The Construction of the IRT Profiles Using Fractures that Store the Average

Shumeyko Oleksandr[1] and Iskandarova-Mala Anastasiia[2]

[1] Dniprovsk State Technical University, Dniprobudivska 2, Kamensk, 51900, Ukraine
shumeiko_a@ukr.net
[2] Dniprovsk State Technical University, Dniprobudivska 2, Kamensk, 51900, Ukraine
anastasia.iskandarova@gmail.com

Abstract. IRT profiles scheme using average interpolating polygons. The article deals with the construction characteristics of the aggregate quality of tests using average interpolating linear splines. It was found that the use of splines with free node allows to build an integral characteristic quality of compilation of tests task.

Keywords: IRT profiles of tests task, complexity of the task, differential ability, characteristic curves, splines that interpolate on average, automatic testing.

1 Introduction

1.1 Item Response Theory (IRT)

Student knowledge and skills control is one of the main elements of the learning process. The effectiveness of managing educational work and the quality of the training of specialists depends on the correct organization of the control. Through control, a "feedback" is established between the teacher and the student, which allows assessing the dynamics of learning the learning material, the actual level of knowledge, skills and abilities, and, accordingly, makes appropriate changes to the organization of the learning process. Testing is an important part of knowledge control methods. The testing system is a versatile tool for identifying students' knowledge at all stages of the learning process. In modern conditions, knowledge of testing techniques and the creation of test-bench bases is a necessary component of the teacher's work.

The use of tests as a tool for measuring knowledge implies the presence of certain quality characteristics arising from the theory of test control [3,6]. The theoretical basis for test control is the classical theory of tests and the modern theory of Item Response Theory (IRT). These theories began to emerge in the studies of the late 19th and early 20th centuries in the scientific works of F. Galton, J. Cattell, A. Binet, T. Simon, , E. Thorndike, C. Spearman, H. Gulliksen, L. Guttman, L. Crocker, J. Algina, G. Rasch, A. Birnbaum and others. The steady increase in the number of publications seeking and improving the IRT model indicates the relevance of choosing these models and their widespread use.

Classic model for profile questions (the probability of a respondent with a level of knowledge θ and a correct answer to question with the complexity no higher βj) is considered a two-parameter model of Birnbaum:

$$P(\theta_i, \beta_j) = \frac{e^{D \cdot a_j (\theta_i - \beta_j)}}{1 + e^{D \cdot a_j (\theta_i - \beta_j)}},$$

where D=1.7 is constant, is item discrimination parameter, which determines the slope of characteristic curve. The disadvantage of the model in its practical application is its non-linear dependence on the parameters, and limited "flexibility". IRT is based on mathematical models that differ in visible function $P(\theta_i, \beta_j)$. Based on these models, profiles of the complexity of the questions and the level of students' preparedness are constructed - characteristic curves. Characteristic curves of the test are the main source of information in the IRT, since all other test's scores are derived from them. Characteristic profile of the task are inherent:

1. The complexity of the task, which is determined by the student's preparedness scale at the level of probability of the correct answer $P(\theta) = 0.5$. So the complexity of the task is the median distribution of the probability of the correct answer.

2. Differential ability, which shows how good the task can distinguish students of different levels of knowledge. Differential ability is estimated by the values of the lower and upper limits. The boundaries are determined by the profile: the bottom is at the level of $P(\theta) = 0.25$ and the upper is $P(\theta) = 0.75$. This property is the level of inclination of the characteristic curve of the task in the middle part. Therefore, the higher the inclination, the better the task of the test will be able to distinguish pupils' knowledge levels.

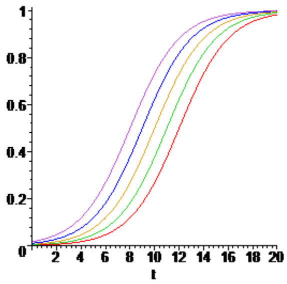


Fig.1. Characteristic curve tasks with the same differential ability, but with different levels of complexity

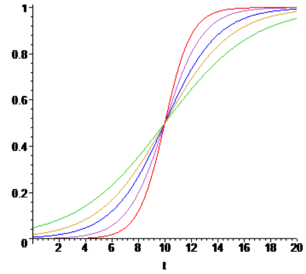


Fig.2. Characteristic curves of tasks with the same level of difficulty, but with

