

CONFIDENCE INTERVALS FOR TEST INFORMATION AND RELATIVE EFFICIENCY

by

Sebie Oosterloo *

Abstract

In latent theory the measurement properties of a mental test can be expressed in the test information function. The relative merits of two tests for the same latent trait can be described by the relative efficiency function, i.e. the ratio of the test information functions. It is argued that these functions have to be estimated if the values of the item difficulties are unknown. Using conditional maximum likelihood estimation as indicated by Andersen (1973), pointwise asymptotic distributions of the test information and relative efficiency function are derived for the case of dichotomously scored Rasch homogeneous items. Formulas for confidence intervals are derived from the asymptotic distributions. An application to a mathematics test is given and extensions to other latent trait models are discussed.

Key words & Phrases: *latent trait theory, test information.*

1. Introduction

Consider an individual responding to a test consisting of k dichotomously scored items. Let X_j ($j=1, \dots, k$) describe the individual's response to item j , which may be correct ($X_j=1$) or mistaken ($X_j=0$). Rasch (1960) postulates that X_1, \dots, X_k are independent and that the logit of $P(X_j=1)$ is equal to $\theta - \alpha_j$, where the latent scores θ and α_j are interpreted as the individual's ability and the difficulty of item j . More precisely:

$$\log[P(X_j=1; \theta, \alpha_j) / P(X_j=0; \theta, \alpha_j)] = \theta - \alpha_j, \quad (1)$$

or

$$P(X_j=x; \theta, \alpha_j) = \exp[(\theta - \alpha_j)x - \psi(\theta - \alpha_j)], \quad (2)$$

where

$$\psi(\theta - \alpha_j) = \log[1 + \exp(\theta - \alpha_j)], \quad (3)$$

The joint distribution of X_1, \dots, X_k is thus given by

$$P(X_1=x_1, \dots, X_k=x_k; \theta, \alpha) = \exp[\theta \sum_{j=1}^k x_j - \sum_{j=1}^k \alpha_j x_j - \psi(\theta; \alpha)], \quad (4)$$

where $\alpha = (\alpha_1, \dots, \alpha_k)'$ (t denoting transposition) and

* Twente University of Technology

$$\psi(\theta; \alpha) = \sum_{j=1}^k \psi(\theta - \alpha_j). \tag{5}$$

If α is regarded as a known vector, then (4) expresses that, as θ varies, the distribution of $X = (X_1, \dots, X_k)$ belongs to an exponential family with canonical parameter θ and complete sufficient statistic $T = \sum_{j=1}^k X_j$. The complete sufficient statistics T is the sum of the independent Bernoulli variables X_1, \dots, X_k and has expectation

$$E_{\theta}(T) = \sum_{j=1}^k \exp(\theta - \alpha_j) [1 + \exp(\theta - \alpha_j)]^{-1} = \sum_{j=1}^k [1 + \exp(\alpha_j - \theta)]^{-1}. \tag{6}$$

Furthermore, the variance of T equals Fisher's information

$$I(\theta; \alpha) = E_{\theta}[(d/d\theta) \log P(X_1, \dots, X_k; \theta, \alpha)]^2$$

on θ contained in the item scores X_1, \dots, X_k :

$$\text{var}_{\theta}(T) = I(\theta; \alpha) = \sum_{j=1}^k [2 + \exp(\alpha_j - \theta) + \exp(\theta - \alpha_j)]^{-1}, \tag{7}$$

Note that the information on θ contained in all item scores together is the sum of the informations

$$i(\theta; \alpha_j) = [2 + \exp(\alpha_j - \theta) + \exp(\theta - \alpha_j)]^{-1} \quad (j = 1, \dots, k) \tag{8}$$

contained in the item scores separately.

The Lindeberg-Feller Central Limit Theorem (Rao, 1965, p. 108) implies that, under weak regularity conditions, the random variable T , having exact expectation (6) and exact variance (7), is also asymptotically normal. This result can be used to construct a 95% confidence interval for θ by inverting the test for $H: \theta = \theta_0$ which accepts H if

$$|T - E_{\theta_0}(T)| \leq 1.96 [\text{var}_{\theta_0}(T)]^{\frac{1}{2}}.$$

A second method to construct a confidence interval for θ is obtained by considering the asymptotic distribution of the maximum likelihood estimator $\hat{\theta} = \hat{\theta}(T, \alpha)$. Note that $\hat{\theta}$ is degenerate if $T = 0$ or $T = k$, but otherwise given by the equations

$$T = \sum_{j=1}^k [1 + \exp(\alpha_j - \hat{\theta})]^{-1}, \tag{9}$$

on account of (6), and general theory on maximum likelihood estimators for parameters in exponential families. The equations (9) have a unique solution; $\hat{\theta}$ is an increasing function of T with $\hat{\theta}(0) = -\infty$ and $\hat{\theta}(k) = +\infty$. A simple linearization of (6) in the neighbourhood of the solution $\hat{\theta}$ of (9) shows that $\hat{\theta}$ is asymptotically normal with expectation θ and variance $I^{-1}(\theta; \alpha)$, which is a general property of maximum likelihood estimators. It follows that, for $1 \leq T \leq k - 1$, $\hat{\theta} \pm 1.96 I^{-\frac{1}{2}}(\hat{\theta}; \alpha)$ defines an interval for θ with confidence coefficient

asymptotically equal to .95.

These results show that the performance of a test is determined, at least asymptotically, by the test information function $I(\cdot; \alpha)$.

The exact representation

$$I(\theta; \alpha) = [(d/d\theta)E_{\theta}(T)]^2 / \text{var}_{\theta}(T) \tag{10}$$

is sometimes used in measurement theory to give an alternative interpretation of $I(\theta; \alpha)$ (Lord, 1980). Note that $I^{\frac{1}{2}}(\theta; \alpha)$ expresses how fast the "true score" $E_{\theta}(T)$ changes, relative to the "standard error of measurement" $[\text{var}_{\theta}(T)]^{\frac{1}{2}}$, as θ varies.

