

AUTHOR Engelen, Ronald J. H.; Jannarone, Robert J.
 TITLE A Connection between Item/Subtest Regression and the Rasch Model. Research Report 89-1.
 INSTITUTION Twente Univ., Enschede (Netherlands). Dept. of Education.
 PUB DATE Dec 89
 NOTE 17p.
 AVAILABLE FROM Bibliotheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
 PUB TYPE Reports - Evaluative/Feasibility (142)
 EDRS PRICE MF01/PC01 Plus Postage.
 DESCRIPTORS *Bayesian Statistics; Comparative Analysis; Equations (Mathematics); *Estimation (Mathematics); *Goodness of Fit; Item Analysis; Item Response Theory; *Mathematical Models; *Regression (Statistics); Test Items
 IDENTIFIERS *Item Subtest Regression; *Rasch Model

ABSTRACT

The purpose of this paper is to link empirical Bayes methods with two specific topics in item response theory--item/subtest regression, and testing the goodness of fit of the Rasch model--under the assumptions of local independence and sufficiency. It is shown that item/subtest regression results in empirical Bayes estimates only if the Rasch model holds. Additionally, it is shown that a newly-derived exploratory goodness-of-fit test for the Rasch model, which does not need item and person parameter estimates, can be seen as an empirical Bayes test. This test compares the observed proportions of correct answers to one specific item, given any pattern that leads to a number-right score. These proportions should be about equal. (RLC)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

ED322199

A Connection Between Item/Subtest Regression and the Rasch Model

Research Report
89-1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. NELISSEN

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) "

Ronald J.H. Engelen
Robert J. Jannarone



ment of
CATION



University of Twente

Division of Educational Measurement and Data Analysis



2
BEST COPY AVAILABLE

Colofon:
Typing: L.A.M. Bosch-Padberg
Cover design: Audiovisuele Sectie TOLAB Toegepaste
Onderwijskunde
Printed by: Centrale Reproductie-afdeling

**A Connection Between Item-Subtest Regression
and the Rasch Model**

Ronald J.H. Engelen

Bob J. Jannarone

A connection between item-subtest regression and the Rasch model / Ronald J.H. Engelen and Robert J. Jannarone - Enschede : University of Twente, Department of Education, December, 1989. - 9 pages

Abstract

Empirical Bayes methods are linked with item response theory under the assumptions of local independence and sufficiency. It is shown that item-subtest regression results in empirical Bayes estimates if and only if the Rasch model holds. Additionally, it will be shown that a newly derived, exploratory, Rasch model test can be seen as an empirical Bayes test.

Key-words: Empirical Bayes estimation, Rasch model, item subtest regression, sufficiency.

Introduction

The purpose of this paper is to link empirical Bayes methods with two specific topics in item response theory (IRT): item-subtest regression and testing the goodness of fit of the Rasch model. It will be shown that item-subtest regression results in empirical Bayes estimates if and only if the Rasch model (Rasch, 1980) holds. Additionally, it will be shown that a Rasch model (RM) test can be derived as an empirical Bayes test from first principles.

To briefly introduce empirical Bayes estimation, let us assume that an observation z of a random variable Z is made, where the distribution function of Z depends on a parameter θ , i.e., the distribution function of Z is given by $F(z|\theta)$. In empirical Bayes estimation, just as in "classical" Bayes estimation, the assumption of a prior distribution function $G(\theta)$ for the unknown parameter θ is essential. In addition, however, "previous" data is used to obtain a (in some way) reasonable estimator of $G(\theta)$, and hence of θ . In some cases it is possible to obtain an empirical Bayes estimator of θ without explicit estimation of $G(\theta)$. For a more detailed introduction to empirical Bayes methods, see, for instance, Robbins (1955, 1964) or Maritz (1970).

Sample item-subtest regressions, which are used by Lord (1980) to construct estimators for item-response functions, are obtained by computing the proportion of examinees that answered one specific item correctly, given their number-correct scores on the test minus the item concerned.

Empirical Bayes Methods in IRT Models

As an introduction to the concept of empirical Bayes estimation in IRT models, we will first explain how the concept works in the simplest IRT model, the binomial model.

In this model, all K items have the same item characteristic curve (ICC), i.e. $\tau_m(\theta) = p(X_m=1|\theta) = \theta$ for all $m = 1, \dots, K$, where θ is the (latent) ability. So, the total score Y is binomially distributed with parameters K and θ . The Bayes point estimator $\delta_G(y)$ of the mean ICC, which is denoted by $\bar{\tau}(\theta) = \sum_{m=1}^K \tau_m(\theta)/K$, under squared error loss is given by

$$(1) \quad \delta_G(y) = \frac{\int \theta p_K(y; \theta) dG(\theta)}{\int p_K(y; \theta) dG(\theta)},$$

where $p_K(y; \theta)$ is the density or discrete probability function of y given θ , and $G(\theta)$ the prior distribution function of θ (Lehmann, 1983). Note that $\bar{\tau}(\theta) = \theta$. For the binomial case (1) leads to

$$(2) \quad \delta_G(y) = \frac{(y+1) n_{K+1}(y+1)}{(K+1) p_K(y)},$$

where $p_{K+1}(z)$ denotes the probability of obtaining a total score of z on a $(K+1)$ -item test (Meredith and Kearns, 1973).

