DOCUMENT RESUME

ED 067 407 TM 001 802

AUTHOR Boldt, Robert F.
TITLE An Estimation Procedure for the Rasch Model Allowing Missing Data.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO RM-72-5
PUB DATE Apr 72
NOTE 16p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Data Analysis; *Item Analysis; *Mathematical Models; *Measurement Techniques; *Statistical Studies; *Test Interpretation
IDENTIFIERS *Rasch Model

ABSTRACT

The Rasch model and other latent trait models encounter some difficulty when faced with an appreciable amount of missing data or omitting behavior. The present note assumes that some reasonable missing data model has been formulated which does not involve the parameters associated with the latent ability of interest. A maximum likelihood function is used that is based on probabilities which are conditional on the occurrence of a response. (Author/DB)

RM-72-5

RESEARCH MEMORANDUM

AN ESTIMATION PROCEDURE FOR THE RASCH MODEL ALLOWING MISSING DATA

Robert F. Boldt

This Memorandum is for interoffice use. It is not to be cited as a published report without the specific permission of the author.

Educational Testing Service
Princeton, New Jersey
April 1972

AN ESTIMATION PROCEDURE FOR THE RASCH MODEL

ALLOWING MISSING DATA

Robert F. Boldt

The Rasch model and other latent trait models encounter some difficulty when faced with an appreciable amount of missing data or omitting behavior. When faced with this fact the response of Wright and Panchapakesan (1969) is to urge that all examinees be forced to respond to all items. Possibly only data from those who answer all questions should be used. However, it is probably desirable to have an estimation procedure which does not depend on complete data. The present note assumes that some reasonable missing data model has been formulated which does not involve the parameters associated with the latent ability of interest. A maximum likelihood function is used that is based on probabilities which are conditional on the occurrence of a response. Let

i be the subscript for items; $i = 1, \quad , I$

t be the subscript for persons; $t = 1, \quad , T$

E_i be an easiness parameter for item i ; $E_i \geqq 0$

A_t be an ability parameter for person t ; $A_t \geqq 0$

λ be a LaGrange multiplier used to impose norming on the A_t 's;

α_{it} be one (1) if the t<u>th</u> examinee responds to the i<u>th</u> item, zero otherwise;

δ_{it} be one (1) if the t<u>th</u> examinee responds to the i<u>th</u> item correctly, zero otherwise.

Then the joint probability function of the observations (δ 's) given the pattern of items attempted (α 's) is

$$P = \prod_i \prod_t \frac{(E_i A_t)^{\delta_{it}}}{(1 + E_i A_t)^{\alpha_{it}}} \quad . \qquad (1)$$

Examination of Equation (1) indicates that an item attempted by no one, or a person who attempts no items, has exponents equal to zero over the entire range of the subscripts and hence cannot affect the value of p. Equation (1) also indicates that for a person who attempts some items and gets none correct, the value of P is at a maximum if E is zero since the denominators are at the least unity. Similarly, an item which is attempted by some but answered correctly by none would receive a parametric value of zero. However, an item correctly answered by all who attempted it or a person who correctly answers all items attempted would have infinite values associated with their parameters, since for those situations the quantity $(AE)/(1 + AE)$ must be unity. Finally, one may note that if $(AE)/(1 + AE)$ is unity, then $1/(1 + AE)$ is zero since the two must add to unity. Hence A may not be taken as infinite for a person who misses any item as that will minimize P, making it zero due to the multiplication by a zero. Similarly A may not be taken as zero for any person who gets an item right for, again, P would become zero and hence not be maximized. Therefore, the value of A assigned to a person who gets items both correct and incorrect would be neither zero nor infinite and by a similar argument the value of E assigned to an item which is answered both correctly and incorrectly would be neither zero nor infinite. The remaining extreme condition, that where A is zero and E is infinite, or vice versa, cannot occur since it requires that the examinee both succeed <u>and</u> fail on the item. However, what <u>can</u> occur that would confuse matters is that an item could be missed by all who fail to answer every item correctly. Then, without using the data for those who got all items right, one would estimate the item parameter to be zero. But we do not believe that the item parameter should be zero, nor that the people who get all the items correct should get infinity as their ability parameters. The fact is that the data do not support estimation of item or people parameters in some cases. To provide for this alternative, a preliminary procedure is introduced which eliminates items and people whose parameters do not fit the criteria that the items are tried by people who both pass and fail, and that the people try items and each person gets some and misses some. This procedure searches for people who miss all or get all, or items that are never missed or always missed and throws them out and records that the throwout occurs on the first search. Then a second search is conducted for perfectly good or bad performances and the items found to be thus are thrown out indicating that the throwout occurred on the second search. The searches continue always recording the number of the search on which an item or person is eliminated. Thus, when told that an item was perfect on the first trial and that a person was a perfect failure on the fourth search, one would know that the person got the item correct if he attempted it. More precise estimation for items and persons such as these seems impossible.

