DOCUMENT RESUME

ED 419 004 TM 028 301

AUTHOR	van der Linden, Wim J.
TITLE	A Procedure for Empirical Initialization of Adaptive Testing Algorithms.
INSTITUTION	Twente Univ., Enschede (Netherlands). Faculty of Educational Science and Technology.
REPORT NO	RR-97-02
PUB DATE	1997-00-00
NOTE	20p.
AVAILABLE FROM	Faculty of Educational Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Ability; *Adaptive Testing; *Algorithms; *Computer Assisted Testing; Foreign Countries; *Item Response Theory; *Test Construction
IDENTIFIERS	*Constraints; Two Parameter Model

ABSTRACT

In constrained adaptive testing, the numbers of constraints needed to control the content of the tests can easily run into the hundreds. Proper initialization of the algorithm becomes a requirement because the presence of large numbers of constraints slows down the convergence of the ability estimator. In this paper, an empirical initialization of the algorithm is proposed based on the statistical relation between the ability variable and background variables known prior to the test. The relation is modeled using a two-parameter logistic version of an item response theory (IRT) model with manifest predictors discussed in A. H. Zwinderman (1991). An empirical example shows how an (incomplete) sample of response data and data on background variables can be used to derive an initial ability estimate or an empirical prior distribution for the ability parameter. An appendix gives the derivation of an equation for the estimator. (Contains 12 references.) (Author/SLD)

A Procedure for Empirical Initialization of Adaptive Testing Algorithms

Research Report 97-02

Wim J. van der Linden

A Procedure for Empirical Initialization of Adaptive Testing Algorithms

Wim J. van der Linden

Abstract

In constrained adaptive testing, the numbers of constraints needed to control the content of the tests can easily run into the hundreds. Proper initialization of the algorithm becomes a requirement because the presence of large numbers of constraints slows down the convergence of the ability estimator. In this paper, an empirical initialization of the algorithm is proposed based on the statistical relation between the ability variable and background variables known prior to the test. The relation is modeled using a two-parameter logistic version of an IRT model with manifest predictors discussed in Zwinderman (1991). An empirical example shows how an (incomplete) sample of response data and data on background variables can be used to derive an initial ability estimate or an empirical prior distribution for the ability parameter.

A Procedure for Empirical Initialization of Adaptive Testing Algorithms

Item response theory (IRT) models the probability of a response to a test item as the result of an interaction between the properties of the item and the ability of the examinee. Typically, this interaction is mapped on a parameter structure for the probability function for the response with separate parameters for the examinee and the item. One of the main advantages of separate parameterization of examinees and items is that it is possible to select items to match the abilities of examinees. If a conventional linear test has to be assembled, a standard approach is to set a target for the information function of the test with optimal values for the part of the ability scale where the examinees are expected to be and select a combination of items that meets the target best in some sense (Birnbaum, 1968). A more powerful application of the principle is found in computerized adaptive testing (CAT) where each individual item in the test is selected to match the current estimate of the ability of the examinee. A popular implementation of the principle of adaptive testing calculates the maximum-likelihood (ML) estimate of ability from the updated response vector of the examinee and selects the next item to have maximum statistical information at the estimate. An alternative Bayesian procedure is to select the items to optimize the posterior distribution of the ability parameter.

Under general conditions, both the MLE and the Bayesian estimator defined on the posterior ability distribution are known to converge to the true ability (Chang & Ying, 1996; Gelman, Carlin, Stern, & Rubin, 1995, Appendix B). The speed of convergence of the algorithm depends on the initialization of the algorithm. Generally, the farther the initial ability estimate or prior distribution away from the true ability of the examinee, the slower the algorithm converges to an estimator with prescribed precision. On the other hand, a perfect initialization does not imply an immediate stop of the algorithm. IRT models define a stochastic relation between the ability and the responses, and even with a perfect start a CAT algorithm needs some time to accept the true ability value with enough certainty.

In constrained adaptive testing the objective is to select items from the pool to maximize the statistical precision of the ability estimator subject to constraints on item or test attributes, or constraints needed to deal with a possible item-set structure or to control item exposure. For a realistic testing program, the number of constraints can easily run into the hundreds (van der Linden & Reese, 1997). Generally, the presence of such constraints slows down the convergence of the ability estimator reinforcing the need for optimal initialization of the algorithm.

