

В.М. Карнаухов

Московский государственный университет природообустройства, Москва, Россия

МОДЕЛЬ РАША КАК ИГРОВАЯ МОДЕЛЬ

Известная модель тестирования Раша построена как игровая модель, в которой участники тестирования, выполняющие задания, и задания теста играют равноправную роль. Адекватность модели Раша подтверждена на примере шахматной классификации А. Эло. Дано «игровое» обоснование известного метода первичных баллов, метода оценивания латентных параметров тестирования по результатам наблюдения тестирования, который обладает наивысшей скоростью вычисления оценок.

Ключевые слова: модель Раша, классификация Эло, адекватность модели, метод первичных баллов, латентные параметры, уровень подготовленности, уровень трудности.

Предпосылки возникновения модели Раша

Перед тем, как рассматривать модель тестирования Раша [1, 2], хотелось бы поговорить о некоторых предпосылках возникновения этой модели. Эти предпосылки связаны с построением спортивных классификаций в таких видах спорта, как шахматы, большой теннис. В этих видах спорта принята не разовая оценка выступления спортсменов, а интегральная, по результатам прошедших состязаний. Остановимся на системе классификации на основе рейтинга.

Оставим пока открытым вопрос о том, какие соображения учитываются при присвоении игрокам, входящим в классификацию, исходного рейтинга (исходного места в классификации). Предположим, что этот вопрос тем или иным способом решен.

Рассмотрим встречу двух игроков. Обозначим оценки класса, т.е. рейтинги игроков через $r(U)$ и $r(V)$ соответственно. Введем в рассмотрение переменную величину t , характеризующую различие в классе игроков U и V . Величину t можно предположить зависящей, например, от отношения $r(U)/r(V)$ рейтингов игроков U и V или же от их разности $r(U) - r(V)$.

В теории классификации шахматистов известна таблица, предложенная американским

профессором А. Эло, по которой пересчитываются текущие рейтинги игроков. Она выглядит так, как показано в таблице.

В этой таблице t – разность рейтингов; h_0 и h_m – проценты выигрыша шахматистов с большим и соответственно с меньшим рейтингом. Полная таблица приведена, например, в [4] и [5]. Данная таблица построена на богатом статистическом материале различных шахматных состязаний.

Пересчет рейтинга после очередного состязания осуществляется по формуле:

$$r_n = r_{cm} + \mu(N - N_{ож}),$$

r_{cm} – старый рейтинг участника соревнований,

r_n – новый рейтинг участника соревнований,

N – число очков, фактически набранных участником,

$N_{ож}$ – ожидаемое число очков, которое ему «полагается» в силу его квалификации,

μ – коэффициент, равный числу, на которое возрастает рейтинг участника в случае, если он набирает на очко больше сверх ожидаемого количества (в системе Эло этот коэффициент равен 10).

Ожидаемое количество $N_{ож}$ рассчитывается по формуле

$$N_{ож} = \sum_{k=1}^{n-1} \frac{h_k}{100},$$

Таблица А. Эло для пересчета рейтинга шахматистов

t	h_0	h_m	t	h_0	h_m	t	h_0	h_m
0–3	50	50	180–188	74	26	518–559	97	3
4–10	51	49	189–197	75	25	560–619	98	2
11–17	52	48	198–206	76	24	620–735	99	1
...	...	...	...	...	...	>735	100	0

где n – число участников однокругового турнира по шахматам,

h_k – проценты выигрыша данного участника у k -го шахматиста турнира.

Можно построить математическую модель встречи двух шахматистов в одной партии, подтверждающую накопленный статистический опыт, выраженный в таблице Эло. Предположим, что отношение m/n среднего числа побед игрока U над игроком V к среднему числу его поражений в серии из N встреч находится в экспоненциальной зависимости от разности рейтингов игроков U и V .

Итак, принимаем, что

$$m/n = a^t,$$

где $a > 1$, $t = r(U) - r(V)$.

