Comparative Analysis of Mathematical Models of User Pesponse Time in Knowledge Testing System

Ella Hovhannisyan
NPUA
Yerevan, Armenia
e-mail:
hovhannisyanella@polytechnic.am

***Abstract*— The paper provides a description of the mathematical models of the user's response time to test tasks. The basis for the comparative analysis of the main mathematical models of the user's response time to the tasks was lognormal V. van der Linden's model.**

A description of the Discrete Response Time Model is given, when test tasks are performed by a group of testees. A discrete response time model and a gamma-distributed time model are proposed, since in the case of a log-normal distribution of response time to tasks for a group of testees, it is difficult to obtain a close to reliable estimate of the total time to complete the entire test.

An algorithm is proposed for estimating the parameters of the gamma distribution for each task, assuming that the complexity of the tasks is determined either by an expert or using appropriate algorithms based on the Rash model.

***Keywords*— Knowledge Testing system, test task, universal user (testee), individual characteristics, user's response time.**

I. INTRODUCTION

Nowadays, the introduction of knowledge assessment technologies into the educational process is an important and urgent task. The use of modern knowledge assessment systems, including adaptive testing systems [1], requires mathematical support based on the processing of statistical data resulting from user's work.

Theoretical research in the field of mathematical models describing the performance of testees allows to estimate characteristics such as difficulty of the tasks, and the abilities of the testees. In this paper, the models of the time of response to tasks, which will allow solving the problem of creating time-limited tests, are considered.

The main issues studied in knowledge assessment systems include the dependence between time and the correctness of the user’s response to the task.

With the development of computing systems and the advent of the possibility of their application for the formation of test models and the design of actual test methods, the Item Response Theory (IRT) has emerged. The ideas of IRT were developed by G. Rasch [2-3].

To use these theories in practice, a study of users' behavioral characteristics, such as the speed of completing tasks, the time taken to do so, the skills of the students, etc. There are various approaches to constructing and evaluating the characteristics of tests and users.

II. MODELS OF USER RESPONSE TIME

Modern models of testing theory make it possible to assess the complexity of tasks, the abilities of test takers, and other characteristics. The above-mentioned models were studied, including the models of response time to tasks. Contemporary models of testing theory allow solving the problem of designing tests with time constraints.

One of the main mathematical models of the user’s response time to a task is the Van der Linden Lognormal Model [4]. Its advantage lies in the possibility of obtaining the distribution of the response time of each user to each task according to the available disparate statistics, since not every user solved each task. A well-known disadvantage of this model is the inability to obtain the exact value of the quantile of the total test execution time by the user, which can be represented as the sum of random variables corresponding to the user’s response time to the test tasks.

Using Van der Linden's model to determine the response time of the entire test (not for each task separately) due to the properties of the (random) lognormal distribution, the complexity of determining the total random time of the test arises, so it is recommended, along with the Van der Linden model, to also use models with discrete time distribution and gamma distribution.

In addition to Van der Linden's model, a simplified random-time model of the universal user's response (ϑ_i) to the i-th task is considered.

III. DISCRETE RESPONSE TIME MODEL

The use of a discrete distribution assumes a random response time model ϑ of a certain universal user for a task i (i-th task). In such a model, the individual characteristics of all users are integrated into a single model (universal user) that reflects the characteristics of the entire test group, which allows to propose effective methods for creating a time-limited test for this group.

Under the assumption that an i number of tasks were completed by j number of testees, histograms depicting the time (expressed in seconds) required for those users to answer

https://doi.org/10.51408/csit2023_67

all i tasks were created. On the basis of the histograms, a discrete model of the distribution of response time to the i-th task was created according to statistical data. Histograms are shown in Fg. 1 and Fg.2. The horizontal x-axis shows the time to complete the task in seconds. Given the Pearson's criterion, the Hypothesis of the log-normal distribution of the response time of the universal user for the i-th task was confirmed. The corresponding probability densities are also shown in Fg. 1. and Fg. 2.