An obvious way to improve initialization of a CAT procedure is to use empirical predictors of the ability of the examinee. These predictors may be available in the form of information on background variables at hand when administering the test. For example, to register for a CAT sessions examinees usually have to fill out a form with biographical information, and a useful statistical relation may be present between the data in the form and the abilities of the examinee. Also, most CAT sessions begin with the examinees reading the instruction to the test and responding to a few exercises before the actual test starts. The time needed to work through the instruction and/or the responses given to the exercises may already contain statistical information on the ability of the examinee. As a final example, knowledge of scores on previous attempts to pass the same test, possibly in combination with the amount of time elapsed since the last attempt and/or information on intermediate coaching could be used to predict the ability of the examinee.

The point of view that additional sources of information on ability available at the time of testing should be exploited is certainly not new. The principle forms a standard belief in Bayesian statistics. However, application of the principle has been inhibited by the perception that in testing the abilities of the examinees should "speak for themselves" and that it may be unfair to let (possibly unfavorable) background information creep into score test scores. This concern is unfunded because it does not make the important distinction between the experiment of selecting the items and the one of generating responses to the items once they are selected. The key question is whether the former can be ignored when estimating the ability of the examinee from the responses obtained through the latter (for a formal definition of ignorability, see Little and Rubin, 1987). As shown in Mislevy and Wu (1988), in adaptive testing the item selection mechanism can be ignored under maximum-likelihood estimation if the interest is in the value of the ability estimate and not in inferences with respect to the sampling distribution of the estimator. Bayesian inference is legitimate provided the knowledge of the background variables has been incorporated into the prior. More formally, the condition means that if data on P background variables X_p, p=0,...,P with a statistical relation to the ability variable θ have been inspected by the statistician, the conditional distribution of θ given $X_1=x_1,...,X_P=x_P$ is the correct prior in Bayesian inference. This paper addresses the question of how to obtain an <u>empirical</u> estimate of the prior from data on the background variables X_p. The estimate can be used to initialize an empirical Bayes algorithm in adaptive testing. Also, the mean of the prior provides an initial point estimate of the ability of the examinee and can be used to select the first item if the interest is in point estimation of ability, for example, in combination with maximum-information item selection.

Models and Procedures

It is assumed that the probability of a successful response to item i=1,...,I in the pool can be described by the 2-parameter logistic (2-PL) model:

$$p_i(\theta) \equiv \mathrm{Prob}\{U_i = 1|\theta\} \equiv \frac{\exp[a_i(\theta - b_i)]}{1+\exp[a_i(\theta - b_i)]}, \quad (1)$$

where $\theta \in (-\infty,\infty)$ is the parameter representing the ability of the examinee, and $b_i \in (-\infty,\infty)$ and $a_i \in [0,\infty)$ are the parameters for the difficulty and discrimination of the item, respectively.

Further, it is assumed that the P predictor or background variables, X_p, p=0,...,P, have the following statistical relation to the ability parameter:

$$\theta = \beta_0 + \beta_1 X_1 + ... + X_P \beta_P + \varepsilon, \quad (2)$$

with error term ε distributed as

$$\varepsilon \_ N(0,\sigma^2). \quad (3)$$

Recall that the model in (2) only has to be linear in the parameters β_p. The model therefore covers the wide class of relations that can be brought into linear form by a monotonic transformation of the original predictor variables (for some examples, see Neter, Wasserman & Kutner, 1990, chap. 4). Also, the variables in (2) can be chosen to represent higher-order predictors in an original polynomial model. Generally, the assumption of normality in (3) is a better approximation to reality, the larger the selection of predictor variables in (2).

From (1)-(2) it follows that

$$\theta | X_1 = x_1, ..., X_P = x_P \_ N(\beta_0 + \beta_1 x_1 + ... + \beta_p xP, \sigma^2), \quad (4)$$

is the empirical distribution of the ability of an examinee randomly sampled from a subpopulation with X_1=x_1, ..., X_p=x_p. Before introducing the distribution in (4) as an empirical prior in adaptive testing, the adaptive procedures considered in this paper will be introduced.

