PROBABILISTIC ARTIFACT FILTRATION IN ADAPTIVE TESTING

Lev S. Kuravsky, Grigory A. Yuryev

Moscow State University of Psychology and Education, 29 Sretenka Street, 127051 Moscow, Russia
l.s.kuravsky@gmail.com

Abstract

Computerized testing is widely used for diagnostics and estimation of professional skills including CM education quality control. Both quality of testing and reliability of its results depend on a selected test technology that was a scientific research object during last years. Numerous problems following applications of traditional testing techniques inspired creation of the adaptive testing technology under consideration, which is based on application of trained structures in the form of continuous-time Markov 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 a test procedure. The main subject under consideration is elimination of artifacts conditioned by certain forms of illegal purposeful interference in testing procedure. It is carried out on the basis of comparing observed and expected subject responses with the aid of the Kalman filter adapted to the peculiarities of the problem in question.

―――――――――――

Keywords: Adaptive testing, Markov models, Kalman filter.

1. INTRODUCTION

Nowadays computer testing is widely used in medicine, psychology and education for diagnostics and estimation of professional skill level for different activities. That includes also important applications in education quality control. Both quality of testing and reliability of its results strongly depend on a selected technology, which became an object of active scientific research last years.

At first tests were developed on the basis of the classical test theory[2,15,18,19] borrowed from physics. This theory assumes that measured characteristics have some "true" values distorted by random and system errors. This approach was quite popular but has a series of significant defects preventing its practical applications:

 − The problems appear when comparing similar peculiarities of tested subjects, which are revealed with the aid of different methods
 − The validity problem is not solved
 − Test points become insufficiently reliable in the ranges of extreme values
 − The method itself is insufficiently reliable and universal.

A new testing technology was developed to overcome the stated problems. It is based on the latent-structural analysis[4] and called the *Item Response Theory* (*IRT*)[15,17]. The *adaptive testing concept* is implemented in this theory, according to which a tested subject gets the calculated tasks of certain difficulty that depends on his current skill level estimation at each step of the testing procedure. The main concept of this new theory was proposed by G. Rash in 1960[26]. It assumes that

probability of the correct answer for a task is determined by the difference of skill level and test difficulty. Depending on the application problem conditions, practically used are also other models which are more complicated and built on the given concept basis[1,26,28,29].

Application of this technology results in some problems:

- "Static" nature of estimates, viz.: ignoring the essential changes of subject characteristics under study during testing procedure, which may become apparent owing to influence of tiredness and other factors
- Time spent on solving tasks has no influence on skill estimation
- A subject needs to solve rather great number of test tasks to get sufficient accuracy of skill estimation
- Difficulty of calculating probability distributions of possible test results that is necessary for evaluating their reliability
- Procedure of estimating result accuracy, which entails application of both the maximum likelihood method and calculation of confidence intervals and, as a result, is comparatively difficult for practical implementation.

The indicated problems make development of new testing technologies topical. Under consideration in this paper are new aspects of application of the approach to adaptive testing [4-11,20-25], which was developed by the authors before and is based on application of *trained structures in the form of discrete- and continuous-time Markov models*. Its main features are:

- Revealing and using temporal dynamics of ability to solve test tasks in constructing estimates
- Possibility of taking the time spent on solution of test tasks into account in constructing estimates
- Possibility of studying temporal skill dynamics in both discrete and continuous time scales
- The number of tasks, which should be put to the subjects to obtain estimates of their skills with a given accuracy, are less than in case of other approaches, with the testing process being accelerated
- Obtaining the probability distribution of possible test results as the outcome
- Advanced technique of identifying model parameters.

The indicated facilities provide for advantages of the new approach over currently used similar testing techniques.

One of the most serious problems arising from the testing process is appearance of *artifacts* in history of subject responses, which distort testing results. These artifacts are conditioned by outer hints, guessing and other forms of illegal purposeful interference in testing procedure. The presented adaptive testing technology makes it possible to fight against these effects, with eliminating artifacts on the basis of comparing observed and expected subject responses for different skill levels. The *Kalman filter*[14, 16] that is a non-stationary feedback system including a forming filter reproducing the ideal behavior model is suggested here as a tool for such a comparison.

