**New version of CAT algorithm by maximum likelihood estimation**

Loc Nguyen

Sunflower Soft Company, Vietnam

ng\_phloc@yahoo.com

http://www.locnguyen.net

**Abstract**

Computer-based testing with support of internet and computer is better than traditional paper-based testing. Computerized Adaptive Testing (CAT) is the branch of computer-based testing but it improves the accuracy of test core when CAT system tries to choose items such as tests, exams, and questions that are suitable to examinees’ abilities. I propose an advanced CAT algorithm based on two mathematical findings such as equation to estimate ability of a given examinee and equation to estimate ability variance among examinees. Such two mathematical equations are derived from maximum likelihood estimation of Item Response Function. The advanced CAT algorithm aims to classify examinees by the best way according to these equations.

*Keywords*: Computerized Adaptive Testing, CAT, Item Response Function, Maximum Likelihood Estimation, ability estimate, ability variance.

**1. Introduction to Computerized Adaptive Testing**

Computer-based testing has more advantages than traditional paper-based testing when there is a boom of internet and computer. Computer-based testing allows students to perform the tests at any time and any place and the testing environment becomes more realistic. Moreover, it is very easy to assess students’ ability by using the Computerized Adaptive Testing (CAT). The CAT is considered as the branch of computer-based testing but it improves the accuracy of test core when CAT systems try to choose items (tests, exams, questions, etc.) which are suitable to students’ abilities; such items are called adaptive items.

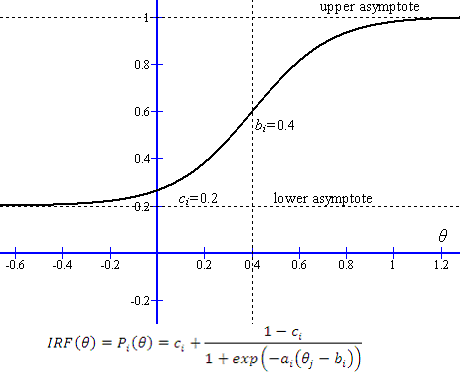
The important problem in CAT is how to estimate students’ abilities so as to select the best items for students. There are some methods to solve this problem such as Bayesian approach (Linden & Pashley, 2002, pp. 3-7) but I propose a new method to compute ability estimates by maximization likelihood estimation (MLE). Based on the proposed MLE method for estimating examinee’s ability, the new version of CAT algorithm is also invented. Such new version of CAT algorithm is called *advanced CAT algorithm*. Section 2 describes the proposed MLE method for estimating examinee’s ability. Section 3 describes advanced CAT algorithm in detailed. First, following is the overview of CAT theory.

*Item Response Theory* (IRT) is defined as a statistical model in which examinees can be described by a set of ability scores that are predictive. Based on mathematical models, IRT links together examinee’s performance on test items, item statistics, and examinee abilities (Rudner, 1998). Note that the term “*item*” indicates test, exam, question, etc. Users in IRT context are examinees. Examinee’s ability is often represented by variable *θ*. Given an examinee and item *i,* IRT is modeled as a function of a true ability *θ* of given examinee with three parameters of item *i* such as *ai, bi,* and *ci*. This function so-called Item Response Function (IRF) or Item Characteristic Curve (ICC) function computes the probability of a correct response of given examinee to item *i*. IRF is specified by equation 1 as follows:

|  |  |
| --- | --- |
|  | **(1)** |

Where *exp*(.) or *e*(.) denotes exponent function. Your attention please, IRF is function of examinee’s ability and it is essentially the probability of a correct response of a given examinee to an item. Suppose that *ai* is greater than 0.

IRF, a variant of logistic function, is plotted as the curve in following figure 1 with *ai=*6*, bi=*0.4*, ci=*0.2.



**Figure 1.** Item Response Function curve

The horizontal axis *θ* is the scale of examinee’s ability (Rudner, 1998). The vertical axis is the probability of correct response to the item specified by three parameters: *ai=*6*, bi=*0.4, *ci=*0.2. As seen in figure 1, the more the IRF shifts right, the more difficult item is. The lower asymptote at *ci=*0.2 indicates the probability of correct response for examinee with lowest ability and otherwise for the upper asymptote at 1.

IRF measures examinee’s proficiency based on her/his ability and some properties of item. Every item *i* has three parameters *ai*, *bi*, *ci* which are specified by experts or statistical data.