The items in the pool are denoted by index i=1,...,I, and index k=1,...,n is used to represent the position of the items in the test. Thus, i_k is the index of the item in the pool administered as the kth item in the test. The set $S_{k-1} \equiv \{i_1,,,,i_{k-1}\}$ is used to denote the first k-1 items in the test. The set of remaining items in the pool is denoted as $R_k \equiv \{1,...,I\} \backslash S_{k-1}$. Finally, $\hat{\theta}^{(k-1)}$ represents the estimator of θ after k-1 items have been administered.

In adaptive testing with the maximum-information criterion, $\hat{\theta}^{(k)}$ is chosen to be the ML estimate of θ and the next item is selected according to

$$i_k \equiv \max_h \{I_h(\hat{\theta}^{(k-1)});\ h \in R_k\}, \tag{5}$$

where $I_h(\theta)$ is defined as Fisher's information in U_h on θ. In a Bayesian approach, the response $U_{i_{k-1}} = u_{i_{k-1}}$ is used to update the posterior distribution of θ by an application of Bayes theorem:

$$f(\theta | u_{i_1},\ldots,u_{i_{k-1}}) = \frac{f(u_{i_{k-1}}|\theta)\, f(\theta|u_{i_1},\ldots,u_{i_{k-2}})}{\int f(u_{i_{k-1}}|\theta)\, f(\theta|u_{i_1},\ldots,u_{i_{k-2}})\, d\theta}, \tag{6}$$

where $f(u_{i_{k-1}}|\theta)$ is the probability function of $U_{i_{k-1}} = u_{i_{k-1}}$ defined by (1). Several Bayesian item selection criteria are possible. For example, a well-known practice is to choose $\hat{\theta}^{(k-1)}$ to be the expected a posteriori (EAP) estimator, which is the expected value of θ over the distribution in (6), and use this estimator to select the item with maximum information. Another example is the minimum expected posterior variance criterion which for each remaining item in the pool, $h \in R_{k-1}$, predicts the posterior variance of the ability estimator both after a correct and an incorrect response and selects the item with minimum expected variance over the responses; that is,

$$\begin{aligned} i_k \equiv \max_h \{ & 1 - p_h(\hat{\theta}^{(k-1)})\, \mathrm{Var}(\theta|u_{i_1},\ldots,u_{i_{k-1}}, U_h = 0) \\ & + p_j(\hat{\theta}^{(k-1)})\, \mathrm{Var}(\theta|u_{i_1},\ldots,u_{i_{k-1}}, U_h = 1); h \in R_{k-1}\}. \end{aligned} \tag{7}$$

Other examples of Bayesian item selection criteria are given in van der Linden (1996).

Initialization of the Algorithm

Both the maximum-information and Bayesian item selection criteria need proper initialization to accelerate the convergence of the algorithm. If ML estimation is used, the initial item often is chosen to be informative near the middle of the ability distribution expected for the population of examinees, but this choice may be suboptimal for a substantial portion of the examinees. In a Bayesian procedure, a typical choice of the prior distribution is a flat prior or a normal prior with large variance located at the middle of the expected ability distribution. The former ignores possible information available about the examinee and the latter may also be suboptimal for a substantial portion of the examinees.

In this paper, an empirical estimate of (4) as the prior in the adaptive procedure is proposed, that is, the density $f(\theta|x_1,...,x_p)$ belonging to the normal distribution in (4) is used to initialize the procedure. This prior defines the following EAP estimate

$$\hat{\theta}^{(0)} = \beta_0+\beta_1 x_1+...+\beta_p x_p, \qquad (8)$$

which can be used as an initial point estimate for the ability parameter in the maximum-information criterion. In a full Bayesian procedure, $f(\theta|x_1,...,x_p)$ can be used as an empirical prior, yielding as the first posterior:

$$f(\theta|u_h,x_1,...,x_P) = \frac{f(u_h|\theta)\,f(\theta|x_1,...,x_P)}{\int f(u_h|\theta)\,f(\theta|x_1,...,x_P)\,d\theta},\ h\in R_0, \qquad (9)$$

where $f(u_h|\theta,x_1,...,x_p)=f(u_h|\theta)$ due to conditional independence of U_h and $X_1,...,X_p$ given θ.

