DIFFERENTIAL WEIGHT PROCEDURE OF
THE CONDITIONAL P.D.F. APPROACH
FOR ESTIMATING THE OPERATING
CHARACTERISTICS OF
DISCRETE ITEM RESPONSES

FUMIKO SAMEJIMA

UNIVERSITY OF TENNESSEE
KNOXVILLE, TENN. 37996-0900

JUNE, 1990

Prepared under the contract number N00014-87-K-0320,
4421-549 with the
Cognitive Science Research Program
Cognitive and Neural Sciences Division
Office of Naval Research

Approved for public release; distribution unlimited.
Reproduction in whole or in part is permitted for
any purpose of the United States Government.

R01-1069-11-001-91
A new procedure of nonparametric estimation of the operating characteristics of discrete item responses has been proposed, and it is called Differential Weight Procedure of the Conditional P.D.F. Approach. Some examples have been given, and sensitivities of the resulting estimated operating characteristics to irregularities of the differential weight functions have been observed and discussed. Usefulness of the method have also been discussed. These outcomes suggest the importance of further investigation of the weight function in the future.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Common Backgrounds and Differences among Different Procedures</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Simple Sum Procedure of the Conditional P.D.F. Approach Combined with the Normal Approach Method</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>Differential Weight Procedure</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>Examples</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>Sensitivities to Irregularities of Weight Functions</td>
<td>16</td>
</tr>
<tr>
<td>7</td>
<td>Usefulnesses of the Differential Weight Procedure</td>
<td>24</td>
</tr>
<tr>
<td>8</td>
<td>Discussion and Conclusions</td>
<td>24</td>
</tr>
</tbody>
</table>

REFERENCES
The research was conducted at the principal investigator's laboratory, 405 Austin Peay Bldg., Department of Psychology, University of Tennessee, Knoxville, Tennessee. Those who worked as assistants for this research include Christine A. Golik, Barbara A. Livingston, Lee Hai Gan and Nancy H. Domm.
I Introduction

In the past couple of decades the author has been engaged in the nonparametric estimation of the operating characteristics of discrete item responses in the context of latent trait models (cf. Samejima, 1981b, 1988). As early as in 1977 the author proposed Normal Approximation Method (Samejima, 1977b) which can be used for the item calibration both in computerized adaptive testing and in paper-and-pencil testing. She also discussed the effective use of information functions in adaptive testing (Samejima, 1977a). Since then, with the support by the Office of Naval Research, she has developed several approaches and methods for the same purpose (cf. Samejima, 1977c, 1978a, 1978b, 1978c, 1978d, 1978e, 1978f, 1980a, 1980b, 1981a; Samejima and Changas, 1981). For convenience, they can be categorized as follows.

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Bivariate P.D.F. Approach</td>
<td>(1) Pearson System Method</td>
</tr>
<tr>
<td>(2) Histogram Ratio Approach</td>
<td>(2) Two-Parameter Beta Method</td>
</tr>
<tr>
<td>(3) Curve Fitting Approach</td>
<td>(3) Normal Approach Method</td>
</tr>
<tr>
<td>(4) Conditional P.D.F. Approach</td>
<td>(4) Lognormal Approach Method</td>
</tr>
<tr>
<td></td>
<td>(4.1) Simple Sum Procedure</td>
</tr>
<tr>
<td></td>
<td>(4.2) Weighted Sum Procedure</td>
</tr>
<tr>
<td></td>
<td>(4.3) Proportioned Sum Procedure</td>
</tr>
</tbody>
</table>

Here by an approach we mean a general procedure in approaching the operating characteristics of a discrete item response, and by a method we mean a specific method in approximating the conditional density of ability, given its maximum likelihood estimate. Thus a combination of an approach and a method provides us with a specific procedure for estimating the operating characteristic of a discrete item response.

These approaches and methods are characterized by two features, i.e.,

(1) estimation is made without assuming any mathematical forms for the operating characteristics of discrete item responses, and

(2) estimation is efficient enough to base itself upon a relatively small set of data of, say, several hundred to a few thousand examinees.

