

# Теория

## КОМПЬЮТЕРНАЯ ПРОГРАММА ДЛЯ ПОДБОРА ПАРАМЕТРОВ ОСНОВНЫХ МОДЕЛЕЙ ITEM RESPONSE THEORY (IRT)<sup>1</sup>

**Олег Деменчёнок**

Восточно-Сибирский институт МВД России  
AskSystem@yandex.ru

Рассмотрено программное обеспечение для подбора параметров моделей IRT. Выявлена целесообразность совершенствования этого программного обеспечения. Проведён анализ задачи оценки параметров модели измерения, по результатам которого в качестве метода поиска экстремума предложен квазиньютоновский метод BFGS. Разработана компьютерная программа, реализующая поиск параметров основных моделей IRT.

**Ключевые слова:** тест, модель Г. Раша, двух- и трёхпараметрическая модель, метод максимального правдоподобия, метод наименьших квадратов, поиск экстремума, метод BFGS.

### 1

На русский язык IRT В.С. Аванесов переводит как математическую теорию измерений (МТИ). См.: Педагогические измерения, 2007 г. № 3. С. 3.

Существенным препятствием на пути внедрения в педагогическую практику Item Response Theory (IRT) является относительная сложность математического аппарата этой теории. Применение IRT фактически невозможно без использования компьютера на этапе оценки параметров математических моделей. Программному обеспечению решения этой задачи посвящена настоящая статья.

## Формулировка задачи оценки параметров

Задача заключается в подборе параметров, при которых достигается наилучшее совпадение результатов конкретного тестирования и математической модели. В данной работе ограничимся тремя базовыми моделями IRT — моделью Г. Раша, двух- и трёхпараметрической моделью.

По трёхпараметрической модели вероятность правильного ответа  $i$ -го тестируемого на  $j$ -е задание равна<sup>2</sup>:

$$P_{ij} = c_j + (1 - c_j) \frac{e^{a_j(\theta_i - \beta_j)}}{1 + e^{a_j(\theta_i - \beta_j)}}, \quad (1)$$

где  $\theta_i$  — уровень подготовленности  $i$ -го тестируемого;  $b_j$  и  $a_j$  — уровень трудности и дифференцирующая способность  $j$ -го задания;  $c_j$  — параметр коррекции на угадывание правильного ответа в  $j$ -м задании.

Уравнение (1) можно записать в алгебраически эквивалентной форме<sup>3</sup>

$$P_{ij} = c_j + (1 - c_j) \frac{1}{1 + e^{-a_j(\theta_i - \beta_j)}}. \quad (2)$$

При  $c_j = 0$  уравнения (1)–(2) соответствуют двухпараметрической модели, а при  $c_j = 0$  и  $a_j = 1$  — модели Раша. Таким образом, выражения (1) и (2) можно считать общими для всех трёх моделей.

Найдём вероятность неправильного ответа на задание  $j$ :

$$P_{ij}(x_{ij} = 0) = 1 - P_{ij} = 1 - c_j - (1 - c_j) \cdot$$

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

$$\left( 1 - \frac{e^{a_j(\theta_i - \beta_j)}}{1 + e^{a_j(\theta_i - \beta_j)}} \right) =$$

$$= (1 - c_j) \left( \frac{1 + e^{a_j(\theta_i - \beta_j)}}{1 + e^{a_j(\theta_i - \beta_j)}} - \frac{e^{a_j(\theta_i - \beta_j)}}{1 + e^{a_j(\theta_i - \beta_j)}} \right),$$

$$P_{ij}(x_{ij} = 0) = (1 - c_j) \frac{1}{1 + e^{a_j(\theta_i - \beta_j)}}, \quad (3)$$

где  $x_{ij}$  — результат выполнения задания,  $x = 1$  при правильном ответе,  $x = 0$  при неправильном ответе.

Из уравнений (1) и (3) получим формулу вероятности результата  $x$  при решении  $i$ -м испытуемым  $j$ -го задания:

$$P_{ij}(x_{ij}) = c_j x_{ij} + (1 - c_j) \frac{e^{x_{ij} a_j (\theta_i - \beta_j)}}{1 + e^{a_j (\theta_i - \beta_j)}}. \quad (4)$$

При  $x = 1$  (правильный ответ) выражение (4) преобразуется в формулу вероятности правильного ответа (1), а при  $x = 0$  (неправильный ответ) — в формулу вероятности неправильного ответа (3).

Степень близости результатов тестирования и модели может оцениваться по методу максимального правдоподобия или методу наименьших квадратов.

Метод максимального правдоподобия требует достижения максимума произведе-

## Теория

2

Partchev I.

A visual guide to item response theory. — Jena: Friedrich-Schiller-Universität, 2004. 61 p.

3

Baker F.B.