To implement (8), the values of the parameters $\beta_0,...,\beta_P$ should be known. To implement (9), in addition the value of the parameter σ^2 is needed. This value can be interpreted as an <u>empirical</u> measure of the prior uncertainty about θ. A method for estimating the values of these parameters is discussed in the following section.

Estimation of Regressions Parameters

A seemingly straightforward approach to estimating the regression parameters in (2) ould be to regress $\hat{\theta}$ on the predictors $X_1,...,X_P$ using the minimum least-squares criterion.

However, though the approach might yield satisfactory estimators of the parameters $\beta_0,...,\beta_p$, it does not give sound results for the estimation of the variance of the prior, σ^2, due to confounding of the prediction error ε with the estimation error in $\hat{\theta}$. Thus, though this approach would allow empirical prediction of the initial ability estimate in a CAT procedure, it would not give us a proper empirical estimate of the prior uncertainty needed for the update in (9). Therefore, a better alternative is direct estimation of $\beta_0,..,\beta_p$ and σ from response data.

Substitution of the regression equation in (2) into the 2-PL model in (1) for an examinee with $X_1=x_1,...,X_p=x_p$ gives the following logistic regression model:

$$p_i(\theta) \equiv \text{Prob}\{U_i = 1|\theta\} \equiv \frac{\exp[a_i(\beta_0+\beta_1 x_1+...+\beta_P x_P+\varepsilon - b_i)]}{1+\exp[a_i(\beta_0+\beta_1 x_1+...+\beta_P x_P+\varepsilon - b_i)]}. \qquad \textbf{(10)}$$

For $a_i=1$, $i=1,...,I$, the model in (10) was discussed by Zwinderman (1991, 1997) as a generalized Rasch model with manifest predictors. Following an approach in Rigdon and Tsutakawa (1983), Zinderman presents an EM algorithm for joint estimation of the item difficulty and the regression parameters. The algorithm is adapted here to the case of the 2-PL with item parameters known from previous item pool calibration. In addition, a discussion is included on how to use the algorithm to estimate the regressions parameters from data collected in an operational adaptive testing program for which it holds that responses to some of the items in the pool are missing.

Let u_{ij}, $i=1,...,I$, $j=1,...,N$, denote the response of examinee j on item i. For each examinee, there is an unknown realization ε_j of the error term ε in (3) which is treated as missing data in the EM algorithm. The values of the predictor variables $X_{1j}=x_{1j},...,X_{Pj}=x_{Pj}$ are treated as known parameters. For examinee j, let $p(\underline{u}_j|x_j,\underline{\beta},\varepsilon_j)$ be the probability of the observed response vector $\underline{u}_j=(u_{j1},...,u_{jI})$ given predictor values $\underline{x}_j=(x_1,...,x_P)$, parameter vector $\underline{\beta}=(\beta_0,...,\beta_P)$, and missing datum ε_j. In addition, $p(\varepsilon_j|\sigma)$ is the (normal) probability density of ε_j. It follows that

$$p(\underline{u}_j|x_j,\underline{\beta},\varepsilon_j) = \prod_{k=1}^{n} p_i(\theta_j)^{u_{ij}}(1-p_i(\theta_j))^{(1-u_{ij})} \qquad \textbf{(11)}$$

and

$$p(\varepsilon_j|\sigma) = (2\pi\sigma^2)^{-1/2}\exp(-\varepsilon_j^2/2\sigma^2). \quad \textbf{(12)}$$

The log-likelihood function associated with the complete data is given by

$$L(\underline{\beta},\sigma;\underline{u}_j|x_j,\varepsilon_j) = -\frac{1}{2}\ln(2\pi\sigma^2) - \varepsilon_j^2/2\sigma^2 \sum_{i=1}^{I} u_{ij}\ln(p_i(\theta_j)) + (1-u_{ij})\ln(1-p_i(\theta_j)). \quad \textbf{(13)}$$

The expectation of the complete-data log-likelihood over the posterior predictive distribution of ε_j is calculated. The density of this distribution is given by

$$p(\varepsilon_j|\underline{u}_j,x_j,\underline{\beta},\sigma) \propto p(\underline{u}_j|x_j,\underline{\beta}_j,\varepsilon_j)\,p(\varepsilon_j|\sigma)$$

$$\propto \frac{\exp[\varepsilon_j \sum_{i=1}^{I} a_i u_{ij} - \varepsilon_j^2/2\sigma^2]}{\prod_{i=1}^{I}\{1 + \exp[a_i(\beta_0+\ldots+\beta_P x_{jP} + \varepsilon_j - b_i)]\}}. \quad \textbf{(14)}$$

