ADAPTIVE TESTING AS A MARKOV PROCESS: MODELS AND THEIR IDENTIFICATION

L. S. Kuravsky
Department of Computer Science
Moscow State University of Psychology and Education
Moscow, Russia
l.s.kuravsky@gmail.com

G. A. Yuriev
Department of Computer Science
Moscow State University of Psychology and Education
Moscow, Russia
nezdeshni@gmail.com

Abstract — **Presented is a new technology of adaptive testing, which is based on application of trained structures in the form of discrete- and continuous-time models. Its peculiarities, in particular, are revealing and using test solution capability changes in quantitative evaluation of their time-domain dynamics as well as taking into account timetable of testing process. The approach suggested has certain advantages over the testing techniques which were used before owing to its greater information capability and acceleration of test procedure.**

Keywords - Markov models; item response theory; classical test theory; test design; adaptive testing.

I. Introduction

Presented is a new technology of adaptive testing, which is based on application of trained structures in the form of discrete- and continuous-time models. Its peculiarities, in particular, are revealing and using test solution capability changes in quantitative evaluation of their time-domain dynamics as well as taking into account timetable of testing process. The approach suggested has certain advantages over the testing techniques which were used before owing to its greater information capability and acceleration of test procedure.

Computer testing widely used in medicine, psychology and education, for diagnostic, professional skill level estimation and capability of many functions executing, include education quality control. Quality of testing and reliability of its results strongly depends on selected test technology, which became object of scientific researches in last dozens of years. In the beginning of test usage, specialists apply classical test theory, for test design, it was borrowed from physics and based on suggestion that measured characteristics has true mean which are distorted by some system errors. This approach used very widely but have many disadvantages. In 1960 year, Danish mathematician G.Rash, proposed new model for optimal test design, it called one-parametric Rash model[8,12]. Rash model was a great step in test development and application. The main idea of new model is that probability of right answer to test question depends on knowledge or skill level distance for every person and difficulty of test items. Some problems in such models using, like:

-ignoring the fact, that person could tired of the testing process which may cause essential change of test result;

-time spent on item solving doesn't influent to next question selection;

-person need to solve many numbers of test items for correct ability estimation;

-hardness of possible test results probability distribution calculation, which necessary for estimation of its reliability;

-comparatively difficult for practical application procedure of estimation result accuracy, which related with using maximum likelihood method and calculation of confidence intervals.

inspired creation of new adaptive testing technology. Features of this technology is under consideration.

A new model of adaptive testing, based on using trainable discrete and continuous time Markov models. It's main features is:

- Identification and use in constructing estimates of temporal dynamics of ability to solve test item;

- The possibility of accounting in the construction of estimates of time spent on the test items solution;

- Possibility to research the temporal dynamics of knowledge or skills in discrete and continuous time scale;

- Compared to other approaches, the number of items, which should bring the subject to obtain estimates of knowledge or skills with a given accuracy, are lower, it speeds up the testing process;

- Obtaining the probability distribution of possible test results as the outcome;

- Advanced technique of models parameter identification.

II. Logit scale to assess the difficulty of tasks and subject skills

Items difficulty, and the ability of subject measured in a single dimensionless scale of logit [20], which expresses the ratio of proportion for correct and incorrect answers. Convertation to the logit scale is calculated as follows:

$$C = ln\frac{r}{1-r},$$

where C - the value of the logit scale, r - the probability of a correct answer. In case of difficulty assessing this parameter characterizes the performance of specific item for the entire set of subjects and in case of measuring ability - the results of a specific subject for the entire set of admissible items. The real values of logit scale are ranged from -6 to +6.

Statistical approximations of these quantities are obtained after replacing in this formula the probability of r to its sample estimate.

