МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2009
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2009
Proceedings of the Thirty Eighth Spring Conference of
the Union of Bulgarian Mathematicians
Borovetz, April 1–5, 2009

ANALYSIS OF A TEST ON PROBABILITY THEORY*

Vessela Stoimenova, Dimitar Atanasov

In this paper we perform an analysis of a test on "Probability Theory" in order to find out which is the most difficult part of the taught material. The study is based on the Item Response Theory and Least Squared Distance Method for estimating the difficulty parameters of the test items and the cognitive attributes.

1. Introduction. In the present paper we consider the results from a test on Probability theory, given to 107 students, studying third year Informatics at the Faculty of Mathematics and Informatics, Sofia University. This speciality studies the course "Probability theory and Mathematical statistics" during the third educational year, winter term. The horarium is 45 hours lectures and 45 hours seminar classes. The course covers a wide range of subjects starting with classical probability, geometric probability, total probability law and the Bayes formula, random variables – discrete and absolutely continuous, their distribution properties, characteristics and moments, conditional mathematical expectation, functions of random variables, transformations, probability generating functions, inequalities for random variables, the Law of Large Numbers and the Central Limit Theorem. The learning and understanding of the covered material is additionally complicated by the fact that except the above stated themes the following elements of the mathematical statistics are introduced: descriptive statistics, estimators and their properties (checking for unbiased, consistent, efficient estimators), methods for achieving estimators (the method of moments, maximum likelihood method, the Rao-Cramer efficiency), confidence intervals, hypotheses testing and nonparametric tests. If time and abilities of the auditory allow, elements of regression analysis are included, as well Markov chains as a simple introduction to stochastic processes. Many years of observation are indicative of the fact that students face difficulties when learning this material and often do not show enviable results when being examined. The exam on Probability theory and Mathematical statistics consists of two parts: in the first part the students do practical exercises covering the above stated material and in the second part they write on two concrete topics (one on probability and another on statistics), given by the lecturer, and after that they defence their knowledge and understanding orally. The tendency to forget and lose the adopted abilities quickly after passing the exam can be seen even for students, who are notable for their excellent performace. Students, who

---

***2000 Mathematics Subject Classification:** 60J80.

Key words: Student assessment, item response theory, probability theory.

The paper is partially supported by the National Science Fund of Bulgaria, Grant No VU-MI-105/2005.

extend their education in some of the proposed by the faculty master programmes, are often constrained to learn anew the basic items and techniques of the probability theory and the curriculum on mathematical statistics should be practically taught again from the same beginning.

In order to understand the basic abilities, adopted by the students during their education, as well to introduce the test examination on this subject, a test is proposed, during the winter term of the academic year 2007–2008, consisting of twelve multiple choice items with four possible answers. There are two running controls during the term – one on probability theory and one on mathematical statistics. The test is given 30 minutes before the second control and students have up to half an hour to complete the test.

Up to this moment there exists no practice to carry out tests on Probability and Statistics. The proposed test is the first of its kind. The students are not prepared to answer such questions and they are explained that the test is obligatory, but the marks will be taken into account only if they are greater than the marks from the first running control.

The items in the test can be described as follows

1. Combinatorics - knowledge of independence of trials and order.
2. The students have to recognize a hypergeometrical distribution and its parameters. The question can also be answered using combinatorial reasoning.
3. The Bayes formula.
4. Conditional probability.
5. Event classification.
6. Mathematical expectation of a function of two independent discrete random variables.
7. Mathematical expectation of another function of two independent discrete random variables.
8. Covariance of two random variables.
9. Finding a concrete value of the normal probability distribution function using tables.
10. Finding a concrete value of the normal probability distribution function using given probability.
11. Finding a marginal distribution of a multidimensional absolutely continuous distribution.
12. Abilities to work with probability generating functions.

In this paper we study the parameters of the test items using the Item Response Theory (IRT) model. The parameters of the "studied abilities" are examined using the Least Squared Distance (LSDM) method. The cognitive attributes, needed for the correct performance on the test items, to be studied are five: A1 – combinatorial abilities, A2 – events and operations with events, A3 – independence of events and random variables, A4 – definition and probability distribution functions of random variables, A5 – moments of random variables.

2. The model. Item Response Theory (IRT) model of test item response is widely used for studying the characteristics of the items in the tests. According to IRT model, the probability for correct item response from student with abilities θ is defined as

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

where D represents the item difficulty and a represents the characteristics of the item to discriminate the students with lower and higher abilities (see f.e. [5], and [6]). The parameter c gives the probability of guessing the correct answer. The constant K can be arbitrarily set, but usually it is set to $K = 1.7$, because in this way P fits the normal ogive curve. When the parameters a and c are fixed to 1 and 0 respectively, the so called Rasch model is reached (see f.e. [2]).