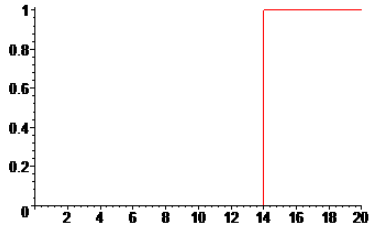


Fig. 3. Characteristic of the task with ideal differential ability

The number of mathematical models in the IRT is constantly increasing, their review appears in scientific periodicals. The reason for this is, first of all, considerable interest in the issues of assessing the quality and reliability of tests in education, as well as the need for the most accurate, reliable and easy to use model. Justifying the disadvantages of parametric models, J. Ramsey, M. Abrahamovich, S. Winsberg, D. Thyssen and G. Weiner (J.O. Ramsay, M. Abrahamowicz, S. Winsberg, D. Thissen, H. Wainer) proposed methods for evaluating characteristic curves that based on the use of spline model [1, 2]. It should be noted that the use of interpolation splines does not always correctly reflect the real characteristics, therefore, it would be advisable to consider spline models based on other approximation models, as in works I. Shelevitsky proposed to use spline regression models [4, 5]. In this paper, as an IRT model, it is suggested to use splines that store the mean of functions, that is, interpolate on average.

2 Methodological Approach

2.1 Preliminaries

Consider the supporting results that are needed in the future.

We denote by $S_r(\Delta_n)$ the set of all polynomial splines of the order r of the minimum defect by partition of $\Delta_n = \{0 = t_0 < t_1 < \ldots < t_n < T\}$, that is, the set of all functions with a continuous $r-1$- th derivative that coincide on each of the intervals (t_{i-1}, t_i) with an algebraic polynomial of degree not higher r.

If for a continuous function $x(t)$ such that there exist

$x^{(\nu)}(z)(z=0,T,\nu=0,1,\ldots,(r-1)/2)$ the spline $s_r(x,\Delta_n,t)\in S_r(\Delta_n)$ is such that for odd r

$$s_r(x,\Delta_n,t_i)=x(t_i),(i=0,1,\ldots,n),$$

$s_r^{(\nu)}(x,\Delta_n,z)=x^{(\nu)}(z)(z=0,T,\nu=0,1,\ldots,(r-1)/2)$, and for paired r, if $t_{i+1/2}=(t_i+t_{i+1})/2,(i=0,1,\ldots,n-1)$ fulfills the conditions of

$$s_r(x,\Delta_n,t_{i+1/2})=x(t_{i+1/2}),(i=0,1,\ldots,n-1),$$

$s_r^{(\nu)}(x,\Delta_n,z)=x^{(\nu)}(z)(z=0,T,\nu=0,1,\ldots,(r-2)/2)$, then it is assumed that the spline interpolates the function $x(t)$ in the nodes of the partition Δ_n, if the equality

$$\frac{1}{h_i}\int_{t_{i-1}}^{t_i}\tilde{s}_r(x,\Delta_n,t)dt=\frac{1}{h_i}\int_{t_{i-1}}^{t_i}x(t)dt$$

is performed, such a spline interpolates on average or is one that preserves the average value of the function $x(t)$.

Theorem 1. Let $x(t)$ be a function that integrates on R^1 and $X(t)$, the antiderivative $x(t)$ is such that $X(0)=0$. In addition, let $s(X,t)$ be a spline that interacts $X(t)$ in nodes t_i $0\le t_0<t_1<\ldots<t_{n-1}<t_n$, that is

$$X(t_i)=s(X,t_i)(i=0,1,\ldots,n) \quad (1)$$

Then spline $s'(X,t)$ preserves the average value of the function $x(t)$ on the $[t_i,t_{i+1}]$ interval.