Отсюда получаем, что вероятность $p(t)$ выигрыша игрока U у игрока V равна

$$p(t) = \frac{Np(t)}{N} \approx \frac{m}{m+n} = \frac{m/n}{1+m/n} = \frac{a^t}{1+a^t}.$$

Теперь вся проблема построения классификации сводится к выбору числового значения параметра a . Статистика шахматных турниров свидетельствует о том, что если один из соперников на один разряд (на одну ступень) в шахматной иерархии стоит выше другого, то первый выигрывает у второго в среднем 75 очков из 100 возможных, т.е. с вероятностью, равной 0,75. Если учесть, что в системе Эло различие между игроками, принадлежащими к двум соседним разрядам шахматной иерархии, составляет $\lambda = 200$ единиц рейтинга, то получим следующее равенство для определения числа a :

$$\frac{a^\lambda}{1+a^\lambda} = 0,75.$$

Откуда находим, что $a_0 = 1,0055$.

Теперь имеем конкретную математическую модель встречи двух шахматистов, выраженную в формуле вероятности P выигрыша игрока U у V :

$$P = \frac{1}{1 + a_0^{-(r(U)-r(V))}}.$$

Если по этой формуле вычислить вероятности для различных значений t , то получим таблицу Эло (см. таблицу). Это совпадение говорит о том, что статистические данные подтверждают справедливость построенной математической модели шахматной игры.

Модель Раша как модель игры

Система игры в шахматы тесно связана с системой решения задания теста участником тестирования. Действительно, решение задания можно интерпретировать как состязание двух соперников с заданными рейтингами (см. ниже). Более того, теория тестирования (Раш Г., 1968–1970 гг.) развивалась параллельно с теорией шахматной классификации (Эло А., 1963–1970). Перейдем к изложению теории тестирования.

Очевидно, что вероятность того, что участник тестирования с уровнем подготовленности s правильно решит определенное тестовое задание с уровнем трудности t , по крайней мере зависит от двух аргументов: $P = P(s, t)$. Эту функцию будем называть функцией успеха. Переменные s и t принято называть латентными (ненаблюдаемыми) параметрами, поскольку они призваны описывать некоторые скрытые характеристики участников тестирования и тестовых заданий.

Можно заменить систему тестирования «участник – задание – решение» системой игры «участник 1 – участник 2 – игра», т.е. представить участника и задание полноправными участниками 1 и 2 соответственно некоторой парной игры (например, шахматы). Можно считать задание участником парной игры, который выигрывает (в случае неверного решения задания) или проигрывает (в случае верного решения задания) своему оппоненту.

Можно определить уровень подготовленности участника следующим образом.

Определение. Число $t > 0$ называется уровнем подготовленности участника, если для любых двух участников с уровнями подготовленности t_1 и t_2 соответственно выполняется соотношение

$$\frac{t_1}{t_2} = \frac{P}{1-P},$$

где P – вероятность выигрыша участника 1 у участника 2.

Из этого определения вытекают следующие свойства:

$$1) \frac{t_1}{t_2} = \lim_{N \rightarrow \infty} \frac{m}{N-m},$$

где m – число выигранных партий участника 1 у участника 2 в общей серии из N партий.

Действительно,

$$\frac{t_1}{t_2} = \frac{P}{1-P} = \frac{\lim_{N \rightarrow \infty} \frac{m}{N}}{1 - \lim_{N \rightarrow \infty} \frac{m}{N}} = \lim_{N \rightarrow \infty} \frac{m/N}{1 - m/N} = \lim_{N \rightarrow \infty} \frac{m}{N-m}.$$

2) Игровой смысл параметра t : значение параметра t приближенно равно отношению числа выигранных партий у участника с уровнем подготовки $t=1$ (стандартного участника) к числу проигранных партий этому участнику.

3) $t \in (0, \infty)$.
 Действительно, $t = \frac{P_1}{1 - P_1}$,

где P_1 – вероятность выигрыша у стандартного участника.

Тогда свойство вытекает из графика зависимости t от P_1 (рис. 1).

4) $P = \frac{1}{1 + t_2 / t_1}$.

5) $t_1 = t_2$ тогда и только тогда, когда $P = 0,5$,
 $t_1 > t_2$ тогда и только тогда, когда $P > 0,5$,
 $t_1 < t_2$ тогда и только тогда, когда $P < 0,5$.

