

Теория

ПОИСК НАИЛУЧШЕГО ЗНАЧЕНИЯ ПАРАМЕТРА ДИФФЕРЕНЦИРУЮЩЕЙ СПОСОБНОСТИ ТЕСТОВОГО ЗАДАНИЯ ДЛЯ МОДЕЛИ БИРНБАУМА

Александр Рыбанов

Волжский политехнический институт (филиал) ФГОУ ВПО
«Волгоградский государственный технический университет»

Рассмотрены задачи поиска параметров модели Г. Раша и задачи поиска наилучшего значения параметра дифференцирующей способности тестового задания для модели А. Бирнбаума. В качестве наилучшего значения параметра дифференцирующей способности тестового задания предлагается значение, обеспечивающее минимум суммы квадратов отклонений, рассчитанных значений по модели А. Бирнбаума от эмпирических данных, полученных по результатам тестирования. Предлагаемый в работе подход к определению характеристик тестовых заданий для модели А. Бирнбаума позволяет получать тесты с более низкой дифференцированной ошибкой измерения.

Ключевые слова: IRT, Rasch Measurement, педагогические измерения, модель Раша, модель Бирнбаума, дифференцирующая способность задания, тестовое задание, логит, дифференцированная ошибка измерения теста.

Введение

Основными составляющими информационного обеспечения автоматизированной системы контроля и мониторинга за качеством учебного процесса¹ являются база тестовых заданий и база результатов тестирования испытуемых. Качественные и количественные характеристики тестовых заданий непосредственным образом влияют на объективную интерпретацию² результатов тестирования испытуемых. Количественные и качественные характеристики каждого тестового задания лучше определять опытным путем.

Рассмотрим задачи получения количественных характеристик тестовых заданий на моделях IRT (*Item Response Theory*). Модель Г. Раша³ описывает вероятность правильного решения тестового задания как функцию, зависящую от уровня подготовленности испытуемого и трудности тестового задания. В модели Бирнбаума⁴ появляется новый параметр — *дифференцирующая способность тестового задания*.

Дифференцирующая способность тестового задания является индикатором условия, определяющего корректировку тестового задания или его исключение из теста. Подбор тестовых заданий сбалансирован-

ной дифференцирующей силы позволяет добиться оптимальной валидности и надёжности теста. Погрешность измерения дифференцирующей силы тестового задания служит причиной формирования некорректных тестов и, как следствие, приводит к увеличению дифференцированной ошибки измерения для всего теста.

Постановка задачи

Результаты тестирования испытуемых представлены бинарной матрицей ответов (эмпирических данных):

$$X = (x_{i,j} | x_{i,j} \in \{0,1\}, i = \overline{1,n}, j = \overline{1,k}),$$

где n — количество участников тестирования; k — количество заданий в тесте.

Интерпретация свойств заданий (теста) и выставление тестовых баллов испытуемым предполагают предварительное решение ряда задач.

Задача поиска параметров модели Г. Раша. Дано:

$$X = (x_{i,j} | x_{i,j} \in \{0,1\}, i = \overline{1,n}, j = \overline{1,k}).$$

Найти уровень подготовленности испытуемых $\theta_i : i = \overline{1,n}$ и трудность для каждого тестового задания $\delta_j : j = \overline{1,k}$, соответствующие дихотомической модели Г. Раша:

$$P_j \{x_{i,j} = 1 | \theta_i, \delta_j\} = \frac{\exp(\theta_i - \delta_j)}{1 + \exp(\theta_i - \delta_j)},$$

Теория

1

Рыбанов А.А.

Алгоритмическое и тематическое обеспечение автоматизированной системы оценки качества учебного процесса по контрольным картам // Вестник компьютерных и информационных технологий. 2009. № 2. С. 30–36.

2

Аванесов В.С.

Проблема объективности педагогических измерений // Педагогические Измерения. № 3, 2008. С. 3–40.

3

Wright B.D.,

Masters G.N.

Rating scale analysis. Rasch measurements. Chicago: MESA Press, 1982.

4

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, 1968, Reading MA: Addison-Wesley.

ПЕД
измерения

где P_j – вероятность того, что i -й испытуемый выполнит правильно j -е задание (и, таким образом, получит 1 балл за выполнение этого задания).