Selection of the Kalman filter for testing artifact elimination is optimal choice since it meets both the accepted adaptive testing concept and relevant context of use better than competitive approaches. In particular, this filter:

- in contrast to the Wiener filter, can process current information on subject response in real time, forming its output after getting next answer immediately without full testing protocol that is not available until completing all the answering procedure
- in contrast to the Stratonovich filter, applies linear estimation methods only, which meet the linear differential adaptive testing model in use best of all and do not complicate the solution process groundlessly
- in contrast to the Luenberger filter, takes into account observation errors and ensures optimal estimation.

Presented in this paper are a new adaptive testing approach based on application of trained structures in the form of continuous-time Markov models, formulation of the artifact Kalman filtration problem and peculiarities of its solution.

2.   MARKOV MODELS OF ADAPTIVE TESTING

2.1. Structure and mathematical description of Markov models in use. Procedure of skill level estimation

Estimation of probabilities for various skill levels is performed basing on test results obtained with the aid of parametric mathematical models described *by Markov random processes with discrete states and continuous or discrete time*[12,13]. Further discussion applies to the models with continuous time only. Directly observable quantity is the difficulty of task being executed, 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 a testee may be with certain probability, transferring from one state to another according to certain rules. The length of these intervals determines the discrimination of estimates obtained in the testing process. In turn, the number of states is determined by the desired discrimination of estimates and available sample size[1].

Both task difficulties and subject' skills are measured in a common dimensionless *logit scale* that represents the ratio of correct and incorrect answers. Conversion to the logit scale is determined by the formula:

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

where C is the required value of the logit scale, r is probability of correct task performance. In case of difficulty assessing this parameter describes the possibility of performance of a certain task for the entire set of subjects and in case of measuring skills - the results of a certain subject for the entire set of admissible tasks. Statistical approximations of the given quantities are obtained after replacing probability **r** by its sample estimates in this formula.

If we denote the upper and lower limits of possible difficulty range as D_{bot} and D_{top}, the state x_0 will correspond to the interval from D_{bot} to $D_{bot}+(D_{top}-D_{bot})/(n+1)$, state x_1 – will correspond to the interval from $D_{bot}+(D_{top}-D_{bot})/(n+1)$ to $D_{bot}+2(D_{top}-D_{bot})/(n+1)$, and so on.

Models which describe the dynamics of these transitions are directed graphs, where nodes[2] correspond to states and arcs[3] correspond to transitions.

In 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 following the arc's directions. These transitions are instantaneous and occur at random moments.

It is assumed that they meet the following two properties of Poisson *flows of events*:
- *Ordinary* (the flow is called ordinary if the probability for occurrence of two or more events during a short time interval is much less than the probability for occurrence of one event during the same period)
- *Independence of the increments* (this property means that numbers of events falling into two disjoint intervals are independent on each other).

---

[1] Considering continuously changing parameter as discrete value we loss a portion of information (it occurs at any idealization). But these losses are insignificant in case of large sample sets when we can set state interval length to be less than measurement error.
[2] Denoted as rectangles.
[3] Denoted as arrows.

It can be shown that the number of events X that fall at any time interval of length τ, beginning at the time t, is distributed according to the Poisson law in the considered flows:

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

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

The aforementioned assumptions concerning the nature of event flows are usual for applications as these flows (or flows which close to them) are frequently take place in reality because of limit theorems for events[12,13].

For continuous-time models, unknown (free) model parameters are events' flow rates. Their values are identified by means of comparing the observed and expected histograms describing the distributions of frequencies for being in the model states, viz.: computed are the rates that provide the best fit for observed and expected frequencies of falling into certain model states at the given time points, in which observed data are available. Expected state probabilities are calculated by integrating the sets of Kolmogorov equations numerically.

Continuous-time Markov models in which free parameters are identified using observation data are called *Markov networks*[5,7,21-23].