Свойство 4 послужило основой для математической модели тестирования датского математика Георга Раша:

$$P = \frac{1}{1 + \frac{t}{s}}.$$

Далее датский математик решил изменить единицы измерения, заменив действительную полуось $(0, \infty)$ на симметричную относительно 0 действительную ось $(-\infty; \infty)$. Для этого он использовал формулы перехода:

$$\theta = \ln(s), \quad \delta = \ln(t).$$

Новая единица измерения была названа логитом. Прежнюю единицу в силу игрового смысла параметра t предлагается назвать игровым логитом.

Формула для функции успеха тоже изменилась:

$$P = \frac{1}{1 + \frac{e^\delta}{e^\theta}} = \frac{1}{1 + e^{-(\theta - \delta)}}.$$

Формулу

$$P = \frac{1}{1 + e^{-(\theta - \delta)}} \quad (1)$$

называют основной логистической моделью Раша. Вероятность успеха зависит, по существу, от одно-

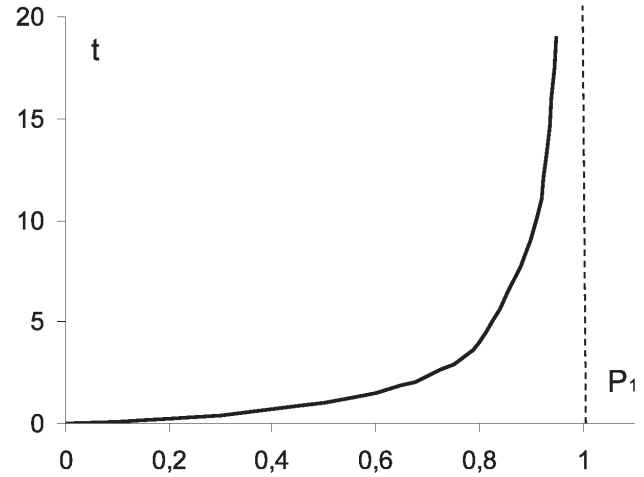


Рис. 1. Зависимость уровня трудности t от вероятности выигрыша у стандартного участника

го параметра – разности $(\theta - \delta)$, и поэтому эта модель является однопараметрической. При этом

1) $\lim_{(\theta - \delta) \rightarrow +\infty} P = 1$.

2) $\lim_{(\theta - \delta) \rightarrow -\infty} P = 0$.

3) $P = 0,5$, если $\theta = \delta$.

Приведем рассуждения по поводу реального диапазона изменения единицы измерения логита. Исходя из практической целесообразности, вероятность наступления некоторого события вычисляется с точностью до 3 знаков после запятой. Поэтому вероятность, равную 0,9999, можно принять за вероятность достоверного события. Решая уравнение

$$0,9999 = \frac{1}{1 + \exp[-(\theta - \delta)]}$$

относительно величины $(\theta - \delta)$, получим, что $\theta - \delta = 9,21$. Задача будет решена достоверно, если учащийся имеет максимально возможный уровень подготовки $\theta = \delta_{\max}$, а задание – минимально возможный уровень трудности $\delta = \delta_{\min}$. Учитывая, что измерение уровней подготовки и уровней трудности осуществляется на одной симметричной шкале логитов ($\delta_{\min} = -\theta_{\max}$), приходим к равенству: $2\theta_{\max} = 9,21$. Отсюда получаем $\theta_{\max} = 4,6$, что позволяет считать, что значения латентных параметров реально меняются в пределах от $-4,6$ до $4,6$. В работе [2] этот

интервал расширен до интервала $(-5; 5)$. Поэтому далее будем считать, что $\theta_{\max} = 5$ и $\theta_{\max} \in [-\theta_{\max}; \theta_{\max}]$. В силу этих рассуждений можно ввести в рассмотрение третью единицу измерения уровня подготовленности, более понятную с точки зрения пользователя, так называемый процентный логит. Связь между шкалой логитов и 100%-ной шкалой может быть осуществлена по формуле

$$T = \frac{\theta + \theta_{\max}}{2 \cdot \theta_{\max}} \cdot 100\%, \quad (2)$$

где θ – уровень подготовленности по шкале логитов, T – уровень подготовленности по 100%-ной шкале (или тестовый балл), $\theta_{\max} = 5$.