* The *ai* parameter called *discriminatory parameter* (Rudner, 1998) tells how well the item discriminates between examinees whose abilities are not different much. It defines the slope of the curve at the inflection point. The higher is the value of *ai*, the steeper is the curve. In case of steep curve, there is a large difference between the probabilities of a correct response for examinees whose ability is slightly below of the inflection point and examinees whose ability is slightly above the inflection point (Rudner, 1998).
* The *bi* parameter called *difficult parameter* (Rudner, 1998) indicates how difficult the item is. It specifies the location of inflection point of the curve along the *θ* axis (examinee’s ability). Higher value of *bi* shifts the curve to the right and implicates that the item is more difficult.
* The *ci* parameter called *guessing parameter* (Rudner, 1998) indicates that the probability of a correct response to item of low-ability examinees is very close to *ci*. It determines the lower asymptote of the curve. This parameter is called guessing parameter because it is the random probability that low-ability examinees guess a correct response to an item when they do not master the item. The upper asymptote always goes through value 1 because the probability that high-ability examinees give right response to an item is 1 (Rudner, 1998).

In general, IRF is used by computerized adaptive testing for choosing the best item which is given to examinee and estimating examinee’ true ability *θ*. Computerized adaptive testing is described right after.

Computerized Adaptive Testing (CAT) (Rudner, 1998) is the iterative algorithm which begins providing examinee an (test) item so as to be best to her/his initial ability; after that the ability is estimated again and the process of item suggestion is continued until stopping criterion is met. This algorithm aims to make a series of items which are evaluated to become chosen items that suitable to examinee’s ability. The set of items from which system picks ones up is called as item *pool*. The items chosen and given to examinee compose the adaptive test. CAT includes following steps (Rudner, 1998) as shown in table 1:

|  |
| --- |
| 1. The initial ability of examinee must be defined and items in the pool that have not yet been chosen are evaluated. The best one among these items is the most suitable to examinee’s current ability estimate. Such best item will be given to examinee in the step 2. IRF is applied into evaluating items. 2. The best item is chosen and given to examinee and the examinee responds. Such item is moved from pool to adaptive test. 3. A new ability estimate of examinee is computed based on responses to all of the chosen items. IRF is applied into computing the ability estimate. It is explained that ability estimate is the estimated value of true ability *θ* of examinee at current time point. 4. Steps 1 through 3 are repeated until stopping criterion is met. |

**Table 1.** Computerized adaptive testing (CAT) algorithm

Note that the chosen item is also called the *administered item* and the process of choosing best item is also called the *administration process*. The ability estimate is the value of *θ* which is fit best to the model and reflects current proficiency of examinee in item but it is not imperative to define precisely the initial ability because the final ability estimate may not be closed to initial ability. The stopping criterion could be time, number of administered items, change in ability estimate, maximum-information of ability estimate, content coverage, a precision indicator (standard error), a combination of factors, etc. (Rudner, 1998).

In step 1, there is the question: “how to evaluate the items so as to choose the best one”. So each item *i* is qualified by the amount of information or entropy at given ability *θ*; such *information function* is denoted *Ii*(*θ*). The best next item is the one that gets most informative or provides highest value of *Ii*(*θ*). Equation 2 specifies information function for item *i* (Rudner, 1998).

|  |  |
| --- | --- |
|  | **(2)** |

Where *Pi*(*θ*) is the probability of a correct response to item *i* and so it is the IRF function specified by previous equation 1 and is the first-order derivative of *Pi*(*θ*). According to equation 1, we have:

The information function *Ii*(*θ*) reflects how much the item *i* matches examinee’s ability. The item should not be too easy or too difficult. In the step 1 of CAT algorithm, the best item is the one that maximizes the information function *Ii*(*θ*). It is easy to find out such best item by brute-force technique that browses all items.

In step 3 of CAT algorithm, it is required to compute the ability estimate. Newton-Raphson method is proposed to find out the ability estimate (Rudner, 1998). Bayesian approach is also the good method for estimating examinee’s ability (Linden & Pashley, 2002, pp. 3-7). However, I propose a new method to calculate ability estimate based on maximization likelihood estimation (MLE). The successive section 2 describes a modified MLE method for CAT.

**2. Maximum likelihood estimation (MLE) for CAT**

Let be the ability estimate of examinee, the goal of this section 2 is to calculate Recall that the ability estimate is very important to step 3 of CAT algorithm; please see table 1 for more details about CAT algorithm.

Suppose there are *N* items given to an examinee; in other words, the size of item pool is *N*. Each item *i* has *qi* optional responses. For example, we have *qi*=4 when item is question with four possible answers such as *A*, *B*, *C*, and *D*. We have *qi*=10 when item is an exam whose resulted grade ranges from 1 to 10. Suppose the number of *correct responses* of given examinee to item *i* is *ri*.

For example, a given examinee do exam *i* whose grade ranges from 1 to 10 and she/he gains grade 9 then, we have *qi*=10 and *ri*=9. Let *P*(*θ*) be the cumulative probability of a correct response to given item. Exactly, *P*(*θ*) is the probability that examinee’s ability is less than or equal to *θ*. Note that the probability *P*(*θ*) is IRF function specified by equation 1. For convenience, let guessing parameter be zero (*c*=0). It means that the probability that examinee guesses correct response equals 0. Equation 3 specifies *P*(*θ*) with *c*=0.

|  |  |
| --- | --- |
|  | **(3)** |

Where *a* and *b* are discriminatory parameter and difficult parameter, respectively. Without loss of generality, equation 3 implicates that guessing parameter is fixed.