The expected complete-data log-likelihood for N examinees is equal to

$$E(L(\underline{\beta},\sigma;\underline{u}|x) = -\frac{N}{2}\ln(2\pi\sigma^2) - (1/2\sigma^2)\sum_{j=1}^{N}\int \varepsilon^2 p(\varepsilon|\underline{u}_j,x_j,\underline{\beta},,\sigma)\,d\varepsilon$$

$$+ \sum_{j=1}^{N}\sum_{i=1}^{I}\int[u_{ij}\ln p_i(\theta_j) + (1-u_{ij})\ln(1-p_i(\theta_j))]\,p(\varepsilon|\underline{u}_j,x_j,\underline{\beta},\sigma)\,d\varepsilon, \quad \textbf{(15)}$$

where $\underline{u} = (u_{ij})$.

The algorithm consists of repeated application of expectation and maximization steps until convergence. At step t the calculations are:

E-step. Calculation of the expected complete-data log-likelihood in (15) given the values of $\sigma^{(t-1)}$ and $\beta_p^{(t-1)}$, p=0,...,P calculated at the step t-1.

M-step. Calculation of the values of the estimates $\sigma^{(t)}$ and $\beta_p^{(t)}$, p=0,...,P, maximizing the expected complete-data log-likelihood from the E-step.

As shown in the Appendix, the two steps boil down to iterative use of the following recursive relations:

$$\sigma^{2(t)} = N^{-1}\sum_{j=1}^{N}\int \varepsilon^2 p(\varepsilon|\underline{u}_j,\underline{x}_j,\underline{\beta}^{(t-1)},\sigma^{(t-1)})\,d\varepsilon \tag{16}$$

$$\sum_{j=1}^{N} x_{jp}\sum_{i=1}^{I} a_i u_{ij} =$$

$$\sum_{j=1}^{N} x_{jp}\sum_{i=1}^{I} a_i \int \frac{\exp[a_i(\beta_0^{(t)}+\ldots+\beta_p^{(t)}x_{jp}+\ldots+\beta_P^{(t)}x_{jP}+\varepsilon-b_i)]}{1+\exp[a_i(\beta_0^{(t)}+\ldots+\beta_p^{(t)}x_{jp}+\ldots+\beta_P^{(t)}x_{jP}+\varepsilon-b_i)]}\,.$$

$$p(\varepsilon|\underline{u}_j,\underline{x}_j,\underline{\beta}^{(t-1)},\sigma^{(t-1)})\,d\varepsilon, \qquad p=0,\ldots,P, \tag{17}$$

where $x_{j0} = 1$. Note that in the first equation the prior variance σ^2 is equated to the average posterior predicted variance. Likewise, the left-hand sums in (17) are equated to their posterior predicted expected values. Standard numerical procedures can be used to solve the equations. For a possible choice of procedure, see the empirical example below.

Parameter estimation in operational CAT. The notation in the preceding section assumes that the examinees respond to all of the items in the pool. However, as already discussed, missing responses to items in an adaptive test can be ignored in ML estimation if no inferences are made with respect to the sampling distribution of the estimator. Thus, if the interest is only in point estimates of the parameters $\beta_1,\ldots,\beta_P$ and σ^2 and not in estimating such quantities as their standard errors, data from an operational CAT program can be used to calculate these point estimates. In doing so, for each examinee the items not administered are simply omitted in the equations in (16)-(17).

In addition, it is possible to use new responses to update previous estimates of the parameters by extending the sums in (16)-(17) over the new examinees and using the old estimates as the starting values in new iterations of the EM algorithm.

Empirical Example

Adaptive versions of Dutch translations of four subtests of the General Aptitude Test Battery (GATB) were studied in Schoonman (1989). In this study, the subtest Name Comparison was first administered to the examinees in the sample (N=306) and total response time was recorded. Responses on the next subtest, Vocabulary, were used to estimate the ability of the examinees measured by this subtest. Schoonman reported a linear correlation between the log response times on the first test and the estimated abilities on the second test equal to -.46.