Описанная выше задача может быть решена методом наибольшего правдоподобия, позволяющим получить модель Г. Раша, наилучшим образом описывающую бинарную матрицу ответов \mathbf{X} . О несоответствии тестового задания модели Г. Раша можно судить по значениям отклонений эмпирических данных от характеристической кривой. С другой стороны, А. Бирнбаум предложил ввести в модель Г. Раша параметр угла наклона характеристической кривой, варьируя значением которого можно добиться наилучшего соответствия тестового задания модели измерения. Т.о. задача выбора наилучшего значения дифференцирующей силы тестового задания (угла наклона характеристических кривых) является актуальной.

Задача поиска наилучшего значения параметра дифференцирующей способности тестового задания для модели А. Бирнбаума. Дано:

$$X = (x_{i,j} | x_{i,j} \in \{0,1\}, i = \overline{1,n}, j = \overline{1,k})$$

$\theta_i : i = \overline{1,k}, \delta_i : i = \overline{1,k}$. Найти для каждого тестового задания значение угла наклона характеристической кривой (параметр дифференцирующей силы за-

дания) $d_i : i = \overline{1,k}$, обеспечивающего максимальное соответствие эмпирических данных двухпараметрической модели А. Бирнбаума:

$$P_j \{X_{i,j} = 1 | \theta_i, \delta_j, d_j\} = \frac{\exp(d_j (\theta_i - \delta_j))}{1 + \exp(d_j (\theta_i - \delta_j))}$$

Математическое описание

Рассмотрим решение задачи поиска параметров модели Г. Раша методом наибольшего правдоподобия. Метод заключается в построении функции правдоподобия⁵ $L(x_{i,j}; \theta_i, \delta_j)$ дискретной случайной величины $x_{i,j}$.

$$L(x_{i,j}; \theta_i, \delta_j) = \prod_{i=1}^n \prod_{j=1}^k P_j \{x_{i,j} | \theta_i, \delta_j\} = \frac{\exp \left[\sum_{i=1}^n \sum_{j=1}^k x_{i,j} (\theta_i - \delta_j) \right]}{\prod_{i=1}^n \prod_{j=1}^k (1 + \exp(\theta_i - \delta_j))}$$

где $P_j \{x_{i,j} | \theta_i, \delta_j\} =$

$$\frac{\exp \{x_{i,j} (\theta_i - \delta_j)\}}{1 + \exp(\theta_i - \delta_j)}$$

В качестве точечных оценок латентных параметров принимают значения θ_i^* и δ_i^* , при которых $L(x_{i,j}; \theta_i^*, \delta_j^*) = \max L(x_{i,j}; \theta_i, \delta_j)$. Следует отметить, что $L(x_{i,j}; \theta_i, \delta_j)$ и

5

Нейман Ю.М.,
Хлебников В.А.
Введение в теорию моделирования и параметризации педагогических тестов. М.: Прометей, 2000. 168 с.

$\ln L(x_{i,j}; \theta_i, \delta_j)$ достигают максимума при одних и тех же значениях θ_i^* и δ_j^* .

Логарифмическая функция правдоподобия в данном случае имеет вид:

$$\ln L(x_{i,j}; \theta_i, \delta_j) = \sum_{i=1}^n a_i \theta_i - \sum_{j=1}^k b_j \delta_j - \sum_{i=1}^n \sum_{j=1}^k \ln [1 + \exp(\theta_i - \delta_j)],$$

где $a_i = \sum_{j=1}^k x_{i,j}$ — первичные баллы i -го испытуемого; $b_j = \sum_{i=1}^n x_{i,j}$ — первичные баллы j -го задания.

Для нахождения максимума функции правдоподобия необходимо решить следующую систему уравнений:

$$\begin{cases} \frac{\partial \ln L(\theta_i, \delta_j)}{\partial \theta} = a_i - \sum_{j=1}^k \frac{\exp(\theta_i - \delta_j)}{1 + \exp(\theta_i - \delta_j)} = 0, i = \overline{1, n} \\ \frac{\partial \ln L(\theta_i, \delta_j)}{\partial \delta_j} = -b_j + \sum_{i=1}^n \frac{\exp(\theta_i - \delta_j)}{1 + \exp(\theta_i - \delta_j)} = 0, j = \overline{1, k} \end{cases}$$

Представленная выше система уравнений называется *системой правдоподобия* и содержит $(n + k)$ уравнений с $(n + k)$ неизвестными латентными параметрами: $\theta_i : i = \overline{1, n}$ и $\delta_j : j = \overline{1, k}$. Система правдоподобия имеет единственное решение, соответствующее максимуму логарифмической функции правдоподобия.

