TITLE

Expected A Posteriori Estimation of Multiple Latent Traits

ABSTRACT

The purpose of this study was to present formulas for estimating multiple latent traits and a latent trait composite using the expected a posteriori (EAP) approach. Formulas for the error associated with the estimates of the latent traits and the composite were provided. Implications for computer-adaptive testing are discussed.

PRESS PARAGRAPH

The expected a posteriori (EAP) ability estimation approach was generalized to multidimensional item response theory. It was shown how this non-iterative, Bayesian method can be used to estimate multiple latent traits, $\theta = [\theta_1, \theta_2, \ldots, \theta_k]$, or a weighted composite of latent traits, θ_{comp}. The formulas for estimating the error associated with the EAP estimates of θ and θ_{comp} were provided. Implications for computer-adaptive testing are discussed.

With advances in computer technology, Item Response Theory (IRT) has moved from a theoretical discussion to a practical framework for designing cognitive ability tests. It has been shown that greater measurement efficiency can be obtained with fewer items and in less test administration time when using IRT as compared to Classical Test Theory (CTT), especially when computer-adaptive testing is involved (Hambleton, Swaminathan, & Rogers, 1991; Segall, 1996; Weiss, 1982). The improvement in efficiency can be attributed to the development of several aspects of IRT, i.e., item parameter estimation (Bock & Aitken, 1981), item selection (Birnbaum, 1968; Bryant, in press; van der Linden, 1999), and latent trait (θ) estimation (Bock & Mislevy, 1982). For the purpose of this study, attention is focused on trait estimation.

Several ability estimation methods are available, which include maximum likelihood (ML), maximum a posteriori (MAP), and expected a posteriori (EAP) (Bock & Aiken, 1981; Embretson & Reise, 2000; Hambleton et al., 1991; Kim & Nicewander, 1993). Of the approaches mentioned above, MAP and EAP are Bayesian methods that incorporate information about the prior ability distribution in order to better approximate the posterior distribution of the latent trait (Bock & Mislevy, 1982). One advantage of using prior information is that an estimate of θ exists when examinees have either all correct or all incorrect responses; this advantage is invaluable in the early stages of a computer-adaptive test (Segall, 1996; Weiss, 1982). Despite the benefits of Bayesian methods over ML, there are distinctions between MAP and EAP that should be noted.

First, the EAP estimate is relatively easy to compute, i.e., it does not require the first and second partial derivatives of the likelihood function that are employed to compute a solution as in MAP and ML approaches. Second, "the EAP estimator has minimum mean square error over the population of ability and, in terms of average accuracy, cannot be improved upon" (Bock & Mislevy, 1982, p. 439). Third, the method is non-iterative, which permits fast calculations of the provisional ability estimate in computer-adaptive testing. These advantages have been demonstrated in empirical research.

In a study investigating ability estimation procedures, Kim and Nicewander (1993) found evidence in support of the minimum standard error of EAP estimates relative to other methods, which included MAP, ML, and weighted likelihood estimation. Using the generalized partial credit model, Wang and Wang (2001) also demonstrated the superiority of EAP over other trait estimation methods in terms of minimum standard error. However, in light of the benefits of EAP in the unidimensional context, little research has thoroughly developed EAP estimation procedures and estimates of the covariance matrix of the posterior distribution of the latent traits for multidimensional IRT (MIRT) models.

In a multidimensional context, Carlson (1987) created a joint ML method of estimating MIRT item parameters and multiple latent traits, $\theta = [\theta_1, \theta_2, \ldots, \theta_k]$. Segall (1996) developed a MAP ability estimation approach for calculating θ while also deriving an item selection algorithm based on the covariance matrix of the posterior distribution of latent traits. Notwithstanding progress made in MIRT (van der Linden, 1999), the potential advantages of EAP estimation of θ have not been thoroughly investigated, perhaps due to the fact that formulas for θ based on EAP estimation and the covariance matrix of the posterior distribution of θ have not been fully developed.

Thus, the purpose of this study is to present EAP formulas for estimating θ and θ_{comp} along with the appropriate standard errors. First, there is a brief presentation of the item response model and the underlying assumptions. Second, the multidimensional extension of EAP ability estimation (θ_{EAP}) is provided, and the formula for computing the covariance matrix of the posterior distribution of latent traits is shown. Third, formulas for the latent trait composite (θ_{comp}) and the standard error of the composite are given. Finally, a numerical example of both θ_{EAP} and θ_{comp} are provided along with the associated standard errors.