The present paper proposes a method which increases accuracies of estimation of the operating characteristics of discrete item responses, especially when the true operating characteristic is represented by a steep curve, and also at the lower and upper ends of the ability distribution where the estimation tends to be inaccurate because of smaller numbers of subjects involved in the base data. Tentatively, it is called the Differential Weight Procedure, and it belongs to the Conditional P.D.F. Approach. This procedure costs more CPU time than the Simple Sum Procedure, which has been used frequently (cf. Samejima, 1981b, 1988), but the advantage of handling more than one item, say, fifty, together in the Conditional P.D.F. Approach is still there.

II Common Backgrounds and Differences among Different Procedures

Let $\theta$ be ability, or latent trait, which assumes any real number. We assume that there is a set of test items measuring $\theta$ whose characteristics are known. This set of test items is called Old Test, whose meaning is somewhat close to the original itempool in the adaptive testing situation.
Let \( h \) denote an item of the Old Test, \( k_h \) be a discrete item response to item \( h \), and \( P_{k_h}(\theta) \) be the operating characteristic of \( k_h \), or the conditional probability assigned to \( k_h \), given \( \theta \). We assume that \( P_{k_h}(\theta) \) is three-times differentiable with respect to \( \theta \). We have for the item response information function, \( I_{k_h}(\theta) \), (Samejima, 1969, 1972)

\[
I_{k_h}(\theta) = -\frac{\partial^2}{\partial \theta^2} \log P_{k_h}(\theta),
\]

and the item information function, \( I_h(\theta) \), is defined as the conditional expectation of \( I_{k_h}(\theta) \), given \( \theta \), such that

\[
I_h(\theta) = E[I_{k_h}(\theta) | \theta] = \sum_{k_h} I_{k_h}(\theta) P_{k_h}(\theta).
\]

Let \( V \) be a response pattern such that

\[
V = \{ k_h \}^t, \quad h = 1, 2, \ldots, n.
\]

The operating characteristic, \( P_V(\theta) \), of the response pattern \( V \) is defined as the conditional probability of \( V \), given \( \theta \), and by virtue of local independence we can write

\[
I_V(\theta) = \prod_{h \in V} P_{k_h}(\theta).
\]

The response pattern information function, \( I_V(\theta) \), is given by

\[
I_V(\theta) = -\frac{\partial^2}{\partial \theta^2} \log P_V(\theta) = \sum_{h \in V} I_{k_h}(\theta),
\]

and the test information function, \( I(\theta) \), is defined as the conditional expectation of \( I_V(\theta) \), given \( \theta \), and we obtain from (2.1), (2.2), (2.3), (2.4) and (2.5)

\[
I(\theta) = E[I_V(\theta) | \theta] = \sum_V I_V(\theta) P_V(\theta) = \sum_{h=1}^n I_h(\theta).
\]

For the sake of simplicity in handling mathematics, the tentative transformation of \( \theta \) to \( \tau \) is made by

\[
\tau = C_1^{-1} \int_{-\infty}^\theta [I(t)]^{1/2} dt + C_0,
\]

where \( C_0 \) is an arbitrary constant for adjusting the origin of \( \tau \), and \( C_1 \) is an arbitrary constant which equals the square root of the test information functions, \( I^*(\tau) \), of \( \tau \), so that we can write

\[
C_1 = [I^*(\tau)]^{1/2},
\]

for all \( \tau \). This transformation will be simplified if we use a polynomial approximation to the square root of the test information function, \( [I(\theta)]^{1/2} \), in the least squares sense which is accomplished by using the method of moments (cf. Samejima and Livingston, 1979) for the meaningful interval of \( \tau \). Thus (2.7) can be changed to the form
\[ \tau \doteq C_1^{-1} \sum_{k=0}^{m} \alpha_k (k+1)^{-1} g^{k+1} + C_0 \]

\[ = \sum_{k=0}^{m+1} \alpha_k^* g^{k}, \]

where \( \alpha_k \ (k = 0, 1, \ldots, m) \) is the \( k \)-th coefficient of the polynomial of degree \( m \) approximating the square root of \( I(\theta) \), and \( \alpha_k^* \) is the new \( k \)-th coefficient which is given by

\[
\begin{align*}
\alpha_k^* &= C_0 & k &= 0 \\
\alpha_k^* &= (C_1 k)^{-1} \alpha_{k-1} & k &= 1, 2, \ldots, m+1.
\end{align*}
\]