If the practitioner has some general idea about the scale of ability, he can get an impression of the quality of a test by visually inspecting the graph of $I(\cdot; \alpha)$. Sometimes, however, the utility of a test can be summarized in a single number. An example thereof is offered by Birnbaum (1968), who argues that in the case of mastery testing, where a decision has to be made whether or not the ability reaches some minimum level θ_0 , the measurement properties of a test are well expressed by the quantity $I(\theta_0; \alpha)$. Note that $I(\theta_0; \alpha) \leq k/4$ and that the maximum $k/4$ is attained if $\alpha_1 = \dots = \alpha_k = \theta_0$, or in other words if the probability of responding correctly is equal to $\frac{1}{2}$ for all items for an individual with ability θ_0 .

Often the practitioner has to choose between two or more tests. If the vectors $\alpha = (\alpha_1, \dots, \alpha_k)'$ and $\beta = (\beta_1, \dots, \beta_k)'$ of two tests are known, then the relative efficiency function $RE(\cdot; \alpha, \beta)$, defined by $RE(\theta; \alpha, \beta) = I(\theta; \alpha) / I(\theta; \beta)$, is a useful instrument in comparing the tests. If, for example, for some value θ , $RE(\theta; \alpha, \beta) = 2$, then the number of items in the second test (with difficulty vector β), would have to be doubled in order to be as informative as the first test.

Many examples of the use of information and relative efficiency functions in constructing, revising, or choosing a test can be found in Lord (1980). Particularly interesting are the applications in the area of tailored testing.

2. Test information and relative efficiency as unknown parameters

The introduction was based on the assumption that the item difficulties are known. This may approximately be true for some "standard" items, but is not true in general. Usually the difficulties have to be estimated from the outcome of a matrix X_{ij} ($j = 1, \dots, k; i = 1, \dots, n$) of independent dichotomous variables describing the response of individual i with ability θ_i , on item j with difficulty α_j . Hence

$$\begin{aligned} P(X_{11} = x_{11}, \dots, X_{nk} = x_{nk}; \theta_1, \dots, \theta_n, \alpha_1, \dots, \alpha_k) \\ = \exp\left[\sum_{i=1}^n \theta_i \sum_{j=1}^k x_{ij} - \sum_{j=1}^k \alpha_j \sum_{i=1}^n x_{ij} - \sum_{i=1}^n \sum_{j=1}^k \psi(\theta_i - \alpha_j) \right]. \end{aligned} \tag{12}$$

Note that the parameter vector $(\theta_1, \dots, \theta_n, \alpha_1, \dots, \alpha_k)$ is not identifiable: adding a constant to all coordinates does not change the joint distribution of the X_{ij} . This

indeterminacy can be removed in different ways. The approach to be chosen should depend on the situation at hand. The following possibilities are mentioned:

- (1) If one item, say item k , can be considered to be "standard", the restriction $\alpha_k = 0$ should be imposed.
- (2) If the n testees are regarded as "representative", then $\sum \theta_i = 0$ might be imposed; analogously if the k item are representative for the item pool, then $\sum \alpha_j = 0$ seems appropriate.
- (3) One might also postulate that $\theta_1, \dots, \theta_n$ is a random sample from some distribution (e.g. a logistic one) with a few unknown parameters. A location parameter of this distribution could then be fixed at a given value, e.g. 0 or 100 (as in the case of standard intelligence). The same procedure might be applied to $\alpha_1, \dots, \alpha_k$.

If no extraneous information is available, it is arbitrary which restriction is placed on the parameters.

Whatever restriction is chosen, the unknown difficulty parameters have to be estimated. As a consequence, $I(\theta; \alpha)$ (θ fixed) and $RE(\theta; \alpha)$ (θ fixed) appear also as unknown parameters to be estimated. Interpretation of estimates of these parameters should take into account the corresponding uncertainties, at least unless these uncertainties are negligible.

Suppose, for example, that an educational institute considers a proposal to replace a standard mastery test, with "known" difficulty vector α , by a new test, because the estimate of the information $I(\theta_0; \beta)$ of the new test exceeds $I(\theta_0; \alpha)$, where θ_0 is the minimum level of ability required. This replacement should take place, however, only if it is reasonably certain that the true value $I(\theta_0; \beta)$ exceeds $I(\theta_0; \alpha)$. It thus becomes interesting to test the hypothesis $H: I(\theta_0; \beta) \leq I(\theta_0; \alpha)$ against its negation, or, almost equivalently, to construct a confidence interval for $I(\theta_0; \beta)$.

3. Asymptotic distributions

In this section a theorem due to Andersen (1973) will be applied to derive the asymptotic distributions of $\log I(\theta; \hat{\alpha})$ and $\log RE(\theta; \hat{\alpha}, \hat{\beta})$, as the size(s) of the sample(s) of persons tend(s) to infinity. Here $\hat{\alpha}$ and $\hat{\beta}$ stand for the so-called conditional maximum likelihood estimators for α and β , to be defined below. In the next section these asymptotic distributions will be used to construct confidence intervals for $I(\theta; \alpha)$ and $RE(\theta; \alpha, \beta)$, respectively. Basing these confidence intervals on the distributions of the logarithms of the estimators $I(\theta; \hat{\alpha})$ and $RE(\theta; \hat{\alpha}, \hat{\beta})$, and not simply on the distributions of these estimators themselves, has two advantages:

- (i) the confidence intervals obtained for $I(\theta; \alpha)$ and $RE(\theta; \alpha, \beta)$ will always have positive lower bounds,
- (ii) the confidence interval for $RE(\theta; \alpha, \beta)$ will be "invariant", in the sense that the upper (lower) bound of the interval for $RE(\theta; \alpha, \beta)$ coincides with the reciprocal of the lower (upper) bound for $RE(\theta; \alpha, \beta) = [RE(\theta; \alpha, \beta)]^{-1}$.

From now on it will be postulated that $\alpha_k = 0$. At the end of this section it will be indicated how to proceed if one of the other restrictions, mentioned in the preceding section, is placed on the parameters.