Item Response Model

Let a set of *k* latent traits be represented by the vector $\theta = [\theta_1, \theta_2, \ldots, \theta_k]$, and assume that each of the *k* traits influences performance on one or more items. The response function for item *i* is given by the Multidimensional 3-Parameter Logistic model (M3PL; Reckase, 1997),

$$P_i(u_i = 1 \mid \theta) = c_i + (1 - c_i)\,\{1 + \exp[-D(\boldsymbol{a}_i{}'\theta + d_i)]\}^{-1}, \qquad (1)$$

where u_i is the dichotomous response to item *i* ($u_i = 1$, if the item is correct, and $u_i = 0$, otherwise), *D* is equal to a scaling constant 1.7, $\boldsymbol{a}_i$ is a vector of *k* discrimination parameters for item *i*, $[a_{1i}, a_{2i}, \ldots, a_{ki}]'$, and d_i is a scalar related to item difficulty. The probability of an incorrect response is $Q_i(\theta) = 1 - P_i(\theta)$. Under the assumption of local independence, the likelihood of a set of responses, $\boldsymbol{u} = [u_1, u_2, \ldots, u_n]$, given θ is

$$L(\boldsymbol{u} \mid \theta) = L(u_1, u_2, \ldots, u_n \mid \theta_1, \theta_2, \ldots, \theta_k) = \prod_{i=1}^{n} P_i(\theta)^{ui}\, Q_i(\theta)^{1-ui}. \qquad (2)$$

Under the assumption of monotonicity, increases in one or more of the k latent traits result in a greater chance of getting an item correct.

EAP Estimation of θ.

The estimate of the latent trait vector, θ, using the Bayesian estimation procedure is an extension of the method proposed by Bock and Aitken (1981). In the unidimensional context, a prior density, $f(\theta)$, is specified and is used to compute the posterior density, $f(\theta \mid \boldsymbol{u})$, which is proportional to the product of the prior density and the likelihood function, $L(\boldsymbol{u} \mid \theta)$. In the multidimensional case, the posterior density of θ is

$$f(\theta \mid \boldsymbol{u}) = L(\boldsymbol{u} \mid \theta)\; f(\theta) \,/\, f(\boldsymbol{u}), \tag{3}$$

where $f(\theta)$ is the prior density given θ, $L(\boldsymbol{u} \mid \theta)$ is provided in Equation 2, and $f(\mathbf{u})$ is the marginal probability of $\boldsymbol{u}$. The prior density can be empirically derived from the data or can be considered as multivariate normal with a mean of μ and covariance of Σ; i.e.,

$$f(\theta) = (2\pi)^{-k/2}\, |\Sigma|^{-1/2} \exp[-.5(\theta - \mu)'\, \Sigma^{-1}(\theta - \mu)]\,, \tag{4}$$

where $\pi \approx 3.1459$ and $|\Sigma|$ is the determinant of Σ.

The EAP estimate of the latent trait vector, θ_{EAP}, is given by

$$E(\theta|u_1, \ldots, u_n) = \theta_{EAP} = \frac{\int \ldots \int \theta\, L(\boldsymbol{u} \mid \theta)\, f(\theta)\, d\theta_1 \ldots d\theta_k}{\int \ldots \int L(\boldsymbol{u} \mid \theta)\, f(\theta)\, d\theta_1 \ldots d\theta_k}\,. \tag{5}$$

Consistent with the meaning of θ_{EAP} in the unidimensional case, θ_{EAP} is the mean (vector) of the posterior multivariate distribution. The covariance matrix of the posterior distribution is

$$\boldsymbol{C}(\theta \mid u_1, \ldots, u_n) = \boldsymbol{C} = \frac{\int \ldots \int (\theta - \theta_{EAP})(\theta - \theta_{EAP})'\, L(\boldsymbol{u} \mid \theta)\, f(\theta)\, d\theta_1 \ldots d\theta_k}{\int \ldots \int L(\boldsymbol{u} \mid \theta)\, f(\theta)\, d\theta_1 \ldots d\theta_k}\,. \tag{6}$$

It should be noted that θ_{EAP} in Equation 5 is a *k* x 1 vector. The covariance matrix of the posterior distribution, ***C***, in Equation 6 is a *k* x *k* matrix. The variance of the k^{th} latent trait is the k^{th} element in the main diagonal of ***C***, and the corresponding standard error for each of the latent traits is computed as the square root of each of the respective elements in the main diagonal of ***C***.