Proving. Consider the relationship

$$\int_{t_i}^{t_{i+1}}x(t)dt-\int_{t_i}^{t_{i+1}}s'(X,t)dt=\int_{t_i}^{t_{i+1}}(x(t)-s'(X,t))dt=\int_{-\infty}^{\infty}(x(t)-s'(X,t))\chi(t,t_i,t_{i+1})dt, \quad (2)$$

where

$$\chi(t,a,b)=\begin{cases}1, & t\in[a,b],\\ 0, & t\notin[a,b],\end{cases}$$

Heaviside step function. Apply to (1) integration by parts

$$\int_a^b udv=uv\Big|_a^b-\int_a^b vdu,$$

let

$$u(t)=\chi(t,t_i,t_{i+1}) \text{ and } dv(t)=(x(t)-s'(X,t))dt,$$

then

$$du(t)=(\delta(t-t_i)-\delta(t-t_{i+1}))dt$$

Where $\delta(t)$ delta function of Dirac and

$$v(t)=\int_0^t(x(\tau)-s'(X,\tau))d\tau=\int_0^t x(\tau)d\tau-\int_0^t s'(X,\tau)d\tau=X(t)-s(X,t).$$

Then

$$\int_{t_i}^{t_{i+1}} x(t)dt - \int_{t_i}^{t_{i+1}} s'(X,t)dt = \left(X(t) - s(X,t)\right)\chi(t, t_i, t_{i+1})\Big|_{t_i}^{t_{i+1}} -$$

$$- \int_{t_i}^{t_{i+1}} \left(X(t) - s(X,t)\right)\left(\delta(t - t_i) - \delta(t - t_{i+1})\right)dt.$$

Given that

$$\left(X(t) - s(X,t)\right)\chi(t, t_i, t_{i+1})\Big|_{t_i}^{t_{i+1}} = \left(X(t_{i+1}) - s(X, t_{i+1})\right) - \left(X(t_i) - s(X, t_i)\right)$$

then from condition (1) we have

$$\left(X(t) - s(X,t)\right)\chi(t, t_i, t_{i+1})\Big|_{t_i}^{t_{i+1}} = 0 .$$

In addition, because

$$\int_{t_i}^{t_{i+1}} \left(X(t) - s(X,t)\right)\delta(t - t_i)dt = X(t_i) - s(X, t_i)$$

and

$$\int_{t_i}^{t_{i+1}} \left(X(t) - s(X,t)\right)\delta(t - t_{i+1})dt = X(t_{i+1}) - s(X, t_{i+1})$$

then from the condition of interpolation (1) we have

$$\int_{t_i}^{t_{i+1}} \left(X(t) - s(X,t)\right)\left(\delta(t - t_i) - \delta(t - t_{i+1})\right)dt = 0.$$

Thus, $\int_{t_i}^{t_{i+1}} x(t)dt = \int_{t_i}^{t_{i+1}} s'(X,t)dt$, that is, spline $s'(X,t)$ preserves the average value of the function $x(t)$ in the interval $[t_i, t_{i+1}]$ in other words, is interpolated on average.

From the results of work [7] it is not difficult to get the next result.

Theorem A. Let $\alpha = 2/7, \gamma = 0.2\alpha + 3$, then for an arbitrary function $x \in C^4$, the sequence $\left\{\Delta_n^*\right\}_{n=1}^{\infty} = \left\{\left\{t_{i,n}^*\right\}_{i=0}^{n}\right\}_{n=1}^{\infty}$ defined by the conditions

$$\int_0^{t_{i,n}^*} \left(\left|x'''(t)\right| + \frac{1}{n^\gamma}\right)^\alpha dt = \frac{i}{n}\int_0^T \left(\left|x'''(t)\right| + \frac{1}{n^\gamma}\right)^\alpha dt\left(i = 0,1,\ldots,n\right) \tag{3}$$

will be asymptotically optimal for interpolation parabolic splines

$$\left\|x - s_2\left(x, \Delta_n^*\right)\right\|_2 = \inf_{\Delta n}\left\|x - s_2\left(x, \Delta_n^*\right)\right\|_2 \left(1 + o(1)\right) =$$

$$= \frac{\Theta}{n^3}\left\|x'''\right\|_\alpha \left(1 + o(1)\right), \Theta = \left(\int_0^1 \left(\theta(t)\right)^2 dt\right)^{1/2},$$

$$\theta(t) = \frac{1}{24}t^2(1-t)^2.$$

In the case when x''' can equal zero to only the finite number of segments (which is quite a natural condition for many real tasks, including for the purpose of our study), the conditions for choosing nodes can be simplified

$$\int_0^{t_{i,n}^*} |x'''(t)|^\alpha dt = \frac{i}{n} \int_0^T |x'''(t)|^\alpha dt (i = 0,1,\ldots,n).$$

Using the theorem and theory 1, we immediately get the following relation.