With this transformation of \( \theta \) to \( \tau \) and by virtue of (2.8), we can use the asymptotic normality with the two parameters, \( \tau \) and \( C_1^{-1} \), as the approximation to the conditional distribution of the maximum likelihood estimator \( \hat{\theta} \), given its true value \( \tau \) (cf. Samejima, 1981b). Then the first through fourth conditional moments of \( \tau \), given \( \hat{\theta} \), can be obtained from the density function, \( g^*(\hat{\theta}) \), of \( \hat{\theta} \) and from the constant \( C_1 \) by the following four formulae (cf. Samejima, 1981b):

\[ E(\tau I \hat{\theta}) = \hat{\tau} + C_1^{-2} \frac{d}{d\hat{\theta}} \log g^*(\hat{\theta}), \]

\[ \text{Var.}(\tau I \hat{\theta}) = C_1^{-2} [1 + C_1^{-2} \frac{d^2}{d\hat{\theta}^2} \log g^*(\hat{\theta})], \]

\[ E[(\tau - E(\tau | \hat{\theta}))^3 I \hat{\theta}] = C_1^{-6} [\frac{d^3}{d\hat{\theta}^3} \log g^*(\hat{\theta})] \]

and

\[ E[(\tau - E(\tau | \hat{\theta}))^4 I \hat{\theta}] = C_1^{-4} [3 + 6C_1^{-2} (\frac{d^2}{d\hat{\theta}^2} \log g^*(\hat{\theta})) + 3C_1^{-4} (\frac{d^2}{d\hat{\theta}^2} \log g^*(\hat{\theta}))^2 \]

\[ + C_1^{-4} (\frac{d^4}{d\hat{\theta}^4} \log g^*(\hat{\theta}))]. \]

This density function, \( g^*(\hat{\theta}) \), can be estimated by fitting a polynomial, using the method of moments (cf. Samejima and Livingston, 1979), as we did in the transformation of \( \theta \) to \( \tau \), based upon the empirical set of \( \hat{\theta} \)'s. Note that in the above formulae the first moment is about the origin, while the other three are about the mean.

The two coefficients, \( \beta_1 \) and \( \beta_2 \), and Pearson's criterion \( \kappa \) are obtained by

\[ \beta_1 = \mu_3 \mu_2^{-3}, \]

\[ \beta_2 = \mu_4 \mu_2^{-2}. \]
and

\begin{equation}
(2.17) \quad \kappa = \beta_1 (\beta_2 + 3)^2 [4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)]^{-1},
\end{equation}

by substituting \( \mu_2, \mu_3 \) and \( \mu_4 \) by Var.(\( r | \hat{r} \)), \( E[(r - E(r | \hat{r}))^3 | \hat{r}] \) and \( E[(r - E(r | \hat{r}))^4 | \hat{r}] \), respectively, which are obtained by formulae (2.12), (2.13) and (2.14).

In the Bivariate P.D.F. Approach, we approximate the bivariate distribution of the transformed latent trait \( r \) and its maximum likelihood estimate \( \hat{r} \) for each subpopulation of examinees who share the same discrete item response to a specified item. Thus the procedure must be repeated as many times as the number of discrete item response categories for each separate item. It is rather a time-consuming approach, and the CPU time for the item calibration increases almost proportionally to the number of new items.

In contrast to this, Conditional P.D.F. Approach deals with the total population of subjects, and all the items together. Effort is focused upon the approximation of the conditional distribution of \( r \), given \( \hat{r} \), for the total population of examinees, and then the result is branched into separate discrete item response subpopulations for each item.

If we compare the two approaches with each other, therefore, we can say that Bivariate P.D.F. Approach is an orthodox approach, while Conditional P.D.F. Approach needs an assumption that the conditional distribution of \( r \), given \( \hat{r} \), is unaffected by the different subpopulations of examinees. While this assumption can only be tolerated in most cases, the latter approach has two big advantages in the sense that the CPU time required in item calibration is substantially less, and that it does not have to deal with subgroups of small numbers of subjects in approximating the joint bivariate distributions of \( r \) and \( \hat{r} \).