If items of this class can be formally ascribed a substantial set of relevant quantitative indicators of t_i, $i=1,\ldots,M$, we should try to derive the formula of linear regression or other dependencies that allow to estimate the difficulties of items using these parameter. The simplest form of regression dependence is as follows:

$$C = \alpha_0 + \sum_{i=1}^{M} \alpha_i t_i,$$

where α_i, $i=0,1,\ldots,M$, - identified by the regression coefficients, t_i, $i=1,\ldots,M$, - used quantitative parametrs. Estimation of the coefficients α_i performed by least squares method with associated test for significance using statistical measures of fit [16].

III. The structure and mathematical description of applied Markov models with discrete and continuous time. The procedure of knowledge or ability evaluation

A Estimation of probabilities for various ability levels performed on test results using a parametric mathematical models described by Markov processes with discrete states and continuous or discrete time. Directly observable quantity is the difficulty of test executing, measured in logit. The valid range of this quantity is divided into several intervals, each of them is considered as a separate state x_i, $i=0,1,\ldots,n$, in which subject may be with certain probability, transferring from one state to another according to certain rules. The length of these intervals determines the resolution of estimates obtained in the testing process. In turn, the number of states is determined by the desired resolution of estimates and available sample size.

If we denote the upper and lower limit of possible difficulty values range as D_{bot} and D_{top} , the state x_0 will correspond to the interval from D_{bot} to $D_{bot} + \frac{D_{top} - D_{bot}}{n+1}$ state x_1 - will correspond to the interval from $D_{bot} + \frac{D_{top} - D_{bot}}{n+1}$ to $D_{bot} + 2\frac{D_{top} - D_{bot}}{n+1}$ and so on.

Models wich describe the dynamics of these transitions is a directed graphs, where nodes correspond to states and arcs correspond to transitions.

In the case of models with continuous time the testing process can be regarded as a random walk on a graph with transitions from one state to another according to the directions of the arcs. These transitions are instantaneous and occur at random moments.

It is assumed that they satisfy the following two properties of Poisson flows of events:

- ordinary (the flow is called ordinary if the probability of two or more events occurrence during a short interval of time is much smaller than the probability of occurrence one event during the same period);

- independence of the increments (this property means that the number of events falling into two disjoint intervals are independent of each other).

We can show that in these flows the number of events X that fall in any time interval of length τ, beginning at the time t, distributed according to Poisson law:

$$P_{t,\tau}(X = m) = \frac{(a(t,\tau))^m}{m!} e^{-a(t,\tau)},$$

where $P_{t,\tau}(X=m)$ - the probability m of events during the interval, $a(t,\tau)$- the average number of events falling in the interval of length τ, beginning at the moment t. Next we will consider only steady flows (which, $a(t,\tau)=\eta\tau$, $\eta=const$). Parameter η is called the intensity of the stationary flow. It is equal to the average number of events per unit of time. The average time between two adjacent events in this case is equal to $1/\eta$.

The above-mentioned assumptions about the properties of flows of events are common in applications, since these streams (or streams that are close to them on properties) is frequently encountered in practice because of the limit theorems for the flow of events [19].

Under these assumptions the time dynamics of probability of staying in different states described by a system of ordinary differential Kolmogorov equations, in which each state corresponds to the equation

$$\frac{dp_k(t)}{dt} = -\sum_j \eta_{kj} p_k(t) + \sum_i \eta_{ik} p_i(t)$$

where k - number of state, p_k and p_i - the probability to be in k and i states, η_{kj} - intensity of the flows going out of state k; η_{ik} - the intensity of flows going in the state k. For the integration of this system, we must specify the initial conditions: $p_0(0),p_1(0),\ldots,p_n(0)$. The normalizing condition

$$\sum_{k=0}^{n} p_k(t) = 1$$

is satisfied at any time moment.

For models with continuous time, unknown (free) model parameters, are the intensity of the flow of events. Their values are determined by comparing the observed and predicted histograms describing the frequency distribution of stay in the states of the model, namely: compute the values that provide the best match of observed and expected frequencies falling into a certain state of the system at given time points. Predicted probabilities of the states obtained by numerical integration of systems of Kolmogorov equations.

Markov models with continuous time and the free parameters, which are identified from observational data, are called Markov networks [5,6, 7,17,18].