Далее вероятность P как функция δ при фиксированном значении уровня подготовленности $\theta = \theta_0$ полностью описывает потенциальные возможности индивидуума с уровнем подготовленности θ_0 при выполнении заданий всевозможных трудностей δ и потому называется характеристической функцией уровня подготовленности θ_0 . Вероятность P как функция θ при фиксированном значении $\delta = \delta_0$ с той же полнотой характеризует возможности участников тестирования с различным уровнем подготовленности при решении задания трудности δ_0 и называется характеристической функцией трудности δ_0 . Графики характеристических функций называются соответствующими характеристическими кривыми.

В рамках логистической модели (1) абсциссы $\delta = \theta_0$ на рис. 2 и $\theta = \delta_0$ на рис. 3 являются корнями уравнений $\frac{\partial^2 P}{\partial \delta^2} = 0$ и $\frac{\partial^2 P}{\partial \theta^2} = 0$ соответственно

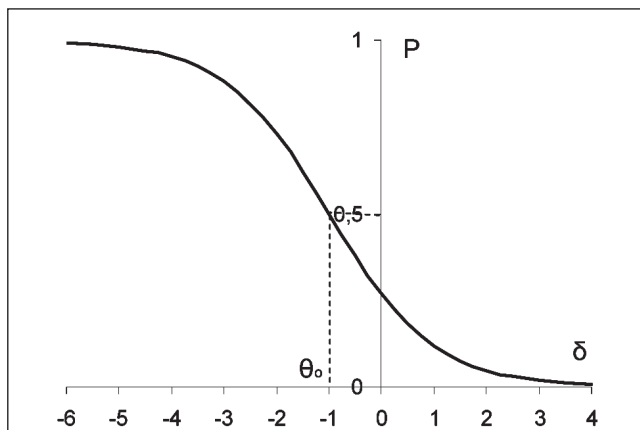


Рис. 2. Характеристическая кривая уровня трудности задания δ_0

и определяют точки перегиба характеристических кривых. Ординаты точек перегиба равны 0,5.

Заметим, что характеристическая кривая уровня подготовленности θ_0 представляет собой стандартную характеристическую кривую уровня подготовленности 0, сдвинутую (без деформации) вдоль оси абсцисс на θ_0 единиц. Это означает увеличение (если $\theta_0 > 0$) или уменьшение (если $\theta_0 < 0$) вероятностей успешного выполнения индивидуумом с $\theta = \theta_0$ заданий всевозможных трудностей. Следовательно, характеристические кривые, соответствующие различным уровням подготовленности, не пересекаются. Это выгодно отличает функцию (1) от других функций успеха, где пересечение возможно.

Аналогичная ситуация имеет место и при параллельном переносе характеристической кривой трудности δ_0 .

Функция успеха (1) принадлежит известному в теории вероятностей семейству логистических функций распределения вероятностей вида

$$F(x) = (1 + e^{-dx})^{-1}$$

с дисперсией $D(x) = \frac{1}{3} \left(\frac{\pi}{d} \right)^2$, где d – параметр.

Точка перегиба всех соответствующих кривых имеет координаты $(0; 0,5)$, но крутизна кривых в точке перегиба пропорциональна величине параметра d , поскольку $F'(0) = d/4$. Этот факт делает целесообразным заменить однопараметрическую функцию успеха (1) двухпараметрической моделью вида

