MARGIE: Marginal maximum likelihood estimation of the parameters of the one-, two-, and three-parameter logistic models

ROBERT L. McKINLEY
The University of Toledo, Toledo, Ohio

Bock and Aitkin (1981) provided a powerful estimation technique for use with item response theory. Marginal maximum likelihood estimation (MMLE) treats the examinee ability parameter as a random nuisance parameter that is eliminated from the estimation process by specifying a form for the ability distribution, and integration over that distribution. Thus, item-parameter estimates are obtained by maximum likelihood estimation (MLE) in the marginal distribution. In this way, problems such as inconsistent parameter estimates which arise in joint MLE are avoided.

The MARGIE program is designed to perform MMLE of the item parameters of the one-parameter logistic (1PL), two-parameter logistic (2PL), and three-parameter logistic (3PL) item response theory (IRT) models. In addition, ability estimates are produced using either MLE or Bayesian expected a posteriori estimation (EAPE). Also, a chi-square test of the goodness of fit of the IRT model to the data is performed, using the procedure described by Yen (1981).

The Models. The program estimates the parameters of the 1PL, 2PL, and 3PL IRT models. The 1PL model is given by

$$P_{ij} = \frac{\exp[D(\theta_j - b_i)]}{1 + \exp[D(\theta_j - b_i)]}, \qquad (1)$$

where P_{ij} is the probability of a correct response to item i by examinee j; θ_j is the ability parameter for examinee j; b_i is the item difficulty parameter for item i; and $D = 1.7$. The 2PL model is given by

$$P_{ij} = \frac{\exp[Da_i(\theta_j - b_i)]}{1 + \exp[Da_i(\theta_j - b_i)]}, \qquad (2)$$

where a_i is the item discrimination parameter for item i, and the remaining terms are as previously defined. The 3PL model is given by

$$P_{ij} = g_i + (1 - g_i) \frac{\exp[Da_i(\theta_j - b_i)]}{1 + \exp[Da_i(\theta_j - b_i)]}, \qquad (3)$$

The author's mailing address is College of Education, the University of Toledo, 2801 W. Bancroft Street, Toledo, OH 43606.

where g_i is the lower asymptote (pseudo-guessing) parameter for item i, and the remaining terms are as previously defined.

MMLE Procedure. The MMLE procedure used in the program is an adaptation of a procedure proposed by Bock and Aitkin (1981) that was designed for use with a two-parameter normal ogive model. The algorithm was modified to make it appropriate for use with logistic models. In the initial step of the estimation procedure, provisional item-parameter estimates are selected, as are quadrature points and weights for the standard normal distribution. Item-test point biserial correlations are used as initial item discrimination parameter estimates. For initial estimates of item difficulty, logistic transformations of item proportion correct scores are used. Quadrature points and weights are selected using the Gauss-Hermite procedure (see Stroud & Sechrest, 1966). At each quadrature point, the provisional item-parameter estimates are used to compute the expected number of examinees having the ability represented by the quadrature point and the expected number of correct responses to each item by examinees with the ability represented by the quadrature point. These expected sample sizes and number-correct scores are used to perform marginal maximum-likelihood item-parameter estimation. This estimation is performed by solving the following set of equations for the indicated parameters:

$$c_i: \quad 0 = \sum_{j=1}^{k} (R_{ij} - P_{ij}N_j)W_j, \qquad (4)$$

$$a_i: \quad 0 = \sum_{j=1}^{k} (R_{ij} - P_{ij}N_j)W_jX_j, \qquad (5)$$

$$g_i: \quad 0 = (1 - g_i)^{-1} \sum_{j=1}^{k} (R_{ij} - P_{ij}N_j)/P_{ij}, \qquad (6)$$

where N_j is the expected sample size at quadrature point j; R_{ij} is the expected number-correct score for item i at point j; X_j is quadrature point j; c_i is equal to $-a_ib_i$; P_{ij} is computed from Equation 1, 2, or 3; and W_j is given by

$$W_j = \frac{(1 - g_i)P_{ij}*(1 - P_{ij}*)}{P_{ij}(1 - P_{ij})}, \qquad (7)$$

where $P_{ij}*$ is given by $[P_{ij}/(1 - g_i)] - g_i$. These equations are solved simultaneously using the Newton-Raphson technique. The resulting item-parameter estimates are used as provisional estimates for step two, and the entire process is repeated. This iterative process continues until the procedure converges, or until a maximum number of steps is reached. For a more detailed discussion of the

MMLE procedure, see either Bock and Aitkin (1981), or Reckase and McKinley (1982).