Let *θ*0 be the examinee’s initial ability. When examinee’s initial ability is not determined yet, *θ*0 can be initialized by zero, as specified by equation 4.

|  |  |
| --- | --- |
|  | **(4)** |

Note that it is possible to initialize *θ*0 by arbitrary number.

According to Bernoulli trial (Montgomery & Runger, 2003, p. 72), the probability that examinee provides *ri* correct responses for given item *i* is:

Where probability *P*(*θ*0) is specified by equation 3 and *θ*0 is examinee’s initial ability specified by equation 4.

The likelihood function (Czepiel, 2002, pp. 4-5) of examinee’s ability when she/he responses *N* items in the pool is specified by equation 5 as follows:

|  |  |
| --- | --- |
|  | **(5)** |

Note, *a* and *b* become variables of the likelihood function *L*(*a*, *b*). The notation denotes combination taken *ri* of *qi* elements and so we have .

It is required to estimate the parameters *a* and *b* so that the likelihood function takes maximum value. Let and be parameter estimates of *a* and *b*, respective; of course, is the maximum value of likelihood function *L*(*a*, *b*). Thus, this method is called maximum likelihood estimation (MLE) and the goal of MLE is to find out parameter estimates and .

Because it is too difficult to work with the likelihood function in the form of product of condition probabilities, it is necessary to take logarithm of *L*(*a*, *b*) so as to transform the likelihood function from repeated multiplication into repeated addition. The natural logarithm of *L*(*a*, *b*) so-called log-likelihood function is denoted *LnL*(*a*, *b*). We have:

Briefly, equation 6 specifies the log-likelihood *LnL*(*a*, *b*).

|  |  |
| --- | --- |
|  | **(6)** |

Where *ln*(.) denotes natural logarithm function, *θ*0 is examinee’s initial ability and *ri* is examinee’s response. The notation denotes combination taken *ri* of *qi* elements and so we have .

Maximizing the likelihood function is equivalent to maximizing *LnL*(*a*, *b*).

The maximization can be done by setting first-order partial derivatives of *LnL*(*a*, *b*) with respect to parameters *a* and *b* to 0 and solving these equations to find out parameter estimates and . Two first-order partial derivatives of *LnL*(*a*, *b*) with respect to parameters *a* and *b* are:

Setting these first-order partial derivatives to 0, we have following equations for solving estimates and .

Given *a* > 0, if *b*=*θ*0 then,

Because the condition is not always true in all situations, it is possible to ignore the case *b*=*θ*0. Without loss of generality, we have equation 7 for finding out two parameter estimates and as follows:

|  |  |
| --- | --- |
|  | **(7)** |

Because *b* is difficult parameter reflecting directly examinee’s ability, it is possible to suppose that discriminatory parameter *a* is arbitrary. Deriving from equation 7, we have:

In general, two possible parameter estimates and are specified by equation 8 as follows:

|  |  |
| --- | --- |
|  | **(8)** |

Where *qi* is the number of possible responses of item *i*, *ri* is the correct response of examinee to item *i* and *θ*0 is the examinee’s initial ability. There is a convention that if examinee responds correctly all items, then, gains maximum value and you can define the maximum value by positive infinity (+∞) or any pre-defined very large number.

The probability *P*(*θ*) (IRF function) specified by equation 3 is essentially cumulative distribution function (CDF). Let *p*(*θ*) be the probability density function of ability *θ*, we have:

|  |  |
| --- | --- |
|  | **(9)** |

Equation 9 indicates that the probability density function of examinee’s ability is the first-order derivative of IRF function. Substituting parameter estimates and specified by equation 8 into equation 9, we get the optimal density function of examinee’s ability, specified by equation 10.

|  |  |
| --- | --- |
|  | **(10)** |

The ability estimate is the expectation of *θ* given optimal density function . We have:

Let and , we have:

(using L’Hôpital’s rule by taking derivatives of both numerator and denominator (Wikipedia, Indeterminate form, 2014))

We have:

(using L’Hôpital’s rule by taking derivatives of both numerator and denominator (Wikipedia, Indeterminate form, 2014))

(using L’Hôpital’s rule by taking derivatives of both numerator and denominator (Wikipedia, Indeterminate form, 2014))

In general, we have a very important result in which the ability estimate equals the difficult parameter estimate . Equation 11 represents this result, as follows:

