А.П. Попов, Т.Ю. Попова, С.Ю. Акулов, Д.И. Цхяев
Педагогический институт Южного федерального университета (г. Ростов-на-Дону)

О ПРИНЦИПИАЛЬНО НОВОМ НАПРАВЛЕНИИ В ТЕОРИИ ТЕСТИРОВАНИЯ

Математика, естественные науки и методика их преподавания

Большинство систем компьютерного тестирования в настоящее время используют в качестве теоретической основы параметрические модели Раша и Бирнбаума [1; 2; 3; 4]. Однако, как показывает анализ [5; 6], эти модели внутренне противоречивы, что не позволяет использовать их для адекватного описания процесса тестирования.

В качестве альтернативы существующим параметрическим моделям предлагается принципиально новая модель [7; 8; 9], в которой тестирование изначально рассматривается как протекающий во времени реальный процесс. В основе модели лежит единственное, но довольно естественное предположение о том, что поиск решения задания является однородным во времени случайным процессом.

1. Краткое описание модели. Однородные во времени случайные процессы относятся к классу безгранично делимых распределений [10; 11], к которому принадлежит, в частности, гамма-распределение с плотностью вероятности:

$$f(\alpha,\lambda,t)=\frac{(\lambda t)^{\alpha-1}}{\Gamma(\alpha)}\cdot e^{-\lambda t}\,\lambda\,. \tag{1}$$

В нашем случае безразмерный параметр α отождествляется с трудностью задания, а имеющий размерность обратного времени параметр λ — с уровнем подготовленности испытуемого. На рис. 1 показаны графики плотности гамма-распределения для нескольких значений трудности заданий при фиксированном уровне подготовленности. По оси абсцисс откладывается время выполнения заданий (в секундах). Как видно из графика, увеличение сложности заданий (при фиксированном уровне подготовленности испытуемых) приводит к смещению максимума распределения в сторону больших времен, при этом максимум становится более размытым.

$$\lambda=0.2\ сек^{-1} \text{ и } \alpha=1,2,3,4\,.$$

Гамма-распределение удовлетворяет интегральному уравнению типа свертки (дельта-функция введена в уравнение, чтобы придать ему наиболее симметричный вид):

$$f(\alpha'+\alpha'',\lambda,t)=\int_0^t\int_0^t f(\alpha',\lambda,t')f(\alpha'',\lambda,t'')\delta(t-t'-t'')dt'dt'' \tag{2}$$

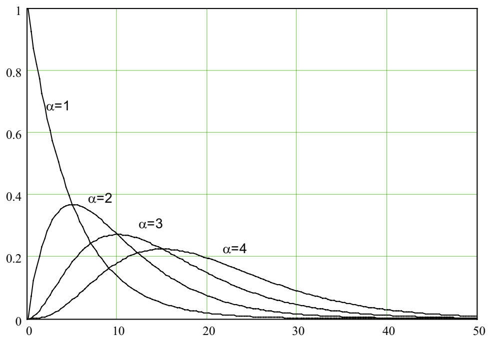

Рис. 1. Графики плотности гамма-распределения при значениях параметров

Смысл уравнения состоит в том, что если t' — время, необходимое для правильного решения задания трудности α', а t'' — время, необходимое для правильного решения задания трудности α'', то сумма этих величин t = t' + t'' совпадает со временем, которое требуется для решения задания суммарной трудности α = α' + α''.

2. Процедура обработки данных тестирования. Опишем основанную на предлагаемой модели процедуру обработки, позволяющую по результатам тестирования оценить сложность тестовых заданий и уровень подготовленности испытуемых.

Введем следующие обозначения:

N — общее число студентов, прошедших тестирование;
n — число заданий в индивидуальных тестах;
$t_{i,j}$ — время, затраченное студентом на выполнение задания;
$\chi_{i,j}$ — функция правильности, принимающая значение 1 или 0 в зависимости от того, правильно или неправильно студент j выполнил задание i.

