolute scaling model and the Lord model in the field of intelligence. *Third Report of the Psychological Laboratory on the Hiyoshi Campus*, Keio University, Japan, 1966.

Kendall, M. G., & Stuart, A. *The advanced theory of statistics*. Vol. 2. London: Griffin, 1961.

Lazarsfeld, P. F. Latent structure analysis. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 3. New York: McGraw-Hill, 1959.

Lazarsfeld, P. F. Latent structure analysis and test theory. In H. Gulliksen and S. Messick (Ed.), *Psychological scaling: Theory and applications*. New York: Wiley, 1960.

Lord, F. M. A theory of test scores. *Psychometric Monographs*, 1952, No. 7.

Lord, F. M. An application of confidence intervals and of maximum likelihood to the estimation of an examinee's ability. *Psychometrika*, 1953, **18**, 57–76.

Lord, F. M., & Novick, M. R. *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley, 1968.

Noryoku-Kaihatsu-Kenkyujo (Educational Test Research Institute) *Manual for scholastic aptitude test*. Tokyo: Noryoku-Kaihatsu-Kenkyujo, 1966. (In Japanese)

Samejima, F. Development and application of Lord's theory of test scores. *Bulletin of Nippon Research Center*, 1962, **1**, No. 1, 60–69. (In Japanese) (a)

Samejima, F. Scale construction when grades are introduced to the responses to each item. *Bulletin of Nippon Research Center*, 1962, **1**, No. 2, 61–74. (In Japanese) (b)

Samejima, F. A general model for the operating characteristic of graded item response. *The L. L. Thurstone Psychometric Laboratory Report*, 1967, **55**. Chapel Hill, North Carolina: The University of North Carolina.

Samejima, F. Estimation of latent ability using a response pattern of graded scores. *Research Bulletin*, 1962, 2. Princeton, New Jersey: Educational Testing Service. (a)

Samejima, F. General considerations on the homogeneous case of graded item response. *The L. L. Thurstone Psychometric Laboratory Report*, 1968, **57**. Chapel Hill, North Carolina: The University of North Carolina. (b)

APPENDIX

Item C_2 , $A_{2.2}$ and $D_{2.2}$ of LIS Measurement Scale for Non-verbal Reasoning Ability (translated from Japanese into English)

(i) Item C_2 (time limit: 10 min.)

This problem is to classify figures according to a certain principle, and you are asked to discover that principle.

Example:

Figures in the same set

Figures disqualified to join that set

In the above example, the five figures on the left-hand side are in one and the same set, while the other five figures on the right-hand side are disqualified to join it. If you watch and compare these two groups of figures carefully, you can find out the principle underlying the classification. In this example, the qualification for a figure to belong to that set is:

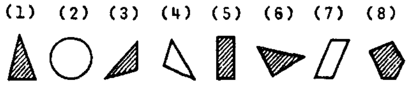
(a) that it is a triangle,

and

(b) that it is shadowed.

Examples:

(1) (2) (3) (4) (5) (6) (7) (8)



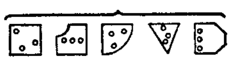
If you examine each of the above eight figures carefully as to whether it is qualified to join the set shown above when they are classified according to that principle, you will get the answer listed as below:

	1	2	3	4	5	6	7	8
being a triangle	*		*	*		*		
being shadowed	*		*		*	*		*

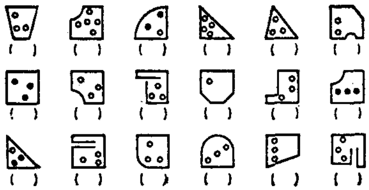
The mark, *, in the above list indicates that the figure with the number shown above meets the corresponding requirement written on the left end of the list. In this example, a figure is disqualified to join the set unless it meets both of these two requirements, so only three figures, (1), (3) and (6), are qualified to join the set among the eight.

In the problem presented below, you are to watch and compare carefully the first five figures "belonging to Group K" with the succeeding eight figures "not belonging to Group K," and to discover the principle of classification. Then classify the figures given below according to the principle you have found, by inserting a circle in the parentheses below each of the figures if you consider it pertains to Group K and by inserting a cross in it if you think it does not, and state briefly the requirements for a figure to join Group K.

Examples of figures belonging to Group K:



Examples of figures disqualified to pertain to Group K:



Requirements for a figure to join Group K:

(ii) <u>Item ($A_{2.1}$ and) $A_{2.2}$.</u> (time limit: 18 min.)

This is a problem of multiplication. Alphabetical letters, A, B, C, etc. are used instead of numbers. You are to discover which letter represents which number.

Example:

$$\begin{array}{r} K\ I \\ \times\quad 2 \\ \hline A\ I \end{array} \qquad (0,\ 3,\ 6)$$

In the above multiplication, letters used are A, I and K, and the numbers represented by them are those shown in the parentheses. The answer of this example is A = 6, I = 0, and K = 3, because, if so, the above equation becomes:

$$\begin{array}{r} 3\ 0 \\ \times\quad 2 \\ \hline 6\ 0 \end{array}$$

and gives a reasonable calculation.

In this way, in each of the following test items, you must find out which letter represents which one of those numbers shown in the parentheses and write the corresponding number under each of the alphabetical letters arranged below the parentheses.

You must keep in mind that only one letter signifies only one number. In the above example, for instance, it never happens that "I" represents "6" in one place and "0" in the other, or both "A" and "K" signify "3". Also you must watch never to use other numbers than those shown in the parentheses.

[1]

```
  I K A
x     3        (0 1 5 6 7)
-------
  E B I         A B E I K
                ‖ ‖ ‖ ‖ ‖
```

[2]*

```
  A H I R U
x         5    (0 1 2 4 5 8)
-----------
H I B A R I     A B H I R U
                ‖ ‖ ‖ ‖ ‖ ‖
```

*Item $A_{2.2}$

(iii) Item ($D_{2.1}$ and) $D_{2.2}$ (time limit: 18 min.)

(Instructions similar to those of items $A_{2.1}$ and $A_{2.2}$ are given first, and then the following two items of addition are presented.)

[1]

```
  K A
+ K A          (1 2 5 6 7)
-----
S H I           A H I K S
                ‖ ‖ ‖ ‖ ‖
```

[2]*

```
  Y A M A
+   U M I      (0 3 4 6 7 8 9)
---------
  K A W A       A I K M U W Y
                ‖ ‖ ‖ ‖ ‖ ‖ ‖
```

*Item $D_{2.2}$