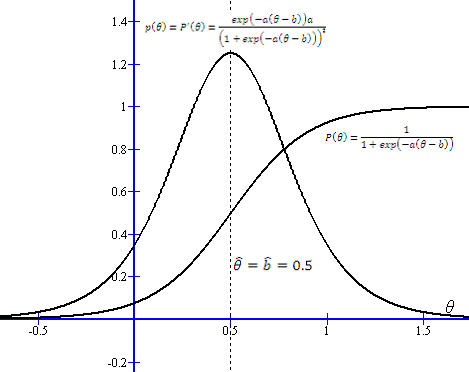
|  |  |
| --- | --- |
|  | **(11)** |

Where *qi* is the number of possible responses of item *i*, *ri* is the correct response of examinee to item *i* and *θ*0 is the examinee’s initial ability. Recall that if examinee responds correctly all items, then, gains maximum value and you can define the maximum value by positive infinity (+∞) or any pre-defined very large number.

Additionally, discriminatory parameter estimate is arbitrary; in practice, can be assigned by fixed discriminatory parameter *a* or by the average over all discriminatory parameters of existing items. For example, let *a*1, *a*2,…, *an* be discriminatory parameters of *n* existing items then, the estimate can be assigned by mean of such *ai* (s).

When parameter estimate was determined, it is easy to perform step 3 (calculating ability estimate of examinee) of the advanced CAT algorithm which will be shown in the next table 2. The equation 8 implicates that each examinee is modeled virtually as a test or exam; please pay attention to this interesting thing because it is the core of the advanced CAT algorithm for multi-user test.

Figure 2 is an example of IRF function *P*(*θ*) specified by equation 3, density function of examinee’s ability specified by equation 8 together with ability given discriminatory parameter *a=*5.



**Figure 2.** IRF function and density function of examinee’s ability together with ability estimate (0.5) given discriminatory parameter *a=*5

Based on the proposed MLE method for estimating examinee’s ability (equation 11), the new version of CAT algorithm is proposed in successive section 3.

**3. New version of CAT algorithm based on MLE**

As aforementioned in equation 11, examinee’s ability estimate is equal to difficult parameter estimate when is essential ability mean given probability density function specified by equation 10. There is a demand of how to specify ability variance of examinee. The new version of CAT algorithm, so-called *advanced CAT algorithm*, is based on such ability variance. Let be the ability variance, we have:

|  |  |
| --- | --- |
|  | **(12)** |

Where is optimal density function of examinee’s ability, specified by equation 10.

It is very easy to infer from equation 12 that the square root of is sample standard error according to study of statistics (Montgomery & Runger, 2003, p. 225). The variance is totally determined by calculation of We have:

We have following indefinite integral:

We have:

Given > 0 and let , we have:

And

Where *Li*2(*x*) is dilogarithm function (Wikipedia, Polylogarithm, 2014). Equation 13 expresses dilogarithm function.

|  |  |
| --- | --- |
|  | **(13)** |

Where *Li*2(0) = 0.

You can also find out equation 13 in (WolframAlpha) and (Weisstein, Dilogarithm). We have

It is easy to infer that

It implies that

|  |  |
| --- | --- |
|  | **(14)** |

Equation 14 expresses how to calculate the expectation used to determine ability variance . It is easy to get an equation similar to the equation 14 for calculating by using the mathematics engine (Wolfram|Alpha, n.d.). According to equation 14, it is necessary to calculate limits of expressions as *θ* approaches +∞ and –∞. We have:

(by using L’Hôpital’s rule by taking derivatives of both numerator and denominator (Wikipedia, Indeterminate form, 2014))

We have:

(by using L’Hôpital’s rule by taking derivatives of both numerator and denominator (Wikipedia, Indeterminate form, 2014))

(by using L’Hôpital’s rule by taking derivatives of both numerator and denominator (Wikipedia, Indeterminate form, 2014))

(by using L’Hôpital’s rule by taking derivatives of both numerator and denominator (Wikipedia, Indeterminate form, 2014))

We have:

It implies

We have:

We have:

(by using L’Hôpital’s rule by taking derivatives of both numerator and denominator (Wikipedia, Indeterminate form, 2014))

According to (Wikipedia, Polygamma function, 2014), given *x* < 0 equation 15 is an *inversion property* of dilogarithm, as follows:

|  |  |
| --- | --- |
|  | **(15)** |

It implies

We have:

We have:

According to equation 14 for calculating the expectation , we have

According to equation 12 for calculating the ability variance , we have

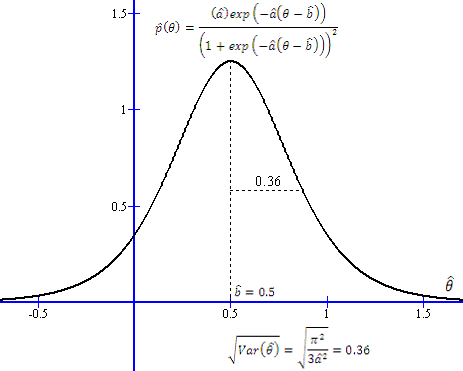
Briefly, equation 16 indicates how to compute the variance of examinee’s ability estimate. Please pay attention to the important result of equation 16 because it is used to compute estimate of discriminatory parameter.

|  |  |
| --- | --- |
|  | **(16)** |

The standard deviation of that is square root of is:

For example, given and , by applying equation 16, we have:

Figure 3 shows the example with , , and standard deviation with note that standard deviation is square root of variance and optimal density function is specified by equation 10.