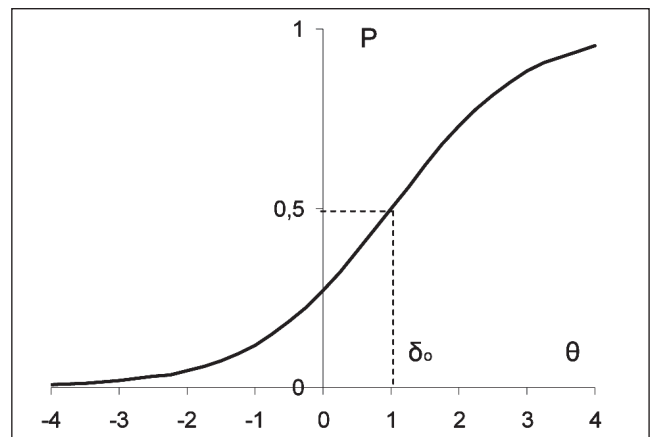


Рис. 3. Характеристическая кривая уровня подготовленности θ_0

$$P = \frac{1}{1 + e^{-d(\theta - \delta)}}. \quad (3)$$

В рамках этой модели каждому тестовому заданию определенной трудности $\delta = \delta_0$ соответствует не единственная характеристическая кривая, как для однопараметрической модели (см. рис. 3), а семейство кривых, пересекающихся в единственной точке перегиба ($\theta = \delta_0; 0,5$) (рис. 4). В этой точке крутизна каждой кривой семейства, т.е. величина угла наклона касательной, равна $d/4$. Поэтому значение параметра d является дополнительной характеристикой задания. Большее значение d является предпочтительным. Дело в том, что при малом d характеристическая кривая трудности δ_0 является пологой. На практике это означает, что участники тестирования с хорошим и плохим уровнем подготовки выполняют задание с приблизительно равным успехом. Наоборот, если d велико, то шансы успешного выполнения данного задания участниками с $\theta < \delta_0$ и $\theta > \delta_0$ отличаются существенно. Поэтому параметр d принято называть дифференцирующей, или дискриминационной, способностью тестового задания.

Двухпараметрическая модель (3) (латентные параметры $\theta - \delta$ и d) носит имя Бирнбаума [3]. Понятно, что при $d = 1$ логистическая модель Бирнбаума совпадает с логистической однопараметрической моделью Раша (1).

При $d = 1,7$ функция (3) хорошо аппроксимирует функцию распределения вероятностей нормального закона с нулевым математическим ожиданием и единичной дисперсией (см. [1–3]). А именно, если $X = N(0,1)$, то функция распределения этой случайной величины имеет вид

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-x^2/2} dx.$$

$$\text{Если } P(\theta, \delta_0) = \frac{1}{1 + e^{-1,7(\theta - \delta_0)}},$$

тогда справедливо неравенство

$$|F_X(1,7(\theta - \delta_0)) - P(\theta, \delta_0)| < 0,01$$

для любого $\theta \in (-\infty; +\infty)$

Этот факт позволяет различные вычисления с логистической моделью интерпретировать, при необходимости, с позиции детально изученного нормального закона распределения вероятностей.

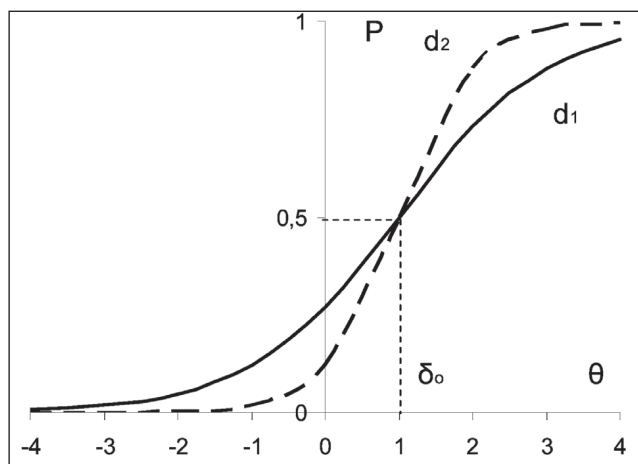


Рис. 4. Характеристические кривые трудности δ_0 задания с разными значениями параметра d ; $0 < d_1 < d_2$

Двухпараметрическая модель Бирнбаума имеет все основания для существования в теории тестирования, поскольку переход от шкалы игрового логита к шкале логитов может быть осуществлен при помощи логарифма по любому основанию, большему 1 (функция перехода должна быть возрастающей). Если осуществить такой переход, мы придем к исследованию двухпараметрической модели Бирнбаума. Модель (3) имеет право на существование хотя бы потому, что, как было показано выше, в шахматной классификации Эло использовалось основание $a_0 = 1,0055$, а не $a = e$.