Now suppose that a random sample of examinees of size N is available. The empirical Bayes estimate of θ is then obtained by simply substituting observed frequencies for theoretical probabilities. Since scores on a $(K+1)$ -item test can not be observed however, (2) must be modified in such a way that only observable quantities remain. Apart from some technical details, this is done by computing empirical Bayes estimates on the subtests in which the m -th ($m=1, \dots, K$) item has been deleted, and evaluating the mean of these estimates to get a more stable result. It can be shown that these estimates are consistent and that the empirical Bayes risk converges to the Bayes risk. More details can be found in Cressie (1982), Jannarone (1979), Meredith and Kearns (1973), Kearns and Meredith (1975) and Robbins (1955).

It should be mentioned that some sort of smoothing of empirical Bayes estimates may be useful, such as constraining the estimates to increase monotonically in the total score y (van Houwelingen, 1977; Jannarone, 1979).

With the binomial model as a guide-line, it is now easy to evaluate empirical Bayes estimates for the ICC's for general IRT models, without making any assumptions whatsoever on the form of the ICC (i.e., non-parametrically).

We want to compute the empirical Bayes estimate of the m -th ICC, $\tau_m(\theta)$. As in the binomial model, this will be done by using the responses on subtest $S_{(m)}$ containing all but the m -th item. In contrast with the binomial model, where all items are equivalent, this estimate has to be computed for each subtest response vector separately. Hence, the empirical

Bayes estimate will be based on $x_{(m)}$, denoting the response vector on $S_{(m)}$. In the following we will use subscripts (m) to denote operations where the m -th item is excluded.

We will need the following Assumption:

$$(3) \quad P(X_m = 1, X_{(m)} = x_{(m)} | \theta) = P(X_m = 1 | \theta) P(X_{(m)} = x_{(m)} | \theta),$$

for all m and all vectors $x_{(m)}$. For that matter, it can be easily shown that (3) is equivalent with the assumption of local independence.

With (3) and the above notations, the Bayes estimate of $\tau_m(\theta)$ based on $x_{(m)}$ is defined by

$$(4) \quad \delta_{(m)}(x_{(m)}) = E[\tau_m(\theta) | X_{(m)} = x_{(m)}].$$

Rewriting (4) leads to

$$(5) \quad \delta_{(m)}(x_{(m)}) = \frac{\int \tau_m(\theta) P(X_{(m)} = x_{(m)} | \theta) dG(\theta)}{\int P(X_{(m)} = x_{(m)} | \theta) dG(\theta)}.$$

Now suppose that

$$(6) \quad P(X_{(m)} = x_{(m)} | \theta) = q(s(x_{(m)}); \theta) h(x_{(m)})$$

for some functions q and h and for all $x_{(m)}$, or equivalently suppose that the sufficient factorization criterion holds on the subtest $S_{(m)}$. Then

$$(7) \quad \delta_{(m)}(x_{(m)}) = \frac{\int \tau_m(\theta) q(s(x_{(m)}); \theta) dG(\theta)}{\int q(s(x_{(m)}); \theta) dG(\theta)}$$

$$= \delta_{(m)}(y_{(m)})$$

provided that $s(x_{(m)}) = s(y_{(m)})$. Hence, in this case, the Bayes estimates for $\tau_m(\theta)$ based on two response patterns $x_{(m)}$ and $y_{(m)}$ are equal if the sufficient statistic for these patterns are equal. But we have sufficiency of the item-deleted test score in the RM as well as in the two-parameter logistic model with known (not necessary equal) discrimination parameters only. So, for the RM, the Bayes estimate of the m -th ICC can now be based on the sufficient statistic $R_{(m)} = \sum_{n \neq m} X_n$. This estimate is easily obtained by substituting $r_{(m)}$ for $x_{(m)}$ in formula (4), which leads to

$$(8) \quad \delta_{(m)}(x_{(m)} | r_{(m)}) = \frac{P(X_m=1, R_{(m)}=r_{(m)})}{P(k_{(m)}=r_{(m)})}$$

which is the usual item true-score regression (Lord, 1980, p. 251-252).

The empirical Bayes estimate of $\tau_m(\theta)$ based on $R_{(m)}$ is now obtained by substituting observed proportions for theoretical probabilities, i.e.,

$$(9) \quad \hat{\delta}_{(m)}(x_{(m)} | r_{(m)}) = \frac{\hat{P}(X_m=1, R_{(m)}=r_{(m)})}{\hat{P}(R_{(m)}=r_{(m)})}$$

Note that our results point towards a possible goodness-of-fit test for the m -th item, given that the RM holds for the subtest $S_{(m)}$: If the observed proportion of examinees answering the m -th item correctly for a subtest score is roughly equal for each pattern that leads to the subtest score, then the RM should also hold for the m -th item, otherwise it should not. Such a test would be explorative, in a way that resembles the 'splitter' technique proposed by Stelzl (1979) and also described by Molenaar (1983).