В табл. 1 приведен учебный пример⁶ с результатами тестирования испытуемых, исполь-

зуемых для определения параметров θ_i и δ_j модели Г. Раша методом наибольшего правдоподобия.

Решение системы правдоподобия выполнено в системе MathCAD с помощью функции Find() (рис. 1). Полученные в результате решения латентные параметры $\theta_i : i = \overline{1, n}$ и $\delta_j : j = \overline{1, k}$ приведены в табл. 1.

На рис. 2-а приведены монотонно возрастающие графики заданий №2, 3, 8. Чем выше логит подготовленности, тем вероятнее получить правильный ответ. На рис. 2-б приведены монотонно убывающие графики испытуемых №1, 6, 12, характеризующие вероятность

правильного ответа на тестовое задание с заданным логитом трудности. Например, вероятность

$p_{i,8}$ правильного ответа на задание с логитом трудности $\delta_8 = 1,108$, для i -го испытуемого: $p_{1,8} = 0,858$, $p_{6,8} = 0,249$, $p_{12,8} = 0,0333$.

Для проверки соответствия эмпирических данных модели Г. Раша испытуемые делятся на группы (в данном случае, на три группы: $Z = 3$) вдоль шкалы θ : 1 — сильная группа, 2 — средняя группа испытуемых, 3 — слабая группа испытуемых. Всего внутри группы с номером k окажутся m_z испытуемых, $z = \overline{1, Z}$.

⁶ Аванесов В.С. Метрическая система Георга Раша // Педагогические измерения. №2. 2010. С. 57–80.

ПЕД
измерения

Таблица 1

Учебный пример таблицы результатов тестирования

| i | $x_{1,1}$ | $x_{1,2}$ | $x_{1,3}$ | $x_{1,4}$ | $x_{1,5}$ | $x_{1,6}$ | $x_{1,7}$ | $x_{1,8}$ | $x_{1,9}$ | $x_{1,10}$ | a_i | θ_i | Группа |
|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|-------|------------|--------|
| 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 9 | 2,908 | 1 |
| 2 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 8 | 1,931 | |
| 3 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 7 | 1,223 | |
| 4 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 6 | 0,604 | 2 |
| 5 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 6 | 0,604 | |
| 6 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 5 | 0,004 | |
| 7 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 5 | 0,004 | |
| 8 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 5 | 0,004 | |
| 9 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 4 | -0.632 | 3 |
| 10 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 4 | -0.632 | |
| 11 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | -1.366 | |
| 12 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | -2.282 | |
| 13 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -3.530 | |
| b_j | 12 | 11 | 9 | 7 | 6 | 6 | 5 | 4 | 3 | 1 | | | |
| δ_j | -3,615 | -2,539 | -1,220 | -0,262 | 0,18 | 0,18 | 0,629 | 1,108 | 1,653 | 2,326 | | | |

В пределах каждой группы r_z испытуемых отвечают правильно на j -е задание теста. Таким образом, для логита подготовленности θ_z группы с номером z эмпирическое значение вероятности $p_{z,j}$ правильного ответа на данное задание представлено в табл. 2:

$$p_{z,j}(\theta_z) = \frac{r_z}{m_z}$$

Критерий соответствия эмпирических данных модели Г. Раша определим как сумму квадратов отклонений рассчитанных значений модели от эмпирических данных $p_{z,j}$:

$$S_j = \sum_{z=1}^Z \left(p_{z,j} - \frac{\exp(\theta_z - \delta_j)}{1 + \exp(\theta_z - \delta_j)} \right)^2$$

Рассмотрим решение задачи поиска наилучшего значения параметра дифференцирующей способности задания модели А. Бирнбаума. В двухпараметрической модели А. Бирнбаума характеристикой тестового задания j является не только логит трудности δ_j , но и дифференцирующая способность δ_j . Дифференцирующая способность определяет как тангенс угла наклона касательной, проведённой к графику тестового задания в точке $\theta = \delta_j$. Чем выше значение δ_j , тем лучше задание подходит для теста. На практике рекомендуется, как правило, оставлять задания со значениями $d_j \in [0.5; 2.5]$

Теория
 12.02.2013

ORIGIN := 1
 Матрица ответов испытуемых:

$$X := \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Количество испытуемых: $n := \text{rows}(X)$
 Количество тестовых заданий: $k := \text{cols}(X)$