Based on the foregoing considerations the following procedures should be followed before initiating the likelihood maximization:

(a) incorporate some provision in the missing data analysis for those items answered by no one and delete them from the present analysis;

(b) incorporate some provision in the missing data analysis for those persons who answer no items and delete them from the present analysis;

(c) assign a parametric value of zero to all items not eliminated in (a) which are answered incorrectly by anyone attempting them, indicate search number one, and delete them from the present analysis in step (g);

(d) assign a parametric value of zero to all examinees not eliminated in (b) who get no items correct, indicate search number one, and delete them from the present analysis in step (g);

(e) assign a parametric value which is infinite to those items not eliminated in (a) but answered correctly by all who attempt them, indicate search number one, and delete them from the present analysis in step (g);

(f) assign a parametric value which is infinite to those examinees not eliminated in (b) but who answer correctly all items attempted, indicate search number one, and delete them from the present analysis in step (g);

(g) accomplish the deletions indicated. If there are none, go to the optimization procedure.

When steps (a)-(g) are completed, a reduced collection of people and items will be left. Some of these people may have gotten all of the remaining items correct, or some of the items may have been answered correctly by all of the remaining people. Deletion of items may leave some people with no responses, or deletion of some people may leave some unattempted items; that is, unattempted by those who were not deleted in steps (b) and (g). Similarly, some people may have been deleted who were the only ones to respond to certain items. With the remaining data carry out the following:

(h) delete those items for whom none of the remaining examinees have made a response and indicate the search number (2 or more)--missing data provisions have already been made so no additional provisions are needed;

(i) delete those persons who made no response to the remaining items and indicate the search number (2 or more)--missing data provisions have already been made so no additional provisions are needed;

(j) repeat steps (c)-(g) indicating the appropriate search numbers. Continue cycling from (c) through (j) until no deletion is indicated for step (g), as indicated. When this iterative cycling is completed, it is hoped that most examinees and most items will be left for entry into the optimization procedure.

Assuming that the subscripts i and t range over values defined only for the remaining data, the likelihood function[1] (L) with norming of A's imposed to eliminate a multiplicative indeterminacy is

$$L = \sum_i \sum_t \delta_{it} \ln A_t E_i - \sum_i \sum_t \alpha_{it} \ln(1 + A_t E_i) + \lambda(A_{\cdot} - V) \quad . \qquad (2)$$

Then

$$\frac{\partial L}{\partial E_i} = \frac{\delta_{i\cdot}}{E_i} - \sum_t \alpha_{it} \frac{A_t}{1 + A_t E_i} \qquad (3)$$

$$\frac{\partial L}{\partial A_t} = \frac{\delta_{\cdot t}}{A_t} - \sum_i \alpha_{it} \frac{E_i}{1 + A_t E_i} + \lambda \qquad (4)$$

$$\frac{\partial L}{\partial \lambda} = (A_{\cdot} - T) \qquad (5)$$

are the derivatives needed for the optimization. By the iterative procedures terminating at step (g) above, it is ensured that the E's and A's sought in this procedure are neither zero nor infinite. Hence the derivatives are not trivially zero due to infinite parameters, nor does multiplication though Equations (3) and (4) produce zeros trivially. Rather, at the solution the derivatives are zero because the parameters found are indeed optimum for the data. Further, the LaGrange multiplier is zero since, from (3) and (4),

$$\sum_i E_i \frac{\partial L}{\partial E_i} = \sum_t A_t \frac{\partial L}{\partial A_t} = \lambda$$