Under the assumption of multivariate normality, Equation 5 can be reasonably approximated using the Gauss-Hermite quadrature nodes (θ_q) and weights, $W(\theta_q)$, that correspond to probabilities at the quadrature nodes across the discrete multivariate distribution. The approximation of Equation 5 is

$$\theta_{EAP} \approx \frac{\sum_{x=1}^{q} \theta_q L(\boldsymbol{u} \mid \theta_q)\, W(\theta_q)}{\sum_{x=1}^{q} L(\boldsymbol{u} \mid \theta_q)\, W(\theta_q)}, \tag{7}$$

where θ_q is one of the finite quadrature nodes in the discrete *k*-dimensional multivariate distribution, $L(\boldsymbol{u} \mid \theta_q)$ is the likelihood of the response pattern given θ_q, and $W(\theta_q)$ is the weight, usually the prior density of the multivariate normal distribution, $f(\theta_q)$. An additional requirement is that the weights are normed so that the sum of $W(\theta_q)$ across the discrete multivariate distribution is one. The numerator in Equation 7 is the sum of the products of the quadratures, likelihoods, and weights at the quadratures across the discrete multivariate distribution, and the denominator is the sum of the products of the likelihoods and weights of the quadratures across the discrete multivariate distribution. The covariance matrix in Equation 6 is approximated by the following expression:

$$\boldsymbol{C} \approx \frac{\sum_{x=1}^{q} (\theta_q - \theta_{EAP})(\theta_q - \theta_{EAP})'\, L(\boldsymbol{u} \mid \theta_q)\, W(\theta_q)}{\sum_{x=1}^{q} L(\boldsymbol{u} \mid \theta_q)\, W(\theta_q)}, \tag{8}$$

where θ_{EAP} is given by Equation 7.

Estimation of θ_{comp}.

In educational and employment settings, decisions are often made on the basis of a composite of scores on a battery of subtests. The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) recommend that when composites are used in decision-making, information related to the standard errors of the composite and subtests should be documented (Standard 2.1). Thus, it is necessary that latent trait composites have the associated standard errors. Formulas for both the composite and the standard error are presented next.

Assume that *k* subtest scores correspond to *k* latent traits, and the subtests are weighted by a vector, $\beta = [\beta_1, \ldots, \beta_k]'$, to create a composite score. This composite corresponds to a composite of latent traits, θ_{comp}. Also, a requirement is made that $\beta_{k \times 1}$ is a unit vector, i.e., $\beta' \beta = 1$. The composite of the latent traits using the EAP estimate of θ is

$$\theta_{comp} = \beta' \theta_{EAP}. \tag{9}$$

The variance of θ_{comp} is

$$Var(\theta_{comp}) = \beta' \boldsymbol{C} \beta , \tag{10}$$

and the standard error of the EAP composite of latent traits is

$$SE(\theta_{comp}) = (\beta' \boldsymbol{C} \beta)^{1/2}. \tag{11}$$

Now that formulas for computing θ_{EAP} and θ_{comp} have been presented with the associated variance and standard errors of estimation, a numerical example is presented that demonstrates the use of the formulas.

Numerical Example of Estimating θ_{EAP} and θ_{comp}.

The item response model used for this numerical example is the M3PL model in Equation 1. The simulated test contains 20 items; each item measures two latent traits, $\theta = [\theta_1, \theta_2]'$. The prior multivariate distribution is assumed to be normal with $\mu = [0, 0]'$ and a $\Sigma = \mathbf{I}$. The identity matrix, **I,** is a 2 x 2 matrix with ones in the main diagonal and zeros in the off-diagonals. The lower and upper bounds of the discrete multivariate distribution are –4 and +4, respectively, on each trait with an equally spaced step size of .5 (i.e., the first quadrature node is [–4.0, –4.0], the second node is [–4.0, –3.5], the third is [–4.0, –3.0], … , etc). The weighting vector used to create the composite is $\beta = [.7071, .7071]'$, which is a measurement direction of approximately 45 degrees from the first latent trait and 45 degrees from the second latent trait. The parameters for the 20 items are listed in Table 1. The simulated examinee used in this example has a true latent trait vector of $\theta_{true} = [.90, .40]'$, and a true latent composite of $\theta_{true\ comp} = \beta'\ \theta_{true} = .919$. The first author wrote a program using the Python programming language to estimate θ_{EAP} and θ_{comp}.

The results of the simulation are provided in Table 2. The simulated responses of the examinee to the 20-item test are given in Column 4 of Table 2. The provisional estimates of θ_{EAP} and θ_{comp} were computed after each item with the associated standard errors; θ_{EAP} is in Columns 5 and 6 with the standard error for each latent trait in parentheses. θ_{comp} is provided in Column 7 with the standard error of the latent composite in parentheses. For each item, the multidimensional extension of the point of maximum item information (θ_{max}; Bryant, in press) is given in Columns 2 and 3.