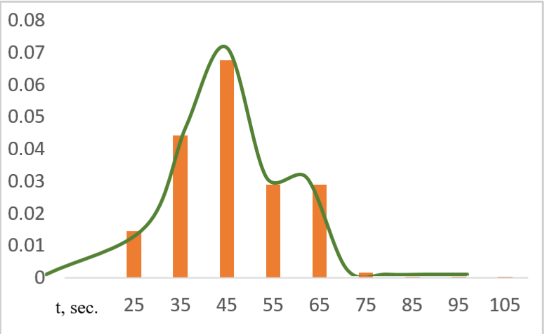

Fig. 1.1 Histogram of task (low level of complexity) execution time

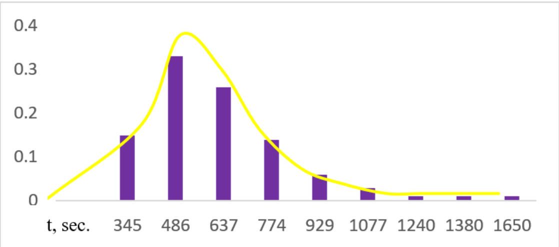

Fig. 1.2 Histogram of task (high level of complexity) execution time

Based on the histograms, discrete response time models were selected.

IV. Gamma Distribution Model

Considering the advantages and disadvantages of the studied models, the time gamma distribution model was taken into account.

The parameters of this model for each task are estimated on the basis of a sample consisting of the values of the time spent on solving this task by specific users. Thus, a model of the response time of a universal user for each task is obtained. The probability densities of lognormal and gamma distributions have similar structures, but it is known that the sum of random variables with a gamma distribution is a gamma-distributed random variable if these random variables have the same parameter ϑ. This provides the ability to find accurate quantiles of the total time spent by the user on the test, as it will have a known gamma distribution.[5-6]

Therefore, the key value for the proposed model will be provided by the original algorithm for selecting the parameters of the gamma distribution of the response time of the universal user to the task proposed below. This algorithm is formulated, so that the distribution parameter ϑ is the same for all tasks, and the value of the second parameter d is determined using the maximum likelihood method.

For each task, the parameters of this model are estimated on the basis of samples (trials), which consist of the values of the time spent by specific test takers on the solution of the given task. In such a way, the time model of the universal user response for each task is obtained. It is known that the sum of random variables with gamma distribution is a gamma-distributed random variable if these random variables have the same parameter ϑ. This makes it possible to find the exact value of the quantile of the total time spent by the user on the test, since it will have a known gamma distribution.

Thus, the basic value for the proposed model will be provided by the algorithm described below, which selects the parameters of the gamma distribution of the universal user task response time.

V. Determination of Gamma Distribution Parameters

The proposed algorithm is to select the parameters of gamma distributions for each task so that the parameter ϑ would be common for all tasks and, at the same time, for the maximum number of tasks with the found estimates of the distribution parameters, the hypothesis of the gamma distribution of the user's response time to this task would be accepted.

Let t_i, i = 1, . . . , I, is the response time of the universal testee to the i-th task, where I is the number of tasks from which the test is formed; t_i^j, j = 1, . . . , I_i, is the implementation of the response time of the j-th testee, spent by him on solving i-th task, where Ii is the number of testees who solved the i-th task.

What follows in this paragraph is a detailed, systematic description of the algorithm for selecting the parameters of gamma distributions of random variables t_i, i = 1, . . . , I.

First, the values of the required parameters and some variables of the algorithm should be reset, followed by assigning $\vartheta^* = 0$, where ϑ^* is the desired value of the gamma distribution parameter, which is the same for all tasks. Set d∗i = 0, where d∗i is the desired value of the second distribution parameter for task i.

Let Q = 0, where Q is the number of tasks for which the hypothesis of a gamma distribution of user response times is accepted.