In each of these two approaches, we can choose one of the four methods listed earlier in estimating the bivariate density of \( r \) and \( \hat{r} \), or the conditional density of \( r \), given its maximum likelihood estimate \( \hat{r} \). In so doing, in the Pearson System Method, we use all four conditional moments of \( r \), given \( \hat{r} \), which are estimated through the formulae (2.11) through (2.14), and, using Pearson's criterion \( \kappa \), which is given by (2.17), one of the Pearson System density functions is selected. In the Two-Parameter Beta Method two of the parameters of the Beta density function, i.e., the lower and upper endpoints of the interval of \( r \) for which the Beta density is positive, are a priori given, and the other two parameters are estimated by using the first two conditional moments of \( r \), given \( \hat{r} \), which are provided by (2.11) and (2.12), respectively. In the Normal Approach Method, again we use only the first two conditional moments of \( r \), given \( \hat{r} \), as the first and second parameters of the normal density function.

If we compare these three methods, it will be appropriate to say that both Two-Parameter Beta Method and Normal Approach Method are simpler versions of Pearson System Method. And yet the latter two methods have an advantage of using only the first two estimated conditional moments of \( r \), given \( \hat{r} \), whereas the former requires the additional third and fourth conditional moments, whose estimations are less accurate compared with those of the first two conditional moments. If we compare the Two-Parameter Beta Method with the Normal Approach Method, we will notice that the former allows non-symmetric density functions, while the latter does not. This is an advantage of the Two-Parameter Beta Method over the Normal Approach Method, and yet the former has the disadvantage of the requirement that two of the four parameters should a priori be set.

Lognormal Approach Method was developed later, which uses up to the third conditional moment and allows more flexibilities in the shape of the conditional distribution of \( r \), given \( \hat{r} \), than the Normal Approach Method. It was intended that a happy medium between the Pearson System Method and the Normal Approach Method would be realised, in the effort of ameliorating the disadvantages of these
two methods and of keeping their separate advantages.

III Simple Sum Procedure of the Conditional P.D.F. Approach Combined with the Normal Approach Method

It is obvious from the discussion given in the preceding section that the Conditional P.D.F. Approach combined with the Normal Approach Method is the simplest and one of the most economical procedures in CPU time. Out of the three procedures of the Conditional P.D.F. Approach the Simple Sum Procedure is the simplest one (cf. Samejima, 1981b). For this reason, the combination of the Simple Sum Procedure of the Conditional P.D.F. Approach and the Normal Approach Method has most frequently been applied for simulated and empirical data. Fortunately, in spite of the simplicity of the procedure, the results with simulated data in the adaptive testing situation and with simulated and empirical data in the paper-and-pencil testing situation indicate that we can estimate the operating characteristics fairly accurately by using this combination (cf. Samejima, 1981b, 1984). This seems to prove the robustness of the Conditional P.D.F. Approach. For one thing, there is a good reason why Normal Approach Method works well, for the conditional distribution of \( r \), given \( \hat{r} \), is indeed normal if the (unconditional) distribution of \( r \) is normal, and it is a truncated normal distribution if the (unconditional) distribution of \( r \) is rectangular, and the truncation is negligible for most of the conditional distributions.

In the Simple Sum Procedure of the Conditional P.D.F. Approach, the operating characteristic, \( P_k(\theta) \), of the discrete item response \( k_g \) of an unknown item \( g \) is estimated through the formula

\[
P_k(\theta) = P_k^*(\theta) = \sum_{s \in k_g} \phi(\tau | \hat{\tau}_s) \sum_{s=1}^N \phi(\tau | \hat{\tau}_s)^{-1},
\]

where \( s (= 1, 2, \ldots, N) \) indicates an individual examinee, and \( \phi(\tau | \hat{\tau}_s) \) denotes the conditional density of \( \tau \), given \( \hat{\tau}_s \). This conditional density is estimated by using the estimated conditional moments of \( \tau \), given \( \hat{\tau}_s \), using one of the four methods, as was described in the preceding section.