Setting $\alpha_k = 0$, reduces (12) to the exponential family

$$\exp\left[\sum_{i=1}^n \theta_i \sum_{j=1}^k x_{ij} - \sum_{j=1}^{k-1} \alpha_j \sum_{i=1}^n x_{ij} - \chi(\theta_1, \dots, \theta_n, \alpha_1, \dots, \alpha_{k-1})\right], \quad (13)$$

where

$$\chi(\theta_1, \dots, \theta_n, \alpha_1, \dots, \alpha_{k-1}) = \sum_{i=1}^n \sum_{j=1}^{k-1} \psi(\theta_i - \alpha_j) + \sum_{i=1}^n \psi(\theta_i). \quad (14)$$

From (13) it is clear that $T_1, \dots, T_n, V_1, \dots, V_{k-1}$, defined by

$$T_i = \sum_{j=1}^k X_{ij}; \quad V_j = \sum_{i=1}^n X_{ij} \quad (15)$$

form a set of complete sufficient statistics. A drawback of the maximum likelihood estimators for the item parameters obtained by maximizing (13) with respect to both $\alpha_1, \dots, \alpha_{k-1}$ and $\theta_1, \dots, \theta_n$, is their inconsistency. Situations where maximum likelihood estimators are inconsistent are described by Neyman and Scott (1948). In their terminology the item parameters are the structural parameters and the abilities are called the incidental parameters. Theoretically satisfactory estimators for the item parameters are obtained by using the so-called conditional maximum likelihood method. The idea is to drive away the incidental parameters $\theta_1, \dots, \theta_n$ from (13) and then maximizing the resulting conditional likelihood with respect to $\alpha_1, \dots, \alpha_{k-1}$. Note that, conditionally given $(T_1, \dots, T_n) = (t_1, \dots, t_n)$, the random vectors are independently $(X_{i1}, \dots, X_{i,k-1})$ ($i = 1, \dots, n$), are independent, having exponential distributions given by

$$\begin{aligned} P(X_{i1} = x_{i1}, \dots, X_{i,k-1} = x_{i,k-1} | T_i = t_i; \alpha_1, \dots, \alpha_{k-1}) \\ = \exp\left[-\sum_{j=1}^{k-1} \alpha_j x_{ij} - \log \gamma(t_i; \alpha_1, \dots, \alpha_{k-1})\right], \end{aligned} \quad (16)$$

with

$$\gamma(t_i; \alpha_1, \dots, \alpha_{k-1}) = \sum_{(x_i)} \exp\left(-\sum_{j=1}^k \alpha_j x_{ij}\right), \quad (17)$$

the summation taken over all $(x_{i1}, \dots, x_{i,k-1})$ with $\sum_{j=1}^{k-1} x_{ij} = t_i$ and over all $(x_{i1}, \dots, x_{i,k-1})$ with $\sum_{j=1}^{k-1} x_{ij} = t_i - 1$. It follows from (16) that, given $(T_1, \dots, T_n) = (t_1, \dots, t_n)$, the random vector (V_1, \dots, V_{k-1}) is a complete sufficient statistic. Its conditional distribution belongs to an exponential family with probability mass function proportional to

$$\exp\left[-\sum_{j=1}^{k-1} \alpha_j v_j - \sum_{i=1}^n \log \gamma(t_i; \alpha_1, \dots, \alpha_{k-1})\right]. \quad (18)$$

The covariance matrix of (V_1, \dots, V_{k-1}) is the Fisher information matrix $I(\alpha; t_1, \dots, t_n)$, which has

$$\sum_{i=1}^n (\partial^2 / \partial \alpha_j \partial \alpha_h) \log \gamma(t_i; \alpha_1, \dots, \alpha_{k-1})$$

as its (j, h) element $(j, h = 1, \dots, k-1)$. If m_{in} describes the number of persons in the sample with number-right score t , and $I(\alpha; t)$ is the matrix which has

$$(\partial^2 / \partial \alpha_j \partial \alpha_h) \log \gamma(t; \alpha_1, \dots, \alpha_{k-1})$$

as its (j, h) element $(j, h = 1, \dots, k-1)$, then

$$I(\alpha; t_1, \dots, t_n) = \sum_{t=1}^{k-1} m_{in} I(\alpha; t) \tag{19}$$

Assuming that the item difficulties are two by two unequal, the matrix $I(\alpha, t)$ is positive definite.

The conditional maximum likelihood estimator for $(\alpha_1, \dots, \alpha_{k-1})$ is obtained by solving the system of $k-1$ equations

$$V_j = - \sum_{i=1}^n (\partial / \partial \alpha_j) \log \gamma(t_i; \alpha_1, \dots, \alpha_{k-1}) |_{\hat{\alpha}_1, \dots, \hat{\alpha}_{k-1}} \tag{20}$$

$(j = 1, \dots, k-1)$ with $\hat{\alpha}_1, \dots, \hat{\alpha}_{k-1}$ as unknowns. Andersen (1973) adapts general maximum likelihood theory to prove the asymptotic normality of the conditional maximum likelihood estimator. He postulates $\sum_{j=1}^k \alpha_j = 0$. For the restriction $\alpha_k = 0$ his result is reformulated as follows:

Theorem 1. If the set $\{\theta_1, \theta_2, \dots\}$ is bounded, the (conditional) distribution of the conditional maximum likelihood estimator $(\hat{\alpha}_1, \dots, \hat{\alpha}_{k-1})'$, given $T_1 = t_1, T_2 = t_2, \dots$, is asymptotically normal with expectation $(\alpha_1, \dots, \alpha_{k-1})'$ and covariance matrix

$$I^{-1}(\alpha; t_1, \dots, t_n),$$

except for a set of sequences (t_1, t_2, \dots) which has probability zero.

To obtain the asymptotic distributions of $\log I(\theta; \hat{\alpha})$ and $\log RE(\theta; \hat{\alpha}, \hat{\beta})$ the asymptotic distribution of $\hat{\alpha}$ (or $\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}$) should be transformed. However, these transformations cannot be justified by standard theory on functions of asymptotically normal vectors (see, e.g., Rao, 1965, p.321). The reason is that the asymptotic covariance matrix $I^{-1}(\alpha; t_1, \dots, t_n)$ (see Theorem 1) is not of the usual form $n^{-1} \Sigma$. By postulating that the persons are randomly selected from a population, the matrix $I^{-1}(\alpha; t_1, \dots, t_n)$ can, asymptotically, be brought into the standard form $n^{-1} \Sigma$.

However, the appendix contains a slight generalization of the standard theory (see

Theorem A1), which makes it possible to prove the (conditional) asymptotic normality of $\log I(\theta; \hat{\alpha})$ and $\log RE(\theta; \hat{\alpha}, \hat{\beta})$ without any additional assumption. As the proof of this asymptotic normality will also utilize Theorem 1, the following results will not be valid for a set of sequences (t_1, t_2, \dots) of probability zero.