For the total test, θ_{EAP} and θ_{comp} are estimated reasonably well given the response pattern and item parameters with θ_{EAP} = [.90, .40]' and θ_{comp} = .922; these estimates are nearly identical to the true values. Also, ability estimates and standard errors exist for the first nine correct responses; these estimates do not exist using the ML approach. Thus, EAP estimation may have utility in situations when all correct or all incorrect responses are present in the early stages of a multidimensional, computer-adaptive test.

In Table 2, observe that as more items are administered, both the standard errors associated with θ_{EAP} and θ_{comp} decrease. For example, θ_{comp} has standard errors of .48, .35, and .29 after the administration of 10, 14, and 17 items, respectively. It appears that the standard errors can be managed with the increase in the number of items, thus making it possible to employ the standard error as a stopping rule in an adaptive testing context. For θ_{comp}, if a .5 standard error criterion were used as a stopping rule, testing in this simulated example would have ended after 10 items were administered. If the commonly used .3 standard error criterion were used, then testing would terminate after 17 items. For θ_{EAP} , if a .45 standard error on each latent trait were used, then testing would terminate after 17 items; the estimate of the trait vector using the .45 stopping rule is θ_{EAP} = [.923 , .446], which is not far from the true ability vector, θ_{true} = [.90, .40]'.

Discussion

The purpose of this study was to present the formulas for estimating θ_{EAP} and θ_{comp} . Formulas for approximating the standard errors of the latent trait vector and the composite were given. A numerical example using the formulas was provided. The main results suggest that the formulas may have utility in estimating ability in a multidimensional adaptive testing context. So, what are some of the specific implications of this research?

Because formulas for EAP estimation of multiple traits are now provided, a comparison of the different estimation procedures in a multidimensional context may begin. Given the results of other studies from the unidimensional perspective, it is expected that some of the advantages of EAP in the unidimensional context will extend to the multidimensional setting. Does EAP estimation in a multidimensional setting produce scores with the minimum mean square error across the population of ability for which the prior distribution is specified? Future research could perhaps speak to this issue.

The results of this study also showed that θ_{EAP}, θ_{comp}, and the standard errors associated with each estimate exist when all correct responses are present. This is an invaluable advantage in a computer-adaptive setting where estimates of ability in the initial stages are used to select the next item. Once again, ML is unable to produce ability estimates when all correct or all incorrect responses are present, thus making it a poor choice as an ability estimator in a multidimensional setting. However, with the incorporation of prior information, EAP estimates of a latent trait vector may be improved dramatically in the selection of the next item in an adaptive testing environment.

Evidence in this study also suggests that the standard error of θ_{EAP} and θ_{comp} may be used as a stopping criterion in a computer-adaptive test. By specifying the amount of error that a decision-maker is willing to accept in scores, the testing session may be terminated once an acceptable amount of error associated with an ability estimate is reached. Future research could perhaps focus on the similarity and differences of using a minimum error criterion, a fixed length stopping criterion, or a fixed time limit criterion in a multidimensional, computer-adaptive test.

The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) state that the standard error associated with both composite and subtest scores should be reported. This study provides standard errors for both the latent traits and the composite of the latent traits and allows researchers and practitioners to meet the requirement with respect to EAP estimation. In particular, an advantage of being able to compute the standard error associated with the composite is that if the weighting vector used to create the latent composite changes, then the standard error of θ_{comp} will change. In a CTT context, when combining subtest scores to create a composite, the standard error does not change as a function of the weighting of the composite. IRT definitely provides a benefit in this respect.

In summary, the estimation of θ_{EAP} and θ_{comp} along with the standard errors may prove useful in advancing MIRT. First, the formulas are relatively easy to compute as compared to the multidimensional extensions of ML and MAP ability estimation procedures. Second, EAP estimates and standard errors exist for all correct and all incorrect response patterns. Third, the error of the ability estimates may have utility as a termination criterion in computer-based assessments. Finally, the formulas, used in conjunction with each other, provide researchers and practitioners the necessary tools to meet professional testing standards.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bock, R. D., & Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of the EM algorithm. *Psychometrika, 46,* 443-459.

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 6,* 431-444.

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* (pp. 453-479). Reading, MA: Addison-Wesley.

Bryant, D. U. (in press). A note on item information in any direction for the multidimensional three-parameter logistic model. *Psychometrika*.