In the case of discrete-time models test process is described by a Markov chain. Transition in each subsequent moment associated with performing the next item. The output of this test is described by the transition probabilities matrix **P**. For this type of models, matrix has the form:

$$\mathbf{P} = \begin{bmatrix} 0 & \pi_0 & 0 & 0 & \dots & 0 & 0 \\ 1-\pi_1 & 0 & \pi_1 & 0 & \dots & 0 & 0 \\ 0 & 1-\pi_2 & 0 & \pi_2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 0 & \pi_{n-1} \\ 0 & 0 & 0 & 0 & \dots & 1-\pi_n & 0 \end{bmatrix}$$

and π_0=1, π_n=0.. Free parameters of the model in this case is the transition probability π_i ($i=0,1,\dots,n-1$).

The probability distribution of residence in the states of the Markov chain in k-th moment of time is expressed by the following equation:

$$p_k = Pp_{k-1} = P^k p_0,$$

where $p_k = (p_{k0}, p_{k1}, \dots, p_{kn})^T$, and the vector p_0 corresponds to the probability distribution at the initial time.

To describe how the probability to be in a given state, changes over time, applied Markov networks and chains, organized by the so-called scheme of "life and death"[1] (**Fig**. 1 and 2). This scheme is a finite chain with $n+1$ states in which transitions from state x_k $(k\neq 0, k\neq n)$ is only possible in the previous state x_{k-1} or the next highest state x_{k+1}. From the states x_0 and x_n are only available state x_1 and x_{n-1}, relatively.

In the case of models with continuous time, probability dynamics of beeing in various states of this scheme is described by the following system of ordinary differential equations:

$$\frac{dp_0(t)}{dt} = -\lambda_0 p_0(t) + \mu_1 p_1(t);$$

....

$$\frac{dp_k(t)}{dt} = -(\lambda_k + \mu_k) p_k(t) + \lambda_{k-1} p_{k-1}(t) + \mu_{k+1} p_{k+1}(t)$$

$$(k = 1,2,\dots,n-1);$$

....

$$\frac{dp_n(t)}{dt} = -\mu_n p_n(t) + \lambda_{n-1} p_{n-1}(t),$$

where $p_k(t)$ is the probability to be in state x_k at time t.

To simplify the problem, as well as providing an acceptable solution to the problem of identification numbers of degrees of freedom for statistics X^2, the intensity of the flow is often assumed to depend on the index k according to certain rules, including a trivial variant: $\lambda_0 = \lambda_1 = \dots = \lambda_{n-1} = \lambda$ and $\mu_1 = \mu_2 = \dots = \mu_n = \mu$. The optimal choice of these dependences based on a technique of statistical hypotheses testing. In case of discrete-time models, similar dependences studing for the transition probabilities.

The procedure of adaptive testing is a consistent presenting items to subject, the difficulty of which determined by the state of Markov network or chain, where he or she is at the moment. If the subject being in state x_i, solves the item, he goes into a state x_{i+1}, otherwise - in the state x_{i-1}. Upon completion of testing, he is in a state x_*, which is the best way appropriate to his level of ability. The principle of selecting the next item is to select the task, Difficulty of which approximately corresponds to the ability of subject. According to observations and results of modern test theory[8,11,12,13], this provides the best differentiation of the subjects in terms of their abilities [10].

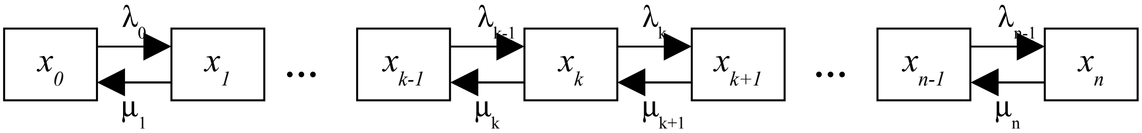
IV. Identification of Markov models with discrete and continuous time

Identification of Markov models are made separately on samples of subjects for each of the considered levels of abilities. Every level of ability C_i, $i=1,\dots,I$, is assigned its own unique set of estimates of model parameters, which allows further identify the value of this characteristics which have the best agreement with observations (see Section 3). Thus, the probability and transitions intensity are functions of two characteristics: the level of ability and difficulty of task. Number of ability levels - is a discrete parameter, which specifies the resolution of characteristics evaluation and setting for each application task according to the sample size of subjects , and the desired result accuracy.