To describe how the probabilities of being at the given states vary with time, Markov networks and chains organized according to the so-called "death and propagation" scheme[4] are applied (Figure 1). This scheme is a finite succession of $n+1$ states, in which transitions from state x_k ($k \neq 0$, $k \neq n$) are possible only to the previous state x_{k-1} or to the next state x_{k+1}. From states x_0 and x_n only states x_1 and x_{n-1} are available, correspondingly.
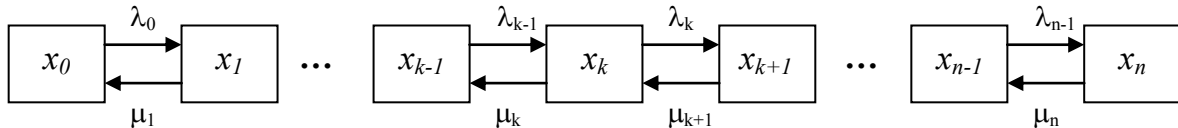

Figure 1. **Markov network representing a continuous-time testing process:** x_i ($i=0,1,…,n$) **are the states,** λ_i ($i=0,1,…,n-1$) **и** μ_i ($i=1,2,…,n$) **are flow rates.**

Probability dynamics of being in various model states for the given scheme is described by the following set of the *Kolmogorov ordinary differential equations*:

---

[4] This scheme was applied firstly in biology to analyze the dynamics of population growth.

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

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

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

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

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

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

where $p_*(t)$ are probabilities of being in states x_* at time t; $*$ is a state number; λ_i ($i=0,1,\ldots,n-1$) and μ_i ($i=1,2,\ldots,n$) are flow rates between the states, which are determined separately for each of the skill levels under consideration. To integrate these equations, one has to assign initial conditions: $p_0(0), p_1(0),\ldots, p_n(0)$. The normalization condition $\sum_{k=0}^{n} p_k(t) = 1$ is valid at any time point.

To simplify the problem and ensure acceptable identification procedure, the flow rates are often assumed to depend on index i according to certain rules, including a trivial variant: $\lambda_0=\lambda_1=\ldots=\lambda_{n-1}=\lambda$ and $\mu_1=\mu_2=\ldots=\mu_n=\mu$. Optimal choice of these dependences is based on the technique of statistical hypotheses testing. In case of discrete-time models, similar dependences are revealed for the transition probabilities.

The procedure of adaptive testing for both model types consists in successive presentation of tasks to a subject, with their difficulty being determined by the state of Markov network, where this subject is at the moment. If the subject being in state x_i completes the task successfully, he goes into state x_{i+1}, otherwise - into state x_{i-1}. Upon completion of testing procedure the subject finds himself in one of states x_*, which meet his skill level best of all. The principle of selecting a next task is to take the test item, whose difficulty meets approximately the current estimation of subject skill. According to the performed studies and modern test theory results, this ensures the best *differentiation* of subjects in terms of their abilities.

2.2. Identification of continuous-time Markov models

Identifications of Markov models are carried out separately for samples of subjects for each of the considered skill levels. Every skill level C_i, $i=1,\ldots,I$, has its own unique set of model parameters estimates. These sets allow later to identify the skill level that meets observations data best of all. Thus, transition probabilities and flow rates are functions of two characteristics: skill level and task difficulty. Number of skill levels is a discrete parameter, which specifies the estimation discrimination for a given characteristic. It is selected for each application problem according to an available subjects' sample size and desired result accuracy.

A discrete-state Markov process is attributed to each time-varying histogram of being in model states. *The Pearson statistics*

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

where N is sample size, p_k is expected probability of being at the *k-th* model state, and F_k is observed frequency of being in *k-th* model state, is used as a goodness-of-fit measure in the sense that its large values correspond to bad fit and its small values correspond to good fit. For model identifi-

cation, the sum of these statistics is minimized in those time points, at which the observation data are available. Observed numbers of hits at different intervals of task difficulty result from testing. The values yielding the best fit of observed and expected frequencies of falling into certain system states at the given time points are considered as required estimates of free model parameters.