MLE of Ability. MLE of ability is performed by maximizing, for each examinee, the criterion function

$$L_j = \ln \prod_{i=1}^{n} P_{ij}, \qquad (8)$$

where L_j is the natural logarithm of the response vector likelihood function given the current estimate of ability; P_{ij} is the probability of the observed response to item i by examinee j; and n is the number of items. The criterion function is maximized by setting the first derivative with respect to θ equal to zero and solving. The solution is obtained using an iterative Newton-Raphson procedure that employs the second derivative with respect to θ of the criterion function.

EAPE of Ability. The EAPE procedure uses as an ability estimate the expected value of the posterior ability distribution. For each response vector, the assumed prior distribution of ability is multiplied by the likelihood distribution to obtain a posterior distribution. The expected value of this posterior distribution is the ability estimate for the examinee. The EAPE ability estimates are a natural by-product of the MMLE procedure and, as such, add no additional computational load. Thus, they are somewhat less expensive to produce than are the MLE ability estimates.

Goodness of Fit. The chi-square goodness of fit procedure utilized by MARGIE is based on a procedure described by Yen (1981). This procedure requires that examinees be divided into deciles on the basis of their ability estimates. For each decile, the expected probability of a correct response to a given item is computed as the mean of the probabilities of a correct response computed for the examinees in the decile using the IRT model and parameter estimates. The observed proportion of correct responses to the item within each decile is also computed. The expected and observed values for the deciles are used to compute a chi-square statistic. The degrees of freedom for the statistic are given by $10 - m$, where m is the number of item parameters estimated (1 for the 1PL model, 2 for the 2PL model, 3 for the 3PL model). The chi-square statistic is given by

$$\chi_i^2 = N \sum_{j=1}^{10} \frac{(E_{ij} - O_{ij})^2}{E_{ij}(1 - E_{ij})}, \qquad (9)$$

where χ_i^2 is the statistic for item i; N is the number of examinees in each decile; E_{ij} is the expected probability of a correct response to item i in decile j; and O_{ij} is the observed proportion correct for item i in decile j. It should be noted that the Yen chi-square test is quite sensitive.

Input. Input to MARGIE consists of the response matrix and a number of input parameters. These parameters include a run name, the number of items, the number of examinees, the calibration model (1 for 1PL, 2 for 2PL, 3 for 3PL), an ability estimation selection (0 for EAPE and 1 for MLE, with 0 being the default), and the input format for the item responses. The program assumes that the item responses are scored, and that an incorrect response is coded as a 0 and a correct response is coded as a 1.

Output. Output includes a listing of the user input parameters, information regarding the number of steps performed, the final item-parameter estimates and chi-square values, and the examinee ability estimates.

Language. The MARGIE program is written entirely in FORTRAN IV. It has been tested on an IBM 370 operating system. No external subroutines are required except those contained in the standard FORTRAN library.

Limitations. Use of the MARGIE program is not recommended for sample sizes less than 200 for the 1PL model, 500 for the 2PL model, and 1,000 for the 3PL model. Smaller sample sizes may result in poor fit of the model to the data.

Availability. A copy of the MARGIE program is available free of charge from the author. Requests for program copies should be accompanied by a computer tape. Otherwise, a tape-purchase fee of $25 will be charged.

REFERENCES

BOCK, R. D., & AITKIN, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika*, **46**, 443-459.

RECKASE, M. D., & McKINLEY, R. L. (1982, July). *Some latent trait theory in a multidimensional latent space*. Paper presented at the Item Response Theory/Computerized Adaptive Testing Conference, Wyzata, MN.

STROUD, A. H., & SECHREST, D. (1966). *Gaussian quadrature formulas*. Englewood Cliffs, NJ: Prentice-Hall.

YEN, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement*, **5**, 245-262.