Theorem 2. Let $x \in C^3$ be such that x''' can equal zero to only a finite number of segments and a sequence $\{\Delta_n^*\}_{n=1}^{\infty} = \{\{t_{i,n}^*\}_{i=0}^{n}\}_{n=1}^{\infty}$, defined by

$$\int_0^{t_{i,n}^*} |x''(t)|^\alpha dt = \frac{i}{n} \int_0^T |x''(t)|^\alpha dt (i = 0,1,\ldots,n) \tag{4}$$

and $s_2(X, \Delta_n^*)$ as a parabolic interpolation spline for $X(t)$, where $X(t)$ is the initial $x(t)$ such that $X(0) = 0$. Then $\tilde{s}_1(x, \Delta_n^*) = s_2'(X, \Delta_n^*)$ is a broken (spline of a minimal defect of order 1), which preserves the mean value of the function $x(t)$, and the partition Δ_n^* is asymptotically optimal.

2.2 Construction of IRT Spline Profiles

Let's turn to the main results of this research. Consider the process of forming respondent responses to test questions. In this case, we have two a priori unknown values that characterize the test question and the respondent, namely, the level of difficulty of the question and the level of knowledge of the respondent. If one of the parameters θ or β is locked, the evaluation task *P(θ,β)* reduces to the determination of dependence on one of the parameters: *P(θ)* - question profile, or *P(β)* - profile of the respondent. Let's assume that *θ* and *β* in the experiment process (testing) remain unchanged, then one can find probability estimates associated with *θ* and *β*. Assume that the result of the response of the *i*-th respondent to the *j*-th task is equal to *ri,j*, where *ri,j=1*, if the answer is correct (but we can use a weighted estimate of *ri,j>0*), in the opposite case, is 0.

$$\hat{\theta}_i = \frac{1}{M} \sum_{j=1}^{M} r_{i,j},$$

and the assessment of the difficulty level of the test is equal

$$\hat{\beta}_i = 1 - \frac{1}{N} \sum_{i=1}^{N} r_{i,j},$$

where M is the number of test tasks and N is the number of respondents.

Given that a respondent with a higher level of knowledge is correct on probabilities with probability no less than a respondent with a lower level of knowledge, we have

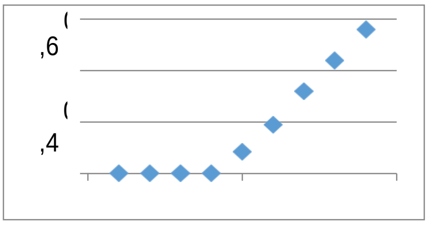
$$P(\theta_i) \geq P(\theta_k),\ if\ \theta_i > \theta_k.$$

This implies the non-declining nature of the $P(\theta_i)$ dependence for a fixed level of complexity of the question *β*. That is, $P(\theta_i,\beta)$ is the characteristic profile of the task. Proceeding from this, $P(\theta_i,\beta)$ is a cumulative probability curve, each point of which corresponds to the probability that the respondent with a knowledge level not greater than θ gives the correct answer to the question with the level of complexity β. Thus, the estimates of the characteristic curve of the question are defined as

$$P_j(\theta) = \frac{1}{N}\sum \{r_{i,j} | \theta_i \le \theta\}.$$

Since we have only a set of $\hat{N}$ ratings θ, we obtain $\hat{N}$ empirical points

$$\hat{P}_j(\theta_k) = \frac{1}{\hat{N}}\sum \{r_{i,j} | \theta_i \le \theta_k\}, k = 1,\ldots,\hat{N}.$$



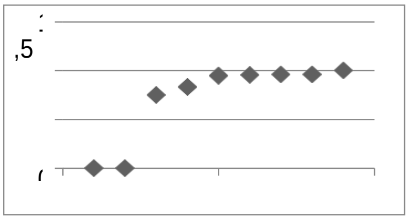


Fig.4.Behavior $\hat{P}_j(\theta_k)$ for various test tasks I and II.

Here is an algorithm for constructing an IRT spline profile, which stores an average value.

Let there be a plural $\hat{P}_j(\theta_k), k = 1,\ldots,\hat{N}$. We denote by

$$I\left(\hat{P}_j(\theta_k)\right) = \frac{1}{\hat{N}}\sum_{\nu=0}^{k} \hat{P}_j(\theta_\nu), k = 1,\ldots,\hat{N}$$

the discrete analogue of the antiderivative $\hat{P}_j(\theta_k)$.