Воспользуемся принципом максимального правдоподобия, согласно которому оценки латентных параметров модели определяются из условия максимума функции правдоподобия:

$$F(\alpha,\lambda)=\prod_{i=1}^{n}\prod_{j=1}^{N}(\chi_{i,j}\cdot f(\alpha_i,\lambda_j,t_{i,j})+1-\chi_{i,j}), \tag{3}$$

или логарифмической функции правдоподобия:

$$\ln(F(\alpha,\lambda))=\sum_{i=1}^{n}\sum_{j=1}^{N}\chi_{i,j}\ln(f(\alpha_i,\lambda_j,t_{i,j}))\,. \tag{4}$$

Необходимые условия максимума функции (4) приводят к системе нормальных уравнений правдоподобия:

$$\psi(\alpha_i)=\frac{\sum_{j=1}^{N}\chi_{i,j}\ln(\lambda_j t_{i,j})}{N_i}\ ,\quad i=1,2,\ldots n \tag{5}$$

$$\lambda_j=\frac{\sum_{i=1}^{n}\chi_{i,j}\alpha_i}{\sum_{i=1}^{n}\chi_{i,j}t_{i,j}}\ ,\quad j=1,2,\ldots N\,, \tag{6}$$

служащих основой определения значений трудности тестовых заданий и уровня подготовленности тестируемых.

В уравнение (5) входит пси-функция Эйлера:

$$\psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}. \tag{7}$$

Помимо этого, в уравнении (5) введено обозначение:

$$N_i = \sum_{j=1}^{N} \chi_{i,j} \tag{8}$$

общего числа студентов, правильно выполнивших задание i.

Смысл уравнения (6) предельно прост: уровень подготовленности испытуемого оценивается как отношение суммарной трудности правильно выполненных заданий к времени, затраченному на их решение.

Назовем безразмерную величину $\tau_{i,j} = \lambda_j \cdot t_{i,j}$ приведенным временем, затраченным студентом j на выполнение задания i. Тогда уравнение (5) означает, что пси-функция трудности задания совпадает со средним значением логарифма приведенного времени, затраченного испытуемыми на поиск его правильного решения.

Уравнения (5—6) допускают решение методом итераций, причем точность решения порядка 0,1 % достигается при 10—25 итерациях. Укажем, как следует выбирать стартовые значения трудности заданий и уровня подготовленности испытуемых. Используя формулы для моментов гамма-распределения [11]:

$$I(\alpha,\lambda,\nu) = \int_0^\infty f(\alpha,\lambda,t)\, t^\nu\, dt = \frac{\Gamma(\alpha+\nu)}{\Gamma(\alpha)\lambda^\nu}, \tag{9}$$

нетрудно получить явные выражения для математического ожидания времени решения тестовых заданий и его дисперсии:

$$\langle t \rangle = \frac{\alpha}{\lambda} \qquad D = \langle t^2 \rangle - \langle t \rangle^2 = \frac{\alpha}{\lambda^2}, \tag{10}$$

откуда следует простая оценка трудности тестового задания:

$$\alpha \approx \frac{\langle t \rangle^2}{D}. \tag{11}$$

Наконец, по известным формулам математической статистики [11] найдем выборочные оценки среднего значения и дисперсии времени правильного решения задания i:

$$\langle t \rangle_i \approx \frac{\sum_{j=1}^{N} \chi_{i,j} t_{i,j}}{N_i} \qquad D_i \approx \frac{\sum_{k=1}^{N_i} \chi_{i,j} \left(t_{i,k} - \langle t \rangle_i\right)^2}{N_i - 1}. \tag{12}$$

Набор стартовых значений параметров получается подстановкой выборочных оценок (12) в формулу (11), с последующей подстановкой полученных значений трудности тестовых заданий в формулу (6).

На каждом шаге итерационного процесса найденные ранее значения латентных параметров обновляются согласно следующим формулам:

$$\lambda_j^{new} = \frac{\sum_{i=1}^{n} \chi_{i,j} \alpha_i^{old}}{\sum_{i=1}^{n} \chi_{i,j} t_{i,j}} \qquad j = 1, 2, \ldots N, \tag{13}$$

$$\alpha_i^{new} = \psi^{-1}\left(\frac{\sum_{j=1}^{N} \chi_{i,j} \ln(\lambda_j^{new} t_{i,j})}{N_i} \right) \qquad i = 1, 2, \ldots n \tag{14}$$

В данной итерационной схеме для вычисления обновленных значений трудности заданий используются уже обновленные значения уровня подготовленности испытуемых. Подобный подход не является единственно возможным, могут быть избраны и другие, быть может, более удобные итерационные схемы. Главным критерием пригодности схемы является только обеспечиваемая этой схемой скорость сходимости итерационного процесса.