$i := 1..n \quad j := 1..k$

Начальные приближения: $\theta_i := \ln\left(\frac{\sum_{j=1}^k X_{i,j}}{k - \sum_{j=1}^k X_{i,j}}\right) \quad \delta_j := \ln\left(\frac{n}{\sum_{i=1}^n X_{i,j}} - 1\right)$

Первичные баллы испытуемых и заданий:

$$a_i := \sum_{j=1}^k X_{i,j} \quad b_j := \sum_{i=1}^n X_{i,j}$$

Given

$$a - \sum_{j=1}^k \frac{e^{\theta - \delta_j}}{1 + e^{\theta - \delta_j}} = 0 \quad -b + \left(\sum_{i=1}^n \frac{e^{\theta_i - \delta}}{1 + e^{\theta_i - \delta}} \right) = 0$$

Result := Find(θ, δ)

| | | | | | | | | |
|----------------------------------|-------|-------|-------|-------|-------|-----------------------|-----------------------|-----|
| Result ₁ ^T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | 2.908 | 1.931 | 1.223 | 0.604 | 0.604 | $3.803 \cdot 10^{-3}$ | $3.803 \cdot 10^{-3}$ | ... |

| | | | | | | | | | | |
|----------------------------------|--------|--------|-------|--------|------|------|-------|-------|-------|-------|
| Result ₂ ^T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | -3.615 | -2.539 | -1.22 | -0.262 | 0.18 | 0.18 | 0.629 | 1.108 | 1.653 | 2.326 |

Рис. 1. Реализация метода наибольшего правдоподобия в MathCAD

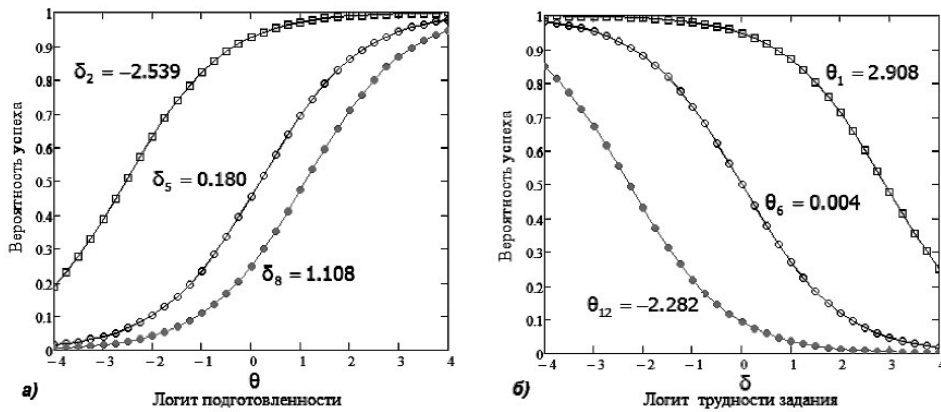


Рис. 2. Графики тестовых заданий и испытуемых

При $\delta_j = 1$ логическая модель А. Бирнбаума совпадает с логической однопараметрической моделью Г. Раша.

ПЕД
измерения

Таблица 2

Вероятность правильного ответа испытуемых z-й группы

| z | $P_{z,1}$ | $P_{z,2}$ | $P_{z,3}$ | $P_{z,4}$ | $P_{z,5}$ | $P_{z,6}$ | $P_{z,7}$ | $P_{z,8}$ | $P_{z,9}$ | $P_{z,10}$ | θ_z |
|----------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|------------|
| 1 | 1 | 1 | 0,667 | 0,667 | 0,667 | 1 | 1 | 0,667 | 1 | 0,333 | 2,021 |
| 2 | 1 | 1 | 0,8 | 1 | 0,6 | 0,4 | 0,4 | 0,2 | 0 | 0 | 0,244 |
| 3 | 0,8 | 0,6 | 0,6 | 0 | 0,2 | 0,2 | 0 | 0,2 | 0 | 0,2 | -1,682 |
| $S_j \times 10^{-3}$ | 5,851 | 13,919 | 133,145 | 237,361 | 49,925 | 36,499 | 47,781 | 31,713 | 207,040 | 53,751 | |

Дифференцирующая способность тестового задания может быть рассчитана на основе бисериального коэффициента корреляции в предположении, что значения латентной переменной, лежащей в основе выполнения заданий, распределены нормально.