It has been proven that under some common conditions the values of the Pearson statistics X^2 obtained after true solution substitution are asymptotically described by χ^2 distribution with $n\text{-}l$ degrees of freedom, where l is the number of parameters to be determined. Moreover, the calculated values of free parameters converge in probability to the target solution with the increase of sample size [3, pp.462-470]. This allows applying the given statistics for verifying the hypothesis that the obtained forecast fits the observation data. The given way of identifying free model parameters is called the *method of χ^2 minimum*[3]. It yields solutions which are close to those obtained by the maximum likelihood technique[3, pp.461-462].

In case of continuous-time models the employed identification procedure consists of two stages. At the preparatory stage, some numerical integration scheme for the aforementioned differential equations is coded with a spreadsheet to calculate the probability functions p_k [5,7,21]. These functions are computed with a given time step. *Runge-Kutta methods* and their equivalents proved to be sufficient to get acceptable accuracy of solution.

At the final stage, a numerical procedure of multidimensional nonlinear optimization[5] [5,7,21] to get required values of free parameters is run. Obtained estimates are considered as model characteristics revealed by the observations. Also the presented criterion allows to compare different variants of Markov models, selecting the best one.

2.3. Calculation of the optimal solution

Knowing a model state, in which the subject is after solving the last task, and calculating the probability of being in this state at the given time for each of the considered skill levels using differential dependencies (see Section 2.1) one can estimate the probabilities of being in the obtained final state with the aid of the Bayes' formula:

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

where C_i is the event of having the i-th subject's skill level ($i=1,\ldots,I$), S is the event of being in the given final model state, $P(C_i)$ is a priori probability of the i-th subject' skill level, $P(S|C_i)$ is probability of being in the given final model state in case of the i-th skill level, $P(C_i|S)$ is probability of the i-th skill level in case of being in the given final model state.

The skill level at which the largest conditional probability is reached

$$P(C_{max}|S) = \max_{i}\{P(C_i|S)\}_{i=1,\ldots,I},$$

yields the required estimate. Probability distribution $\{P(C_i|S)\}_{i=1,\ldots,I}$ obtained as a result of the problem solution makes it possible to evaluate the reliability of this optimal estimate.

As shown in Section 2.1, discrimination of the given estimation is determined by the length of the logit interval between the corresponding adjacent skill levels, which, in its turn, results from the number of skill levels I on the assumption of constancy of such intervals.

---

[5] A lot of software for numerical optimization is available now. In particular, users of the Excel spreadsheet can apply Frontline Systems, Inc. programs.

3. MATHEMATICAL FORMULATION AND SOLUTION OF THE KALMAN FILTRATION PROBLEM IN ADAPTIVE TESTING WITH MARKOV MODELS IN USE

In case of the adaptive testing variant in question an observed process represents the history of being in Markov models' states. It is expressed by the vector $\mathbf{x}(t)=(x_0(t),x_1(t),\ldots,x_n(t))^T$, in which at every time point one and only one of components $x_i(t)$, $i=0,\ldots,n$, that corresponds to the state, where a tested subject is situated, equals to unity, with other components being equal to zero. In its turn, the information process $\mathbf{P}(t)=(p_0(t),p_1(t),\ldots,p_n(t))^T$ under study represents the probability dynamics of being in model states.

Equations of the information and observed processes, which are used for construction of the multivariate continuous Kalman filter for the model type[6] under consideration, can be given in the following way [14, 16]:

$$\frac{d\mathbf{P}}{dt} = \mathbf{FP} \, ,$$

$$\mathbf{x}(t)=\mathbf{P}(t)+\mathbf{v}(t),$$

where the restrictions $\mathbf{E}(\mathbf{v}(t))=\mathbf{0}$ and $\mathbf{E}(\mathbf{v}(t)\mathbf{v}^T(\tau))=\mathbf{R}\delta(t-\tau)$ are imposed on the random observation errors $\mathbf{v}(t)$, the forming filter matrix $\mathbf{F}$ of dimension $(n+1)\times(n+1)$ is

$$\begin{pmatrix} -\lambda_0 & \mu_1 & . & . & . & 0 & 0 \\ \lambda_0 & -(\lambda_1+\mu_1) & \mu_2 & . & . & 0 & 0 \\ . & . & . & . & . & . & . \\ 0 & . & \lambda_{k-1} & -(\lambda_k+\mu_k) & \mu_{k+1} & 0 & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & . & . & \lambda_{n-2} & -(\lambda_{n-1}+\mu_{n-1}) & \mu_n \\ 0 & 0 & . & . & . & \lambda_{n-1} & -\mu_n \end{pmatrix},$$

$\mathbf{R}$ is a symmetric positive-definite matrix, which is supposed further as time-independent one. In practical calculations this matrix can be substituted with one of its sampling estimations $\hat{\mathbf{R}}$ obtained for each of the considered skill levels using observation results.