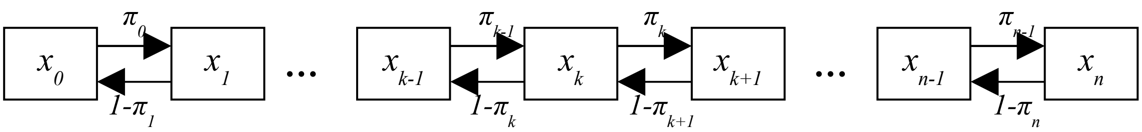
With each changing with time histogram of being in model states relates the Markov process with discrete states. Pearson statistics

$$X^2 = \sum_{k=0}^{n} \frac{(F_k - p_k N)^2}{p_k N},$$

[1] They were first applied in biology to analyze the dynamics of population growth.



<u>Fig. 1.</u> Markov Network representing the testing process with continuous time: x_i ***(i=0,1,…,n)*** – state, λ_i ***(i=0,1,…,n-1)*** and μ_i ***(i=1,2,…,n)*** - the intensity of the transitions.



<u>Fig. 2.</u> Markov Chain representing the testing process with discrete time: x_i ***(i=0,1,…,n)*** – state, π_i ***(i=0,1, …,n-1)*** – the probability of the transitions.

where N - number of elements in the sample, p_k - the predicted probability of hitting the k model state, and F_k - observed frequency of being in k model state, which is used as a measure of conformity in the sense that its higher values mean bad matching predicted and observed results, and small value - a good match. Minimizes the sum of these statistics in those times, for which there are observations. The observed numbers of hits at different intervals of tasks difficulty are determined by results of testing subjects group. As required estimates of free models parameters we use values which provide the best match of observed and predicted frequencies falling into a certain state of the system at given time points.

This method of free parameters identification is called the minimum χ^2 [16] and gives solutions that are close to those obtained by maximum likelihood method [16, pp. 461-462]. Proved that, when a number of general conditions are performed, the values of the Pearson statistic X^2 derived by substituting the true solution is asymptotically described by χ^2 distribution with $n-l$ degrees of freedom where l - the number of parameters to be defined, and the calculated values of free parameters converge in probability to the desired solution with increasing sample size [16, pp.462-470]. This allow to use given statistics for testing hypothesis that the resulting forecast is consistent with observations.

Procedure which is used for parameters calculating consists of two stages. At the preparatory stage we code coded numerical integration scheme with a spreadsheet for the system of differential equations, it allows to calculate the probability function p_k [5,17,18]. These functions are computed by a given time step. To compute solutions with acceptable accuracy were sufficient Runge-Kutta methods or their equivalents.

At the final stage starts the numerical procedure of multidimensional nonlinear optimization[2] [5,17,18], allowing to obtain the desired values of free parameters. The resulting estimates of the free parameters are considered as features of the model revealed by the observations. These criteria also allow you to compare the different versions of Markov models, choosing among them best.

It is shown that the above inverse problem, which boils down to calculating the parameters of differential equations that have solutions which are closest to observed, trainable Markov models in fact act as one of the varieties of neural networks [6,17,18].

V. The possibility of models with discrete and continuous time. Transitions from one type of model to another

It should be noted that the identification of a Markov chain is more convenient than Markov network, because it requires simpler calculations. In turn, the Markov network put on less restriction to testing procedure because of the greater flexibility, using information about the time spent on problem solving during solutions searching, as well as the ability to obtain solutions in case of small number of observation moments.