The Basics of Item Response Theory. 2 ed. Hieneman, Portsmouth, New Hampshire, 2001.

ния вероятностей всех наблюдаемых независимых событий<sup>4</sup>:

$$F = P_1 \cdot P_2 \cdot \dots \cdot P_N = \prod_{i=1}^N P_i \rightarrow \max. \quad (5)$$

Подставив в выражение (5) значение вероятности из формулы (4), получим:

$$F = \prod_{i=1}^n \prod_{j=1}^m P_{ij}(x) = \prod_{i=1}^n \prod_{j=1}^m \left( c_j x_{ij} + (1 - c_j) \frac{e^{x_{ij} a_j (\theta_i - \beta_j)}}{1 + e^{a_j (\theta_i - \beta_j)}} \right) \rightarrow \max, \quad (6)$$

где  $F$  — целевая функция;  $n$  — количество испытуемых;  $m$  — число тестовых заданий.

По методу наименьших квадратов наилучшее согласование расчётных и экспериментальных данных соответствует минимальной сумме квадратов отклонений между ними:

$$F = \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - P_{ij})^2 \rightarrow \min,$$

$$F = \sum_{i=1}^n \sum_{j=1}^m \left( x_{ij} - \left( c_j + (1 - c_j) \frac{1}{1 + e^{-a_j (\theta_i - \beta_j)}} \right) \right)^2 \rightarrow \min, \quad (7)$$

где  $P_{ij}$  — вероятность правильного ответа  $i$ -ого испытуемого на  $j$ -е задание.

Для определённости решения вводится ограничение, например, в виде равенства нулю среднего значения уровня подготовленности всех испытуемых:

$$\bar{\theta} = 0. \quad (8)$$

С точки зрения математики задача оценки параметров модели IRT сводится к известной задаче численного поиска экстремума (максимума или минимума) функции.

## Обзор программных средств

Программные средства для решения поставленной задачи можно разделить на две группы: универсальные и специализированные программы.

Универсальными являются электронные таблицы (прежде всего Microsoft Excel), а также математические пакеты MathCad, Mathematica, Maple, MathLab и др. Эти программы способны находить экстремум функции многих переменных. Основное преимущество универсальных программ — гибкость (возможна тонкая настройка поиска решения, расчёт и визуализация любых дополнительных параметров и т.д.).

Наиболее доступен Microsoft Excel, входящий в состав широко распространённого пакета программ для обработки деловой информации Microsoft Office. Excel можно настроить на поиск параметров любой модели IRT. Однако есть и недостатки:

4 Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. — М.: ЮНИТИ, 1998. 1022 с.

• **поиск экстремума в Excel ограничен по числу переменных** — не более 200. Соответственно, предельным будет относительно небольшой объём результатов тестирования. Например, результаты выполнения 50 тестовых заданий 100 испытуемыми. Большой объём встроенная в Excel надстройка «Поиск решения» обработать не сможет. Ограничение не очень жёсткое, для многих практических задач ресурсы Excel могут смотреться как достаточные;

• **относительная сложность.** В Excel нужно не только ввести данные, но и составить формулы для вычисления целевой функции, ограничений, правильно задать параметры поиска решения, что требует времени и некоторой квалификации;

• **низкая скорость** — поиск решения требует перерасчёта записанных на листе Excel формул при каждом вычислении целевой функции. Это замедляет процесс решения почти на порядок по сравнению с операциями в оперативной памяти; обработка предельного объёма результатов тестирования занимает от 4 до 5 минут при тактовой частоте процессора 2,66 ГГц.

Недостаток математических пакетов (MathCad и др.) — существенная трудоёмкость подготовительных работ и сложность для неспециали-

тов. Поэтому их нельзя рекомендовать для широкого использования в практической оценочной деятельности.

Специализированные программы (например, Winsteps, RUMM) созданы для работы с моделями IRT. Эти программы эффективны: время поиска решения невелико при достаточно высокой точности подбора параметров. Недостатки: ограничения по видам моделей, неудобные настройки (например, в Winsteps заложена только одна модель — модель Раша, а параметры поиска указываются во внешнем текстовом файле), отсутствие русскоязычного интерфейса, сложные правила ввода исходных данных, а также многошаговость подготовительного процесса. Бесплатные версии специализированных программ малоприспособлены для практического применения ввиду ограничений по времени использования: период работы программы RUMM без регистрации — 90 дней и объёму данных. Бесплатная версия Winsteps обрабатывает данные не более чем 75 испытуемых, при этом максимальное число заданий равно 25.

С учётом изложенного представляется целесообразной разработка другой компьютерной программы, которая должна сочетать:

1) удобство ввода результатов тестирования;

Теория

2008

ПЕД  
измерения

- 2) простоту настройки;
- 3) высокую скорость и точность подбора параметров модели педагогических измерений.