In the Weighted Sum Procedure of the Conditional P.D.F. Approach, we have for the estimated operating characteristic of \( k_g \)

\[
P_k(\theta) = P_k^*(\theta) = \sum_{s \in k_g} w(\hat{\tau}_s) \phi(\tau | \hat{\tau}_s) \sum_{s=1}^N w(\hat{\tau}_s) \phi(\tau | \hat{\tau}_s)^{-1}
\]

where \( w(\hat{\tau}_s) \) is the weight function of \( \hat{\tau}_s \). When we combine one of these two approaches with the Normal Approach Method, \( \phi(\tau | \hat{\tau}_s) \) in (3.1) or in (3.2) is approximated by the normal density function, using the first two estimated conditional moments of \( \tau \), given \( \hat{\tau}_s \), which are given by (2.11) and (2.12), respectively, as its parameters, \( \mu_{\hat{\tau}_s} \) and \( \sigma_{\hat{\tau}_s} \), in the formula

\[
\phi(\tau | \hat{\tau}_s) = [2\pi]^{-1/2}[\sigma_{\hat{\tau}_s}]^{-1} \exp\left[-(\tau - \mu_{\hat{\tau}_s})^2/(2\sigma_{\hat{\tau}_s}^2)\right].
\]

IV Differential Weight Procedure

If we accept the approximation of the conditional distribution of \( \hat{\tau} \), given \( \tau \), by the asymptotic normality, as we do in these approaches (cf. Samejima, 1981b), the other conditional distribution, i.e., that of \( \tau \), given \( \hat{\tau} \), will become more or less incidental. Thus in the Bivariate P.D.F. Approach the bivariate distribution of \( \tau \) and \( \hat{\tau} \) is approximated for each separate item score subpopulation of subjects of each unknown test item. In the Conditional P.D.F. Approach, however, the incidentality
of this second conditional distribution is not rigorously considered, and the implicit assumption exists such that for the fixed value of \( \tilde{r} \) the conditional distributions of \( r \) are similar for the different item score subpopulations.

Take the dichotomous response level, for example. On this level, each item is scored "right" or "wrong", "affirmative" or "negative", etc. The above assumption of non-incidentality may be acceptable when the operating characteristic of the correct answer of the item is represented by a mildly steep curve, as in the case with most practical situations, and the questions are asked to subjects whose ability levels are compatible with the difficulty levels of the questions, as in the case with adaptive testing and, though less rigorously, with many cases of paper-and-pencil testing.

This assumption is not acceptable, however, when the operating characteristic of the correct answer is represented by a steep curve. If the operating characteristic follows the Guttman scale, for example, then the conditional distributions of \( r \), given \( \tilde{r} \), for the two separate item score subpopulations are distinctly separated, and they do not even overlap! If we use the Simple Sum Procedure or the Weighted Sum Procedure for an item which nearly follows the Guttman scale, therefore, the resulting estimated operating characteristics of the correct and the incorrect answers will tend to be flatter than they actually are.

This problem can be solved by estimating differential conditional distributions of \( r \), given \( \tilde{r} \), for the separate discrete item responses to an "unknown" item. Let \( \phi_{k_g}(r | \tilde{r}) \) denote the conditional density of \( r \), given \( \tilde{r} \), for the subpopulation of subjects who share the same discrete item response \( k_g \) to an "unknown" item \( g \). We can write

\[
\phi_{k_g}(r | \tilde{r}) = f_{k_g}^*(r) \psi(\tilde{r} | r) [g_{k_g}^*(\tilde{r})]^{-1},
\]

where \( f_{k_g}^*(r) \) indicates the density of \( r \) for the subpopulation of subjects who share \( k_g \) as their common item score of item \( g \), \( \psi(\tilde{r} | r) \) is the conditional density of \( \tilde{r} \), given \( r \), which is approximated by the normal density, \( n[r, C_1^{-1}] \), and \( g_{k_g}^*(\tilde{r}) \) is the marginal density of \( \tilde{r} \), for this subpopulation, and for which we have

\[
g_{k_g}^*(\tilde{r}) = \int_{-\infty}^{\infty} f_{k_g}^*(r) \psi(\tilde{r} | r) \, dr.
\]