**Figure 3.** An example of examinee’s ability variance

Suppose there are *k* examinees *u*1, *u*2,…, *uk* who have *k* ability estimates , ,…, after they have done a number of test items. Let be statistical sample mean (Montgomery & Runger, 2003, p. 190) of examinees’ ability estimates, equation 17 specifies the mean .

|  |  |
| --- | --- |
|  | **(17)** |

The ability variance is now considered as statistical sample variance (Montgomery & Runger, 2003, p. 191), which is calculated by equation 18 as follows:

|  |  |
| --- | --- |
|  | **(18)** |

Where is the ability mean specified by equation 17.

Equation 18 shares the same purpose with the equation 16 but their approaches are different. The variance specified by equation 16 is essentially the *local variance* focusing on individual examinee. The variance specified by equation 18 is essentially the *global variance* across *k* examinees. The reason of using the same notation for both local variance and global variance is explained right later when the discriminatory parameter *a* (see previous section 2) is computed based on these ability variances.

As aforementioned in previous section 2, at the step 1 of CAT algorithm shown table 1, the best item is the one that maximizes the information function *Ii*(*θ*) specified by equation 2. Each test item *i* has individual discriminatory parameter *ai*, difficult parameter *bi*, and guessing parameter *ci*. If there are *k* examinees in multi-user test, it is necessary to choose items so that it is easy to discriminate examinees according to their ability estimates. In other words, such items allow system to classify examinees into distinguishable groups according to examinees’ abilities. These items are called *discriminatory items*. Let *a\** be the so-called *discriminatory estimate*, it is easy to recognize that *a\** is the global estimated value of discriminatory parameter. In formal, discriminatoryitem is defined as the one whose discriminatory parameter *ai* is approximated to *a\**. According to equation 16 we have:

Applying equation 18 into determining , we have:

Briefly, equation 19 specifies discriminatory estimate.

|  |  |
| --- | --- |
|  | **(19)** |

Where , ,…, are *k* ability estimates of *k* examinees *u*1, *u*2,…, *uk* and is the ability mean specified by equation 17 with assumption that there are *k* examinees *u*1, *u*2,…, *uk* who have *k* ability estimates , ,…, . Based on the discriminatory estimate, the new version of CAT algorithm for multi-user test is described in table 2. Such new CAT algorithm is called *advanced CAT algorithm*.

|  |
| --- |
| 1. Suppose there are *k* examinees *u*1, *u*2,…, *uk* who have *k* ability estimates *θ*1, *θ*2,…, *θk* after they have finished a number of items in previous test. Suppose there are *n* items in the test pool and each item *i* has individual parameters such as *ai*, *bi*, and *ci*. Items available in test pool must be evaluated. Let be ability mean of such *k* examinees.   Let *a\** be discriminatory estimatethat is calculated at step 3 in previous test. If *a\** is not determined, it is initialized by the mean of discriminatory parameters *ai* (s).  Following code is evaluation process to choose the best items recommended to examinees.  For each item *i* not recommended to examinee *uj* yet  Let  be deviation between difficult parameter *bi* of item *i* and the modified ability mean . Note that the mean is modified by so that examinee’s ability conforms to item’s difficult parameter. Your attention please, examinee’s ability shares the same meaning with difficult parameter according to equation 11. Let *C* be set of items whose deviations are less than a pre-defined threshold *δ*. The *δ* is called informative threshold and *C* is called informative set. Thus, items in *C* are called *informative items*, which are ones whose difficult parameters are approximate to the average ability . Instead of using threshold *δ*, we can construct *C* by limiting its size. For example, given size 10, the set *C* will consist of 10 items whose deviations are smaller than remaining items.  For each item in *C*, the best item is the one whose parameter *ai* is nearest to discriminatory estimate *a\** with note that *a\** is specified by equation 19. In other words, if the best item is item *v* then, the deviation is relatively small and the deviation is smallest. It is easy to infer that the best item is the informative and discriminatory item.  End For  The best items are the most suitable to examinees’ current ability estimates. Not like traditional CAT mentioned in table 1, it is not necessary to calculate information function specified by equation 2 for selecting the best item. For this reason, the computation cost is decreased significantly.   1. Such best items are given to examinees and examinees make responses to these items. Following code describes process of test performance.   For each examinees *uj* among *k* examinees  The best item *vj* is given to *uj*. The examinee *uj* performs the test item and her/his response (result) is collected. Please see previous section 2 for more details about the example of response. For example, if item is an exam whose grade ranges from 1 to 10 then, the response is the resulted grade of examinee.  End For   1. New ability estimates of examinees are computed based on responses to all of the chosen items and current abilities *θ*1, *θ*2,…, *θk*. Concretely, the *k* ability estimates , ,…, are re-calculated according to equation 11. Moreover, discriminatory estimate *a\** is re-computed based on new estimates , ,…, according to equation 19. Let *θ*1, *θ*2,…, *θk* are current abilities of *k* examinees *u*1, *u*2,…, *uk*. We assign , ,…, in order to update current abilities *θ*1, *θ*2,…, *θk*. 2. Algorithm terminates if stopping criterion is met; otherwise going back step 1. |