Предлагаемая в работе Lord F.M. формула для оценки параметра дифференцирующей способности j -го задания теста имеет вид⁷:

$$d_j = \frac{r_{bis_j}}{\sqrt{1 - (r_{bis_j})^2}},$$

где r_{bis_j} бисериальный коэффициент корреляции j -го задания.

В табл. 3 приведены значения бисериальных коэффициентов и дифференцирующих способностей тестовых заданий учебного примера таблицы результатов тестирования.

Результаты моделирования

Критерий соответствия эмпирических данных модели А. Бирнбаума определим как сумму квадратов отклонений рассчитанных значений модели от эмпирических значений $p_{z,j}$, полученных по результатам тестирования:

$$S_j = \sum_{z=1}^z \left(p_{z,j} - \frac{\exp(d_j(\theta_z - \delta_j))}{1 + \exp(d_j(\theta_z - \delta_j))} \right)^2.$$

Найдем оптимальные значения $d_j^* : j = \overline{1, k}$, при которых функция $S_j(d_j^*)$ достигает минимального значения:

$$S_j^* = S_j(d_j^*) =$$

$$= \min \sum_{z=1}^z \left(p_{z,j} - \frac{\exp(d_j^*(\theta_z - \delta_j))}{1 + \exp(d_j^*(\theta_z - \delta_j))} \right)^2.$$

Полученные значения дифференцирующей способности

Таблица 3

Дифференцирующая способность тестового задания

| j | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| r_{bis_j} | 0,132 | 0,488 | 0,305 | 0,495 | 0,495 | 0,707 | 0,652 | 0,534 | 0,752 | 0,293 |
| d_j | 0,133 | 0,559 | 0,321 | 0,569 | 0,569 | 0,998 | 0,859 | 0,632 | 1,142 | 0,306 |

d_j^* будут являться наилучшими с точки зрения соответствия эмпирически данным.

На рис. 3 приведены характеристические кривые для тестовых заданий, построенные по модели А. Бирнбаума с параметрами $d_j = 1$ (соответствует модели Г. Раша), d_j' (рассчитан на основе бисериального коэффициента корреляции), d_j^* (получен в результате минимизации функции $S_j(d_j^*)$). Поиск значений d_j^* осуществлялся в MathCAD с помощью встроенной функции Minimize().

Количественные результаты моделирования, приведённые в табл. 4, показывают, что для тестовых заданий №1, 2, 4, 7, 9, 10 значения S_j^* и S_j' имеют существенные различия.

Обсуждение результатов

Для сравнительного анализа результатов, представленных в табл. 4, воспользуемся понятием информационной функции⁸, которая позволяет оценить меру точности измерения каждым отдельным заданием или тестом в целом.

Для тестовых заданий информационная функция модели А. Бирнбаума описывается как:

$$I_j(d_j, \theta, \delta_j) = d_j^2 P(\theta, \delta_j, d_j) Q(\theta, \delta_j, d_j)$$

Информационная функция всего теста вычисляется как сумма информационных функций тестовых заданий:

$$I(\theta) = \sum_j^k I_j(d_j, \theta, \delta_j) = \sum_{j=1}^k d_j^2 P(\theta, \delta_j, d_j) Q(\theta, \delta_j, d_j)$$

Таблица 4

Значения параметра дифференцирующей способности тестового задания d_j , d_j^* , d_j' и критерия оптимальности S_j , S_j^* , S_j'

| j | d_j | S_j | d_j^* | S_j^* | d_j' | S_j' |
|-----|-------|--------------------------|---------|--------------------------|--------|--------------------------|
| 1 | 1 | $5,851 \times 10^{-3}$ | 0,807 | $2,619 \times 10^{-3}$ | 0,133 | $298,325 \times 10^{-3}$ |
| 2 | 1 | $13,919 \times 10^{-3}$ | 0,903 | $12,995 \times 10^{-3}$ | 0,559 | $35,874 \times 10^{-3}$ |
| 3 | 1 | $133,145 \times 10^{-3}$ | 0,328 | $58,056 \times 10^{-3}$ | 0,321 | $58,078 \times 10^{-3}$ |
| 4 | 1 | $237,361 \times 10^{-3}$ | 7,853 | $111,000 \times 10^{-3}$ | 0,569 | $292,838 \times 10^{-3}$ |
| 5 | 1 | $49,925 \times 10^{-3}$ | 0,553 | $16,910 \times 10^{-3}$ | 0,569 | $16,980 \times 10^{-3}$ |
| 6 | 1 | $36,499 \times 10^{-3}$ | 1,157 | $34,552 \times 10^{-3}$ | 0,998 | $36,541 \times 10^{-3}$ |
| 7 | 1 | $47,781 \times 10^{-3}$ | 1,888 | $10,157 \times 10^{-3}$ | 0,859 | $68,826 \times 10^{-3}$ |
| 8 | 1 | $31,713 \times 10^{-3}$ | 0,787 | $28,569 \times 10^{-3}$ | 0,632 | $31,379 \times 10^{-3}$ |
| 9 | 1 | $207,040 \times 10^{-3}$ | 15,000 | $0,020 \times 10^{-3}$ | 1,142 | $185,386 \times 10^{-3}$ |
| 10 | 1 | $53,751 \times 10^{-3}$ | 2,287 | $40,031 \times 10^{-3}$ | 0,306 | $140,798 \times 10^{-3}$ |

