Finding the Maximum Likelihood Estimate of a Parameter in an Item Response Model

Cheah Piang Kew
Singapore Examinations and Assessment Board

A paper presented at the 32nd Annual Conference of the
International Association for Educational Assessment, Singapore, May 2006.

Abstract

A crucial step in applying item response theory to practical testing problems is the estimation of ability and item parameters in the chosen item response model. It is usually not possible to estimate the parameters analytically. In the situation where there is only one parameter, two fast iterative numerical procedures for finding the maximum likelihood estimate of the parameter are presented in this paper. A numerical example is given to illustrate their use. Scope for further research is then discussed.

Contact details:
Address: 298 Jalan Bukit Ho Swee, Singapore 169565

Email address: cheah_piang_kew@seab.gov.sg

Tel: 65-63777721
Fax: 65-62729439

Introduction

Whether item response theory can be successfully applied to provide solutions to practical testing problems depends very much on the availability of procedures for estimating the examinee's ability parameter and the item parameters, for example, item difficulty, item discrimination and the guessing parameter. An approach that is often adopted in the estimation of parameters is the maximum likelihood method. An advantage of the maximum likelihood method is that the asymptotic sampling distributions of maximum likelihood estimates are known. This knowledge facilitates the application of item response theory.

It is assumed hereafter in this paper that there is only one unknown parameter in an item response model to be estimated.

Newton –Raphson Procedure

A common and powerful method of finding the maximum likelihood estimate (mle) of a parameter θ when the likelihood equation:

$$l'(\theta) = 0$$

cannot be solved analytically is the Newton-Raphson iterative procedure (Burden and Faires 1993). However, this procedure entails the existence of a continuous non-vanishing second derivative of the likelihood function (or log-likelihood function), $l(\theta)$, in the likelihood equation. The second derivative, if it exists, may be tedious to compute. Furthermore, the initial point must be chosen sufficiently close to the root of the likelihood equation for the procedure to converge.

Two Procedures Proposed

This paper presents two iterative procedures for finding the mle of a parameter that require only a continuous first derivative of the likelihood function and the identification of two arbitrary points between which the maximum point of the likelihood function lies. The proximity of the two points to the maximum point is, however, not crucial for the convergence of the procedures.

In the following two procedures, we assume that l is a likelihood function with a continuous first derivative and a unique maximum point and no other turning points. A and B are two arbitrary points to the left and the right of the maximum point respectively, with $l'(A) > 0$ and $l'(B) < 0$.

Procedure 1

The sub-interval contained in the interval (A, B) over which l' changes sign (and hence contains the maximum point) is identified and is progressively made finer until the pre-determined accuracy is achieved using the following algorithm:

Step 1 Fix two points A and B with A to the left and B to the right of the maximum point of the likelihood function

Step 2 Compute the value of $l'(\theta)$ at A and B to verify that $l'(A) > 0$ and $l'(B) < 0$.

Step 3 Divide the interval (A, B) into 10 equal sub-intervals with the points P_0, P_1, P_2, ..., P_{10}, where $P_0 = A$ and $P_{10} = B$.

Step 4 Compute the value of $l'(\theta)$ at P_i, i=1 to 10 until l' changes sign or vanishes.

Step 5 Suppose $l'(P_i) < 0$ and $l'(P_j) > 0$, j = 1 to i-1. Sub-divide the interval (P_{i-1}, P_i) into 10 equal sub-intervals.

Step 6 Compute the value of $l'(\theta)$ at each of the end-points of the sub-intervals between P_{i-1} and P_i immediately after P_{i-1} until $l'(\theta)$ changes sign or vanishes.

Step 5 and *Step 6* are repeated until the width of a sub-interval is less than the error margin. The point at which $l'(\theta)$ vanishes or any point in the sub-interval over which $l'(\theta)$ changes sign can then be assigned to be the value of the mle to the required accuracy.

Procedure 2

A sequence of points, alternately smaller and larger than the root of the likelihood equation, is computed so that the sequence converges to the root as follows:

Step 1 Fix two points A and B with A to the left and B to the right of the maximum point of the likelihood function.

Step 2 Compute the value of $l'(\theta)$ at A and B to verify that $l'(A) > 0$ and $l'(B) < 0$.

Step 3 Let C = B – (B – A) / 4.

Step 4 Compute $l'(C)$. If $l'(C) < 0$, shift C to the left gradually each time by $\frac{1}{2}$ of the distance between its current position and A until $l'(C) > 0$. A_1 will be assigned the value of C when $l'(C) > 0$ for the first time.

Step 5 Let D = $A_1 + (B - A_1)/4$.

Step 6 Compute $l'(D)$. If $l'(D) > 0$, shift D to the right gradually each time by $\frac{1}{2}$ of the distance between its current position and B until $l'(D) < 0$. B_1 will be assigned the value of D when $l'(D) < 0$ for the first time.

Step 3 – Step 6 constitute one iteration. In the ith iteration, replace A and B with A_{i-1} and B_{i-1} respectively. The procedure is stopped when the value of $l'(\theta)$ vanishes or the width of either of the two sub-intervals: (A_i, B_{i-1}) or (A_i, B_i) is less than the error margin. The point at which $l'(\theta)$ vanishes or any point in the sub-interval (A_i, B_{i-1}) or (A_i, B_i) can then be assigned to be the value of the mle to the required accuracy.

The choice of the factor $\frac{1}{4}$ in *Step 3* and *Step 5* is arbitrary and hence may not be optimal while the choice of the factor $\frac{1}{2}$ in *Step 4* and *Step 6* is to ensure that C and D converge to A_{i-1} and B_{i-1} for $l'(C)$ and $l'(D)$ to turn positive and negative respectively.

Example

Consider the likelihood function

$$l(\theta) = \theta(1-\theta).$$

Then

$$l'(\theta) = 1-2\theta.$$

Here the mle is exactly equal to 0.5.
Let A = 0.1 and B = 1.
Procedure 1 takes 17 computations and Procedure 2 takes 14 computations of $l'(\theta)$ to estimate the mle with an error of less than 0.001. However, dividing (A, B) into sub-intervals of width 0.001 and estimating the mle by brute force i.e. by computing $l(\theta)$ at each of the end-points would require close to 900 computations of $l(\theta)$.

Discussions

Procedure 1 and Procedure 2 are simple to use and are relatively fast. They only require the likelihood function to have a continuous derivative. Procedure 1 could be made shorter by computing the value of $l'(\theta)$ only at selected end-points of the sub-intervals. The factor of $\frac{1}{4}$ used in *Step 3* and *Step 5* of Procedure 2 is chosen arbitrarily. More research could be conducted to see if other values of the factor can be used to expedite the procedure. Procedure 2 is the faster of the two procedures in finding the mle in the above numerical example. However, it is not known whether this holds for all cases.

References

Burden, R.L. and Faires, J.D. (1993). *Numerical analysis, fifth edition.* PWS Publishing Company.