The Kalman filter differential equation that determines the unbiased estimator $\hat{\mathbf{P}}(t)=(\hat{p}_0(t),\hat{p}_1(t),\ldots,\hat{p}_n(t))^T$ of the process under study[7] with minimum mean square of error $\mathbf{e}(t)=\mathbf{P}(t)-\hat{\mathbf{P}}(t)$, is represented as

$$\frac{d\hat{\mathbf{P}}(t)}{dt} = \mathbf{F}\hat{\mathbf{P}}(t)+\mathbf{K_c}(t)\big(\mathbf{x}(t)-\hat{\mathbf{P}}(t)\big),$$

where $\mathbf{K_c}(t)$ is the matrix amplification factor of a Kalman filter.

In the classical case this factor is given by the equation

---

[6] Peculiarities of these models are: absence of information noise, equality of dimensions for information and observation processes, unity observation matrix.
[7] Kalman filter output.

$$\mathbf{K_c}(t) = \mathbf{U}(t)\mathbf{R}^{-1},$$

in which the error covariance matrix $\mathbf{U}(t) = \mathbf{E}(\mathbf{e}(t)\mathbf{e}^{\mathrm{T}}(t))$ is solution of a certain matrix form of the Riccati equation:

$$\frac{d\mathbf{U}(t)}{dt} = \mathbf{F}\mathbf{U}(t) + \mathbf{U}(t)\mathbf{F}^{T} - \mathbf{U}(t)\mathbf{R}^{-1}\mathbf{U}(t).$$

However, as components of the information process estimation $\hat{\mathbf{P}}(t)$ in the problem under consideration are standardized magnitudes (sum of probabilities of being in the states of a Markov network equals to unity), correction of the amplification factor $\mathbf{K_c}(t)$ is necessary to meet this condition at any time point.

If the normalization condition $\sum_{k=0}^{n} \hat{p}_k(t) = 1$ is valid at the initial time point $t=0$ and the right part of the Kalman filter equation ensures the equality $\sum_{k=0}^{n} \frac{d\hat{p}_k(t)}{dt} = 0$ when $t \geq 0$, the given normalization condition is valid at any time point $t \geq 0$. It is obvious that the condition $\sum_{k=0}^{n} \frac{d\hat{p}_k(t)}{dt} = 0$ is equivalent to zero elements' sum of the vector specified by the matrix expression $\mathbf{F}\hat{\mathbf{P}}(t) + \mathbf{K_c}(t)(\mathbf{x}(t) - \hat{\mathbf{P}}(t))$. Since zero elements' sum of the vector $\mathbf{F}\hat{\mathbf{P}}(t)$ results from the afore-cited structure of the matrix $\mathbf{F}$, zero elements' sum of the vector $\mathbf{K_c}(t)(\mathbf{x}(t) - \hat{\mathbf{P}}(t))$ is necessary and sufficient for zero elements' sum of all the given matrix expression.

Elements' sum of the vector $\mathbf{x}(t) - \hat{\mathbf{P}}(t)$ equals to zero according to conditions of the problem under consideration as components of the above-mentioned vectors are interpreted as probabilities. Taking this fact into account, it is easy to prove that equality of all column elements' sums of the matrix $\mathbf{K_c}(t)$ is sufficient condition for zero elements' sum of the vector $\mathbf{K_c}(t)(\mathbf{x}(t) - \hat{\mathbf{P}}(t))$. Therefore, if one replaces the matrix amplification factor $\mathbf{K_c}(t)$ in the Kalman filter equation by the standardized factor $\mathbf{K_n}(t)$ that is close to it in some metric and has column elements' sums, which are equal to each other, then the condition $\sum_{k=0}^{n} \frac{d\hat{p}_k(t)}{dt} = 0$ is satisfied. The matrix $\mathbf{K_n}(t)$ can be calculated by multiplying the matrix $\mathbf{K_c}(t)$ on the right by the diagonal matrix $\mathbf{D}$, which elements are obtained using the formula

