PLISKA

STUDIA MATHEMATICA BULGARICA

ПЛИСКА

БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

Pliska Stud. Math. Bulgar. **22** (2013), 33–40

PLISKA
STUDIA MATHEMATICA
BULGARICA

GROUP COMPARISONS ON COGNITIVE ATTRIBUTES USING THE LEAST SQUARES DISTANCE MODEL OF COGNITIVE DIAGNOSIS

Dimiter M. Dimitrov, Dimitar Atanasov

Abstract. As the cognitive operations are hypothesized according to a cognitive theory in the context of a study, they are latent (hidden) in nature and cannot be measured and scored directly from the test. The least squares distance model (LSDM) of cognitive diagnosis uses estimates of the item parameters under a specific item-response theory (IRT) model to provide estimates of the probability of a person to process correctly any cognitive attribute given the persons location on the IRT logit scale. In this paper a methodology for comparing two (or more) groups of individuals, according to their performance on a given set of cognitive attributes is presented.

1. Introduction. The cognitive structure of a test in education, psychology, and other behavioral fields is typically defined by a set of cognitive operations, processes or rules, referred to as cognitive attributes, and information about which attributes are required for the correct solution of each test item. As the attributes are hypothesized according to a cognitive theory in the context of study, they are latent (hidden) in nature and cannot be measured and scored directly. Knowledge about cognitive structures can help test developers, psychologists, and educators to better understand the cognitive processes of thinking, learning, and performance. The least squares distance model of cognitive diagnosis ([2], [3]) uses estimates of the item parameters under a specific item-response

2010 *Mathematics Subject Classification:* 62P15
Key words: cognitive diagnosis modeling, item response theory, assessment.

theory (IRT) model to provide estimates of the probability of a person to process correctly any cognitive attribute given the persons location on the IRT logit scale. The graphical depiction of such probabilities along the IRT logit scale for a given cognitive attribute is referred to its attribute probability curve (APC).

One of the most popular and useful in practice methodology for studying the properties of the test items is the so called Item Response Theory model (IRT) ([1], [4]).

In a 3-parametric IRT model the probability for a correct performance on a given item from a student with ability on a logit scale θ can be represented as a 3-parametric logistic curve. The *difficulty* parameter a represents the level of the ability, required for a correct performance on the item. The *discrimination* parameter b indicates how effectively the item discriminates between the examinees who are relatively high on the criterion of interest against those who are relatively low. The *pseudo guessing* parameter c represents the probability that examinees with very low ability can guess the correct answer. Then the probability of a correct performance is given by the following logistic function

$$P(\theta) = c + (1-c)\frac{\exp(Db(\theta - a))}{1+\exp(Db(\theta - a))}, \tag{1}$$

where D is a constant, usually set to $D = 1.7$ when $P(\theta)$ approximates the normal ogive curve. Plotted against θ it gives the so called Item Characteristic Curves (ICC). An example of ICC for two items is presented on Figure 1. The parameters of the *Item 2* are shown. The difficulty of the item is the ability level, giving the probability of a correct performance equal to 0.5 if there is no guessing. The discrimination of the item is presented by the slope of the tangent at the point of difficulty. The guess parameter represents the probability for a correct item response (just by guessing) from a subject with a small level of abilities. In general, in this example, *Item 2* is more difficult, but less discriminative than *Item 1*, having a larger value of the guessing parameter.

LSDM ([2]) provides an interesting approach to a cognitive assessment. Cognitive diagnosis models are widely used in psychology and psychometrics in form of the conjunctive assumption that a correct item response is produced when all attributes required by the item are mastered. LSDM technique can be considered as an extension of the classical models as it can deal with both conjunctive and disjunctive assumptions ([3]). The basic idea behind the LSDM is that the probability for correct performance on the item can be represented by a correct performance on the required attributes. For example a conjunctive model for positive response on the dichotomous item D_l, $i = 1, \ldots, N$ determined from a