Result 1. Conditionally given $T_1=t_1, T_2=t_2, \dots$ $\log I(\theta; \hat{\alpha})$ is asymptotically normal, with expectation $\log I(\theta; \alpha)$ and variance

$$\tau^2(\theta; \hat{\alpha}) = I^{-2}(\theta; \hat{\alpha}) \begin{bmatrix} \partial / \partial \alpha_1 i(\theta; \hat{\alpha}_1) \\ \vdots \\ \partial / \partial \alpha_{k-1} i(\theta; \hat{\alpha}_{k-1}) \end{bmatrix}'$$

$$I^{-1}(\hat{\alpha}; t_1, \dots, t_n) \begin{bmatrix} \partial / \partial \alpha_1 i(\theta; \hat{\alpha}_1) \\ \vdots \\ \partial / \partial \alpha_{k-1} i(\theta; \hat{\alpha}_{k-1}) \end{bmatrix}, \tag{21}$$

i.e. the statistic

$$\tau^{-1}(\theta; \hat{\alpha}) [\log I(\theta; \hat{\alpha}) - \log I(\theta; \alpha)]$$

tends in distribution to a standard normal variable as the sample size n increases.

Result 2. Suppose the vectors $\alpha = (\alpha_1, \dots, \alpha_k)'$ and $\beta = (\beta_1, \dots, \beta_h)'$, with $\beta_h = 0$, are estimated by the conditional maximum likelihood method on the basis of the responses of one sample of persons. Let t_i denote the number-right score of person i on the total test (hence $0 \leq t_i \leq k+h$) and let $I(\alpha, \beta; t_1, \dots, t_n)$ be defined similarly to $I(\alpha; t_1, \dots, t_n)$ (see above). Then, conditionally given $T_1=t_1, T_2=t_2, \dots$, $\log RE(\theta; \hat{\alpha}, \hat{\beta})$ is asymptotically normal with expectation $\log RE(\theta; \alpha, \beta)$ and variance

$$\nu^2(\theta; \hat{\alpha}, \hat{\beta}) = [RE(\theta; \hat{\alpha}, \hat{\beta})]^{-2} \begin{bmatrix} \partial / \partial \alpha_1 RE(\theta; \hat{\alpha}, \hat{\beta}) \\ \vdots \\ \partial / \partial \beta_{h-1} RE(\theta; \hat{\alpha}, \hat{\beta}) \end{bmatrix}$$

$$I^{-1}(\hat{\alpha}, \hat{\beta}; t_1, \dots, t_n) \begin{bmatrix} \partial / \partial \alpha_1 RE(\theta; \hat{\alpha}, \hat{\beta}) \\ \vdots \\ \partial / \partial \beta_{h-1} RE(\theta; \hat{\alpha}, \hat{\beta}) \end{bmatrix} \tag{22}$$

Result 3. Let the last item of test α also be the last item of test β . Suppose the conditional maximum likelihood estimators $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_{k-1}, 0)'$ and $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_{h-1}, 0)'$ are based on the responses of two non-overlapping samples of size m and n respectively. Then, conditionally given $T_1=t_1, T_2=t_2, \dots$, $\log RE(\theta; \hat{\alpha}, \hat{\beta})$ is asymptotically, i.e., if both $m \rightarrow \infty$ and $n \rightarrow \infty$, normal

with expectation $\log RE(\theta; \alpha, \beta)$ and variance

$$\epsilon^2(\theta; \hat{\alpha}, \hat{\beta}) = \tau^2(\theta; \hat{\alpha}) + \tau^2(\theta; \hat{\beta}), \tag{23}$$

where $\tau^2(\theta; \hat{\beta})$ is defined similarly to $\tau^2(\theta; \hat{\alpha})$ (see (21)).

A proof of Result 1 will now be presented; the other results are proved analogously.

Proof of Result 1. Define m and M by

$$m = \inf\{\lambda \mid \lambda \text{ is an eigenvalue of at least one of the matrices } I(\alpha; 1), \dots, I(\alpha; k-1)\}$$

and

$$M = \sup\{\lambda \mid \lambda \text{ is an eigenvalue of at least one of the matrices } I(\alpha; 1), \dots, I(\alpha; k-1)\}.$$

Together with (19) the definitions of m and M imply

$$\left(\sum_{t=1}^{k-1} m_{tn}\right) m I_{k-1} \leq I(\alpha; t_1, \dots, t_n) \leq \left(\sum_{t=1}^{k-1} m_{tn}\right) M I_{k-1},$$

and thus, more importantly,

$$\left(\sum_{t=1}^{k-1} m_{tn}\right)^{-1} M^{-1} I_{k-1} \leq I^{-1}(\alpha; t_1, \dots, t_n) \leq \left(\sum_{t=1}^{k-1} m_{tn}\right)^{-1} m^{-1} I_{k-1} \tag{24}$$

Thus the condition (35) of Theorem A1 is satisfied with $A = B = I_k$ and

$$a_n = \left(\sum_{t=1}^{k-1} m_{tn}\right)^{-1} M^{-1}, \quad b_n = \left(\sum_{t=1}^{k-1} m_{tn}\right)^{-1} m^{-1},$$

and $\log I(\theta; \hat{\alpha})$ is asymptotically normal with mean $\log I(\theta; \alpha)$ and variance

$$\begin{aligned} & \begin{bmatrix} (\partial / \partial \alpha_1) \log I(\theta; \alpha) \\ \vdots \\ (\partial / \partial \alpha_{k-1}) \log I(\theta; \alpha) \end{bmatrix}' I^{-1}(\alpha; t_1, \dots, t_n) \begin{bmatrix} (\partial / \partial \alpha_1) \log I(\theta; \alpha) \\ \vdots \\ (\partial / \partial \alpha_{k-1}) \log I(\theta; \alpha) \end{bmatrix} = \\ & = I^{-2}(\theta; \alpha) \begin{bmatrix} (\partial / \partial \alpha_1) i(\theta; \alpha_1) \\ \vdots \\ (\partial / \partial \alpha_{k-1}) i(\theta; \alpha_{k-1}) \end{bmatrix}' I^{-1}(\alpha; t_1, \dots, t_n) \begin{bmatrix} (\partial / \partial \alpha_1) i(\theta; \alpha_1) \\ \vdots \\ (\partial / \partial \alpha_{k-1}) i(\theta; \alpha_{k-1}) \end{bmatrix} \end{aligned}$$

Finally, the replacement of the α_i ($i = 1, \dots, k-1$) by their estimators $\hat{\alpha}_i$ in the above expression for the asymptotic variance of $\log I(\theta; \hat{\alpha})$ is justified by Slutsky's Theorem (see, e.g., Serfling, 1980). \square

Note that Result 3 should be adapted before it can be applied to situations where it is only for calibration purposes that the standard item, with difficulty zero, is included in (one of the) tests. The nature of the required adaptation is obvious however.