Теория

12-0000

8

Lord F.M.,
Novick M.R.
Statistical Theories of
Mental Test Scores.
Reading, MA: Addison-
Wesley Publishing
Company, 1968.

ПЕД
измерения

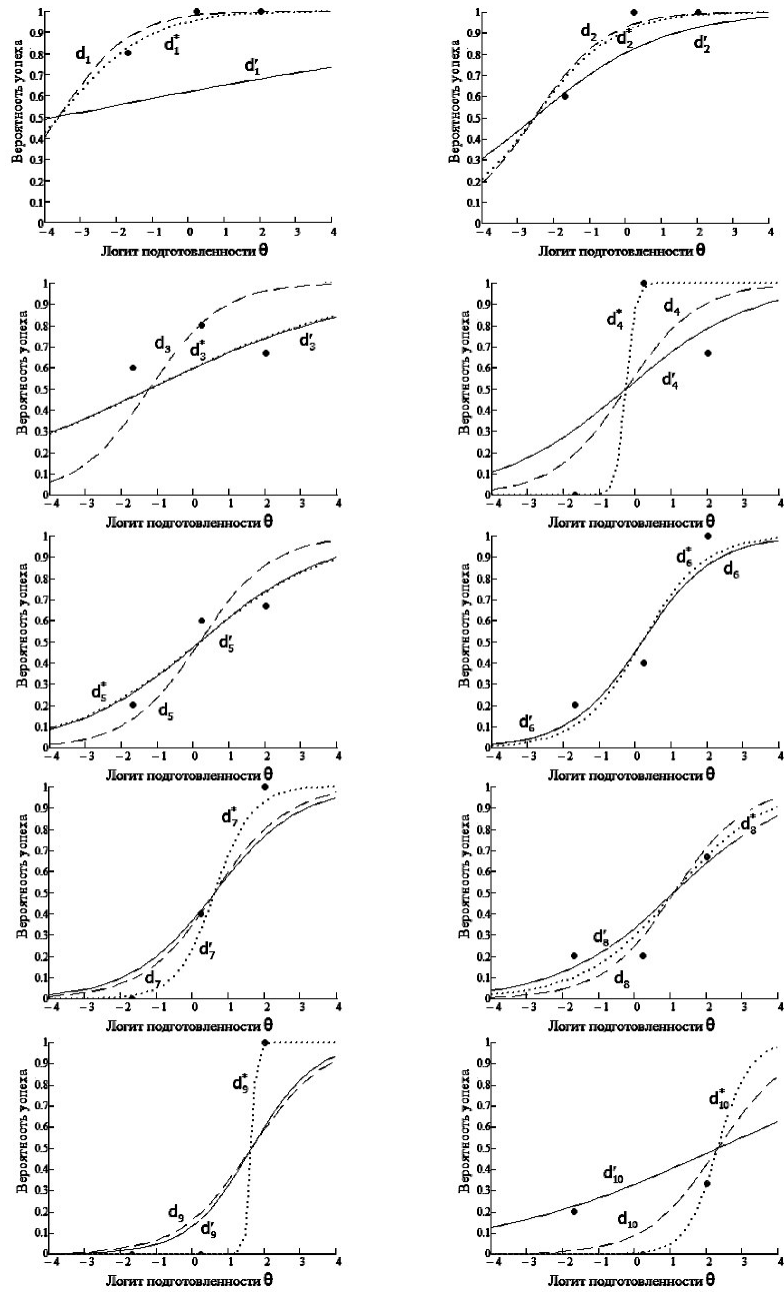


Рис. 3. Графики для заданий с параметрами d_j, d'_j, d_j^*

Дифференцированная ошибка измерения для теста описывается следующей функцией:

$$SEM(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

По значениям дифференцированной ошибки измерения можно судить об эффективности компоновки теста по различным значениям параметра дифференцирующей способности тестового задания.