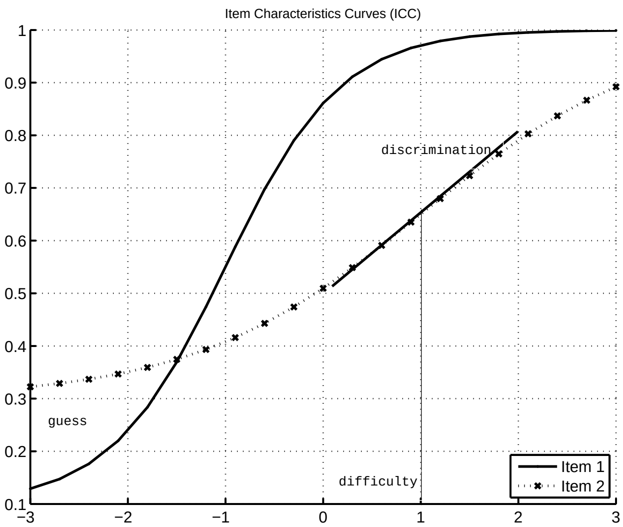

Fig. 1. Item Characteristic Curve

set of attributes $A_1, \ldots, A_k$ can be expressed as

$$P(D_i = 1 \mid \theta) = \prod_{l=1}^{k} P(A_l = 1 \mid \theta)^{q_{il}}, \tag{2}$$

where the values q_{il} form the matrix Q and $q_{il} = 1$, $l = 1, \ldots, k$, if the item D_i requires the attribute A_l, and 0 otherwise.

The purpose of this paper is to propose a LSDM-based probabilistic approach to comparing groups of examinees according to their performance on a set of cognitive attributes. Such groups can be defined, for example, by gender or different treatment conditions (say, experimental and control groups) in educational or psychological research.

For the computations and the representation of the results a MATLAB package is used, available at

`http://www.ir-statistics.net/index.cgi/software-irt`,

under the GNU License.

2. The Algorithm. The proposed algorithm is demonstrated via a particular example, taken from [2]. Let us have an IRT calibration and a Q-matrix of a set of 10 items

Table 1. IRT parameters of test items

Item	Difficulty	Discr.	Q-matrix
1	−2.0300	0.6000	1 0 0 0 0
2	−1.2900	0.8100	0 1 0 0 0
3	−1.0300	0.7500	0 1 0 1 0
4	−1.5800	0.8100	0 1 0 0 0
5	0.5900	0.6200	0 1 1 0 0
6	−1.6500	0.7500	0 0 0 1 0
7	2.2200	0.5400	0 1 0 0 1
8	−1.4600	0.6500	0 0 0 1 0
9	2.5800	0.7500	0 0 0 1 1
10	−0.6600	0.5400	1 0 1 0 0

First, the estimates of the test ability scores of the persons from the groups being compared are obtained by fitting an appropriate IRT model to the sample data (binary item scores, 1 = true, 0 = false). For example, on Table 2 the ability scores for 60 persons, separated in three groups are provided. Let us note that the average ability in these groups is statistically different (with p-value equal to 0.002).

Second, the APCs of attributes (`attributePerformance`) that underlie the success on test items (`itemPerformance`) are obtained through the use of the LSDM. The used functions are given in the following lines:

```
itemPerformance = irt_item_performance(th,itemParameters)
attributePerformance = lsdm(itemPerformance,Q,1)'
opt.legend = 1;
opt.colour = 1;
plot_item(attributePerformance,th,opt)
```

Third, these APCs are described analytically by fitting them to the 3-parameter logistic (3PL) model in IRT.

```
attributeParameters = irt_fit(th, attributePerformance)
```

Fourth, using this 3PL model of cognitive attributes, the probability for each person to process correctly a given attribute (i.e. the person's attribute score) is obtained as a function of the test ability score of that person on the IRT logit scale.

```
for g = [1,2,3] %for each group
```