Метод первичных баллов оценки латентных параметров тестирования

Основным приложением математической модели Раша на данный момент остается электронное тестирование различных уровней: от проведения локальных олимпиад [6, 7] до Централизованного тестирования (ЦТ) и Единого государственного экзамена (ЕГЭ). Правда, в 2011–2013 гг. методика шкалирования, описанная, например, в приложении к приказу ФИПИ [8] (или на сайте официального информационного портала Единого государственного экзамена: <http://ege.edu.ru/ru/main/scaling/>), не использовала модель Раша. Эта методика будет исследована на предмет точности в следующих работах. Здесь же мы изложим основные положения методики, основанной на понятии уровня подготовленности участника тестирования. Эта методика в соответствующей литературе носит название метода первичных

баллов, который имеет наивысшую скорость вычисления оценок.

Предположим ниже, что:

а) тест состоит из M заданий;

б) все задания оцениваются целыми оценками, причем m_j – максимальный балл при решении j -го задания;

в) максимальное число набранных баллов при решении теста равно K , т.е.

$$K = \sum_{j=1}^M m_j;$$

г) в тестировании принимают участие N участников тестирования.

Опишем основные этапы этой методики.

1. Сбор и обработка результатов всех выпускников.

2. Определение общего количества баллов, набранных при решении каждого задания теста ($c_1 \dots c_M$) и общего количества N_i человек, набравших конкретное количество первичных

баллов $i=0, \dots, K$, при этом $N = \sum_{i=0}^K N_i$.

3. Оценивание латентных параметров тестирования (уровней подготовленности участников тестирования $\theta_0 \dots \theta_K$ и уровней трудности заданий $\delta_0 \dots \delta_K$) на основании собранных данных при помощи метода первичных баллов (см. ниже)

4. Перевод значений подготовленности $\theta_0, \dots, \theta_K$, расположенных на шкале логитов, в тестовые баллы $T_0, \dots, T_K$, расположенные на шкале процентных логитов, при помощи линейной функции по формуле (2).

Причем пока всё перечисленное не будет сделано, можно утверждать только, что

$$T_0 = 0 \text{ и } T_K = 100.$$

Основным шагом в этой цепочке, требующим теоретического обоснования, является шаг 3. Опишем метод первичных баллов.

Согласно этому методу оценки $\bar{\theta}_i, \bar{\delta}_j$ латентных параметров тестирования вычисляются по формулам:

$$\bar{\theta}_i = \ln \left(\frac{i}{K-i} K_1 \right), \text{ причем } K_1 = \frac{\sum_{r=0}^K (K-r) \cdot N_r}{\sum_{r=0}^K r \cdot N_r},$$

$$i = 1 \dots K-1, \quad \bar{\theta}_0 = -\theta_{max}, \quad \bar{\theta}_K = \theta_{max},$$

$$\bar{\delta}_j = \ln \left(\frac{N \cdot m_j - c_j}{c_j} \cdot K_2 \right), \text{ причем } K_2 = \frac{\sum_{j=1}^M c_j}{\sum_{j=1}^M N \cdot m_j - c_j},$$

$j = 1, \dots, M$.

В качестве обоснования этих формул можно привести следующие рассуждения. Введем в нашу «игру» (экзамен) помимо стандартных «игроков»: абитуриенты, задания двух дополнительных «игроков»: «тест» и «средний игрок». Под «тестом» понимается совокупность всех заданий теста, при помощи которых данный «игрок» сражается с другими участниками данной «игры». Под «средним игроком» понимается совокупность всех абитуриентов, при помощи которых данный «игрок» сражается с другими участниками игры. Пусть θ_{cp} – уровень подготовленности «среднего игрока» и δ_{cp} – уровень подготовленности «теста». Тогда коэффициент K_1 можно представить как отношение, где $t_{тест}$ – уровень подготовленности «теста», t_{cp} – уровень подготовленности «среднего игрока». Поэтому

$$\begin{aligned} \bar{\theta}_i &= \ln \left(\frac{i}{K-i} K_1 \right) = \ln \left(\frac{i}{K-i} \right) + \ln(K_1) \approx \\ &\approx \ln \left(\frac{t_i}{t_{mecm}} \right) + \ln \left(\frac{t_{mecm}}{t_{cp}} \right) = \end{aligned}$$

$$= \ln(t_i) - \ln(t_{mecm}) + \ln(t_{mecm}) - \ln(t_{cp}) = \theta_i - \theta_{cp}.$$