since at solution the derivatives are equal to zero. Note also that, given the E 's, A_t is the only variable in one of Equations (4) and given the A 's, E_i is the only variable in one of Equations (3). Further, it can be shown that at solution for a fixed set of A 's, the second derivatives with respect to the E 's are negative, as are the second derivatives with respect to the A 's for fixed E 's and hence the optima are maxima. Further, for fixed A 's, which are assumed to be positive, there is only one optimum value of an E on the positive half-axis. To show this, note that if the E^{-1} is factored out of the right-hand side of (3), the resulting expression contains terms of the form $AE/(1 + AE)$ which is monotonically increasing in E. Hence the derivative is zero for only one value of E, that is, when $\delta_{i\cdot} = \sum_i \alpha_{it} \frac{E_i A_t}{1 + E_i A_t}$. Existence of the optimum is assured by steps (a)-(j) since zero and infinity values for E lead to zero values for P, and the function is positive at intermediate values (of either the A 's or the E 's).

Newton iterations are suggested to optimize L. Since the LaGrange multiplier is zero, λ is not represented in the derivatives used and the norming of the A 's is preserved as a part of the iterative procedure (step 4a below). For these iterations the needed second derivatives are as follows:

$$\frac{\partial^2 L}{\partial E_i^2} = -\frac{\delta_{i}}{E_i^2} + \sum_t \alpha_{it} \frac{A_t^2}{(1 + A_t E_i)^2} \qquad (6)$$

$$\frac{\partial^2 L}{\partial A_t^2} = -\frac{\delta_{.t}}{A_t^2} + \sum_i \alpha_{it} \frac{E_i^2}{(1 + A_t E_i)^2} \quad . \qquad (7)$$

Note these derivatives do not include other parameters of a kind. That is, in Equation (6) only one value of the subscript i is involved so that E_j is not involved if $j \neq i$. A similar condition obtains for the A 's. Therefore, given the A 's, the E 's can be found one at a time. Similarly, given the E 's , the A 's can be found one at a time. Therefore the increments needed for the iterations are

$$\Delta_j E_i = -\left(\frac{\partial L}{\partial_j E_i}\right) \div \left(\frac{\partial^2 L}{\partial_j E_i^2}\right) = \qquad (8)$$

$${}_jE_t\left[\delta_{i.} - \sum_t \alpha_{it}\left(\frac{{}_jE_i A_t}{1 + {}_jE_i A_t}\right)\right] \div \left[\delta_{i.} - \sum_t \alpha_{it}\left(\frac{{}_jE_i A_t}{1 + {}_jE_i A_t}\right)^2\right] \quad ,$$

where the prefix j 's are included to indicate a value of E at a particular iteration. In this notation

$${}_{(j+1)}E_i = {}_jE_i + \Delta_j E_i \quad . \qquad (9)$$

Note that no prefixed subscript has been included for the A 's. The A 's do not, of course, remain constant during the entire optimization procedure and subscripts for A 's will be included for equations which describe iterations for finding A 's . However, when E 's are being found the

A 's are not changing. The reader should keep in mind that E 's are found while A 's are held constant, but after the E 's are found, then the A 's are modified. The whole procedure stops when all E 's and A 's satisfy the normal equations.

The iterations described above can yield negative E 's as can be seen by examination of Equation (8). Note that the expression in parentheses in the numerator of (8) contains terms of the form $AE/(1 + AE)$ and since all parameters are positive, these terms are fractions. Note also that for each such term in the numerator there is a corresponding term containing the square of the fraction. Hence when the denominator of (8) is nearly zero, the numerator is relatively large and negative, and the incremented value of E may be negative. Such a situation arises for overly large values of E and so when the incremented E appears to be negative, set the incremented E equal to half the previous E rather than using Equation (9). This halving procedure will eventually yield a denominator which is positive and of appreciable size and a nonnegative value for the incremented E .