### Анализ задачи оценки параметров модели измерения

Известны десятки методов оптимизации, и для выбора наиболее подходящего в каждом конкретном случае целесообразно проанализировать специфику задачи. Каковы же особенности рассматриваемой задачи?

**1. Взаимная независимость результатов тестирования.** В моделях IRT принято, что результат тестирования одного испытуемого не зависит от результатов другого испытуемого. Аналогично независимы и результаты выполнения тестовых заданий. Следовательно, при поиске экстремума по одной из переменных можно рассматривать не всю матрицу результатов, а только одну строку (при определении уровня подготовленности отдельного испытуемого) или столбец (при определении параметров задания). Это означает многократное (в десятки раз) снижение объёма вычислений при поординатном поиске.

**2. Большое количество переменных.** В ходе решения переменными величинами считаются уровень подготовленнос-

ти испытуемых, уровень трудности заданий, а также уровень различающей способности заданий. Количество испытуемых исчисляется, как минимум, десятками, а нередко — сотнями и тысячами. Количество заданий — несколько десятков, а иногда и сотен. Это означает, что задача имеет большую размерность. Такие задачи обычно требуют большой оперативной памяти и длительного времени счёта. Поэтому следует выбрать метод оптимизации, отличающийся высокой скоростью поиска, т.е. метод Ньютона или один из квазиньютоновских методов<sup>5</sup>.

**3. Начальные значения параметров модели близки к оптимальным.** В предыдущей статье<sup>6</sup> на примере решения конкретной задачи показано, что начальные значения — это хорошая отправная точка для подбора параметров модели измерения. Подбор параметров модели Г. Раша даёт возможность увеличить среднюю вероятность результатов тестирования лишь на 5–9%, по сравнению с начальными значениями. Для двухпараметрической модели увеличение составляет от 8 до 15%. Близость начальных значений к оптимальным даёт возможность ограничиться поиском одного экстремума.

Полученные выводы представляются парадоксальными: наиболее эффективным может

5

*Бахвалов Н.С.,  
Жидков Н.П.,  
Кобельков Г.М.*  
Численные методы. —  
М.: Бином, 2004. 640 с.

6

*Деменчёнок О.Г.*  
Подбор параметров модели педагогических измерений // Педагогические измерения, 2008, № 2.

оказаться либо самый простой метод — метод покоординатного спуска, либо один из наиболее сложных — метод Ньютона или квазиньютоновский метод.

### Оценка параметров модели измерения методом покоординатного спуска

Метод покоординатного спуска (также называемый методом Гаусса–Зейделя) — это самый простой метод прямого поиска. Он основан на сравнении значений целевой функции. Частные производные не используются. Поиск заключается в изменении одной переменной, тогда как другие остаются постоянными, пока не будет получен минимум. Затем поочередно изменяют все остальные переменные. Покоординатный спуск продолжается до тех пор, пока улучшается значение целевой функции. Метод легко реализуем. Однако такой алгоритм работает плохо, если в выражение целевой функции входят члены, содержащие произведение искомого параметров модели<sup>7</sup>. Таким образом, нельзя гарантировать эффективность метода для двух- и трёхпараметрической модели, поскольку они содержат произведение дифференцирующей способности и разности уровней подготовленности и трудности  $a(\theta - \beta)$ .

Для модели Раша метод покоординатного спуска вполне пригоден.

### Квазиньютоновский метод оценки параметров модели измерения

Классический метод Ньютона характеризуется высокой сходимостью (т.е. поиск сходится к точке экстремума при минимальном количестве итераций). Метод основан на обращении матрицы Гессе (матрицы вторых производных), которая имеет вид:

$$H(F) = \begin{bmatrix} \frac{\partial^2 F}{\partial x_1^2} & \frac{\partial^2 F}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 F}{\partial x_1 \partial x_k} \\ \frac{\partial^2 F}{\partial x_2 \partial x_1} & \frac{\partial^2 F}{\partial x_2^2} & \dots & \frac{\partial^2 F}{\partial x_2 \partial x_k} \\ \dots & \dots & \ddots & \dots \\ \frac{\partial^2 F}{\partial x_n \partial x_1} & \frac{\partial^2 F}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 F}{\partial x_n^2} \end{bmatrix}, \quad (9)$$

где  $x_i$  —  $i$ -й параметр модели;  $k$  — число оцениваемых параметров.

Необходимость расчёта матрицы Гессе на каждой итерации значительно увеличивает объём вычислений и время расчёта, особенно для задач большой размерности. Так, для тысячи переменных ( $k = 1000$ ) на каждом шаге нужно рассчитать миллион вторых производных ( $k^2 = 1000000$ ). Поэтому были

Теория

Методы

7

Химмельблау Д.  
Прикладное нелинейное программирование: Пер. с англ. — М.: Мир, 1975. 534 с.