Discussion

It has been shown that an empirical Bayes justification for using item-subtest regressions as item response function estimates holds only for the Rasch models. Indeed, (4) and (6) show that item-subtest regression is inappropriate otherwise. Similar conclusions may hold regarding the use of number-correct scores in other IRT models as well.

Additionally, an (exploratory) goodness-of-fit test for the Rasch model, which does not need item and person parameter estimates has been indicated. This test compares the observed proportions of correct answers to one specific item, given any pattern that leads to a number-right score of, say, R' on the remaining items. These proportions should be about equal.

References

- Cressie, N. (1982). A useful empirical Bayes identity. Annals of Statistics, 10, 625-629.
- Jannarone, B. (1979). Empirical Bayes estimates of expected (true) scores for two latent trait models. Unpublished Ph.D. Thesis: University of California, Berkeley.
- Kearns, J., & Meredith, W. (1975). Methods for evaluating empirical Bayes estimates of latent trait scores. Psychometrika, 40, 373-394.
- Molenaar, I.W. (1983) Some improved diagnostics for failure of the Rasch model. Psychometrika, 48, 49-72.
- Lehmann, E.L. (1983). Theory of point estimation. New York: Wiley.
- Lord, F.M. (1980). Applications of item response theory to practical testing problems. Hillsdale: Lawrence Erlbaum.
- Maritz, J.S. (1970). Empirical Bayes methods. London: Methuen.
- Meredith, W., & Kearns, J. (1973). Empirical Bayes point estimates of latent trait scores without knowledge of the trait distribution. Psychometrika, 38, 533-554.
- Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests. Chicago: The University of Chicago Press.
- Robbins, H. (1955). An empirical Bayes approach to statistics. In J. Neyman (Ed.): Proceedings of the Third Berkeley symposium on mathematical statistics and probability, Volume 1. Berkeley: U.C. Press.

- Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. Annals of Mathematical Statistics, 35, 1-20.
- Stelzl, I. (1979). Ist der Modelltest des Rasch-Modells geeignet, Homogenitätshypothesen zu prüfen? Ein bericht über Simulationsstudien mit inhomogenen Daten. [Is the Rasch model sensitive to tests of homogeneity?] Zeitschrift für Experimentelle und Angewandte Psychologie, 26, 652-672.
- Van Houwelingen, J.C. (1977). Monotizing empirical Bayes estimators for a class of discrete distributions with monotone likelihood ratio. Statistica Neerlandica, 31, 95-104.

Titles of recent Research Reports from the Division of
Educational Measurement and Data Analysis,
University of Twente, Enschede,
The Netherlands.

- RR-89-1 R.J.H. Engelen & R.J. Jannarone, *A connection between item/subtest regression and the Rasch model*
- RR-88-18 H.J. Vos, *Applications of decision theory to computer based adaptive instructional systems*
- RR-88-17 H. Kelderman, *Loglinear multidimensional IRT models for polytomously scored items*
- RR-88-16 H. Kelderman, *An IRT model for item responses that are subject to omission and/or intrusion errors*
- RR-88-15 H.J. Vos, *Simultaneous optimization of decisions using a linear utility function*
- RR-88-14 J.J. Adema, *The construction of two-stage tests*
- RR-88-13 J. Kogut, *Asymptotic distribution of an IRT person fit index*
- RR-88-12 E. van der Burg & G. Dijksterhuis, *Nonlinear canonical correlation analysis of multiway data*
- RR-88-11 D.L. Knol & M.P.F. Berger, *Empirical comparison between factor analysis and item response models*
- RR-88-10 H. Kelderman & G. Macready, *Loglinear-latent-class models for detecting item bias*
- RR-88-9 W.J. van der Linden & T.J.H.M. Eggen, *The Rasch model as a model for paired comparisons with an individual tie parameter*
- RR-88-8 R.J.H. Engelen, W.J. van der Linden, & S.J. Oosterloo, *Item information in the Rasch model*
- RR-88-7 J.H.A.N. Rikers, *Towards an authoring system for item construction*
- RR-88-6 H.J. Vos, *The use of decision theory in the Minnesota Adaptive Instructional System*
- RR-88-5 W.J. van der Linden, *Optimizing incomplete sample designs for item response model parameters*
- RR-88-4 J.J. Adema, *A note on solving large-scale zero-one programming problems*

- RR-88-3 E. Epekkooi-Timminga, *A cluster-based method for test construction*
- RR-88-2 W.J. van der Linden & J.J. Adema, *Algorithmic test design using classical item parameters*
- RR-88-1 E. van der Burg & J. de Leeuw, *Nonlinear redundancy analysis*

Research Reports can be obtained at costs from Bibliotheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.

Department of
EDUCATION

Education by
the Department of Education
of the University of Twente