We notice that there is a relationship

\[
f_{k_g}^*(r) = f^*(r) P_{k_g}^*(r) \int_{-\infty}^{\infty} f^*(r) P_{k_g}^*(r) \, dr)^{-1},
\]

where \( f^*(r) \) denotes the density of \( r \) for the total population. Since we have

\[
\phi(r | \tilde{r}) = f^*(r) \psi(\tilde{r} | r) [g^*(\tilde{r})]^{-1},
\]

where \( g^*(\tilde{r}) \) is the density of \( \tilde{r} \) for the total population of subjects, as was described in the preceding section, which is given by

\[
g^*(\tilde{r}) = \int_{-\infty}^{\infty} f^*(r) \psi(\tilde{r} | r) \, dr,
\]

from the above formulae we obtain
\[ (4.6) \quad \phi_{k_0}(r \mid \hat{\tau}) = \phi(r \mid \hat{\tau}) P_{k_0}^*(r) h(\hat{\tau}), \]

where \( h(\hat{\tau}) \) is a function of \( \hat{\tau} \) and constant for a fixed value of \( \hat{\tau} \). Thus \( \phi_{k_0}(r \mid \hat{\tau}) \) is a density function proportional to \( \phi(r \mid \hat{\tau}) P_{k_0}^*(r) \). We notice that \( \phi(r \mid \hat{\tau}) \) in this formula is common to all the item scores and across different unknown items, while \( P_{k_0}^*(r) \) is a specific function of \( r \) for each \( k_0 \). Since \( \phi(r \mid \hat{\tau}) \) can be estimated by one of the four methods described earlier, our effort should be focused on finding an appropriate differential weight function for each \( k_0 \). Let \( W_{k_0}(r) \) denote such a differential weight function, which replaces \( P_{k_0}^*(r) h(\hat{\tau}) \) in (4.6). Thus we can revise (3.1) and (3.2) into the forms

\[ (4.7) \quad \hat{P}_{k_0}(r) = \hat{P}_{k_0}^*[r(\theta)] = \sum_{s \in k_0} W_{k_0}(r) \phi(r \mid \hat{\tau}_s) \left[ \sum_{s=1}^{N} W_{k_0}(r; s) \phi(r \mid \hat{\tau}_s) \right]^{-1} \]

and

\[ (4.8) \quad \hat{P}_{k_0}(\theta) = \hat{P}_{k_0}^*[r(\theta)] = \sum_{s \in k_0} w(\hat{\tau}_s) W_{k_0}(r) \phi(r \mid \hat{\tau}_s) \left[ \sum_{s=1}^{N} w(\hat{\tau}_s) W_{k_0}(r; s) \phi(r \mid \hat{\tau}_s) \right]^{-1}. \]

Since the differential weight function \( W_{k_0}(r) \) involves \( P_{k_0}^*(r) \), which itself is the target of estimation, we may use its estimate, \( \hat{P}_{k_0}^*(r) \), obtained by the Simple Sum Procedure or by the Weighted Sum Procedure, as its substitute. In so doing we may need some local smoothings of \( \hat{P}_{k_0}^*(r) \) where the estimation involves substantial amounts of error because of locally small numbers of subjects in the base data, etc. In some cases we may need several iterations by renewing the differential weight functions on each stage until the resulting estimated operating characteristic converges.

V Examples

We have tried this proposed method on the simulated data provided by Dr. Charles Davis of the Office of Naval Research, using the Simple Sum Procedure of the Conditional P.D.F. Approach combined with the Normal Approach Method with some modifications as the initial estimate of \( P_{k_0}(r) \) in the differential weight function. These data are simulated on-line item calibration data of the initial itempool calibration based upon conventional testing, in which 100 dichotomous items are divided into four subtests of 25 items each, and each subtest has been administered to 6,000 hypothetical examinees, and those of different rounds based upon adaptive testing, in which each of the 50 new binary items has been administered to a subgroup of 1,500 hypothetical subjects out of the total of 15,000. These hypothetical examinees' ability distributes unimodally within the interval of \( \theta \), (-3.0,3.0), with slight negative skewness.