ПЕД  
измерения

разработаны квазиньютоновские методы, основанные на приближенных выражениях для матрицы Гессе. Одним из наиболее известных является метод BFGS, получивший своё название по фамилиям предложивших его авторов: Broyden–Fletcher–Goldfarb–Shanno (Бройден–Флетчер–Гольдфарб–Шанно)<sup>8</sup>. В основе метода лежит идея восстановления квадратичной аппроксимации функции по значениям её градиентов в ряде точек. Тем самым метод объединяет достоинства градиентного метода — не требуется вычисление матрицы вторых производных, и метода Ньютона — быстрая сходимость вследствие использования квадратичной аппроксимации.

Для эффективности метода BFGS желательно иметь аналитические зависимости частных производных целевой функции (разностная аппроксимация производных снижает скорость и точность метода).

### Частные производные целевой функции по методу максимального правдоподобия

Выражение целевой функции (6) представляет собой произведение большого числа сомножителей, каждый из которых — сумма двух слагаемых. Аналитически получить част-

ные производные в этом случае затруднительно. Однако при  $c_j = 0$  выражение (6) значительно упрощается. Выведем частные производные для двухпараметрической модели и модели Раша.

Обозначим  $s_j$  сумму результатов выполнения  $j$ -го задания всеми испытуемыми:

$$s_j = \sum_{i=1}^n x_{ij}. \quad (10)$$

После очевидных преобразований из (6) получим выражение:

$$F = \frac{\sum_{i=1}^n \sum_{j=1}^m x_{ij} a_j \theta_i - \sum_{j=1}^m s_j a_j \beta_j}{\prod_{i=1}^n \prod_{j=1}^m (1 + e^{a_j(\theta_i - \beta_j)})} \rightarrow \max. \quad (11)$$

Максимум функции правдоподобия совпадает с максимумом её логарифма:

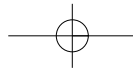
$$F_l = \ln \left( \frac{\sum_{i=1}^n \sum_{j=1}^m x_{ij} a_j \theta_i - \sum_{j=1}^m s_j a_j \beta_j}{\prod_{i=1}^n \prod_{j=1}^m (1 + e^{a_j(\theta_i - \beta_j)})} \right) \rightarrow \max,$$

$$F_l = \sum_{i=1}^n \sum_{j=1}^m x_{ij} a_j \theta_i - \sum_{j=1}^m s_j a_j \beta_j - \sum_{i=1}^n \sum_{j=1}^m \ln (1 + e^{a_j(\theta_i - \beta_j)}) \rightarrow \max. \quad (12)$$

Найдем частные производные логарифмической функции правдоподобия по каждому из аргументов:

$$\frac{\partial F_l}{\partial \theta_i} = \sum_{j=1}^m x_{ij} a_j - \sum_{j=1}^m \frac{a_j e^{a_j(\theta_i - \beta_j)}}{1 + e^{a_j(\theta_i - \beta_j)}} =$$

8  
Гилл Ф.,  
Мюррей У., Райт М.  
Практическая оптимизация: Пер. с англ. — М.:  
Мир, 1985. 509 с.



$$= \sum_{j=1}^m a_j \left( x_{ij} - \frac{e^{a_j(\theta_i - \beta_j)}}{1 + e^{a_j(\theta_i - \beta_j)}} \right), \quad (13)$$

$$\frac{\partial F_l}{\partial \beta_j} = -a_j s_j + a_j \sum_{i=1}^n \frac{e^{a_j(\theta_i - \beta_j)}}{1 + e^{a_j(\theta_i - \beta_j)}} =$$

$$= a_j \left( \sum_{i=1}^n \frac{e^{a_j(\theta_i - \beta_j)}}{1 + e^{a_j(\theta_i - \beta_j)}} - s_j \right), \quad (14)$$

$$\frac{\partial F_l}{\partial a_j} = \sum_{i=1}^n x_{ij} \theta_j - s_j \beta_j -$$

$$- \sum_{i=1}^n \frac{(\theta_i - \beta_j) e^{a_j(\theta_i - \beta_j)}}{1 + e^{a_j(\theta_i - \beta_j)}} = \quad (15)$$

$$= \sum_{i=1}^n \left( x_{ij} \theta_j - \frac{(\theta_i - \beta_j) e^{a_j(\theta_i - \beta_j)}}{1 + e^{a_j(\theta_i - \beta_j)}} \right) - s_j \beta_j.$$

Формулы (13)–(14) пригодны также для модели Раша. Частные производные для трёхпараметрической модели будем определять численно.

### Частные производные целевой функции по методу наименьших квадратов

Запишем также частные производные критерия подбора параметров по методу наименьших квадратов.