Let $A_1, \ldots, A_m$ be the cognitive attributes, required for the correct response of the test items. Then the probability for correct item response on item n can be represented as

$$P_n(\theta) = \prod_{k=1}^{m} P(A_k = 1 \mid \theta)^{q_{nk}}, \tag{2}$$

where $P(A_k = 1 \mid \theta)$ is the probability for correct performance on the attribute A_k from a student with abilities θ. The indexes q_{nk} form the so called Q-matrix, which gives the dependence of the items from the attribute set. The value q_{nk} is set to be equal to 1, if the n-th item depends on the attribute A_k, and 0 otherwise.

The Least Squares Distance Method (LSDM), developed by Dimitrov [7], solves the equation (2) using the least squares fit of the following representation

$$\log P_n(\theta) = \sum_{k=1}^{m} q_{ik} \log P(A_k = 1 \mid \theta). \tag{3}$$

3. Main results. The performance of the students on the test items is calculated as a proportion of the correct answers on a particular item from students with a given level of ability. The ability level is considered to be a representation of the row score of the student. On Figure 1, the observed proportions of the correct answer on the items of the test is shown.

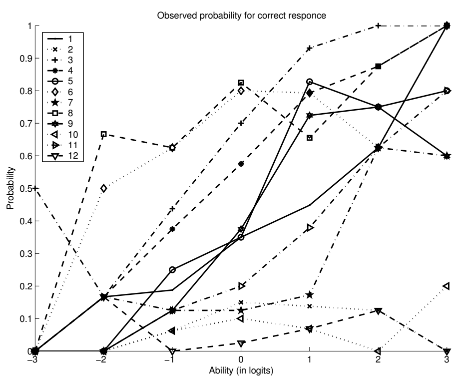

Fig. 1. Observed performance of test items

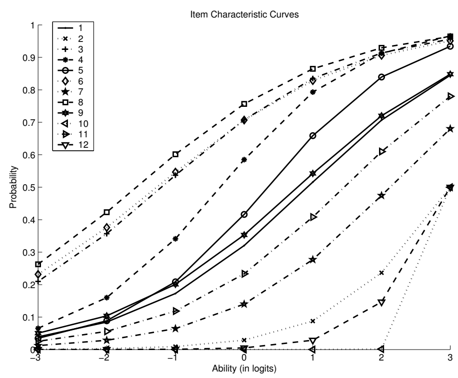

Fig. 2. Item Characteristic Curves (ICC)

This performance is fitted to a 2-parameter IRT model. The pseudo guessing parameter is omitted because of the small number of observations. From another point of view there is a strong evidence that the test items are written having in mind the best practice, including distractors in the answers in order to reduce the possibility of guessing the correct answer.

Obtained values of the difficulty and discrimination parameter are presented in Table 1.

Item	Difficulty	Discrimination
1	0.9221	0.4790
2	3.0000	0.6880
3	−1.2006	0.4328
4	−0.3427	0.5880
5	0.3393	0.5851
6	−1.2699	0.4074
7	2.1196	0.5035
8	−1.5702	0.4243
9	0.7813	0.4556
10	3.0000	3.8071
11	1.4538	0.4808
12	3.0000	1.0352

Table 1. IRT parameters of test items

Obviously Items 2 (recognizing the distribution), 10 (distribution function) and 12 (probability generation function) are the most difficult items in the test. Almost all students have recognized the Hypergeometrical distribution, but not its parameters. They face difficulties in connecting the probability with definition of the distribution function. And finally, the generating functions are usually the most difficult part of the subject. Items 1 (notion of order in combinatorics), 7 (expectation) and 11 (marginal distributions as a correct integral formula) are relatively difficult. Items 4 (conditional probability) is relatively easy. Items 3 (Bayes formula), 6 (expectation) and 8 (covariance) are the easiest items in the test.

On Figure 2, the Item Characteristic Curves (ICC) of the items of the test are presented, giving the probability of correct performance on the items according to the 2-paramater IRT model.

One of the main properties of the test is the test reliability. It may be considered as a measurement of consistency of scores across different evaluators over different time periods. It is used to substantiate elements of accuracy to the test administered on candidates and get the real feel of his ability. For the given test it is estimated as 0.7664, using the techniques given in [3]. The reliability depends on the *variance of the true score* and the *variance of the error score* (estimated here as 5.66 and 1.7275, respectively). The sum of expected item scores, referred to as *expected NR score* for the given test, is 4.49. The score dependability index [1] is presented on Figure 3.