For the purpose of illustration, Figure 5-1 presents the results of the Differential Weight Procedure using the results of the Simple Sum Procedure of the Conditional P.D.F. Approach combined with the Normal Approach Method with some modifications as the initial estimates, for eight items of the initial itempool. They are dichotomous items, and were intentionally selected from those items whose true operating characteristics of the correct answer are non-monotonic, in order to visualize the benefit of the nonparametric estimation of the operating characteristic. In each graph, also presented for comparison is the best fitted operating characteristic of the correct answer following the three-parameter logistic model, which has been given by Dr. Michael Levine. The logistic model on the dichotomous level is represented by
FIGURE 5-1

Eight Examples of the Estimated Operating Characteristic of the Correct Answer Using the Differential Weight Procedure (Dotted Line), in Comparison with the True Operating Characteristic (Solid Line) and the Best Fitted Three-Parameter Logistic Curve (Dashed Line).
FIGURE 5-1 (Continued)
FIGURE 5-1 (Continued)
FIGURE 5-1 (Continued)
ROUND9: ORIG. ITEM POOL: EST. OPER. CHAR. CD1, CD2, 3PM FITTED BY PROG8; 8516 OR 9220, 9004, 9006; 06/13/90

FIGURE 5-1 (Continued)
FIGURE 5-1 (Continued)
FIGURE 5-1 (Continued)
FIGURE 5-1 (Continued)
\( P_g(\theta) = [1 + \exp\{-D\alpha_g(\theta - b_g)\}]^{-1} , \)

where \( P_g(\theta) \) denotes the operating characteristic of the correct answer to item \( g \), \( \alpha_g \) and \( b_g \) are the item discrimination and difficulty parameters, respectively, and \( D \) is a scaling factor which is usually set equal to 1.7. We can see in these graphs that the resulting estimated operating characteristics are fairly close to the true ones, and that they reflect the non-monotonicities.

VI Sensitivities to Irregularities of Weight Functions

As we have proceeded, several factors have been identified and observed which affect the resulting estimated operating characteristics substantially. They are concerned with the differential weight function, and can be itemized as: 1) lower end ambiguities, 2) upper end ambiguities, 3) local irregularities and 4) overall irregularities.

Out of these factors, lower and upper end ambiguities basically come from the fact that we do not usually have sufficiently large numbers of subjects on the lowest and the highest ends of the interval of \( \theta \) of interest upon which the estimation of the operating characteristics is made. Also the fact that the test information function \( I(\theta) \) is used in the transformation of \( \theta \) to \( r \) which is specified by (2.7) may have something to do with these ambiguities. It has been observed (Samejima, 1979b) that in using equivalent items following the Constant Information Model (Samejima, 1979a) the speed of convergence of the conditional distribution of the maximum likelihood estimate \( \hat{\theta} \), given \( \theta \), to the asymptotic normality with \( \theta \) and \( [I(\theta)]^{-1/2} \) as its two parameters substantially differs for different levels of \( \theta \), in spite of the fact that the amount of test information is constant for every level of \( \theta \).

To be more specific, the convergence is observed to be much slower at those levels which are close to either end of the interval of \( \theta \) for which the amount of test information is non-zero and constant, and faster at intermediate levels of \( \theta \). This situation can be ameliorated if we replace the test information function \( I(\theta) \) in (2.7) by one of its two modified forms (cf. Samejima, 1990a). We can write for the Modification Formula No. 1, \( T(\theta) \),

\( T(\theta) = I(\theta) [1 + \frac{\partial}{\partial \theta} B(\hat{\theta}_V \mid \theta)]^{-2} , \)

which is the reciprocal of an approximate minimum bound of the variance of the maximum likelihood estimator, where \( B(\hat{\theta}_V \mid \theta) \) is the MLE bias function of the test consisting of items with any discrete item responses \( k_h \). In the general case of discrete item responses, we can write for the bias function of the maximum likelihood estimate

\( B(\hat{\theta}_V \mid \theta) = E[\hat{\theta}_V - \theta \mid \theta] = -(1/2)[I(\theta)]^{-2} \sum_{h=1}^{n} \sum_{k_h} P'_{kh}(\theta) P''_{kh}(\theta) [P_{kh}(\theta)]^{-1} , \)

where, as before, \( P_{kh}(\theta) \) is the operating characteristic of the discrete response \( k_h \), and \( P'_{kh}(\theta) \) and \( P''_{kh}(\theta) \) denote the first and second partial derivatives of \( P_{kh}(\theta) \) with respect to \( \theta \), respectively. Modification Form. 1's No. 2, \( \Xi(\theta) \), is given by