Throughout this section the restriction $\alpha_k = 0$ was imposed in order to obtain identifiable parameters. In the previous section some alternative restrictions were mentioned. The results obtained above are easily adapted to the case where the restriction $\sum_{j=1}^k \alpha_j = 0$ is imposed. For example, the estimator for $I(\theta_0; \alpha)$ has, under this new restriction, the same distribution as $I(\theta_0 + k^{-1} \sum_{j=1}^k \hat{\alpha}_j; \hat{\alpha})$ had under the old parametrization $\alpha_k = 0$.

However, if a restriction is put on the ability parameters, conditioning with respect to the number-right scores T_1, \dots, T_n does not work. Conditioning should take place with respect to $T_1 - T_n, T_2 - T_n, \dots, T_{n-1} - T_n$ if the restriction $\sum_{i=1}^n \theta_i = 0$ is imposed, and with respect to T_1, \dots, T_{n-1} if the parametrization $\theta_n = 0$ is preferred. This is easily seen by substituting $\theta_n = \sum_{i=1}^{n-1} \theta_i$ or $\theta_n = 0$ in (12).

If it is postulated that $\theta_1, \dots, \theta_n$ is a random sample from some distribution with a few unknown parameters, there is no problem at all: in that case standard (i.e. "unconditional") maximum likelihood theory is applicable.

4. Asymptotic confidence intervals

Based on the asymptotic distributions of the preceding section, interval estimators for $I(\theta; \alpha)$ and $RE(\theta; \alpha, \beta)$, with confidence coefficient approximately equal to 95, will now be presented. It will again be postulated that $\alpha_k = 0$.

Result 1 implies that an approximate 95% confidence interval for $\log I(\theta; \alpha)$ has endpoints $\log I(\theta; \hat{\alpha}) \pm 1.96 \tau(\theta; \hat{\alpha})$. But the events

$$\{\log I(\theta; \hat{\alpha}) - 1.96 \tau(\theta; \hat{\alpha}) < \log I(\theta; \alpha) < \log I(\theta; \hat{\alpha}) + 1.96 \tau(\theta; \hat{\alpha})\}$$

and

$$\{I(\theta; \hat{\alpha}) \exp[-1.96 \tau(\theta; \hat{\alpha})] < I(\theta; \alpha) < I(\theta; \hat{\alpha}) \exp[+1.96 \tau(\theta; \hat{\alpha})]\}$$

coincide (for all $\theta \in \mathbb{R}$); it follows that the endpoints for an approximate 95% interval for $I(\theta; \alpha)$ are given by

$$I(\theta; \hat{\alpha}) \exp[\pm 1.96 \tau(\theta; \hat{\alpha})] \tag{25}$$

Note that it is not true that the probability that the whole graph of the true information $I(\cdot; \alpha)$ is enclosed between the graphs of the two random functions defined by (25) is asymptotically equal to 95%.

However, a conservative method to obtain random functions \underline{I} and \bar{I} , with

$$\lim_{n \rightarrow \infty} P \{ \underline{I}(\theta) \leq I(\theta; \alpha) \leq \bar{I}(\theta) \text{ for all } \theta \in \mathbb{R} \} \geq 0.95, \tag{26}$$

is as follows:

- first construct an approximate 95% confidence region R for α using Theorem 1

- next define \underline{I} and \bar{I} pointwise by

$$\underline{I}(\theta) = \inf_{\alpha \in R} I(\theta; \alpha) \tag{27}$$

and

$$\bar{I}(\theta) = \sup_{\alpha \in R} I(\theta; \alpha). \tag{28}$$

The proof of (26) is obvious: the event $\{\underline{I}(\theta) \leq I(\theta; \alpha) \leq \bar{I}(\theta) \text{ for all } \theta \in \mathbb{R}\}$ is, by the definitions (27) and (28), implied by the event that the 95% confidence region R covers the true parameter vector α . For $k=1$, i.e. if the information curve of one item is studied, the method described above leads to confidence bounds with coefficient converging exactly to .95. This is easily seen by studying graphs of $I(\cdot; \alpha)$, for different values α in the confidence interval R .

The formula to be used to compute endpoints of an interval for $RE(\theta; \hat{\alpha}, \hat{\beta})$ (θ fixed) depends on the situation. In the situations described in Result 2 and Result 3, one should use

$$RE(\theta; \hat{\alpha}, \hat{\beta}) \exp[\pm 1.96 \nu(\theta; \hat{\alpha}, \hat{\beta})] \tag{29}$$

and

$$RE(\theta; \hat{\alpha}, \hat{\beta}) \exp[\pm 1.96 \epsilon(\theta; \hat{\alpha}, \hat{\beta})], \tag{30}$$

respectively.

An Example

To measure ability in manipulating fractions, a test consisting of 10 dichotomously scored items was administered to a sample of 208 Dutch secondary school pupils (grade level 8). With Andersen's likelihood-ratio test (Andersen (1973)) no deviations from the Rasch model could be detected (significance level $p=0.69$). The estimated item difficulties are given in Table 1. Note that the difficulty of the last item, which accidentally is the most difficult one, is set equal to zero.

Table 1. Estimated Difficulties of 10 Mathematics Test Items ($n = 208$)

item	1	2	3	4	5	6	7	8	9	10
difficulty	-3.17	-3.14	-0.73	-0.71	-1.64	-1.26	-1.73	-0.73	-0.99	0

The graph of the estimated information function is given in Figure 1.