The data set was reanalyzed to estimate both the parameters β_0 and β_1 in the regression equation of the true Vocabulary ability on the log response time and the uncertainty parameter σ directly from the data. The equations in (17) were solved for p=0,1 using Newton's method which gives the updates

$$(\beta_0^{(t)},\beta_1^{(t)})' = (\beta_0^{(t-1)},\beta_1^{(t-1)})' - \begin{bmatrix} \frac{\partial^2 \ln E(L)}{\partial \beta_0^2} & \frac{\partial^2 \ln E(L)}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 \ln E(L)}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 \ln E(L)}{\partial \beta_1^2} \end{bmatrix}^{-1} . (\frac{\partial E(L)}{\partial \beta_0}, \frac{\partial E(L)}{\partial \beta_1})' . \quad (18)$$

The first derivatives in (18) are given in (A.5), whereas the second derivatives can be shown to be equal to

$$\frac{\partial^2 E(L)}{\partial \beta_0^2} = -\sum_{j=1}^{N}\sum_{i=1}^{I} a_i^2 \int \frac{\exp[a_i(\beta_0^{(t)}+\beta_1^{(t)}x_j+\varepsilon-b_i)]}{[1+\exp[a_i(\beta_0^{(t)}+\beta_1^{(t)}x_j+\varepsilon-b_i)]]^2} p(\varepsilon|.)\,d\varepsilon, \quad (19)$$

$$\frac{\partial^2 E(L)}{\partial \beta_1^2} = -\sum_{j=1}^{N} x_j^2 \sum_{i=1}^{I} a_i^2 \int \frac{\exp[a_i(\beta_0^{(t)}+\beta_1^{(t)}x_j+\varepsilon-b_i)]}{[1+\exp[a_i(\beta_0^{(t)}+\beta_1^{(t)}x_j+\varepsilon-b_i)]]^2} p(\varepsilon|.)\,d\varepsilon, \quad (20)$$

$$\frac{\partial^2 E(L)}{\partial \beta_0 \beta_1} = -\sum_{j=1}^{N} x_j \sum_{i=1}^{I} a_i^2 \int \frac{\exp[a_i(\beta_0^{(t)}+\beta_1^{(t)}x_j+\varepsilon-b_i)]}{[1+\exp[a_i(\beta_0^{(t)}+\beta_1^{(t)}x_j+\varepsilon-b_i)]]^2} p(\varepsilon|.)\,d\varepsilon, \quad (21)$$

and $p(\varepsilon|.)$ follows from (14). The integrals in (16) and (18) were calculated using Gauss-Hermite quadrature. Several sets of starting values were tried, all resulting in the following estimates: $\hat{\beta}_0$=5.833, $\hat{\beta}_1$=-1.279, and $\hat{\sigma}^2$=.986. Thus, the best way to initialize the adaptive procedure for the Vocabulary test is to use $\hat{\theta}^{(0)}$=5.833-1.279x or to use N(5.833-1.279x,.986) as a prior distribution for θ.

The standard deviation of the log response times was calculated as .433. It follows that the correlation between the true abilities and the log response times can be estimated as -.59. The difference between this result and Schoonman's estimate of -.46 indicates the loss of information incurred when estimating the regression parameters from the estimated abilities rather than the true abilities using the model in this paper.

Appendix

The equation for the estimator of σ^2 is found by setting the derivative of the expected complete-data log-likelihood in (15) to σ equal to zero. As only the first two terms are needed,

$$\frac{\partial}{\partial\sigma}[-N/2\ln(2\pi\sigma^2) - (1/2\sigma^2)\sum_{j=1}^{N}\int\varepsilon^2 p(\varepsilon|\underline{u}_j,x_j,\underline{\beta},\sigma)\,d\varepsilon]$$

$$= -N\sigma^{-1}+\sigma^{-3}\sum_{j=1}^{N}\int\varepsilon^2|\underline{u}_j,x_j,\underline{\beta},\sigma)\,d\varepsilon. \tag{A.1}$$

Multiplying by σ^2 yields

$$\sigma^2 = N^{-1}\sum_{j=1}^{N}\int\varepsilon^2 p(\varepsilon|\underline{u}_j,x_j,\underline{\beta},\sigma)\,d\varepsilon. \tag{A.2}$$