Учитывая «нормальность» распределения уровня подготовленности абитуриента, причем, как правило, с нулевым математическим ожиданием, можно считать, что $\theta_{cp} = 0$. Поэтому получим желаемое равенство $\theta_i \approx \bar{\theta}_i$.

Если последнее предположение неверно, то

$$\theta_i \approx \bar{\theta}_i + \bar{\theta}_{cp}, \text{ где } \bar{\theta}_{cp} = \frac{\sum_{i=1}^{K-1} \bar{\theta}_i \cdot N_i}{N - N_0 - N_K}.$$

Аналогичные рассуждения можно провести и для уровня трудности задания. А именно, коэффициент K_2 можно представить как отношение $\frac{t_{cp}}{t_{mecm}}$, где $t_{тест}$ – уровень подготовленности

«теста», t_{cp} – уровень подготовленности «среднего игрока». Поэтому

$$\begin{aligned}\bar{\delta}_j &= \ln\left(\frac{N \cdot m_j - c_j}{c_j} \cdot K_2\right) = \ln\left(\frac{N \cdot m_j - c_j}{c_j}\right) + \ln(K_2) \approx \\ &\approx \ln\left(\frac{t_j}{t_{cp}}\right) + \ln\left(\frac{t_{cp}}{t_{мечт}}\right) = \ln(t_j) - \ln(t_{cp}) + \ln(t_{cp}) - \ln(t_{мечт}) = \\ &= \delta_j - \delta_{мечт}.\end{aligned}$$

Учитывая «нормальность» распределения уровня трудности теста, причем, как правило, с нулевым математическим ожиданием, можно считать, что $\delta_{тест} = 0$. Поэтому получим желаемое равенство $\delta_j \approx \bar{\delta}_j$.

Если последнее предположение неверно, то

$$\delta_j \approx \bar{\delta}_j + \bar{\delta}_{мечт}, \text{ где } \bar{\delta}_{мечт} = \frac{\sum_{j=1}^M \bar{\delta}_j}{M}.$$

Основные результаты работы

Итак, в данной работе получены следующие результаты:

- 1) Модель тестирования Раша построена как игровая модель, в которой участники тестирования, выполняющие задания, и задания теста играют равноправную роль.
- 2) Адекватность модели Раша действительности подтверждена на примере шахматной классификации А. Эло.
- 3) Обосновано введение двухпараметрической модели Бирнбаума на примере шахматной классификации.
- 4) Введены и рассмотрены различные шкалы измерения латентных параметров тестирования: шкалы игрового логита, обычного логита (логита) и процентного логита.
- 5) Приведено «игровое» обоснование известного метода первичных баллов, метода оценивания латентных параметров тестирования по результатам наблюдения тестирования, который обладает наивысшей скоростью вычисления оценок.

ЛИТЕРАТУРА

1. *Rasch G.* Probabilistic Models for Some Intelligence and Attainment Tests. – Copenhagen Denmark: Danish Institute for Educational Research, 1968.

2. *Нейман Ю.М., Хлебников В.А.* Введение в теорию моделирования и параметризации педагогических тестов. – М., 2000. – 169 с.

3. *Birnbaum A.* Some latent trait models and their use in inferring an examinee's ability/ In Lord F.M. and Novick M.R. Statistical Theories of Mental Test Scores. – Reading MA: Addison-Wesley, 1968.

4. *Гук Е.Я.* Шахматы и математика. – М.: Наука, 1983 (Библиотека «Квант». – Вып. 24).

5. *Садовский Л.Е., Садовский А.Л.* Математика и спорт. – М.: Наука, Главная редакция физико-математической литературы, 1985 (Библиотека «Квант». – Вып. 44).