$$d_{jj} = \frac{\sum\limits_{l,m=0}^{n} k_{lm}}{(n+1)k_{*j}},$$

where d_{jj} is the j-th diagonal element of the matrix $\mathbf{D}$; k_{lm}, $l,m=0,\dots,n$, are elements of the matrix $\mathbf{K_c}(t)$; k_{*j} is the sum of elements of the j-th column of the matrix $\mathbf{K_c}(t)$. Such a replacement is correct when the matrix $\mathbf{K_n}(t) = \mathbf{U}(t)\mathbf{R}^{-1}\mathbf{D}$ is within permissible variation limits of the factor $\mathbf{K_c}(t)$, which are caused by sampling errors of the matrix $\mathbf{R}$ estimation. This fact can be checked with the aid of a relevant statistical hypothesis test.

In particular, to do this one can:
- Generate a set of sampling estimations of the covariance matrix $\mathbf{R}$, which correspond to confidence intervals for the given sample size N
- Calculate a sample of matrixes $\{\mathbf{K_{ni}}(t)\}_{i=1,..,M}$ using these estimations
- Calculate a sampling distribution of the Euclidean norm of differences $\{\|\mathbf{K_{ni}}(t) - \mathbf{K_c}(t)\|_E\}_{i=1,...M}$ for the classical and standardized amplification factors
- Taking into account the fact that the obtained sampling distribution in case of sufficiently great number of elements in matrix amplification factors corresponds approximately to the normal one, calculate expectation and variance sampling estimations for it and evaluate the probability p of exceeding the differences' Euclidean norm $\|\mathbf{K_n}(t) - \mathbf{K_c}(t)\|_E$.

If $p \geq 0,05$ then usage of the standardized factor $\mathbf{K_n}(t)$ is acceptable. The method under consideration can be combined with the clustering procedure using Kohonen self-organizing feature maps[9, 25].

In accordance with the presented adaptive testing procedure, the Kalman filtration is carried out autonomously for each of the skill levels which are taken into account in definition of a problem to be solved.

Finally, it should be noted that there are some interesting analogies between the Kalman filter and hidden Markov models[8, 24], which are partially under consideration in the review[27].

4.   SOFTWARE IMPLEMENTATION

The filtration procedure under consideration has been software implemented using the LabVIEW graphical programming software environment (Figure 2), with integration of both the Riccati matrix equation and the Kalman filter equation having been carried out by numerical methods[8]. To estimate initial state of the error covariance matrix $\mathbf{U}(0)$, about which the observations do not yield usually sufficient useful information, the following assumptions were in use:
- $\mathbf{E}(\mathbf{e}(0)) = \mathbf{0}$
- Components of the $\mathbf{e}(0)$ vector of filtration errors are statistically independent
- Variances for components of the $\mathbf{e}(0)$ vector of filtration errors are proportional to the corresponding variances for components of the observed random noise vector $\mathbf{v}(t)$.