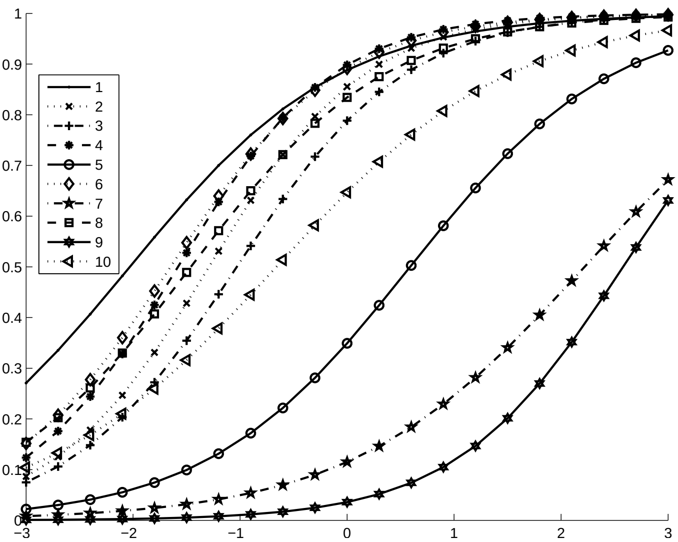

Fig. 2. Characteristic Curves of the test items

```
    performanceForGroup{g} =
     irt_item_performance(ability(:,g),attributeParameters);
end;
```

Fifth, the obtained attribute scores for the persons in the study sample are used with an appropriate statistical method for group comparison (e.g., t-test, ANOVA, ANCOVA, and so forth). Here the MATLAB functions `anova1` and `multcompare` are used.

```
for k = 1:5 % for each of the attributes
    compare = [];
    for g = [1,2,3] % for each group
        compare = [compare performanceForGroup{g}(k,:)']
    end;
[p,a,s] = anova1(compare)
multcompare(s)
end;
```

According to the comparison presented on Figure 2 there is no between group difference based on performance of attribute $A1$. Whereas, a statistically significant difference between group 2 and group 3 can be found considering the performance of the attribute $A3$ (Figure 2).

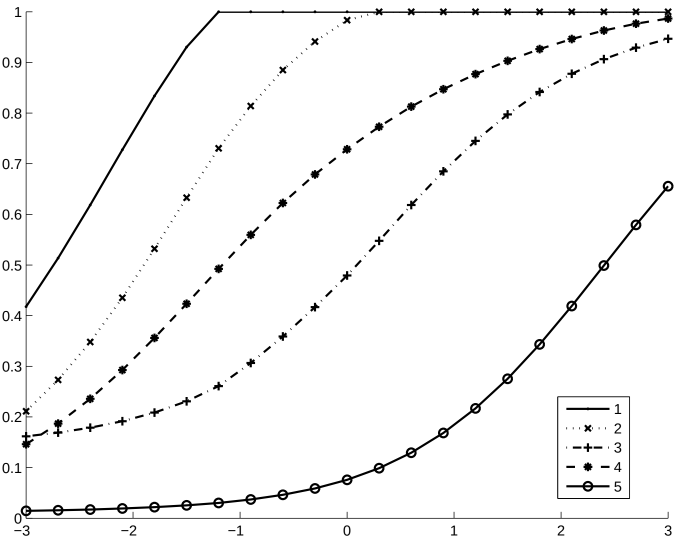

Fig. 3. Characteristic Curves of the attributes (APC)