Likewise, the equations for β_p, p=0,...,P, are found by setting the derivative of the last term of (15) equal to zero:

$$\frac{\partial}{\partial\beta_p}[\sum_{j=1}^{N}\sum_{i=1}^{I}\int[u_{ij}\ln p_{ij}+(1-u_{ij})\ln(1-p_{ij})]\,p(\varepsilon|\underline{u}_j,x_j,\underline{\beta},\sigma)\,d\varepsilon$$

$$= \sum_{j=1}^{N}\sum_{i=1}^{I}\int[(\frac{u_{ij}(1-p_{ij})}{p_{ij}(1-p_{ij})}-\frac{(1-u_{ij})p_{ij}}{p_{ij}(1-p_{ij})})\frac{\partial p_{ij}}{\partial\beta_p}]\,p(\varepsilon|\underline{u}_j,x_j,\underline{\beta},\sigma)\,d\varepsilon = 0. \tag{A.3}$$

Because

$$\frac{\partial p_{ij}}{\partial\beta_p} = a_i x_{jp} p_{ij}(1-p_{ij}), \tag{A.4}$$

it follows that

$$\sum_{j=1}^{N}\sum_{i=1}^{I}\int a_i x_{jp}(u_{ij}-p_{ij})\,p(\varepsilon|\underline{u}_j,x_j,\underline{\beta},\sigma)\,d\varepsilon = 0. \tag{A.5}$$

Hence,

$$\sum_{j=1}^{N} x_{jp} \sum_{i=1}^{I} a_i u_{ij} = \sum_{j=1}^{N} x_{jp} \sum_{i=1}^{I} a_i \int \frac{\exp[a_i(\beta_0 + \beta_1 x_{j1} + \ldots + \beta_P x_{jP} + \varepsilon - b_i)]}{1 + \exp[a_i(\beta_0 + \beta_1 x_{j1} + \ldots + \beta_p x_{jP} + \varepsilon - b_i)]} \cdot$$

$$p(\varepsilon | \underline{u}_j, \underline{x}_j, \underline{\beta}, \sigma)\, d\varepsilon . \quad \textbf{(A.6)}$$

References

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick, Statistical theories of mental test scores. reading, MA: Addison-Wesley.

Chang, H.-H,, & Ying, Z (1996, June). Building a statistical foundation for computerized adaptive testing. Paper presented at the annual meeting of the Psychometric Society, Banff, Alberta, Canada.

Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B (1995). Bayesian data analysis. New York: Chapman & Hall.

Little, R.J.A., & Rubin, D.B. (1987). Statistical analysis with missing data. New York: Wiley.

Mislevy, R.J. & Wu, P.-K. (1988). Inferring examinee ability when some items are missing (RR 88-48). Princeton, NJ: Educational Testing Service.

Neter, J., Wasserman, W., & Kutner, M.H. (1990). Applied linear statistical models (3rd Ed.). Homewood,IL: Irwin.

Rigdon, S.E., & Tstutakawa, R.K. (1983). Parameter estimation in latent trait models. Psychometrika, 48, 567-574.

Schoonman, W. (1989). An applied study on computerized adaptive testing. Amsterdam, The Netherlands: Swets & Zeitlinger.

van der Linden, W.J. (1996). Bayesian item selection criteria for adaptive testing (Research Report 96-1). Enschede, The Netherlands: University of Twente, Department of Educational Measurement and Data Analysis.

van der Linden, W.J., & Reese, L.M. (1997). A model for optimal constrained adaptive testing. Manuscript submitted for publication.

Zwinderman, A.H. (1991). A generalized Rasch model for manifest predictors. Psychometrika, 56, 589-600.

Zwinderman, A.H. (1997). Response models with manifest predictors. In W.J. van der Linden & R.K. Hambleton (Eds.), Handbook of modern item response theory (pp. 245-256). New York: Springer-Verlag.

Author Note

This study received funding from the Law School Admission Council (LSAC). The opinions and conclusions contained in this report are those of the author and do not necessarily reflect the position or policy of LSAC. This report is also published in the LSAC Research Report Series. The author is indebted to Wouter Schoonman for the data set used in the example, Wim M.M. Tielen for his computational assistance, Cees A.W. Glas for lending one of his software procedures, and Lynda M. Reese for her editorial comments. Address all correspondence to: W.J. van der Linden, Department of Educational Measurement and Data Analysis, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands. Email: vanderlinden@edte.utwente.nl.