Transitions from the continuous-time model to the corresponding models with discrete time, and vice versa, are performed in the event that it is expedient to solve an applied problem. Go to the continuous-time model provides by appropriate display of discrete time moments sequence of Markov chain in continuous time axis of Markov network and

[2] It is now proposed a lot of software for solving numerical optimization. In particular, users of the Excel spreadsheet can apply a Frontline Systems, Inc. software [9].

decision identification problem described above. In this case, the modified statistics

$$X_*^2 = N\sum_{k=0}^{n}\frac{(p_{ik}-p_k)^2}{p_k}$$

used as a measure of compliance the forecast and observations, as observed frequencies staying in the model states at the k -th moment of observation time the values $p_{ik}N$ are used, where p_{ik} - components of p_k vector, which describes probability distribution of residence in model state in k -th observation moment, and N - size of subjects sample, which in this case is the parameter of the task.

In the case of the reverse transition, the following procedure is recommended:

1) selection of the sampling interval d for the continuous time axis, which corresponds to a time step of Markov chain;
2) integration the system of differential Kolmogorov equations, which describes the dynamics of changing probability to be in the states of the original Markov network, in order to determine the vectors of probability distribution for the Markov chain $p_i(i = 0,1, ..., l)$, corresponding to specified sequence of $l+1$ discrete time points, taken with a selected sampling interval d;
3) constructing the matrix of transition probabilities $\mathbf{P}=\|p_{ij}\|$ of order $(n+1)\times(n+1)$, where $n+1$ is the number of state for considered Markov chain, which represents the behavior dynamic of this chain and is expressed by free variables, denoting the transition probabilities, and, if necessary, the analytical expressions composed from these variables;
4) numerical solution of multidimensional optimization problem with the least squares criterion

$$\sum_{i=0}^{l-1}(\mathbf{p}_{i+1}-\mathbf{P}\mathbf{p}_i)^T(\mathbf{p}_{i+1}-\mathbf{P}\mathbf{p}_i)$$

with the following restrictions imposed on the unknown free variables are used to express the matrix P:

- $0\le p_{ij}\le 1$ for all i,j
- $\sum_{i=1}^{n} p_{ij} = 1$ for all j.

The calculated values of free variables determine the desired transition probability matrix P. Statistical significance of the solution residual can be estimated using a measure of compliance X_*^2 discussed above, in which the sample size N serves as an additional criterion for deciding significance, namely: for the smallest value X_*^2 we determine N, which corresponds to a significance level $p=0,05$, and estimate how relevant is received value. In the role of the observed frequencies we use estimates obtained using continuous-time model.

VI. Search for optimal solutions

Knowing the state of the model, in which subject is after solving the last item, and calculating the probability of being in this state at a given time for each of the considered levels of abilities we can estimate the probability of staying in specified final state according to Bayes formula:

$$P(C_i \mid S) = \frac{P(C_i)P(S \mid C_i)}{\sum_{k=1}^{I} P(C_k)P(S \mid C_k)},$$

where C_i - an event associated with the presence of the subject i level abilities ($i=1,\dots,I$), S - an event associated with the finding in a given final model state, $P(C_i)$ - a priori probability of the i -th level abilities of subject, $P(S|C_i)$ - the probability of finding a given finite model state in the presence of i -level abilities, $P(C_i|S)$ - the probability of i-level abilities in conditions of stay in a given finite model state.

Ability level at which the largest conditional probability is reached

$$P(C_{max} \mid S) = \max_i\{P(C_i \mid S)\}_{i=1,\dots,I},$$

gives the desired estimate. Probability distribution $\{P(C_i \mid S)\}_{i=1,\dots,I}$, which is the result of solving the item, make possible to evaluate the reliability of this estimate.

As stated in Section 3, the resolution of assessment determined by the length of the interval between the adjacent levels of abilities C_{i+1}-C_i, which, in caes of constancy of these lengths, given by number of ability levels I.