При избранной нами схеме реализации итерационного процесса в соответствии с формулами (13—14) возникает задача эффективного и достаточно точного вычисления функции, обратной (в функциональном смысле) к пси-функции Эйлера. Как показал опыт, в избранной нами схеме достаточно высокая точность решения (около 0.1 %) в большинстве случаев достигается уже при 20—25 итерациях. В настоящее время проблема скорости сходимости итерационного процесса исследована нами на чисто эмпирическом уровне с привлечением конкретных примеров, но в дальнейшем мы рассчитываем получить более глубокие и общие результаты.

3. Простое обобщение модели. Уже при анализе первых результатов тестирования был установлен любопытный факт: распределение времени поиска не только правильного, но и неправильного решения тестового задания хорошо описывается гамма-распределением (разумеется, с другими значениями параметров):

$$f(\beta,\mu,t) = \frac{(\mu t)^{\beta-1}}{\Gamma(\beta)} \cdot e^{-\mu t} \mu. \tag{15}$$

Параметр β можно отождествить с трудностью поиска неверного решения, а параметр μ — с уровнем неподготовленности испытуемого. Нужно, однако, понимать, что предлагаемая нами интерпретация в данном случае носит условный характер. Возможно, что предлагаемая нами терминология и в целом не вполне удачна, поскольку параметр модели, названный нами уровнем подготовленности испытуемого, характеризует скорость поиска решения в знакомой ситуации и совершенно не учитывает того, что в силу разных обстоятельств с некоторыми темами в тестовых заданиях испытуемый вообще может быть незнаком. По сути дела, этот параметр является неким аналогом быстродействия ЭВМ, но он никак не связан со способностью испытуемого к регулярным систематическим занятиям, к усвоению не отдельных фрагментов, а всего предмета в целом.

Для оценки латентных параметров модели по-прежнему воспользуемся принципом максимального правдоподобия, но теперь учтем возможность выхода на неверное решение на чисто феноменологическом уровне, вводя в функцию правдоподобия в качестве новых параметров вероятности выхода на верное или неверное решение для каждого тестового задания.

Введем следующие обозначения:

N – общее число студентов, прошедших тестирование;

n – число заданий в индивидуальных тестах;

$t_{i,j}$– время, затраченное студентом j на выполнение задания i;

$\tilde{\chi}_{i,j}$ – функция, принимающая значение 1 или 0 в зависимости от того, правильно или неправильно студент j выполнил задание *i ;*

$\tilde{\chi}_{i,j}$ – функция, принимающая значение 0 или 1 в зависимости от того, правильно или неправильно студент выполнил задание *i;*

p_i– вероятность верного решения задания i;

q_i – вероятность неверного решения задания i;

$N_i = \sum_{j=1}^{N} \chi_{i,j}$ – число студентов, верно решивших задание i;

$\tilde{N}_i = \sum_{j=1}^{N} \tilde{\chi}_{i,j}$ – число студентов, неверно решивших задание .

Введенные параметры не являются независимыми, поскольку должны выполняться следующие очевидные соотношения:

$$\tilde{\chi}_{i,j} = 1 - \chi_{i,j} \quad \tilde{N}_i = N - N_i \quad q_i = 1 - p_i \quad i = 1 \ldots n,\ j = 1 \ldots N . \quad (16)$$

Вследствие сделанных выше предположений логарифмическая функция правдоподобия разбивается на сумму двух слагаемых:

$$\ln(F(p,\alpha,\lambda,q,\beta,\mu)) = \ln(F(p,\alpha,\lambda)) + \ln(F(q,\beta,\mu)) , \quad (17)$$

которые соответствуют возможности выхода на верное или неверное решение тестовых заданий:

$$\ln(F(p,\sigma,\lambda)) = \sum_{i=1}^{n} N_i \ln(p_i) + \sum_{i=1}^{n}\sum_{j=1}^{N} \chi_{i,j} \ln(f(\sigma_i,\lambda_j,t_{i,j})) , \quad (18)$$

$$\ln(F(q,\beta,\mu)) = \sum_{i=1}^{n} \tilde{N}_i \ln(q_i) + \sum_{i=1}^{n}\sum_{j=1}^{N} \tilde{\chi}_{i,j} \ln(f(\beta_i,\mu_j,t_{i,j})) . \quad (19)$$