Графики стандартной ошибки измерения $SEM^*(\theta)$ и $SEM'(\theta)$ по тесту с различными значениями дифференцирующей способности d_j^* и d_j' для

одних и тех же тестовых заданий (с №1 по №10), приведённые на рис. 4, подтверждают, что при использовании параметра d_j^* стандартная ошибка измерения меньше.

По результатам, представленным в табл. 4, для модели А. Бирнбаума с параметром d_j^* из теста должны быть удалены задания №3, 4, 9, т.к. $d_3, d_4, d_9 \notin [0.5; 2.5]$ а судя по значениям d_j' — из теста должны быть удалены задания №3, 1, 10. Таким образом, тест, скомпонованный по параметру d_j^* , должен содержать задания №1, 2, 5–8, 10, а тест, скомпонован-

Теория

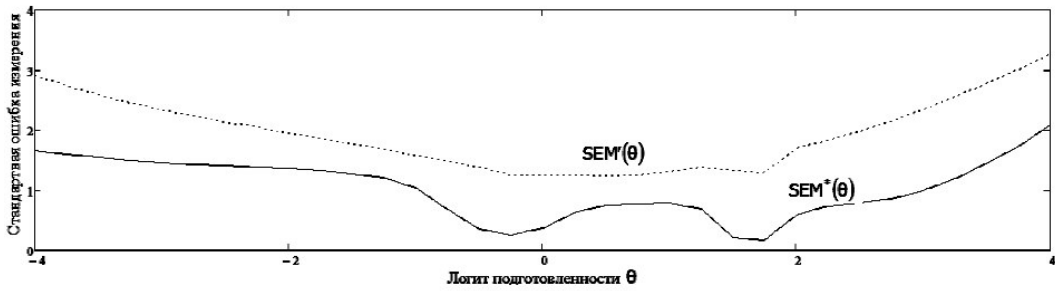


Рис. 4. Графики стандартной ошибки измерения $SEM^*(\theta)$ и $SEM'(\theta)$

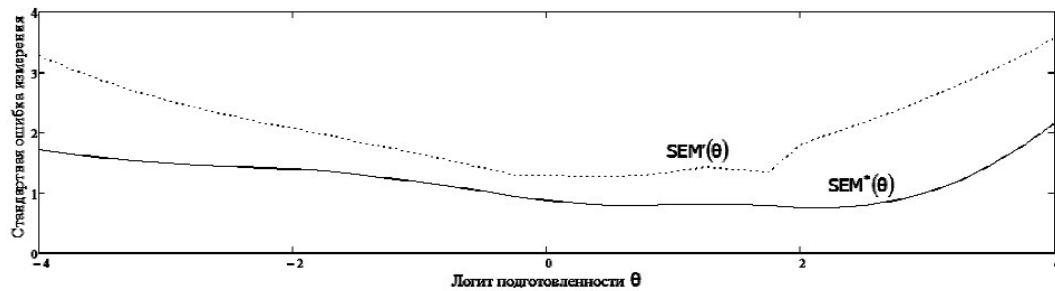


Рис. 5. Графики стандартной ошибки измерения $SEM^*(\theta)$ и $SEM'(\theta)$ для тестов после компоновки

| |
|-----------|
| ПЕД |
| измерения |

ный по параметру d_j^* – задания №2, 4–9.

Графики стандартной ошибки измерения $SEM^*(\theta)$ и $SEM^*(\theta)$ по тестам, после редуцирования множества тестовых заданий, приведённые на рис. 5, также подтверждают, что при использовании параметра d_j^* точность измерений увеличивается. Кроме того, график $SEM^*(\theta)$ свидетельствует о том, что тест, скомпонованный по параметру d_j^* , обеспечивает практически равную точность оценок

испытуемых в заданном интервале оценок на шкале логитов.

Заключение

Предлагаемый в работе подход к решению задачи поиска наилучшего значения параметра дифференцирующей способности тестовых заданий является допустимым и позволяет добиться снижения стандартной ошибки измерения для всего теста в целом.