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

RR-97-02	W.J. van der Linden, *A Procedure for Empirical Initialization of Adaptive Testing Algorithms*
RR-97-01	W.J. van der Linden & Lynda M. Reese, *A Model for Optimal Constrained Adaptive Testing*
RR-96-04	C.A.W. Glas & A.A. Béguin, *Appropriateness of IRT Observed Score Equating*
RR-96-03	C.A.W. Glas, *Testing the Generalized Partial Credit Model*
RR-96-02	C.A.W. Glas, *Detection of Differential Item Functioning using Lagrange Multiplier Tests*
RR-96-01	W.J. van der Linden, *Bayesian Item Selection Criteria for Adaptive Testing*
RR-95-03	W.J. van der Linden, *Assembling Tests for the Measurement of Multiple Abilities*
RR-95-02	W.J. van der Linden, *Stochastic Order in Dichotomous Item Response Models for Fixed Tests, Adaptive Tests, or Multiple Abilities*
RR-95-01	W.J. van der Linden, *Some decision theory for course placement*
RR-94-17	H.J. Vos, *A compensatory model for simultaneously setting cutting scores for selection-placement-mastery decisions*
RR-94-16	H.J. Vos, *Applications of Bayesian decision theory to intelligent tutoring systems*
RR-94-15	H.J. Vos, *An intelligent tutoring system for classifying students into Instructional treatments with mastery scores*
RR-94-13	W.J.J. Veerkamp & M.P.F. Berger, *A simple and fast item selection procedure for adaptive testing*
RR-94-12	R.R. Meijer, *Nonparametric and group-based person-fit statistics: A validity study and an empirical example*
RR-94-10	W.J. van der Linden & M.A. Zwarts, *Robustness of judgments in evaluation research*
RR-94-9	L.M.W. Akkermans, *Monte Carlo estimation of the conditional Rasch model*
RR-94-8	R.R. Meijer & K. Sijtsma, *Detection of aberrant item score patterns: A review of recent developments*
RR-94-7	W.J. van der Linden & R.M. Luecht, *An optimization model for test assembly to match observed-score distributions*
RR-94-6	W.J.J. Veerkamp & M.P.F. Berger, *Some new item selection criteria for adaptive testing*
RR-94-5	R.R. Meijer, K. Sijtsma & I.W. Molenaar, *Reliability estimation for single dichotomous items*

RR-94-4 M.P.F. Berger & W.J.J. Veerkamp, *A review of selection methods for optimal design*

RR-94-3 W.J. van der Linden, *A conceptual analysis of standard setting in large-scale assessments*

RR-94-2 W.J. van der Linden & H.J. Vos, *A compensatory approach to optimal selection with mastery scores*

RR-94-1 R.R. Meijer, *The influence of the presence of deviant item score patterns on the power of a person-fit statistic*

RR-93-1 P. Westers & H. Kelderman, *Generalizations of the Solution-Error Response-Error Model*

RR-91-1 H. Kelderman, *Computing Maximum Likelihood Estimates of Loglinear Models from Marginal Sums with Special Attention to Loglinear Item Response Theory*

RR-90-8 M.P.F. Berger & D.L. Knol, *On the Assessment of Dimensionality in Multidimensional Item Response Theory Models*

RR-90-7 E. Boekkooi-Timminga, *A Method for Designing IRT-based Item Banks*

RR-90-6 J.J. Adema, *The Construction of Weakly Parallel Tests by Mathematical Programming*

RR-90-5 J.J. Adema, *A Revised Simplex Method for Test Construction Problems*

RR-90-4 J.J. Adema, *Methods and Models for the Construction of Weakly Parallel Tests*

RR-90-2 H. Tobi, *Item Response Theory at subject- and group-level*

RR-90-1 P. Westers & H. Kelderman, *Differential item functioning in multiple choice items*

Research Reports can be obtained at costs, Faculty of Educational Science and Technology, University of Twente, Mr. J.M.J. Nelissen, P.O. Box 217, 7500 AE Enschede, The Netherlands.

NOTICE

REPRODUCTION BASIS

- [x] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

- [ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").