Условия максимума функции (17) приводят к системе нормальных уравнений правдоподобия, из которых следуют естественные формулы для вероятности выхода на верное или неверное решение тестового задания:

$$p_i = \frac{N_i}{N} \quad q_i = \frac{\tilde{N}_i}{N} . \quad (20)$$

Остальные латентные параметры модели определяются из условия максимума редуцированных функций правдоподобия, соответствующих возможности выхода на верное или неверное решение тестовых заданий:

$$\ln(F_{true}(\alpha,\lambda)) = \sum_{i=1}^{n}\sum_{j=1}^{N} \chi_{i,j} \ln(f(\alpha_i,\lambda_j,t_{i,j})) , \quad (21)$$

$$\ln(F_{false}(\beta,\mu)) = \sum_{i=1}^{n}\sum_{j=1}^{N} \tilde{\chi}_{i,j} \ln(f(\beta_i,\mu_j,t_{i,j})) . \quad (22)$$

Отметим, что для определения латентных параметров редуцированных функций правдоподобия (21—22), отвечающих возможности выхода на верное или неверное решение тестовых заданий, получаются однотипные системы нормальных уравнений правдоподобия, которые точно совпадают с уравнениями (5—6) из предыдущего раздела.

4. Выбор рейтинговой оценки. Выходными данными описанной в предыдущем разделе процедуры полной обработки данных тестирования являются:

1) вероятность выхода на верное (неверное) решение тестовых заданий;

2) трудности тестовых заданий;

3) уровень подготовленности испытуемых.

Возникает вопрос: что делать с результатами обработки и как использовать их для определения рейтинга испытуемых? И здесь, как и при решении всех других вопросов, мы применили эмпирический подход, путем многочисленных проб и ошибок медленно продвигаясь к поставленной цели. Вначале наиболее естественной казалась идея использовать в качестве основы рейтинговой оценки сам уровень подготовленности испытуемых, но очень быстро стало ясно, что это далеко не единственный и даже, может быть, не самый важный фактор, который следует учитывать при определении рейтинга испытуемых. Выше уже отмечалось, что сам по себе параметр, ассоциированный нами с уровнем подготовленности испытуемых, является аналогом быстродействия ЭВМ и характеризует скорость, с которой испытуемый решает задания на знакомую ему тему. Ясно, что скорость мышления, способность быстро решать задачи известного типа еще не определяют подлинный уровень подготовленности. Ведь часто даже очень способный, от природы одаренный учащийся на поверку оказывается человеком, не способным к регулярному, систематическому освоению учебного материала. Пропуская лекционные и практические занятия, он не усваивает необходимые теоретические знания, не приобретает практических навыков решения задач. При компьютерном тестировании такой студент может получить относительно низкую рейтинговую оценку, которая, тем не менее, является вполне объективной и заслуженной.

Были испробованы самые разнообразные варианты конструирования рейтинговой оценки по результатам тестирования (все попытки и все варианты описывать нет смысла). Долгое время мы использовали две рейтинговые оценки. Основой первой оценки служила суммарная сложность правильно решенных заданий:

$$\gamma_j \approx \sum_{i=1}^{n} \chi_{i,j}\alpha_i \ , \ j = 1,2,\dots N . \qquad (23)$$

В основу второй оценки была положена скорость выполнения тестовых заданий с учетом их сложности:

$$\tilde{\gamma}_j \approx \sum_{i=1}^{n} \chi_{i,j}\alpha_i \frac{\langle t \rangle_i}{t_{i,j}} \ , \ j = 1,2,\dots N \cdot \qquad (24)$$

Обе оценки приводились затем к 100-балльной рейтинговой шкале нормировкой на лучший результат в группе, что, в частности, оказывается удобным при переводе оценок в 5-балльную шкалу. Было установлено, что между этими оценками существует корреляционная связь, иногда довольно тесная, но, в конце концов, был сделан вывод, что наиболее объективным отражением достижений учащегося служит именно оценка за суммарную трудность правильно решенных заданий. К этому выводу мы пришли, анализируя данные деканатов, а также текущие оценки, выставляемые по данному предмету преподавателями, непосредственно работающими в группах.