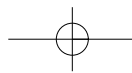
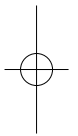
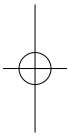
### Теория

$$\frac{\partial F}{\partial \theta_i} = \frac{\partial \left( \sum_{j=1}^m \left( x_{ij}^2 - 2x_{ij} \left( c_j + \frac{(1-c_j)}{1 + e^{-a_j(\theta_i - \beta_j)}} \right) + \left( c_j + \frac{(1-c_j)}{1 + e^{-a_j(\theta_i - \beta_j)}} \right)^2 \right) \right)}{\partial \theta_i} =$$

$$= \frac{\partial \left( \sum_{j=1}^m \left( x_{ij}^2 - 2x_{ij} c_j - \frac{2x_{ij}(1-c_j)}{1 + e^{-a_j(\theta_i - \beta_j)}} + c_j^2 + \frac{2c_j(1-c_j)}{1 + e^{-a_j(\theta_i - \beta_j)}} + \frac{(1-c_j)^2}{(1 + e^{-a_j(\theta_i - \beta_j)})^2} \right) \right)}{\partial \theta_i} =$$

$$= (1-c_j) \sum_{j=1}^m \frac{2x_{ij} \cdot e^{-a_j(\theta_i - \beta_j)} a_j}{(1 + e^{-a_j(\theta_i - \beta_j)})^2} + (1-c_j) \sum_{j=1}^m \frac{2c_j \cdot e^{-a_j(\theta_i - \beta_j)} a_j}{(1 + e^{-a_j(\theta_i - \beta_j)})^2} + (1-c_j)^2 \sum_{j=1}^m \frac{2e^{-a_j(\theta_i - \beta_j)} a_j}{(1 + e^{-a_j(\theta_i - \beta_j)})^3},$$

$$\frac{\partial F}{\partial \theta_i} = (1-c_j) \sum_{j=1}^m \frac{2 \cdot e^{-a_j(\theta_i - \beta_j)} (-a_j)}{(1 + e^{-a_j(\theta_i - \beta_j)})^2} \left( x_{ij} - c_j - \frac{1-c_j}{1 + e^{-a_j(\theta_i - \beta_j)}} \right), \quad (16)$$





ПЕД	
	измерения

$$\frac{\partial F}{\partial \beta_j} = -(1-c_j) \sum_{i=1}^n \frac{2x_{ij} e^{-a_j(\theta_i - \beta_j)} a_j}{(1 + e^{-a_j(\theta_i - \beta_j)})^2} + (1-c_j) \sum_{i=1}^n \frac{2c_j e^{-a_j(\theta_i - \beta_j)} a_j}{(1 + e^{-a_j(\theta_i - \beta_j)})^2} + (1-c_j)^2 \cdot \sum_{i=1}^n \frac{2e^{-a_j(\theta_i - \beta_j)} a_j}{(1 + e^{-a_j(\theta_i - \beta_j)})^3} = (1-c_j) \sum_{i=1}^n \frac{2 \cdot e^{-a_j(\theta_i - \beta_j)} (-a_j)}{(1 + e^{-a_j(\theta_i - \beta_j)})^2} \left( x_{ij} - c_j - \frac{1-c_j}{1 + e^{-a_j(\theta_i - \beta_j)}} \right), \quad (17)$$

$$\frac{\partial F}{\partial a_j} = -(1-c_j) \sum_{i=1}^n (-1) \cdot \frac{2x_{ij} \cdot e^{-a_j(\theta_i - \beta_j)} \cdot (\beta_j - \theta_i)}{(1 + e^{-a_j(\theta_i - \beta_j)})^2} + (1-c_j) \cdot \sum_{i=1}^n (-1) \frac{2c_j \cdot e^{-a_j(\theta_i - \beta_j)} \cdot (\beta_j - \theta_i)}{(1 + e^{-a_j(\theta_i - \beta_j)})^2} + (1-c_j)^2 \sum_{i=1}^n (-2) \frac{e^{-a_j(\theta_i - \beta_j)} \cdot (\beta_j - \theta_i)}{(1 + e^{-a_j(\theta_i - \beta_j)})^3} = (1-c_j) \sum_{i=1}^n \frac{2 \cdot e^{-a_j(\theta_i - \beta_j)} \cdot (\beta_j - \theta_i)}{(1 + e^{-a_j(\theta_i - \beta_j)})^2} \left( x_{ij} - c_j - \frac{1-c_j}{1 + e^{-a_j(\theta_i - \beta_j)}} \right). \quad (18)$$

Формулы (16)–(18) пригодны для всех трёх основных моделей IRT (МТИ). Нетрудно заметить, что второй множитель в частных производных (16)–(18) одинаков и равен разности между результатом выполнения задания и расчётной вероятностью правильного ответа.