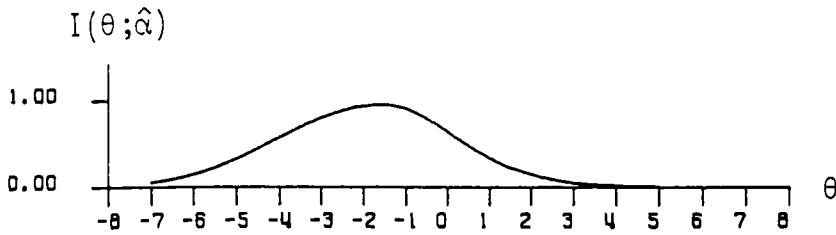


Figure 1. Graph of the estimated information function of the 10 item test from Table 1

To get a first impression of the variability of relative efficiency functions, two subtests were defined: subtest α consists of the items 1 to 5, subtest β of the remaining 5 items. Next the group of 208 subjects was randomly divided into 4 groups of equal size. In Figure 2 an estimated graph of $RE(\theta; \alpha, \beta)$ is given for each of these four subgroups, arbitrarily numbered from I to IV.

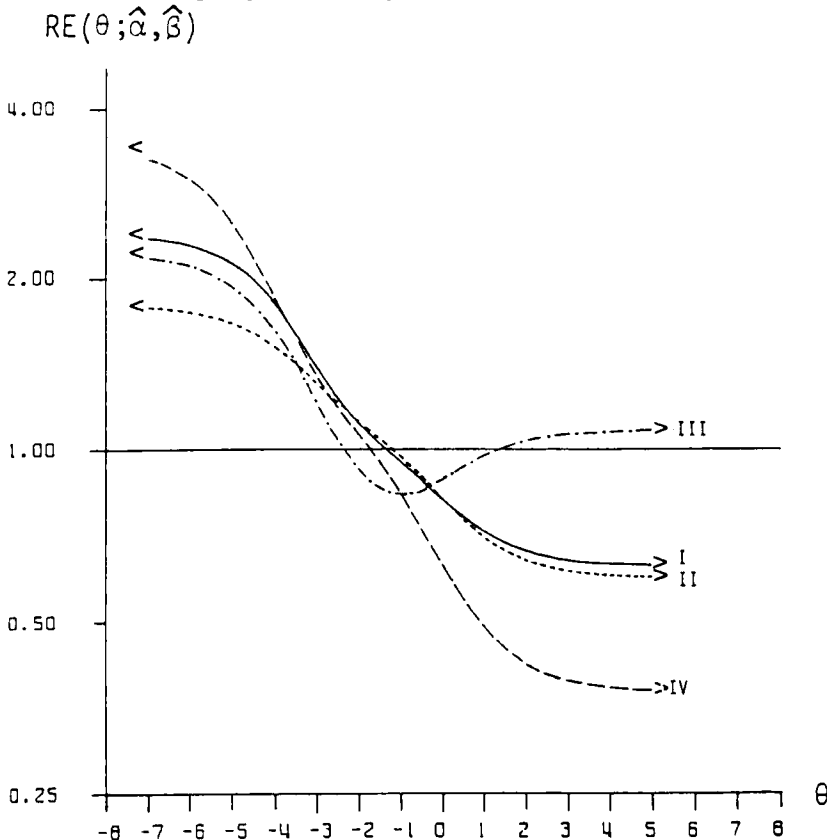


Figure 2. Graphs of estimated relative efficiency function of two subtests based on four random groups of 52 subjects, with $<$ and $>$ indicating limiting values.

In figure 3 the estimated graph of $RE(\theta; \alpha, \beta)$, based on group I is given again, but now it is supplemented by "95% confidence bands". These bands are obtained by concatenating the confidence intervals for $RE(\theta; \alpha, \beta)$ for different values θ .

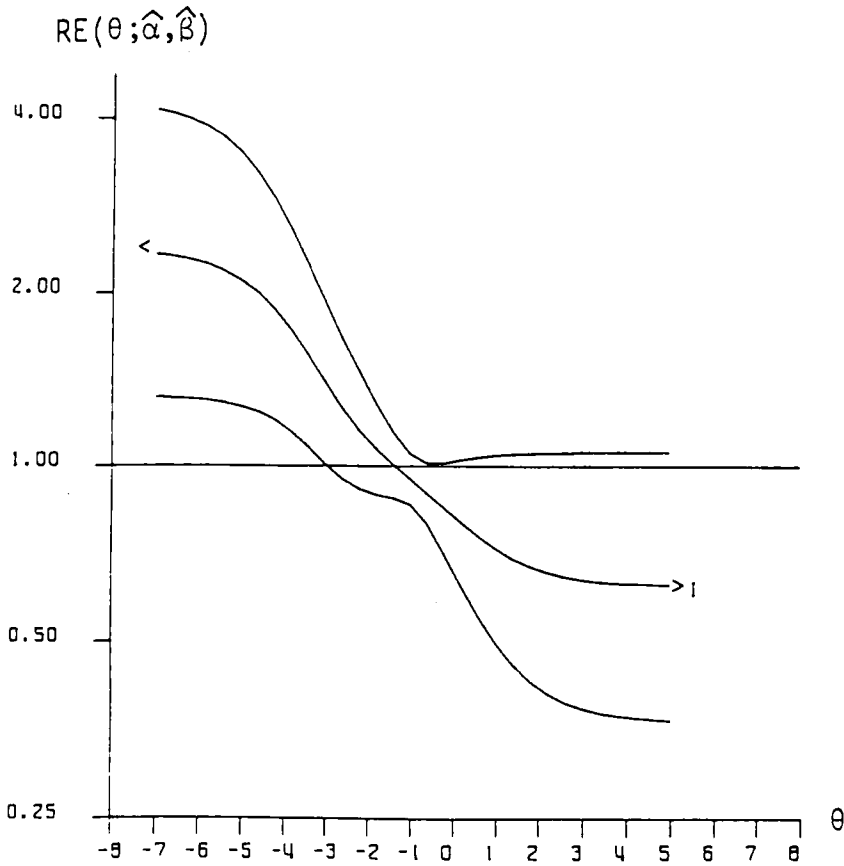


Figure 3. Graph of the estimated relative efficiency function of two subtests based on a random group of 52 subjects, supplemented by "95% confidence bands".

Finally, figure 4 gives bands for $RE(\cdot; \alpha, \beta)$, based on the total sample of 208 subjects.