Carlson, J. E. (1987). *Multidimensional item response theory estimation: A computer program* (Research Report ONR 87-2). Iowa City, IA: The American College Testing Program.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park CA: Sage.

Kim, J. K., & Nicewander, W. A. (1993). Ability estimation for conventional tests. *Psychometrika, 58,* 587-599.

Reckase, M. D. (1997). A linear logistic multidimensional model for dichotomous item response data. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 271-286). New York: Springer-Verlag.

Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika*, 61, 331-354.

van der Linden, W. J. (1999). Multidimensional adaptive testing with a minimum error-variance criterion. *Journal of Educational and Behavioral Statistics*, 24, 398-412.

Wang, S. & Wang, T. (2001). Precision of Warm's weighted likelihood estimates for a polytomous model in computer adaptive testing. *Applied Psychological Measurement, 25,* 317-331.

Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement, 6*, 473-492.

Table 1.

Simulated Item Parameters for the Multidimensional 3-Parameter Logistic Model

Item	a_{i1}	a_{i2}	c_i	$MDIFF_i$	Item	a_{i1}	a_{i2}	c_i	$MDIFF_i$
1	1.5	.5	.10	-1.0	11	1.5	.5	.10	.85
2	.5	1.5	.10	-.7	12	1.5	.5	.10	.9
3	1.5	.5	.10	-.5	13	1.5	.5	.10	.9
4	.5	1.5	.10	-.4	14	1.5	.5	.10	1.0
5	1.5	.5	.10	0	15	.5	1.5	.10	1.0
6	.5	1.5	.10	0	16	.5	1.5	.10	1.5
7	.5	1.5	.10	.4	17	1.5	.5	.10	1.5
8	.5	1.5	.10	.5	18	.5	1.5	.10	1.5
9	.5	1.5	.10	.5	19	.5	1.5	.10	2.0
10	.5	1.5	.10	.6	20	1.5	.5	.10	2.0

Note: a_{i1} is the discrimination parameter for θ_1, a_{i2} is the discrimination parameter for θ_2, $MDIFF_i$ is the multidimensional difficulty, c_i is the pseudo-guessing parameter, and d_i is $-MDIFF_i\,[(a_{i1})^2 + (a_{i2})^2]^{1/2}$.

Table 2.

Simulated Results With True θ = [.9, .4] and Prior Distribution (μ = [0, 0], Σ = **I**)

	θ_{max}			θ_{EAP}		
Item	θ_{1max}	θ_{2max}	Response	θ_1 (*SE*)	θ_2(*SE*)	θ_{comp}(*SE*)
1	-.893	-.298	1	.244 (.880)	.081 (.987)	.230 (.890)
2	-.202	-.608	1	.291 (.886)	.349 (.864)	.453 (.808)
3	-.419	-.140	1	.488 (.811)	.381 (.869)	.615 (.757)
4	-.108	-.234	1	.505 (.821)	.561 (.798)	.754 (.715)
5	.056	.019	1	.719 (.760)	.580 (.810)	.918 (.682)
6	.019	.056	1	.727 (.770)	.757 (.756)	1.05 (.653)
7	.145	.435	1	.739 (.783)	.959 (.711)	1.20 (.631)
8	.177	.530	1	.751 (.793)	1.11 (.676)	1.32 (.614)
9	.177	.530	1	.761 (.800)	1.22 (.651)	1.40 (.601)
10	.208	.625	0	.711 (.764)	.712 (.495)	1.01 (.482)
11	.862	.287	1	1.05 (.712)	.646 (.499)	1.20 (.459)
12	.909	.303	1	1.29 (.655)	.601 (.502)	1.34 (.433)
13	.909	.303	1	1.45 (.614)	.572 (.504)	1.43 (.415)
14	1.00	.335	0	1.01 (.489)	.652 (.487)	1.17 (.348)
15	.335	1.00	0	1.04 (.485)	.467 (.421)	1.07 (.317)
16	.493	1.48	0	1.05 (.484)	.415 (.401)	1.04 (.309)
17	1.48	.493	0	.923 (.441)	.446 (.398)	.968 (.290)
18	.493	1.48	0	.932 (.441)	.406 (.384)	.946 (.284)
19	.651	1.95	0	.934 (.441)	.395 (.381)	.940 (.283)
20	1.95	.651	0	.900 (.428)	.403 (.379)	.922 (.277)

I is a 2 x 2 identity matrix with ones on the main diagonal and zeros on the off-diagonal. The weight of the composite is β = [.7071, .7071].