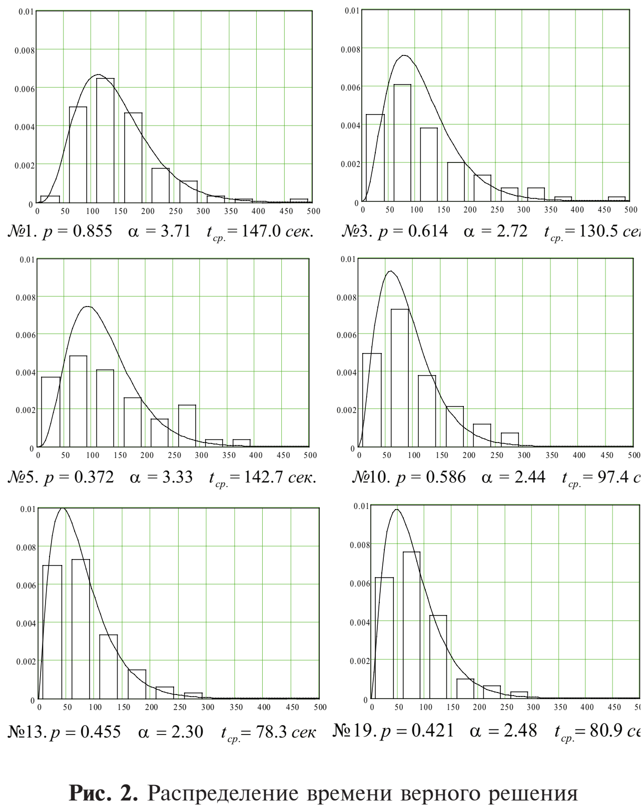

Рис. 2. Распределение времени верного решения тестовых заданий

5. Эмпирическая проверка адекватности модели. Предлагаемая нами модель тестирования допускает прямую опытную проверку. В частности, возможно:

1) сравнение теоретического и эмпирического распределения времени решения тестовых заданий;

2) проверка гипотезы об аддитивности трудности составных тестовых заданий;

3) проверка гипотезы о нормальном законе распределения рейтинговых оценок.

Приведем результаты подобной проверки на примере тестирования по элементарной математике, в котором приняли участие 145 студентов первых четырех курсов отделений физики, математики и информатики ПИ ЮФУ, проведенного в осеннем семестре 2006 г.

Сравнение теоретического и эмпирического распределения времени решения тестовых заданий. Напомним, что в основе модели лежит предположение о том, что распределение времени решения тестовых заданий описывается формулой (1). Построив гистограммы эмпирического распределения времени решения тестовых заданий по результатам тестирования, можно сравнить их с теоретической зависимостью.

На рис. 2 приведены результаты подобного сравнения для нескольких произвольно выбранных тестовых заданий. В подписи под каждым графиком указаны номер тестового задания, вероятность и трудность верного решения задания, а также среднее время решения.

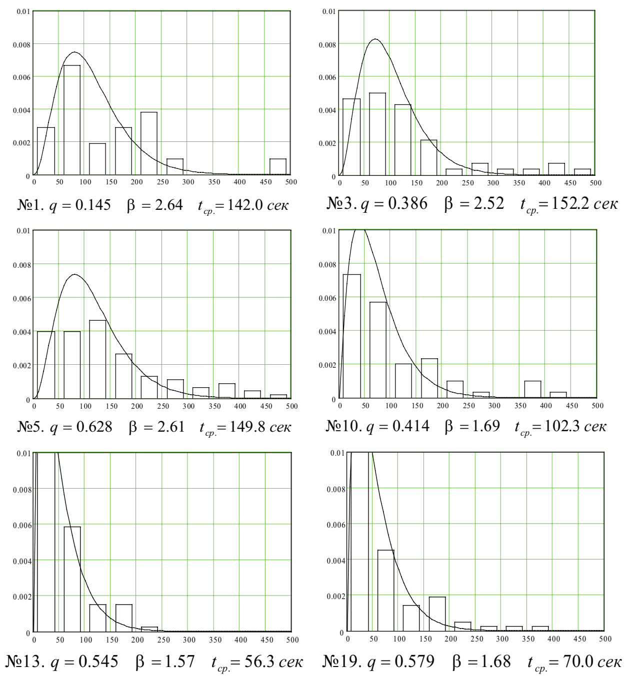

Рис. 3. Распределение времени неверного решения тестовых заданий

Интересно отметить, что распределение времени поиска неверного решения тестовых заданий подчиняется той же закономерности, что и процесс поиска верного решения (с другими значениями параметров).

На рис. 3 представлены результаты сравнения эмпирического и теоретического распределения времени поиска неверного решения для тех же самых тестовых заданий, при этом в подписи под каждым графиком указаны номер тестового задания, вероятность и трудность неверного решения задания, а также необходимое для этого среднее время.