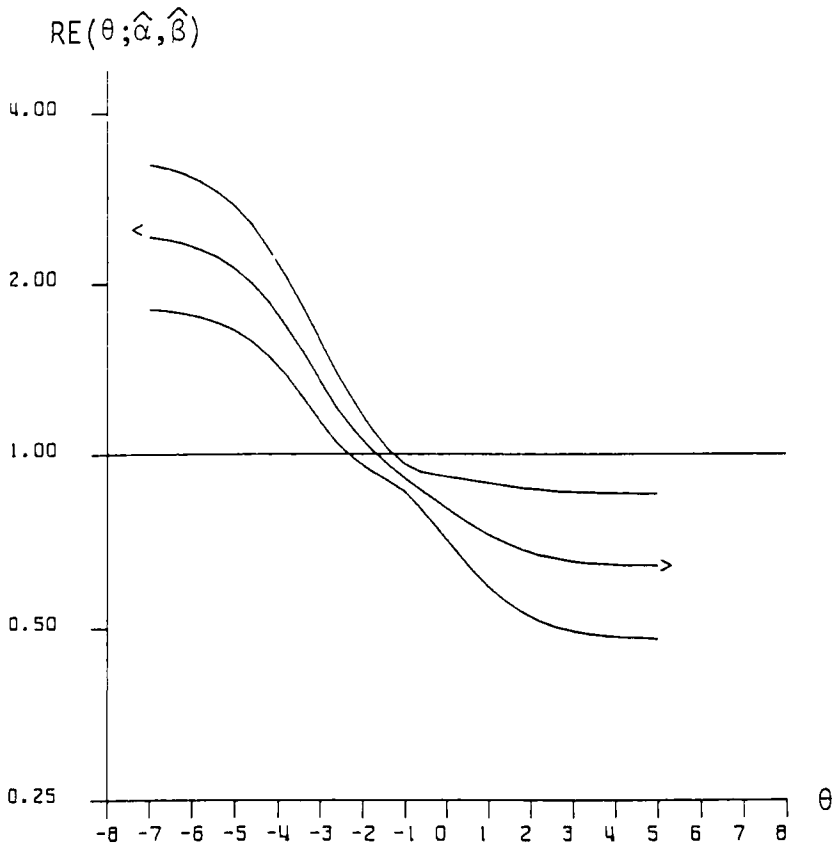


Figure 4. Graph of estimated relative efficiency function of two subtests based on the total group of 208 subjects, supplemented by "95% confidence bands".

From the Figures 2, 3 and 4 it can be seen that the variability of $RE(\theta; \hat{\alpha}, \hat{\beta})$ is very large for extremely low and for extremely high values θ . One should not be impressed too much by this variability however: Figure 1 tells that the information of the total test is small for extreme θ -values. However, a researcher who only has the data of group III (see figure 2) might be tempted to draw the qualitative conclusion that the two subtests are about equally informative for very clever pupils, whereas another researcher, who studies the data of group IV, would probably be convinced that the second test is much more informative for these pupils. This illustrates the need to take into account sampling fluctuations in interpreting estimated relative efficiency functions; confidence bands as found in Figures 3 and 4 thus seem to be indispensable instruments for a careful interpretation.

5. Discussion

In this paper it was shown how asymptotic confidence intervals for test information and relative efficiency can be computed for tests consisting of dichotomously scored Rasch homogeneous items. Two natural questions are raised:

- (1) Can the method used be generalized to more general test models?
- (2) How large should the estimation sample(s) be before the asymptotic results are satisfactory?

The answer to question 1 is: yes, but only to test models for which the asymptotic normality of estimators for the item parameters has been shown. Thus, in principle the method of this paper could be generalized to the case of polychotomously scored Rasch tests (Andersen, 1973). The computations would become very complicated however. As far as is known to the author, asymptotic normality has not been shown for estimators of the item parameters in the two and three parameter logistic models (Lord, 1980). Therefore, a theoretically sound generalization of the method of this paper cannot be offered for these models, unless a prior distribution $g(\theta)$ for the ability parameters is specified. Even if this prior $g(\theta)$ is only known to belong to some parametric family, e.g. the family of logistic distributions, asymptotic distributions for functions of estimators for item parameters can easily be derived, because it then follows from standard statistical theory that the maximum-likelihood estimators for item parameters are asymptotically normally distributed.

An answer to question 2 can only be offered by a simulation study. It is obvious, however, that the results of this paper are applicable if the size(s) of the estimation samples are very large. But then the computation of confidence bands is a waste of time: it is known beforehand that the bands will be very narrow. A possible simulation study should therefore be directed to cases of estimation samples of moderate size.

Appendix: A theorem on asymptotic normality

Let (X_n) be a sequence of k -dimensional random vectors, that are asymptotically normal with mean vector μ and covariance matrix Σ_n (notation: X_n is $AN_k(\mu, \Sigma_n)$). In Serfling (1980, p.122) it is proved that, for smooth functions $g: \mathbb{R}^k \rightarrow \mathbb{R}^h$, the h -dimensional random vector $g(X_n)$ is then also asymptotically normal, provided that the condition

$$\Sigma_n = b_n \Sigma \tag{31}$$

holds for a sequence of positive numbers b_n , with $b_n \rightarrow 0$.

Theorem 1 implies that (31) does not hold for the approximate (conditional) covariance matrix of the c.m.l. estimator for the vector α (or: $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$) of item-difficulties. In this appendix the asymptotic normality of $g(X_n)$ will be proved

under a condition slightly more general than (31) (see Theorem A1 below). In section 3 this result is used to derive the asymptotic distribution of $\log I(\theta; \hat{\alpha})$ and $\log RE(\theta; \hat{\alpha}, \hat{\beta})$.

The notation $A \leq B$ will be used to indicate that the matrix $B - A$ is non-negative definite.

Theorem A1. Suppose

$$X_n \text{ is } ANC(\mu, \Sigma_n) \tag{32}$$

Assume further that there exist non-degenerate covariance matrices A and B and sequences (a_n) and (b_n) of positive numbers tending to zero, with

$$a_n \leq b_n \text{ for all sufficiently large } n \tag{33}$$

and

$$\limsup a_n^{-1} b_n < +\infty \tag{34}$$

such that

$$a_n A \leq \Sigma_n \leq b_n B \text{ for all sufficiently large } n. \tag{35}$$

Let $g: \mathbb{R}^k \rightarrow \mathbb{R}^h$ be a totally differentiable function for which the matrix $D \in \mathbb{R}^{h \times k}$, defined by

$$D_{ij} = \partial g_i / \partial x_j |_{x=\mu}, \tag{36}$$

is unequal zero. Then

$$g(X_n) \text{ is } AN_h(g(\mu), D \Sigma_n D'). \tag{37}$$

For the proof of Theorem A1 the following lemma's are useful.