Having the calculated IRT parameters and ICC, the attribute performance can be calculated, using the LSDM technique. Table 2 gives the Q-matrix, used for the dependence between attributes and items.

Applying the Least Squares method for solving the equation (3) the following attribute performance is obtained (Figure 4).

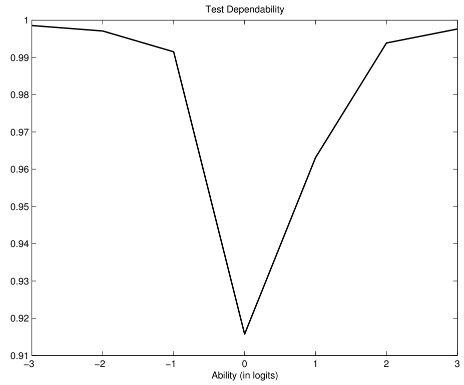

Fig. 3. Test score dependability

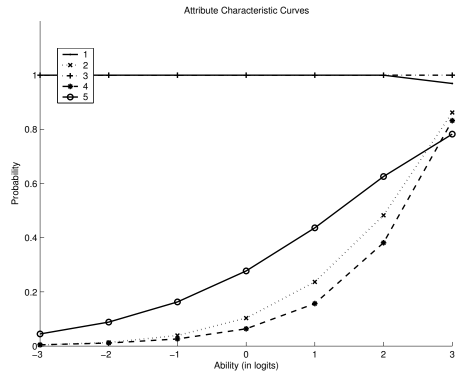

Fig. 4. Attribute Performance Curves

Item	A1	A2	A3	A4	A5
1	1	1	1	1	0
2	1	1	1	1	0
3	1	1	1	0	0
4	0	1	1	0	0
5	0	1	1	0	0
6	0	0	1	0	1
7	0	0	1	0	1
8	0	0	1	1	1
9	0	1	0	1	0
10	0	1	0	1	0
11	0	0	1	1	0
12	0	0	1	0	1

Table 2. Q-matrix

It can be seen that attributes A1 (combinatorics) and A3 (independence) are well performed by students with different ability levels. In practice these two attributes are possessed by all students. On the opposite attributes A2 (events) and A4 (distribution functions) are extremely difficult for the students. Only attribute A5 (moments) is with a middle level of difficulty.

On the basis of this test difficulties in the calculus of distribution functions and connections between random variables and events can be seen. In practice material related to probability generating functions is not learned.

In the future the test will be expanded using more precisely defined cognitive attributes. This will give the opportunity for a better understanding the teaching process in the course of “Probability theory and Mathematical statistics”.

Remark. The calculations for this study are done by MATLAB based software package available at `http://evaluation.nbu.bg`. The package incorporates functions for estimating IRT parameters as well as LSDM model and related techniques for bootstrapping and test evaluation.

REFERENCES

[1] R. L. Brennan, M. T. Kane. An index for dependability for mastery tests. *Journal of Educational Measurement*, **14** (1977), 277–289.

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

[3] D. Dimitrov. Marginal True-Score Measures and Reliability for Binary Items as Function of their IRT Parameters. *Applied Psychological Measurement*, **27**, No 6 (2003), 440–458.

[4] D. Dimitrov. Least Squares Distance Method of Cognitive Validation and Analysis for Binary Itens Using Their Item Response Theory Parameters. Applied Psychological Measurement, 2007.

[5] F. Lord. Applicationas of Item Response Theory to Practical Testing Problems. Lawrence Erlbaum Ass. Inc., 1980.

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

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

Vessela Stoimenova
Sofia University
Faculty of Mathematics and Informatics
5, J. Bourchier Str.
1164 Sofia, Bulgaria
e-mail: stoimenova@fmi.uni-sofia.bg

Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Acad. G. Bonchev Str., Bl. 8
1113 Sofia, Bulgaria

АНАЛИЗ НА ТЕСТ ПО ТЕОРИЯ НА ВЕРОЯТНОСТИТЕ

Весела Стоименова, Димитър Атанасов

В тази статия анализираме тест по Теория на вероятностите, проведен с цел да се установят най-трудните дялове от преподавания материал. Изследването е проведено, използвайки Теорията за отговор на тестов въпрос (Item Response Theory) и Метода на най-малките квадратични разстояния (Least Squared Distance Method) за оценка на трудността на тестовите въпроси и когнитивните атрибути, необходими за правилен отговор.