We denote by $S\left(I\left(\hat{P}_j(\theta_k)\right), t\right)$ the parabolic spline of the minimal defect with two free nodes $t_1, t_2 \in (0,1)$, such that it is determined by the conditions

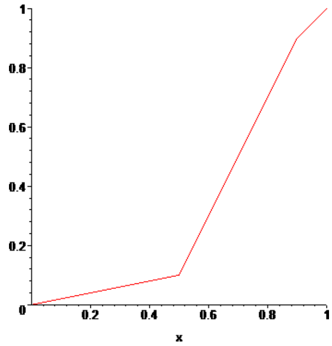
$$S\left(I\left(\hat{P}_j(\theta_k)\right), 0\right) = 0, \quad S'\left(I\left(\hat{P}_j(\theta_k)\right), 0\right) = 0,$$

and the nodes t_1, t_2 are determined by the conditions (4), where $X = \hat{P}_j(\theta_k)$.

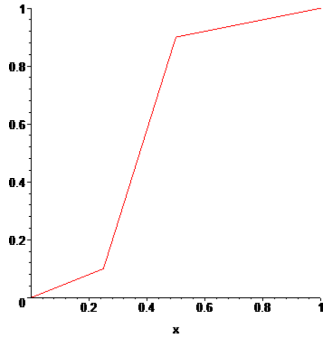
According to Theorem 2, the derivative $S'\left(I\left(\hat{P}_j(\theta_k)\right), t\right)$ s a linear spline, which preserves the mean value.

The nodes of the optimal partition $\left(t_1, S'\left(I\left(\hat{P}_j(\theta_k)\right), t_1\right)\right)$ and $\left(t_2, S'\left(I\left(\hat{P}_j(\theta_k)\right), t_2\right)\right)$ characterize the behavior of the IRT profile. The smaller the value of $|t_2 - t_1|$ and more $\left|S'\left(I\left(\hat{P}_j(\theta_k)\right), t_2\right) - S'\left(I\left(\hat{P}_j(\theta_k)\right), t_1\right)\right|$, the better the differential function of the problem, the higher the value of , the greater the complexity of the task. The ideal task meets the conditions 1t $t_1 = t_2 = 0.75$ and $S'\left(I\left(\hat{P}_j(\theta_k)\right), t_1\right) = 0, S'\left(I\left(\hat{P}_j(\theta_k)\right), t_2\right) = 1.$

So for $\hat{P}_j(\theta_k), k = 1,\ldots,\hat{N}$ we have



For the test task I (see Fig. 4-I)



for the task II (see Fig. 4- II)

Thus, an aggregated characteristic is obtained, by which it is possible to automate the process of assessing the quality of test tasks and to analyze the complexity of a particular test task (as in the example given in Fig. 4, task I is compiled rather qualitatively, and the task II is too simple).

3 Conclusion

Using an IRT model based on splines that interpolate on average allows to obtain an aggregate characteristic of the assessment of the quality of test tasks, which allows for automatic testing of test quality.

References

1. Abrahamowicz M.: Multicategorical spline model for item response. Psychometrika 57, 5-27 (1992).
2. Winsberg, S., Thissen, D., & Wainer, H.: Fitting item characteristic curves with spline functions. Educational Testing Service, 84-52 (1984).
3. Linda, C., and A. James, T.: Introduction to classical and modern test theory. Logos (2010).
4. Duban, R., Shelevytskyy, I.: Spline models in test measuring knowledge. Mother School 7-8 (967-968), 11-14 (2010).
5. Duban, R., Shelevytskyy, I.: Comlpex issues and respondent knowledge profile spline models in knowledge testing. Management Information System and Devices 156, 71-77 (2011).
6. Chelyshkova, M.: Theory and Practice of Designing Pedagogical Tests. Logos 432 (2002).
7. Shumeiko, A.: On the choice of nodes for interpolation parabolic splines. High schools Mat. 4, 67-71 (1990).
8. Shumeyko, A, Knurenko, V.: Building IRT profiles by using splines that retain Average value. Mathematical modeling 1 (32), 28-32 (2015).
9. Shumeiko, A, Iskandarova, A.: The contruction of the IRT profiles by using piecewise linear regression whit free nodes. Modeling in industry, transport and education 11, (2016).