Lemma A1. Suppose X and Y are k -dimensional normal random vectors with mean vectors zero and covariance matrices A and B respectively, where

$$A \leq B. \tag{38}$$

Then the random variable $\|Y\| = (Y'Y)^{\frac{1}{2}}$ is stochastically larger than $\|X\| = (X'X)^{\frac{1}{2}}$.

Proof. Let U be a k -dimensional standard normal vector. Then $\|X\| \sim (U'AU)^{\frac{1}{2}} \leq (U'BU)^{\frac{1}{2}} \sim \|Y\|$, and the result follows.

Lemma A2. Suppose X_n is $AN_k(\mu, \Sigma_n)$ and n satisfies the condition (35). Then for every $\lambda \in \mathbb{R}^k \setminus \{0\}$ the sequence of random variables

$$(\lambda' \Sigma_n \lambda)^{-\frac{1}{2}} \|X_n - \mu\| \tag{39}$$

is bounded in probability, i.e. for every $\epsilon > 0$ there exists $M_{\lambda, \epsilon}$ and $N_{\lambda, \epsilon}$ such that $P((\lambda' \Sigma_n \lambda)^{-\frac{1}{2}} \|X_n - \mu\| \leq M_{\lambda, \epsilon}) \geq 1 - \epsilon$ for all $n > N_{\lambda, \epsilon}$.

Proof. The first inequality of (35) implies that, for sufficiently large n ,

$$\begin{aligned} (\lambda' \Sigma_n \lambda)^{-\frac{1}{2}} \|X_n - \mu\| &\leq (\lambda' A \lambda)^{-\frac{1}{2}} a_n^{-\frac{1}{2}} \|X_n - \mu\| \leq \\ &(\lambda' A \lambda)^{-\frac{1}{2}} 2(\limsup a_n^{-1} b_n)^{\frac{1}{2}} b_n^{-\frac{1}{2}} \|X_n - \mu\|. \end{aligned} \tag{40}$$

As $(\lambda' A \lambda)^{-\frac{1}{2}} 2(\limsup a_n^{-1} b_n)^{\frac{1}{2}}$ is a constant, it suffices to show that $b_n^{-\frac{1}{2}} \|X_n - \mu\|$ is bounded in probability. However, as X_n is $AN_k(\mu, \Sigma_n)$,

$$b_n^{-\frac{1}{2}} (X_n - \mu) \text{ is } AN_k(0, b_n^{-1} \Sigma_n) \tag{41}$$

Now let Z be a random vector which is normally distributed with mean 0 and covariance matrix B . Then (35), (41) and Lemma 1 imply that $b_n^{-\frac{1}{2}} \|X_n - \mu\|$ is asymptotically stochastically smaller than $\|Z\|$, i.e.

$$\lim_{n \rightarrow \infty} P(b_n^{-\frac{1}{2}} \|X_n - \mu\| \geq k) \leq P(\|Z\| \geq k) \text{ for all } k > 0. \tag{42}$$

As the right-hand side of (42) can be made arbitrarily small by choosing k large enough, $b_n^{-\frac{1}{2}} (X_n - \mu)$ is bounded in probability and the lemma is proved.

Proof of Theorem A1. It suffices to show that for any $\lambda \in \mathbb{R}^h$ satisfying

$$D' \lambda \neq 0, \tag{43}$$

the random variables

$$\lambda' [g(X_n) - g(\mu)] (\lambda' D \Sigma_n D' \lambda)^{-\frac{1}{2}} \tag{44}$$

tend in distribution to a standard normal variable (see, e.g., Serfling, 1980). Thus choose a vector λ satisfying (43). Taylor's Theorem implies the existence of a random vector E_n with

$$g(X_n) - g(\mu) = D(X_n - \mu) + E_n \|X_n - \mu\| \quad (45)$$

and $\|E_n\| \rightarrow 0$ if $X_n \rightarrow \mu$ and thus (see Rao, 1965, p.320)

$$\|E_n\| \xrightarrow{p} 0. \quad (46)$$

From (45) it is seen that the random variables (44) can be written

$$\frac{\lambda' [g(X_n) - g(\mu)]}{(\lambda' D \Sigma_n D' \lambda)^{\frac{1}{2}}} = \frac{\lambda' D (X_n - \mu)}{(\lambda' D \Sigma_n D' \lambda)^{\frac{1}{2}}} + \frac{\lambda' E_n \|X_n - \mu\|}{(\lambda' D \Sigma_n D' \lambda)^{\frac{1}{2}}}. \quad (47)$$

On account of (32) the first term on the right-hand side of (47) is asymptotically standard normal. Furthermore (46) and Lemma A2 imply that the second term on the right-hand side of (47) converges to zero in probability. Application of Slutsky's Theorem (Serfling, 1980, p.19) completes the proof.

Acknowledgement

Thanks go to Willem Schaafsma for useful discussions, to Rien Steen for his programming work and to Anita Burchartz for typing the manuscript.

References

- ANDERSEN, E.B. (1973) *Conditional inference and models for measuring*, Mentalhygiejnisk Forlag, Copenhagen.
- BIRNBAUM, A. (1968) Some latent trait models and their use in inferring an examinee's ability, in: F.M. Lord and M.R. Novick, *Statistical Theories of Mental Test Scores*, Addison and Wesley, Reading.
- LEHMANN, E.L. (1959) *Testing statistical hypotheses*, Wiley and Sons, New York.
- LORD, F.M. (1980) *Applications of item response theory to practical problems*, Lawrence Erlbaum Associates, Hillsdale, N.J..
- RAO (1965) *Linear statistical inference and its applications* (1st ed.), Wiley and Sons, New York.
- RASCH, G. (1960) *Probabilistic models for some intelligence and attainment tests*, Danmarks Paedagogiske Institut, Copenhagen.
- SERFLING, R.J. (1980) *Approximation Theorems of Mathematical Statistics*, Wiley and Sons, New York.

Received June 83