---

[8] It is necessary to note that the filtration procedure, in which $\mathbf{U}(t)$ is determined by integration of the Riccati equation, is more correct than the analogous procedures used in significant number of applications, where $\mathbf{U}(t)$ is calculated as a solution of the equation $\mathbf{FU}(t) + \mathbf{U}(t)\mathbf{F}^T - \mathbf{U}(t)\mathbf{R}^{-1}\mathbf{U}(t) = 0$ for the stationary case.
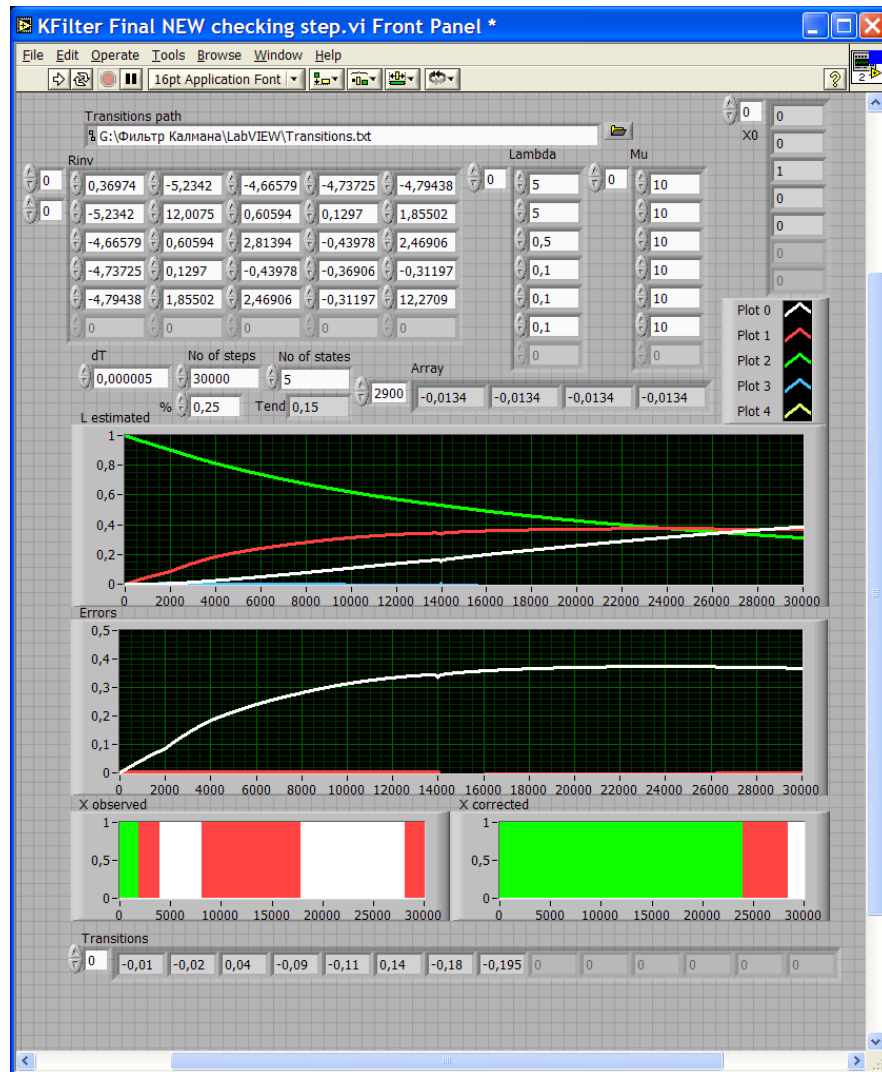
Figure 2. Results of Kalman filtration for the Markov model with 5 states.

5. MAIN RESULTS AND CONCLUSIONS

1. The probabilistic method to filter out artifacts distorted adaptive testing results, which is based on application of trained structures in the form of continuous-time Markov models was developed and software implemented.
2. Elimination of artifacts conditioned by certain forms of illegal purposeful interference in testing procedure is carried out on the basis of comparing observed and expected subject responses for different skill levels with the aid of the Kalman filter adapted to peculiarities of the adaptive testing problem.
3. Selection of the Kalman filter for elimination of artifacts is the optimal approach among the relevant ones since it meets the adaptive testing concept under consideration as well as its application context in the best way.
4. The approach suggested has advantages over the testing techniques, which were used before, owing to its greater information capability, acceleration of a test procedure and firmness with regards to illegal purposeful interference in testing procedure.