6. *Карнаухов В.М.* Электронное тестирование с двумя и более попытками для решения одного задания. – М.: ФГБОУ ВПО МГУП, 2011. – 172 с.

7. *Карнаухов В.М.* Математическая олимпиада на ЭВМ: теория, реализация, опыт. – М.: ФГБОУ ВПО МГУП, 2013. – 177 с.

8. Методика шкалирования результатов ЕГЭ в 2012 году: Приложение к приказу ФИПИ от 18 апреля 2012 г. – № 13-П.

V.M. Karnaukhov

Moscow State University of Environmental Engineering, Moscow, Russia

RASCH'S MODEL AS A GAME MODEL

Key words: Rasch's model, Elo's classification, the model adequacy, the method of primary numbers, latent parameters, qualification level, difficulty level.

The article describes the application of the famous Rasch's model of testing as a game model, in which the participants fulfill the tasks that play an equal role. The adequacy of the Rasch's model is confirmed by the example of A.Elo's chess classification. It gives «a game» justification of famous method of primary numbers, testing's latent parameters method according to testing observation results, which has the highest speed calculation.

In the beginning of the article the basic principles of A.Elo's chess classification are presented (1963-1970 years). We consider the mathematical model of two chess players meeting in one game. The adequacy of this model is confirmed by accumulated statistical material, expressed in Elo's table. The built model coincides with the Rasch's model for solving test tasks. This coincidence makes it possible to associate the testing as a game, in which the equal «players» are the participants of the testing and test tasks.

To build a «game» version of Rasch's model the author introduces the concept of qualification level of the participant (the participant of testing

is either the student solving tasks or test task). Basic properties of qualification level are identified. By means of these properties the main formulas of mathematical model of testing are proved. In particular, a formula is derived to calculate the probability of decision by the participant of testing with a specific qualification level of test task with a specific difficulty level. Further we considered different units of measurement of qualification level: game logit, logit and percent logit. Moreover, the solid change ranges for the units are established.

The article considers the two-parameter Birnbaum's model for testing. One of the variants of this model is used in Elo's chess classification.

Today the primary application of the mathematical Rasch's model is electronic testing of various levels, e.g. local Olympiads or the Unified State Examination. The paper deals with the methods of calculation for the latent parameters of testing: qualification levels of the participants and difficulty levels of test tasks. This method is based on transformation of primary numbers that has the highest rate of calculation. It gives a «game» justification of

primary numbers method by means of two additional players: «the average player» and «test».

REFERENCES

1. *Rasch G.* Probabilistic Models for Some Intelligence and Attainment Tests. – Copenhagen Denmark: Danish Institute for Educational Research, 1968.
2. *Nejman Ju.M., Hlebnikov V.A.* Vvedenie v teoriju modelirovanija i parametrizacii pedagogicheskikh testov. – M., 2000. – 169 s.
3. *Birnbaum A.* Some latent trait models and their use in inferring an examinee's ability/ In Lord F.M. and Novick M.R. Statistical Theories of Mental Test Scores. – Reading MA: Addison-Wesley, 1968.
4. *Gik E.Ja.* Shahmaty i matematika. – M.: Nauka, 1983 (Biblioteka «Kvant». – Vyp. 24).
5. *Sadovskij L.E., Sadovskij A.L.* Matematika i sport. – M.: Nauka, Glavnaja redakcija fiziko-matematicheskoy literatury, 1985 (Biblioteka «Kvant». – Vyp. 44).
6. *Karnauhov V.M.* Jelektronnoe testirovanie s dvumja i bolee popytkami dlja reshenija odnogo zadaniya. – M.: FGBOU VPO MGUP, 2011. – 172 s.
7. *Karnauhov V.M.* Matematicheskaja olimpiada na JeVM: teorija, realizacija, opyt. – M.: FGBOU VPO MGUP, 2013. – 177 s.
8. Metodika shkalirovanija rezul'tatov EGJe v 2012 godu: Prilozhenie k prikazu FIPI ot 18 aprelja 2012 g. – № 13-P.