Формулы (13)–(18) наряду с условием (8) и целевыми функциями (7) и (12) образуют математическую основу для поиска параметров модели методом BFGS. При использовании модели Раша в формулы (13), (14), (16) и (17) подставляем  $a_j = 1$ ; производные целевой функции по дифференцирующей способности заданий равны нулю.

### Концепция компьютерной программы

В качестве основы создаваемой программы выбран Microsoft Excel, что обеспечивает:

1) лёгкость освоения программы (пользователь работает в знакомой среде, с привычными меню и панелями инструментов);

2) удобство ввода и вывода данных;

3) возможность импорта данных разных форматов (Microsoft Word, файлы баз данных, текстовые и HTML-файлы и др.).

Поскольку встроенное в Excel средство поиска реше-

ния ограниченно пригодно для подбора параметров модели педагогических измерений, то функциональность можно расширить с помощью встроенного в программы пакета Microsoft Office языка программирования Visual Basic for Applications (VBA). VBA практически не уступает другим современным средствам программирования, существенно увеличивая и без того огромные возможности Microsoft Office.

Для выбора метода поиска параметров модели IRT протестирован метод покоординатного спуска, BFGS, а также методы наискорейшего спуска и сопряжённых градиентов. Расчёты проведены при числе параметров модели  $k=50$ ; 1000 и 10000 (см. таблицу).

Результаты свидетельствуют о превосходстве метода BFGS: при равной точности затраты времени минимальны. Метод покоординатного спуска

Теория

370000

Таблица 1. Анализ методов поиска экстремума

Метод поиска	Время поиска решения, с					
	модель Раша		двухпараметрическая модель		трёхпараметрическая модель	
	метод НК	метод МП	метод НК	метод МП	метод НК	метод МП
<b>Число параметров модели <math>k = 50</math></b>						
Покоординатный спуск	0,5	0,5	5	10	5	11
BFGS	0,6	0,3	0,7	1,2	2	5
Наискорейший спуск	2	1	8	6	8	8
Сопряженных градиентов	1	1	6	4	6	5
<b>Число параметров модели <math>k = 1000</math></b>						
Покоординатный спуск	10	5	19	99	20	85
BFGS	8	3	15	12	18	90
Наискорейший спуск	24	8	37	96	36	100
Сопряженных градиентов	17	5	22	34	25	35
<b>Число параметров модели <math>k = 10000</math></b>						
Покоординатный спуск	105	37	174	955	181	932
BFGS	64	36	134	176	144	422
Наискорейший спуск	90	254	712	848	814	827
Сопряженных градиентов	74	197	626	787	612	664

ПЕД  
измерения

близок по быстродействию к BFGS при подборе параметров модели Раша, но существенно уступает при двухпараметрической модели. Градиентные методы уступают BFGS по всем позициям.

По результатам испытаний для поиска экстремума принят метод BFGS.

Программная реализация метода BFGS обеспечивает достаточно высокую точность расчётов:

- ошибка вычислений не превышает  $10^{-15}$  (параметры модели и производные заданы десятичными числами двойной точности);
- абсолютная ошибка определения точки экстремума не превышает 0,001 по любой из координат (т.е. для всех параметров модели гарантируется точность до четвертого знака после запятой).

Следует понимать, что точность определения точки экстремума и точность оценки параметра — это

разные величины. Программа выдаёт наиболее вероятные значения параметров модели. Неизвестные реальные значения этих параметров могут отличаться от расчётных. Это отличие оценивается величиной стандартной ошибки  $\sigma$  (чем меньше  $\sigma$ , тем выше точность определения искомого параметра).

Программа определяет стандартные ошибки по формулам<sup>9</sup>:

$$\sigma_{\theta_i} = \frac{1}{\sqrt{\sum_{j=1}^m a_j^2 \left[ \left( \frac{1-P_{ij}}{P_{ij}} \right) \left( \frac{P_{ij}-c_j}{1-c_j} \right)^2 \right]}}, \quad (19)$$

$$\sigma_{\beta_j} = \frac{1}{\sqrt{\sum_{i=1}^n a_j^2 \left[ \left( \frac{1-P_{ij}}{P_{ij}} \right) \left( \frac{P_{ij}-c_j}{1-c_j} \right)^2 \right]}}. \quad (20)$$

где  $\sigma_{\theta_i}$  — стандартная ошибка уровня подготовленности  $i$ -го испытуемого;  $\sigma_{\beta_j}$  — стандартная ошибка уровня трудности  $j$ -го задания.