The iterative sequence is shown graphically in Figure 1. In this figure

---

Insert Figure 1 about here

---

the derivative of L with respect to E is monotonically decreasing with a monotonically increasing slope. Results are drawn in at three places, a, b, and c. If an iteration begins at point a , the effect of the Newton iteration is to make the next point be at $a + \Delta a$, which is to the left of zero due to the flatness of the slope of the function. Note that a point of departure at b also moves to the left but not too far, and also from the shape of the curve it can be seen that movement is to a point smaller than the solution. Finally, a point of departure at the point c leads to an increase but not to more than the solution. Therefore, one expects, except for the halving procedure, at most one decrease in E which does not yield a negative E and after that a series of positive increments which eventually become so small one stops.

A similar situation exists for the A 's, and the halving procedure should be followed if negative A 's are observed following addition of the increments. The formula for the increment of A is as follows.

$$\Delta_j A_t = -\left(\frac{\partial L}{\partial_j A_t}\right) \div \left(\frac{\partial^2 L}{\partial_j A_t^2}\right) = \tag{10}$$

$$_jA_t\left[\delta_{\cdot t} - \sum_i \alpha_{it}\left(\frac{E_i\ _jA_t}{1 + E_i\ _jA_t}\right)\right] \div \left[\delta_{\cdot t} - \sum_i \alpha_{it}\left(\frac{E_i\ _jA_t}{1 + E_i\ _jA_t}\right)^2\right].$$

As with the E 's the incrementing rule is

$$_{(j+1)}A_t \quad _jA_t + \Delta_j A_t \quad . \tag{11}$$

The iteration procedure will next be outlined, thus introducing notation for the stopping rule. That rule will be developed at the end.

Step 1. Set all A 's equal to V/T , all E 's equal to T/V .

Step 2. Choose δ , the maximum tolerable error for a change in a theoretical proportion, ϵ = upper bound on the change on the theoretical probability for correct response, and V , the norming constant for the A 's.

Step 3. Calculate new A 's; initialize t at unity, η at unity.

Step 3a. Given A_t , calculate $\Delta A_t = \Delta$.

Step 3b. If $0 \leqq \Delta \leqq \epsilon A_t$, go to 3d. If not, set $\eta = 0$ and go to

3c. For Δ use Equation (10).

Step 3c. If $\Delta < 0$, copy $\frac{A_t}{2}$ into A_t . If $\Delta > 0$, add Δ to A_t . Then go to 3a.

Step 3d. If $t = T$, go to Step 4. If not, add unity to t and go to 3a.

Step 4. Calculate new E 's. Initialize i at unity.

Step 4a. Copy $V A_t/A.$ into A_t , for all t .

Step 4b. Given E_i calculate $\Delta E_i = \Delta$. For Δ use Equation (8).

Step 4c. If $0 \leqq \Delta \leqq \epsilon E_i$, go to 4e. If not, set $\eta = 0$ and go to 4d.

Step 4d. If $\Delta < 0$, copy $\frac{E_i}{2}$ into E_i . If $\Delta > 0$, add Δ to E_i . Then go to 4b.

Step 4e. If $i = I$, go to Step 4f. If not, add unity to i and go to 4b.

Step 4f. If η equals zero, go to Step 3. If not, exit.

Steps 1 through 4 yield A 's and E 's that introduce change of no more than ϵ into a theoretical probability of a correct response. That is, when a Δ is negative, it is recycled. When Δ is positive and small enough, we want to stop. If an E is being computed, a theoretical probability of correct response using the incremented E is

$$\frac{A(E + \Delta)}{1 + A(E + \Delta)} .$$

The error, ϵ , between the proportion using parameters at a given time and the next time is

$$\frac{A(E + \Delta)}{1 + A(E + \Delta)} - \frac{AE}{1 + AE} . \qquad (12)$$

If the Δ in the denominator of (12) is set equal to zero and the Δ in the numerator is not, we get a quantity which is surely greater than the error and which is

$$\frac{A\Delta}{1 + AE} \ .$$

Further, $AE/(1 + AE)$ is surely less than one; an even larger quantity is Δ/E . Hence if

$$(\Delta/E) < \epsilon \ ,$$

the change is less than ϵ . Similar logic holds for the change on A .

Reference

Wright, B., & Panchapakesan, N. A procedure for sample-free item analysis. Educational and Psychological Measurement, 1969, 29, 23-48.