Acknowledgements

REFERENCES

1. Avanesov V.S., "Pedagogical measurement of latent qualities", Pedagogical Diagnostics, 2003, No 4, pp 69-78 (in Russian).
2. Kardanova E.Yu., "Modeling and test parameterization: theory foundations and applications", Moscow: FGU «Federal Testing Center», 2008 (in Russian).
3. Cramer H., "Mathematical methods of statistics", Moscow: Mir, 1976, pp 648 (in Russian).
4. Kuravsky L.S., Baranov S.N., "Synthesis of Markov networks for forecasting of fatigue failures", Neurocomputers: Development and Application, 2002, No 11, pp 29-40 (in Russian).
5. Kuravsky L.S., Baranov S.N., "Application of neural networks for diagnostics and forecasting of fatigue failures of thin-walled structures", Neurocomputers: Development and Application, 2001, No 12, pp 47-63 (in Russian).
6. Kuravsky L.S., Baranov S.N., Kornienko P.A., "Trained multifactor Markov networks and their application for studying psychological characteristics", Neurocomputers: Development and Application, 2005, No 12, pp 65-76 (in Russian).
7. Kuravsky L.S., Baranov S.N., Malykh S.B., "Neural networks for forecasting, diagnostics and data analysis: Textbook", Moscow: RUSAVIA, 2003, 100 pp (in Russian).
8. Kuravsky L.S., Baranov S.N., Yuryev G.A., "Synthesis and identification of hidden Markov models for fatigue damage diagnostics", Neurocomputers: Development and Application, 2010, No 12, pp 20-36 (in Russian).
9. Kuravsky L.S., Marmalyuk P.A., Panfilova A.S., Ushakov D.V., "Studying factor influences on psychological characteristics development with the aid of a new approach to estimation of goodness-of-fit measure", Information Technologies, 2011, No 11, pp 67-77 (in Russian).
10. Kuravsky L.S., Yuryev G.A., "Adaptive testing as a Markov process: models and their identification", Neurocomputers: Development and Application, 2011, No 2, pp 21-29 (in Russian).
11. Kuravsky L.S., Yuryev G.A., "Using Markov models in processing testing results", Voprosy Psychologii, 2011, No 2, pp 98-107 (in Russian).
12. Ovcharov L.A., "Applied problems of the queuing theory", Moscow: Mashinostroenie, 1969, 324 pp (in Russian).
13. Saaty Th.L., "Elements of queuing theory with applications", Moscow: LIBROCOM, 2010, 520 pp (in Russian).
14. Tikhonov V.I., Shakhtarin B.I., Sizykh V.V., "Random processes. Examples and tasks. /V.5. Estimation of signals, their parameters and spectra. Information theory foundations", Moscow: Goryachaya liniya – Telecom, 2009, 400 pp (in Russian).
15. Tyumeneva Yu.A., "Psychological measurements", Moscow: Aspect-Press, 2007 (in Russian).
16. Shakhtarin B.I., "Random processes in radio frequency engineering. /V.1. Linear transforms", Moscow: Goryachaya liniya – Telecom, 2010, 520 pp (in Russian).
17. Baker F.B., "The basics of Item Response Theory", ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD, 2001.
18. Gregory R.J., "Psychological testing: History, principles, and applications" (5th edition), New York: Pearson, 2007.
19. Gulliksen H., "Theory of mental tests", John Wiley & Sons Inc, 1950.
20. Kuravsky L.S., Malykh S.B., "Application of Markov models for analysis of development of psychological characteristics", Australian Journal of Educational & Developmental Psychology, 2004, Vol 2, pp 29-40.
21. 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.
22. 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.

23. 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.
24. Kuravsky L.S., Baranov S.N. and Yuryev G.A., "Synthesis and identification of hidden Markov models based on a novel statistical technique in condition monitoring", Proc. 7th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies, Stratford-upon-Avon, England, June 2010, 23 pp.
25. Kuravsky L.S., Marmalyuk P.A. and Panfilova A.S., "Estimation of goodness-of-fit measures for identification of unrestricted factor models employing arbitrarily distributed observed data", Proc. 8th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies, Cardiff, UK, June 2011, 19 pp.
26. 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.
27. Roweis S. and Ghahramani Z. "A unifying review of linear Gaussian models", Neural Computation, Vol 11, No 2, 1999, pp 305–345.
28. Wright B.D., Masters G.N., "Rating scale analysis. Rasch measurements", Chicago: MESA Press, 1982.
29. Wright B.D., Stone M.N., "Best test design", Chicago: MESA Press, 1979.