**Table 2.** Advanced CAT algorithm

By focusing on both information function and discriminatory parameter, the advanced CAT algorithm achieves two purposes:

* Best test items given to examinees are adaptive to examinees’ abilities.
* It is easy to classify examinees according to their abilities.

In normal the stopping criterion in step 4 of advanced CAT algorithm is often the number of (test) items, for example, if the test has 10 items then the examinee’s final estimate is specified at 10th item and the test ends. This form is appropriate to examination in certain place and certain time and user is the examinee who passes or fails such examination.

Suppose in situation that user is the learner who wants to gains knowledge about some domain as much as possible and she/he does not care about passing or failing the examination. In other words, there is no test or examination and the learners prefer to study themselves by doing exercise. There is an exercise and items are questions that belong to this exercise. It is possible to use another stopping criterion in which the exercise ends only when the learner cannot do it better or worse (Nguyen, 2013). At that time her/his knowledge becomes saturated and such knowledge is her/his actual knowledge. The *ability error* is used to assess the saturation of learner’s knowledge. The ability error is difference between current ability estimate and previous examinee’s ability *θ*. Given threshold *ξ*, if the ability error is less than *ξ* then the CAT algorithm terminates; hence this is the new stopping criterion for CAT algorithm. Equation 20 specifies ability error denoted *Err*.

|  |  |
| --- | --- |
|  | **(20)** |

If advanced CAT algorithm is applied into multi-user test, there will be *k* ability errors for *k* examinees.

In this case, the advanced CAT algorithm will terminate if all ability errors *Errj* are less than given threshold *ξ*.

Now the advanced CAT algorithm is described comprehensively. It is necessary to give an example for illustrating such algorithm. Suppose there is a multi-user test with 5 items and 4 examinees, as seen in table 3. Each item has 10 optional responses and each examinee has zero initial ability. The test stops when every examinee finishes 5 items.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Items | | | |
| *ai* | *bi* | *ci* | *qi* |
| Item 1 | *a*1=2 | *b*1=2 | *c*1=0 | *q*1=10 |
| Item 2 | *a*2=4 | *b*2=3 | *c*2=0 | *q*2=10 |
| Item 3 | *a*3=3 | *b*3=5 | *c*3=0 | *q*3=10 |
| Item 4 | *a*4=3 | *b*4=2 | *c*4=0 | *q*4=10 |
| Item 5 | *a*5=5 | *b*5=4 | *c*5=0 | *q*5=10 |
|  | | | | |
|  | Examinees | | | |
| *Ability* (*θi*) | | | |
| Examinee 1 | *θ*1=0 | | | |
| Examinee 2 | *θ*2=0 | | | |
| Examinee 3 | *θ*3=0 | | | |
| Examinee 4 | *θ*4=0 | | | |

**Table 3.** A multi-user test with 5 items and 4 examinees

The ability mean is:

The discriminatory estimate *a\** is initialized as follows:

The deviations between difficult parameters of items and the ability mean are:

Suppose we choose two items (item 1 and item 4) whose deviations are the smallest ones according to step 1 of advanced CAT algorithm. So, the informative set *C* is:

The deviations between item 1, item 4 and discriminatory estimate *a\** are:

Because the deviation is the smallest one, item 4 is the best item that is given to four examinees. According to step 2 of advanced CAT algorithm, suppose that responses of examinees 1, 2, 3, and 4 to item 4 are 8, 7, 6, and 6, respectively. Of course, we have *r*1=8, *r*2=7, *r*3=6, and *r*4=6. The ability estimates , , , and discriminatory estimate *a\** will be calculated according to step 3 of advanced CAT algorithm. Given zero initial ability, according to equation 11 we have:

The mean of ability estimates , , , is:

According to equation 19, we have:

The ability estimates , , , are modified based on current discriminatory estimate 13.24 and old discriminatory estimate 3.4. We have:

Examinees’ abilities *θ*1, *θ*2, *θ*3, and *θ*4 are re-assigned as follows:

The item pool now includes four items 1, 2, 3, and 5. Going back step 1 of advanced CAT algorithm and the test is repeated for the second time so as to give examinees new items. The ability mean is:

The be discriminatory estimate *a\** was determined in previous test time:

The deviations between difficult parameters of items and the ability mean are:

Suppose we choose two items (item 1 and item 2) whose deviations are the smallest ones according to step 1 of advanced CAT algorithm. So, the informative set *C* is:

The deviations between item 1, item 2 and discriminatory estimate *a\** are:

Because the deviation is the smallest one, item 2 is the best item that is given to four examinees. According to step 2 of advanced CAT algorithm, suppose that responses of examinees 1, 2, 3, and 4 to item 2 are 1, 6, 2, and 9, respectively. Of course, we have *r*1={8, 1}, *r*2={7, 6}, *r*3={6, 2}, and *r*4={6, 9}; recall that responses of examinees 1, 2, 3, and 4 to item 4 in previous test are 8, 7, 6, and 6, respectively. The ability estimates , , , and discriminatory estimate *a\** will be calculated according to step 3 of advanced CAT algorithm. Given zero initial ability, according to equation 11 we have:

The mean of ability estimates , , , is:

According to equation 19, we have:

The ability estimates , , , are modified based on current discriminatory estimate 34.11 and old discriminatory estimate 13.24. We have:

Examinees’ abilities *θ*1, *θ*2, *θ*3, and *θ*4 are re-assigned as follows:

The item pool now includes four items 1, 3, and 5. Because examinees do not finished five items yet, the advanced CAT algorithm does not stop according to its step 4. Similarly, such four steps of advanced CAT are repeated and items 5, 3, and 1 are given to examinees in turn. Of course, choosing items 5, 3, and 1 in succession is based on deviations and according to step 1 of advanced CAT algorithm. Following table 4 shows results of our multi-user test.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | *θi* | *ri* |  | *a\** |
| Item 4 | Examinee 1 | *θ*1=0 | *r*1=8 |  | 13.24 |
| Examinee 2 | *θ*2=0 | *r*2=7 |  |
| Examinee 3 | *θ*3=0 | *r*3=6 |  |
| Examinee 4 | *θ*4*=*0 | *r*4=6 |  |
| Item 2 | Examinee 1 | *θ*1=0.1047 | *r*1=1 |  | 34.11 |
| Examinee 2 | *θ*2=0.064 | *r*2=6 |  |
| Examinee 3 | *θ*3=0.0306 | *r*3=2 |  |
| Examinee 4 | *θ*4*=*0.0306 | *r*4=9 |  |
| Item 5 | Examinee 1 | *θ*1=–0.0059 | *r*1=4 |  | 78.22 |
| Examinee 2 | *θ*2=0.0181 | *r*2=6 |  |
| Examinee 3 | *θ*3=–0.0119 | *r*3=5 |  |
| Examinee 4 | *θ*4*=*0.0322 | *r*4=9 |  |
| Item 3 | Examinee 1 | *θ*1=–0.0034 | *r*1=7 |  | 194.41 |
| Examinee 2 | *θ*2=0.007 | *r*2=3 |  |
| Examinee 3 | *θ*3=–0.0034 | *r*3=8 |  |
| Examinee 4 | *θ*4*=*0.0177 | *r*4=9 |  |
| Item 1 | Examinee 1 | *θ*1=0 | *r*1=2 |  | 616 |
| Examinee 2 | *θ*2=0.001 | *r*2=7 |  |
| Examinee 3 | *θ*3=0.0005 | *r*3=9 |  |
| Examinee 4 | *θ*4*=*0.008 | *r*4=5 |  |
|  | Examinee 1 | *θ*1=–0.0004 |  | | |
| Examinee 2 | *θ*2=0.0005 |
| Examinee 3 | *θ*3=0.0007 |
| Examinee 4 | *θ*4*=*0.002 |

**Table 4.** Results of multi-user test with 5 items and 4 examinees

After doing final test item 1, four examinees gain final abilities *θ*1=–0.0004, *θ*2=0.0005, *θ*3=0.0007, and *θ*4*=*0.002. As shown in table 4, the final discriminatory estimate *a\**=616 is very high while the initial value of *a\** is 3.4. It implies that the variance of *θ*1, *θ*2, *θ*3, and *θ*4 gets small; please see equation 19 for more details about inverse proportion between *a\** and such variance. Therefore, it is required to use high-value discriminatory parameter to discriminate among examinees.