In the next step let n = 0, where n is a counter. In order to test statistical hypotheses, the confidence level of 1 - β was chosen. For all i = 1, . . . , I using a sample of size Ii, and the maximum likelihood method, the sequential estimates ϑi of the parameter ϑ should be found, ultimately allowing to choose the minimum and maximum values from the set of obtained results. To vary the parameter ϑ, choose the step h=(ϑmax – ϑmin)/L, where L is a preselected number of discretization steps in ϑ.

Let $\vartheta_n = 0$.

Let n = n + 1 and $\vartheta_n = \vartheta_{n-1} + h$:

For each i = 1, . . . , I over the sample t_i^j, i = 1, . . . , I, j = 1, . . . , I_i, an estimate of the second parameter of the gamma distribution ϑn.

$$t_i^j, i = 1,\ldots,I, j = 1,\ldots I_i,$$

$$d_i = \frac{\underline{t_i^j}}{\vartheta_n},$$

where $\underline{t_i^j}$ - is the optional mathematical expectation.

Step three. For all i = 1, . . . , I at the chosen confidence level $1 - \beta$, it is necessary to check the hypothesis H0 - $t_i \sim (d_t, \vartheta_n)$ using the Pearson criterion. If the number of accepted hypotheses Q' is greater than Q, it is assumed that:

$$Q = Q', \vartheta* = \vartheta_n, d*i = d_i, i = 1, . . . , I:$$

$$\vartheta^* = \vartheta_n$$

$$d_i^* = d_i, i = 1, \dots, \mathrm{I}$$

Step four: If $n < P - 1$, then go to step two. After this step, the work of the algorithm ends.

The resulting distribution model allows us to propose an effective algorithm for solving the actual problem of generating a test for a universal user so that its complexity is minimally different from the complexity level specified by the expert and at the same time minimizing the test execution time, which is guaranteed not to be exceeded with a given confidence level. At the same time, it is assumed that the complexity of each task is estimated using the Rasch model [2] based on the processing of statistical data on the work of the testees.

VI. CONCLUSION

The paper proposes the use of a discrete distribution, which involves the construction of a model of a random response time of a universal user to a task. In such a model, the individual characteristics of all users of the tested group are counted integrally when forming the response time model of some universal user. To test the adequacy of the model with the participation of real users, an experiment was conducted.

A model with a gamma distribution of time was considered, the main advantage of which is the distribution density, which is close in structure to the density of the lognormal distribution. As a result, these features of the model will allow us to consider the formulation of the problem of generating a time-limited test with a quantile quality criterion and propose an effective algorithm for solving it.

REFERENCES

[1] Ella Hovhannisyan, “Adaptive Testing in the System of Learning Outcome Assessment”, *Proceedings of International Conference Computer Science and Information Technologies*, Yerevan, Armenia, 2015.

[2] Rasch G. “Probabilistic models for some intelligence and attainment tests“- Chicago, IL, USA: University of Chicago Press, pp.224, 1980.

[3] Hovhannisyan E. H., “A method for Assessing the Learning Outcomes based on the Rush model“, *Bulletin of NPUA*: *Collection of Scientific Papers*, Yerevan, pp. 114-118, 2015.

[4] Van der Linden W.J., Scrams D.J., Schnipke D.L., et al. “Using response-time constraints to control for differential seededness in computerized adaptive testing“. *Appl. Psych. Meas.*, vol. 23. no. 3. pp. 195–210, 1999.

[5] Kan, Yu.S., Application of the Quantile Optimization to Bond Portfolio Selection, in Stochastic Optimization Techniques. Numerical Methods and Technical Applications, Lecture Notes in Economics and Mathematical Systems, Marti, K., Ed., Berlin: Springer, vol. 513, pp. 285–308, 2002.

[6] Kibzun A.I., Kan Yu.S. “Stochastic Programming Problems with Probabilistic Criterions“. Moscow, *Fizmatlit*, pp. 372, 2009.