Кроме того, рассчитывается степень близости модели и исходных данных как средняя вероятность фактически полученных результатов выполнения тестовых заданий:

$$\bar{p} = \frac{\sum_{i=1}^n \sum_{j=1}^m P_{ij}}{n \cdot m} = \frac{\sum_{i=1}^n \sum_{j=1}^m \left( c_j x_{ij} + (1-c_j) \frac{e^{x_{ij} a_j (\theta_i - \beta_j)}}{1 + e^{a_j (\theta_i - \beta_j)}} \right)}{n \cdot m}. \quad (21)$$

### Краткая характеристика компьютерной программы

**Название:** Estimate2PL.

**Назначение:** подбор параметров основных моделей IRT — модели Г. Раша,

9

*Partchev I.*  
A visual guide to item  
response theory. Jena:  
Friedrich-Schiller-  
Universitat, 2004. 61 p.

двух- и трёхпараметрической модели.

**Условия распространения:** freeware (бесплатное программное обеспечение, распространяемое без исходных кодов).

**Операционная среда:** операционная система Windows; Microsoft Excel 97-2007.

**Инсталляция** не нужна, достаточно скопировать программу в любую папку.

**Объём программы:** 308 кб (zip архив 89 кб).

**Автор:** Деменчёнок О.Г.

## Работа с программой Estimate2PL

Рекомендуемая последовательность подбора параметров модели педагогических измерений состоит из следующих шагов.

**1. Копирование программы на компьютер.** Компьютерная программа Estimate2PL выложена на сайт автора этой статьи [www.asksystem.narod.ru](http://www.asksystem.narod.ru) и доступна для скачивания. Полученный файл Estimate2PL.zip нужно распаковать в любую папку.

**2. Ввод данных.** Загрузить Microsoft Excel и открыть файл Estimate2PL.xls. Используя стандартные средства Microsoft Excel можно:

- ввести данные с клавиатуры;
- вставить скопированные данные (стандартными средствами

Excel можно вставить данные из другой книги Excel, из документа Microsoft Word и т.д.);

- импортировать данные из файлов других форматов, для чего выбрать команду *Файл – Открыть* и указать нужный тип файлов (файлы баз данных, текстовые файлы, HTML и др.).

Данные вводятся на лист Пример1, Пример2 или Пример3 в виде таблицы, столбцы которой – результаты выполнения заданий, а строки – результаты испытуемых (рис. 1). На указанных листах имеются примеры исходных таблиц, облегчающие понимание правил ввода данных. Результат может принимать одно из двух значений: 1 («правильно») или 0 («неправильно»). Допускаются пропуски (задание не выполнялось). Любые значения, кроме 0 и 1, интерпретируются как пропуск задания. В первой строке записываются идентификаторы заданий (например, номера заданий); в столбце А – идентификаторы испытуемых (номера или фамилии). В случае использования трёхпараметрической модели дополнительно нужно ввести значения параметра коррекции на угадывание правильного ответа (рис. 2). Эти значения вводятся в строке, следующей за последним результатом тестирования. Форматирование (шрифт, выделение цветом, рамки и т.п.) на процесс подбора параметров не влияет.

Теория

39/0000

**ПЕД**  
измерения

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		Задание 4	Задание 5	Задание 6	Задание 7	Задание 8	Задание 9	Задание 10	Задание 11	Задание 12	Задание 13	Задание 14	Задание 15	Задание 16	Задание 17	
2	Студент 25	0	1	1	0	0	0	0	0	0	0	0	0	0	0	
3	Студент 4	1	0	0	1	0	1	0	0	0	0	0	0	0	0	
4	Студент 33	1	0	0	1	0	0	1	0	0	0	0	0	0	0	
5	Студент 1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	
6	Студент 27	1	1	1	1	0	0	0	0	0	0	0	0	0	0	
7	Студент 11	0	1	1	1	1	1	0	0	0	0	0	0	0	0	
8	Студент 12	1	1	0	1	0	1	1	0	0	0	0	0	0	0	
9	Студент 17	1	0	1	1	1	1	1	0	0	0	0	0	0	0	
10	Студент 19	1	1	1	1	1	1	0	0	0	0	0	0	0	0	
11	Студент 30	1	1	1	1	1	1	0	0	0	0	0	0	0	0	

Рис. 1. Ввод исходных данных

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
28	Студент 32	1	1	1	1	1	1	1	1	0	0	0	0	0	0	
29	Студент 21	1	1	1	1	1	1	1	1	0	1	0	0	0	0	
30	Студент 22	1	1	1	1	1	1	1	1	1	0	0	0	0	0	
31	Студент 34	1	1	1	1	1	1	1	0	1	0	1	0	0	0	
32	Студент 15	1	1	1	1	1	1	1	1	1	1	0	0	0	0	
33	Студент 23	1	1	1	1	1	1	1	1	1	1	1	0	0	0	
34	Студент 7	1	1	1	1	1	1	1	1	1	1	0	1	0	0	
35	Студент 24	1	1	1	1	1	1	1	1	0	1	0	0	1	1	
36		0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	

Рис. 2. Ввод значений параметра коррекции на угадывание правильного ответа для трёхпараметрической модели (строка 36)

**Подбор параметров модели.** Следует разрешить выполнение макросов с помощью команды меню *Сервис-Макрос-Безопасность*. Затем нажать кнопку *Поиск решения* (рис. 3) и выбрать режим работы (рис. 4).