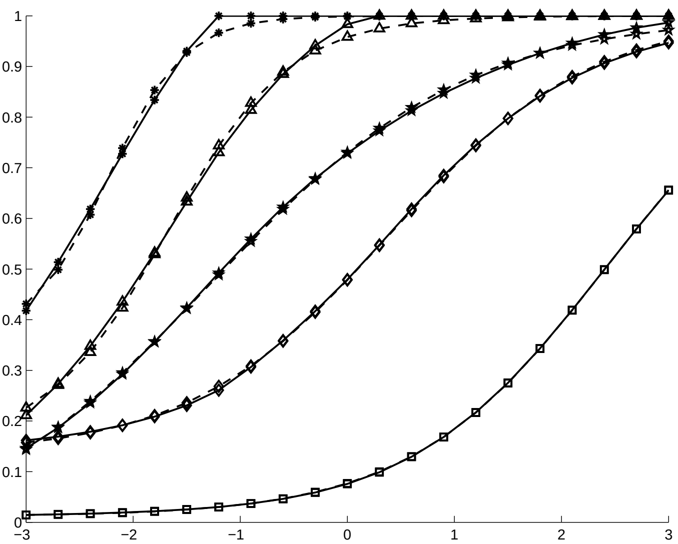

Fig. 4. Fit of the attribute curves

Table 2. Persons ability measures

Person	Group 1	Group 2	Group 3
1	0.5377	0.1715	0.8978
2	1.8339	−1.7075	0.7586
3	−2.2588	0.2172	1.3192
4	0.8622	1.1302	1.3129
5	0.3188	−0.0111	0.1351
6	−1.3077	0.5347	0.9699
7	−0.4336	0.2269	0.8351
8	0.3426	−0.8034	1.6277
9	3.5784	−0.2061	2.0933
10	2.7694	−1.2873	2.1093
11	−1.3499	0.3884	0.1363
12	3.0349	−1.6471	1.0774
13	0.7254	−1.5689	−0.2141
14	−0.0631	−1.3095	−0.1135
15	0.7147	−3.4443	0.9932
16	−0.2050	0.9384	2.5326
17	−0.1241	−0.1748	0.2303
18	1.4897	−1.2549	1.3714
19	1.4090	0.8703	0.7744
20	1.4172	−2.2115	2.1174

3. Conclusions. The proposed methodology shows how a between-group comparison can be achieved, based on the performing on a set of cognitive attributes. The considered example demonstrates that although that there is a statistically significant difference in the average person ability between the groups, this difference can't be considered as statistically significant at the level of the cognitive attribute performance in general. This difference can be caused by a distinct performing on one or more attributes with no difference in the others.

The results from such group comparisons on the latent cognitive attributes provide a valuable feedback to educators and psychologists for improving as-

Table 3. IRT parameters of the attributes

Attribute	Diff.	Discr.	Guess
1	−2.2257	1.6543	0.3669
2	−1.6742	1.0383	0.1518
3	0.3911	0.6280	0.1349
4	−1.2786	0.4939	0.0578
5	2.4264	0.6432	0.0119

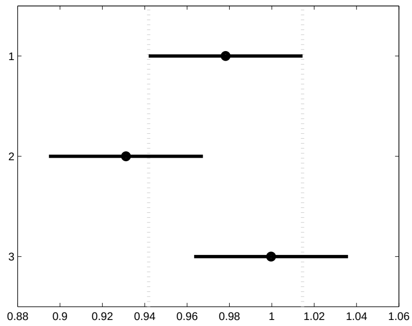

Fig. 5. Between group comparison of attribute $A1$

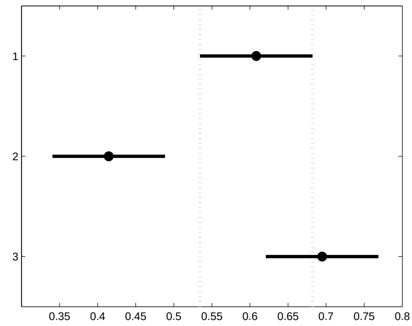

Fig. 6. Between group comparison of attribute $A3$

sessment, teaching strategies, and curriculum development in targeted areas of learning, teaching, and behavioural intervention.

The usage of specially developed software package is demonstrated.

REFERENCES

[1] L. Crocker, J. Algila. Introduction to Classical and Modern Test Theory, Warsworth, 1986.

[2] D. M. Dimitrov. Least squares distance method of cognitive validation and analysis for binary items using their item response theory parameters. *Applied Psychological Measurement* **31** (2007), No 5, 367–387.

[3] D. M. Dimitrov, D. V. Atanasov. Conjunctive and Disjunctive Extensions of the Least Squares Distance Model of Cognitive Diagnosis. *Educational and Psychological Measurement*, 2011.

[4] E. Smith, R. Smith. Introduction to Rasch Measurement. JAM Press, Maple Grove, Minnesota US, 2004.

Dimitar Atanasov
New Bulgarian University
21, Montevideo Blvd
1618 Sofia, Bulgaria
e-mail: `datanasov@nbu.bg`

Dimiter Dimitrov
George Mason University, Fairfax Campus
West Building 2007 4400
University Dr. MS 6D2 Fairfax, VA 22030
e-mail: `ddimitro@gmu.edu`