\( \Xi(\theta) = I(\theta) \{ [1 + \frac{\partial}{\partial \theta} B(\hat{\theta}_V \mid \theta)]^2 + I(\theta) [B(\hat{\theta}_V \mid \theta)]^2 \}^{-1} , \)

which is the reciprocal of an approximate minimum bound of the mean squared error of the maximum likelihood estimator. When the MLE bias function of the test is monotone increasing, as is the case in many situations, it is obvious from (6.1), (6.2) and (6.3) that we have the relationship,
FIGURE 6-1

Seven Examples of the Estimated Operating Characteristic of the Correct Answer Using the Differential Weight Procedure (Dotted Line), in Comparison with the True Operating Characteristic (Solid Line), When the Differential Weight Function (Short Dashed Line) Has Irregularities. The Function Was Also Proportionally Enlarged and Plotted (Long Dashed Line) to Visualize the Angles and Other Irregularities Well.
FIGURE 6-1 (Continued)
FIGURE 6-1 (Continued)
FIGURE 6-1 (Continued)
FIGURE 6-1 (Continued)
FIGURE 6-1 (Continued)
FIGURE 6-1 (Continued)
(6.4) \[ \Xi(\theta) \leq T(\theta) \leq I(\theta) , \]

where the first inequality in (6.4) always holds regardless of the shape of the MLE bias function. Which one of the two modified test information functions is more appropriate to use depends upon the situation, and we need more investigation to answer this question.

By irregularity we mean non-smoothness, which is exemplified by an unnatural angle, etc. It has been observed that for most items the resulting operating characteristic is amazingly sensitive to these irregularities of the differential weight function. In order to observe these sensitivities, Figure 6-1 illustrates how these irregularities, which are involved in the differential weight function, affect the resulting estimated operating characteristic.

The effect of local irregularities is most interesting to observe in these examples presented by Figure 6-1. In each of these graphs, the artificially irregular differential weight function for the correct answer is drawn by a short dashed line, and, in order to emphasize its irregularities, it was proportionally enlarged and shown by a long dashed line. We can see in each graph that, when the differential weight function has an unnatural angle, for example, the resulting estimated operating characteristic of the correct answer also shows an unnatural angle at approximately the same level of \( \theta \). We can also see in these graphs how overall irregularities of the differential weight function affect the resulting estimated operating characteristic, and how sensitive the latter is to the former. This type of sensitivity of the resulting estimated operating characteristic to the irregularities of the differential weight function is encouraging as well as threatening, for it promises success in the estimation provided that we succeed in finding the right differential weight function.

VII Usefulnesses of the Differential Weight Procedure

It is obvious that item analysis in the true sense of the word starts from the accurate estimation of the operating characteristics of the item responses. Thus the nonparametric estimation of the operating characteristic offers a great deal of information about an item, when it is successful. In this sense we can say that the Differential Weight Procedure provides us with promise for the successful item analysis in general.

Following this, we can conceive of many applications of the method for different purposes. To give some examples, it will be especially useful for the on-line item calibration in computerized adaptive testing; also it will be useful in the revision of multiple-choice test items in order to reduce the effect of noise and to make the ability estimation efficient (cf. Samejima, 1990b).

VIII Discussion and Conclusions

A new procedure of nonparametric estimation of the operating characteristics of discrete item responses has been proposed, and it is called Differential Weight Procedure of the Conditional P.D.F. Approach. Some examples have been given, and sensitivities of the resulting estimated operating characteristics to irregularities of the differential weight functions have been observed and discussed. Usefulnesses of the method have also been discussed.

These outcomes suggest the importance of further investigation of the weight function in the future.

To summarize, although Simple Sum Procedure of the Conditional P.D.F. Approach combined with the Normal Approach Method works reasonably well for the on-line item calibration of adaptive testing, and also for the paper-and-pencil testing, especially when the number of subjects is large, if we wish to increase the accuracy of estimation we can use the Differential Weight Procedure. The disadvantage
will be the added CPU time, so we need to consider the balance of the cost and accuracy of estimation before we make our decision. It will be less expensive, however, if we compare the CPU time required for the present procedure with the time required for the Bivariate P.D.F. Approach.

References


[14] Samejima, F. Estimation of the operating characteristics when the test information of the Old Test is not constant I: Rationale. ONR/RR-80-2, 1980a.


[22] Samejima, F. and Changas, P. S. How small the number of test items can be for the basis of estimating the operating characteristics of the discrete responses to unknown test items. *ONR/RR-81-3*, 1981.