VII. Main results and conclusions

1. The new technology of adaptive testing, based on the use of trainable Markov models with discrete and continuous time is developed.
2. Markov network and chains, organized according to "life and death" scheme are used to describe the dynamics of changes probability to be in the state of the model
3. The main features of this technology are:
- Identification and use in estimates constructing, temporal dynamics of ability to solve test item;
- The possibility of accounting in the construction of estimates of time spent on the test items solution;
- Compared to other approaches, the number of items, which should be presented to subject for obtaining estimates of knowledge or skills with a given accuracy is lower, which speeds up the testing process;
- Advanced technique of identification of free model parameters.
4. The methods for identifying the parameters of Markov models with discrete and continuous time, based on the method of minimum χ^2 and numerical procedure of multidimensional nonlinear optimization is developed.
5. In order to simplify procedures of testing and identification the techniques for transformations discrete time models in to corresponding continuous-time model, and vice versa are developed.

6. Presented approach to adaptive testing has advantages over previously because of its more informative associated with the accounting of time factor influence on the test results, and the acceleration of test process, which is less tiring for the subjects.

References

[1] Baker F.B. The Basics of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD, 2001.

[2] Gregory R.J. Psychological testing: History, principles, and applications (5th edition). - New York: Pearson. 2007

[3] Gulliksen H. Theory of Mental Tests. - John Wiley & Sons Inc, 1950.

[4] Haberman S.J. Analysis of qualitative data: new developments. - N.Y., 1988.

[5] Kuravsky L.S. and Baranov S.N. Condition monitoring of the structures suffered acoustic fatigue failure and forecasting their service life. - Proc. Condition Monitoring 2003, Oxford, United Kingdom, pp. 256-279, July 2003.

[6] Kuravsky L.S. and Baranov S.N. Neural networks in fatigue damage recognition: diagnostics and statistical analysis. - Proc. 11th International Congress on Sound and Vibration, St.-Petersburg, Russia, pp. 2929-2944, July 2004.

[7] Kuravsky L.S. and Baranov S.N. The concept of multifactor Markov networks and its application to forecasting and diagnostics of technical systems. - Proc. Condition Monitoring 2005, Cambridge, United Kingdom, pp. 111-117, July 2005.

[8] Rasch, G. Probabilistic models for some intelligence and attainment tests. // Copenhagen, Danish Institute for Educational Research, expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press. 1960/1980

[9] URL: http:// www.solver.com.

[10] URL: http://www.matlab.mgppu.ru/work/0015.htm.

[11] Wright B.D., Stone M.N.. Best Test Desighn. - Chicago: MESA Press, 1979.

[12] Wright B.D., Masters G.N. Rating scale analysis. Rasch measurements. – Chicago: MESA Press, 1982.

[13] Avanesov V.S. Pedagogical dimension of the latent qualities. - Pedagogical Diagnostics, 2003, № 4, p.69-78.(in Russian).

[14] Zaochnikov B.I., Naidenova N.N., Nikiforov S.V., Chelyshkova M.B. Scaling and alignment of educational outcomes measurements. - M.: Logos, 2003.(in Russian)

[15] Cardan E.J. Modeling and parameterization of tests: basic theory and application. - Federal Centre for Testing, 2008. (in Russian)

[16] H. Cramer, Mathematical Methods of Statistics. - M.: Mir, 1976. - 648. (in Russian)

[17] Kuravsky L.S., Baranov S.N., Malyh S.B. Neural networks in problems of prediction, diagnosis and analysis. - M.: Rusava, 2003. p. 100. (in Russian)

[18] Kuravsky L.S., Baranov S.N. Application of neural networks for diagnosis and prediction of fatigue failure of thin-walled structures. - Neurocomputers: Development and Application, 2001, № 12, pp. 47-63. (in Russian)

[19] Ovcharov L.A. Applied problems of queuing theory. - Moscow: Mashinostroenie, 1969. p. 324. (in Russian)

[20] Tyumeneva Y.A. Psychological dimension. - Moscow: Aspekt-Press, 2007. (in Russian)