K	L	M	N	O	P	Q	R	S	T
Задание 13	Задание 14	Задание 15	Задание 16	Задание 17					
0	0	0	0	0					
0	0	0	0	0					
0	0	0	0	0					
0	0	0	0	0					
0	0	0	0	0					
0	0	0	0	0					
0	0	0	0	0					

Поиск решения

Рис. 3. Запуск поиска решения

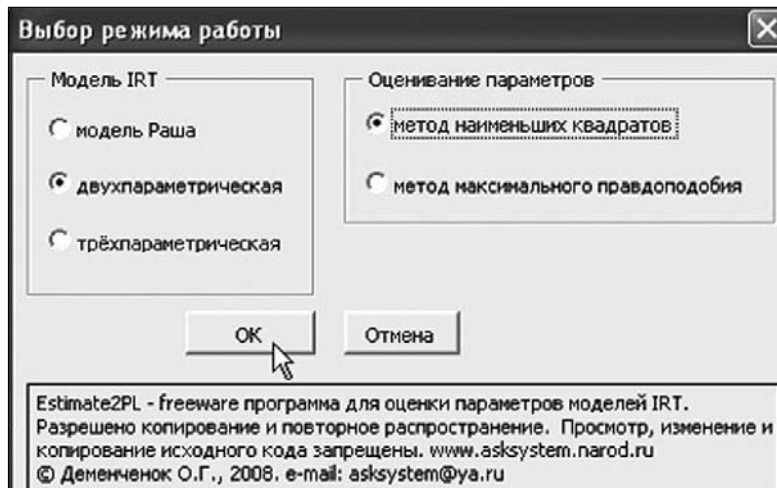


Рис. 4. Выбор режима работы

Теория  
Методы

4. **Получаемые результаты** способность заданий; степень (уровни подготовленности испытуемых, уровни трудности заданий и их стандартные ошибки; дифференцирующая

способность заданий; степень близости модели и исходных данных) записываются в таблицу с исходными данными

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1		Задание 4	Задание 5	Задание 6	Задание 7	Задание 8	Задание 9	Задание 10	Задание 11	Задание 12	Задание 13	Задание 14	Задание 15	Задание 16	Задание 17		$\theta$	$\sigma$	
2	Студент 25	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	-3,7	0,2	
3	Студент 4	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	-4,1	0,2	
4	Студент 33	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	-0,6	0,4	
5	Студент 1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	-3,6	0,3	
6	Студент 27	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	-3,6	0,3	
7	Студент 11	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	-2,6	1,1	
8	Студент 12	1	1	0	1	0	1	1	0	0	0	0	0	0	0	0	-0,4	0,9	
9	Студент 17	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	-0,4	0,9	
10																	-1,2	0,3	
23																	-0,4	0,3	
24	Студент 10	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1,3	0,4	
25	Студент 14	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1,3	0,4	
26	Готово																		
27	Поиск решения завершен, результаты подбора параметров записаны на лист MS Excel.																		
28	OK																		
34	Студент 7	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	4,3	0,3	
35	Студент 24	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	4	0,3	
37	$\beta$	-5	-3,9	-3,9	-5,5	-3,1	-3,3	-0,9	1	3,1	2,7	6,6	4,1	4,6	4,6				
38	$\sigma$	1,1	0,2	0,2	1,2	0,8	0,4	0,2	0,2	0,8	0,2	1,6	0,2	0,8	0,8				
39	$a$	0,6	10	10	0,7	10	10	10	10	0,7	10	0,4	10	1,4	1,4				
40	Степень близости модели и исходных данных равна 91,6%. Время поиска решения 1 с																		

Рис. 5. Вывод результатов подбора параметров

ПЕД	
	измерения

Уровни подготовленности испытуемых  $\theta$  и соответствующие стандартные ошибки  $s$  записываются в столбцах справа от исходных данных (столбцы Q и R на рис. 5). Уровни трудности заданий  $\beta$ , их стандартные ошибки  $\sigma$ , дифференцирующая способность заданий  $a$  и степень близости модели и исходных данных записываются в строках ниже исходных данных (строки 37–40 на рис. 5).

Таким образом, для поиска параметров моделей IRT нужно ввести результаты тестирования и нажать кнопку *Поиск решения*. Всё остальное программа Estimate2PL выполнит автоматически. Это существенно проще, чем в других известных автору компьютерных программах.

Автор надеется, что данная работа будет способствовать внедрению моделей IRT в педагогическую практику.