Проверка гипотезы об аддитивности трудности составных тестовых заданий. Первоначально для проверки гипотезы аддитивности предполагалось создать специальную базу тестовых заданий, в которую должны были входить задания двух типов: достаточно простые задания-кирпичики и сложные, составленные из простых. В силу целого ряда объективных причин эта идея оказалась практически нереализуемой. В связи с этим начались поиски иных способов проверки гипотезы аддитивности.

После довольно длительного периода проб и ошибок, путем прямого экспериментирования с обработанными результатами тестирования, удалось разработать метод проверки гипотезы об аддитивности трудности составных заданий. Метод основан на идее виртуального тестирования. Как проводится подобное тестирование? Вначале выбирается произвольная пара заданий, затем к исходному тесту добавляется сложное задание, составленное из этой пары. Трудность добавленного задания оценивалась двумя независимыми способами: суммированием трудности заданий-кирпичиков и по результатам обработки данных виртуального тестирования.

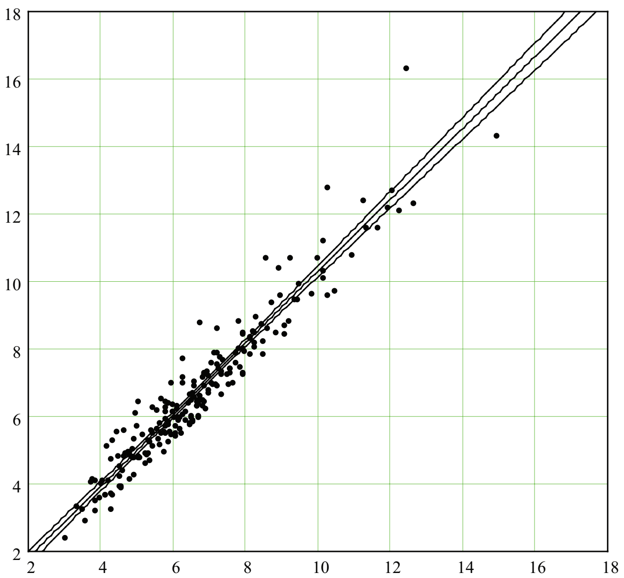

Рис. 4. Регрессионная зависимость между двумя оценками трудности составных заданий

На рис. 4 представлены результаты сравнения двух оценок трудности составных тестовых заданий: по оси абсцисс откладывается оценка x, полученная прямым суммированием трудностей исходных заданий, а по оси ординат – оценка y, полученная в результате обработки данных виртуального тестирования.

Уравнение регрессии в канонической форме:

$$\frac{y-6.7359}{2.2245}=0.9598\cdot\frac{x-6.6398}{2.0112}. \tag{25}$$

Значение коэффициента корреляции очень близко к 1, что указывает на тесную корреляционную связь между двумя независимыми оценками трудности составных заданий.

Запишем уравнение регрессии в виде линейной зависимости:

$$y=ax+b. \tag{26}$$

По хорошо известным формулам математической статистики [11] легко определяются не только средние значения коэффициентов линейной регрессии $a=1.0616$ $b=-0.3127$, но и их среднеквадратичные отклонения $\sigma(a)=0.0215$ $\sigma(b)=0.1494$. Теперь уже нетрудно, задавшись некоторым значением уровня значимости, найти также доверительные интервалы для коэффициентов линейной регрессии.

При уровне значимости $\alpha=0.05$ и числе степеней свободы $n = 210$ можно воспользоваться асимптотическим значением коэффициента Стъюдента $t_{n-2,1-\alpha}=1.96$ и найти доверительные интервалы для коэффициентов линейной зависимости:

$$a=1.0616\pm0.0422 \quad b=-0.3127\pm0.2927\cdot \tag{27}$$

Поскольку значения коэффициентов $\alpha = 1$ и $b = 2$ попадают в данный доверительный интервал, то гипотезу об аддитивности трудности составных заданий можно считать статистически обоснованной с надежностью не ниже $1-\alpha=0.95$.

Проверка гипотезы о нормальном законе распределения рейтинговых оценок. Большинство теоретиков и практиков тестирования твердо уверены, что при корректном выборе рейтинговой шкалы распределение оценок должно описываться нормальным законом. Хотя мы и не вполне разделяем эту уверенность, но от проверки этой гипотезы отказываться все же не стали.