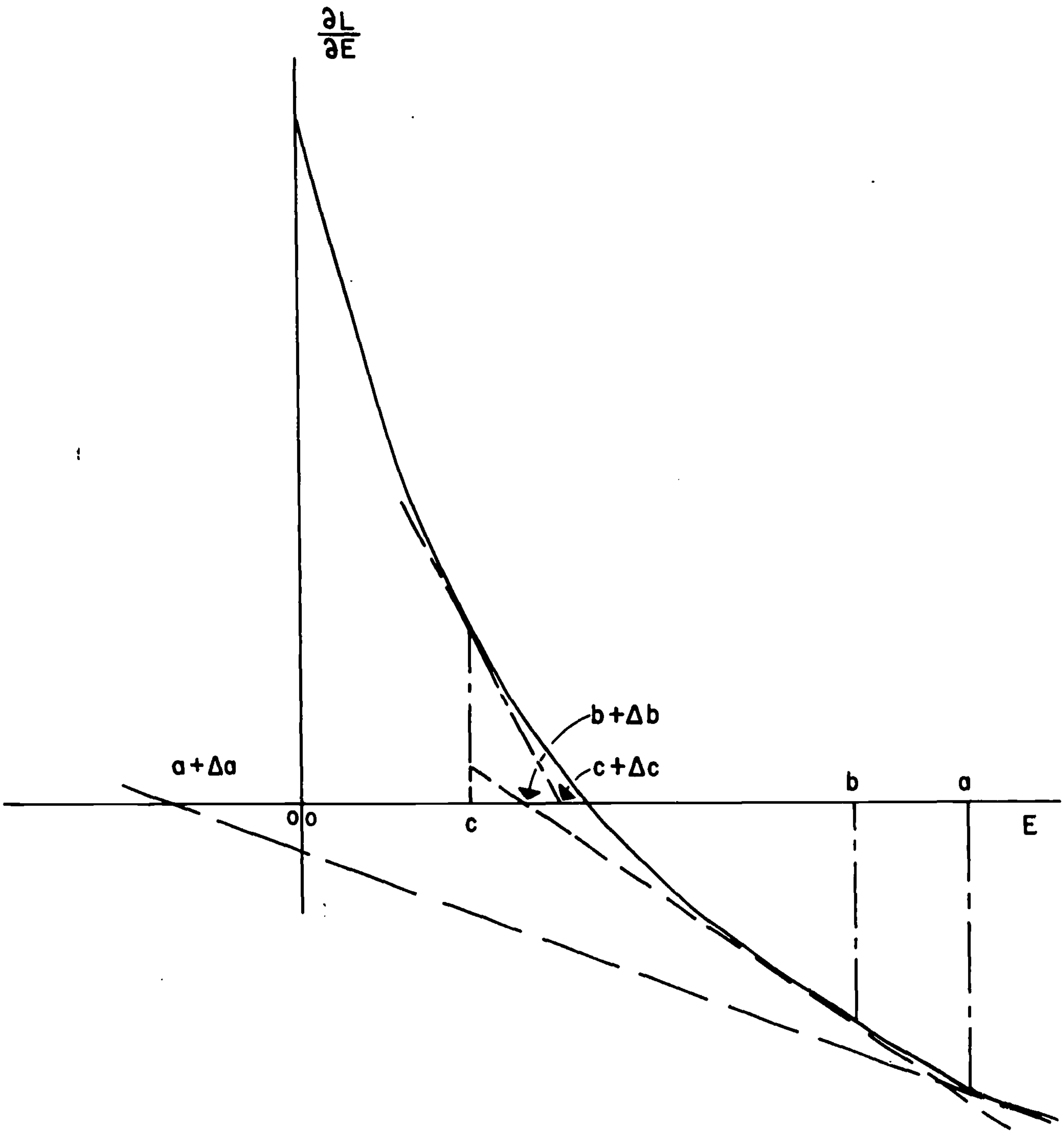
Footnote

[1]The common dot notation will be used for summation, e.g.,

$$x_{ij.\ell} = \sum_k x_{ijk\ell} \quad .$$

Figure Caption

Fig. 1. Second trial values ($a + \Delta a$, $b + \Delta b$, $c + \Delta c$) of E following different first trial values (a, b, c) of E .


∂L/∂E
a+Δa
o
o
c
b+Δb
c+Δc
b
a
E