**4. Conclusion**

In general, we recognized that CAT gives us the excellent tool for assessing student’s ability. The CAT algorithm includes four steps in which step 3 is the most important when student’s ability estimate is determined. I propose a new method to compute the ability estimate based on maximum likelihood estimation (MLE). Thus, the ability estimate is the estimated value of difficult parameter which, in turn, is learned given student’s test results; please equation 11 for more details about ability estimate. Then, the discriminatory parameter is estimated based on the variance of examinee’s ability. Please see equations 16 and 19 for more details about ability variance and discriminatory estimate. When the ability variance and the discriminatory estimate are determined, the advanced CAT algorithm for multi-user test is proposed. In other words, the advanced CAT algorithm is the result of combination of three equations 11, 16 and 19. These equations are essential contributions of this research. The strong point of advanced CAT algorithm is to achieve two purposes:

* Best test items given to examinees are adaptive to examinees’ abilities.
* It is easy to classify examinees according to their abilities.

I had a problem of calculating the variance among examinees’ abilities but finally, I found out the formula to calculate this variance with support of (Wolfram|Alpha, n.d.) engine. Equation 16 is the most important finding in the research, in which the dilogarithm function is key solution. So, ability variance is the main aspect of the paper that makes it worth worldwide. However, both ability estimate and ability variance are based on the condition that guessing parameter *ci* of IRF is zero. Consequently, IRF is considered as cumulative distribution function according to equation 3. When guessing parameter *ci* is non-zero, IRF does not satisfy an axiom of probability distribution because the IRF approaches if the true ability *θ* approaches negative infinity. If we translate the IRF to horizontal axis IRF(*θ*) = 0 then, the IRF will approach 1–*ci* below 1 when the true ability *θ* approaches positive infinity. This is also invalid, which raises a hazardous problem. Therefore, for further research, I will try my best to re-calculate ability estimate and ability variance in the most general case .

**List of equations**

* Equation 1. Item Response Function
* Equation 2. Information function for item *i*
* Equation 3. IRF with zero guessing parameter
* Equation 4. Examinee’s initial ability
* Equation 5. Likelihood function of examinee’s ability
* Equation 6. Log-likelihood function of examinee’s ability
* Equation 7. Equation for finding out parameter estimates
* Equation 8. Discriminatory and difficult estimates
* Equation 9. Probability density function of examinee’s ability
* Equation 10. Optimal probability density function of examinee’s ability
* Equation 11. Examinee’s ability estimate
* Equation 12. Examinee’s ability variance
* Equation 13. Dilogarithm function
* Equation 14. Expectation of square of examinee’s ability
* Equation 15. Inversion property of dilogarithm
* Equation 16. Examinee’s explicit ability variance
* Equation 17. Examinee’s statistical ability mean
* Equation 18. Examinee’s ability variance as statistical sample variance
* Equation 19. Discriminatory estimate
* Equation 20. Ability error used as stopping criterion of CAT algorithm

# References

Czepiel, S. A. (2002). *Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation.* Czepiel's website http://czep.net.

Linden, W. J., & Pashley, P. J. (2002). Item Selection and Ability Estimation in Adaptive Testing. In W. J. Linden, G. A. Glas, W. J. Linden, & G. A. Glas (Eds.), *Computerized Adaptive Testing: Theory and Practice* (p. 323). Kluwer Academic Publishers.

Montgomery, D. C., & Runger, G. C. (2003). *Applied Statistics and Probability for Engineers* (3rd Edition ed.). New York, NY, USA: John Wiley & Sons, Inc.

Nguyen, L. (2013, June 23-24). The Bayesian approach and suggested stopping criterion in Computerized Adaptive Testing. (A. T. Al-Taani, Ed.) *International Journal of Research in Engineering and Technology (IJRET), 2*(1-2103), 36-38. Retrieved from http://journalsweb.org/siteadmin/upload/74977%20IJRET021007.pdf

Rudner, L. M. (1998, November). *An On-line, Interactive, Computer Adaptive Testing Tutorial.* Retrieved 2009, from Lawrence M. Rudner's website: http://edres.org/scripts/cat

Weisstein, E. W. (n.d.). *Dilogarithm*. (Wolfram Research) Retrieved December 17, 2014, from MathWorld--A Wolfram Web Resource: http://mathworld.wolfram.com/Dilogarithm.html

Wikipedia. (2014, October 1). *Indeterminate form*. (Wikimedia Foundation) Retrieved November 11, 2014, from Wikipedia website: http://en.wikipedia.org/wiki/Indeterminate\_form

Wikipedia. (2014, October 12). *Polygamma function*. (Wikimedia Foundation) Retrieved November 10, 2014, from Wikipedia website: http://en.wikipedia.org/wiki/Digamma\_function

Wikipedia. (2014, August 22). *Polylogarithm*. (Wikimedia Foundation) Retrieved December 17, 2014, from Wikipedia webaite: http://en.wikipedia.org/wiki/Polylogarithm

Wolfram|Alpha. (n.d.). *Wolfram|Alpha Mathematics Engine*. (Wolfram Alpha LLC-A Wolfram Research Company) Retrieved December 16, 2014, from Wolfram|Alpha: http://www.wolframalpha.com