На рис. 5 показано распределение оценок за общую трудность правильно выполненных заданий. Для сравнения показан также график плотности нормального распределения с тем же математическим ожиданием и дисперсией. Применяя критерий Колмогорова, можно показать, что гипотеза о нормальном распределении оценок статистически обоснована с надежностью не ниже 0.90.

Приобретенный в течение пяти лет непрерывной и целенаправленной работы опыт убеждает в том, что оценка за суммарную сложность правильно решенных заданий объективно показывает уровень подготовленности и может использоваться при определении рейтинга испытуемых.

Так, в рейтинг-листах тестирования по элементарной математике деканы факультетов «Физика» и «Математика и информатика» сочли сомнительными не более 4 оценок из общего числа (145 — менее 3%). Можно также отметить, что, по мнению преподавателей этих факультетов, во всех группах рейтинг-листы возглавили объективно лучшие студенты.

Совпадение предсказываемых моделью теоретических зависимостей с эмпирическими распределениями дает основания считать, что модель адекватно описывает процесс тестирования и позволяет достаточно объективно оценивать трудность тестовых заданий и уровень подготовленности тестируемых. Работа по обобщению, развитию и совершенствованию модели продолжается.

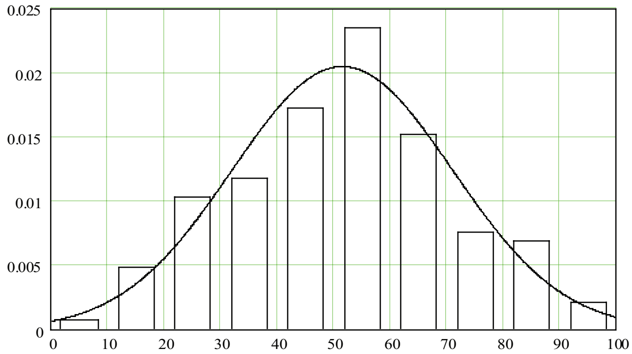

Рис. 5. Распределение оценок за суммарную трудность правильно выполненных заданий

Вместе с тем авторы далеки от мысли, что тестирование может служить универсальным средством оценки знаний, а тем более способностей учащихся. Иной раз приходится слышать, что тестирование способно полностью заменить традиционные формы контроля или что его можно использовать и как инструмент измерения способностей учащихся, и даже как средство обучения и самообучения студентов. Ниже перечислен ряд причин, по которым согласиться с этим мнением трудно.

1. Содержание многих дисциплин практически не допускает представления в форме тестовых заданий (в качестве примера можно привести такие дисциплины, как живопись, дизайн или начертательная геометрия).
2. Все испытуемые должны быть поставлены в равные условия, а поэтому генерируемые клиентской программой индивидуальные тесты должны иметь приблизительно одинаковую трудность. Для этого, в свою очередь, нужно обеспечить одинаковую трудность заданий в пределах одного блока, что представляется весьма проблематичным для большинства дисциплин гуманитарного цикла (история, литература, экономика, философия, педагогика).
3. Можно представить себе обучающую программу со встроенными средствами тестового контроля, и даже написать демоверсию программы, но создать полноценную обучающую программу по всем разделам даже одной дисциплины кажется нам непосильной задачей.
4. Наконец, нужно понимать, что любая, даже самая совершенная система автоматического оценивания уровня подготовленности студентов по данным тестирования оперирует с формальными, чисто внешними признаками (правильность ответов, время решения и т.д.) глубоко скрытого процесса поиска решения. Поэтому здесь уместнее говорить не об объективности полученных оценок, а лишь об объективизации самого процесса оценивания.

Таким образом, на долю тестирования (в какой бы форме оно не проводилось) остается входной, текущий и выходной контроль знаний, причем лишь для круга дисциплин, допускающих достаточно глубокую формализацию излагаемого материала. Это, прежде всего, математика, физика, информатика, дисциплины естественнонаучного цикла, некоторые разделы лингвистики и психологии. Но и здесь тестирование следует использовать лишь как некое сито, служащее для предварительного отсева студентов, не усвоивших необходимого минимума знаний по данному предмету.

Литература

1. Rasch G. Probabilistic Model for Some Intelligence and Attainment Tests. Chicago: Univ. of Chicago Press, 1980.
2. Wright B. D., Stone M. N. Best Test Design. Chicago: MESA Press, 1979.
3. Wright B. D., Masters G. N. Rating scale analysis. Rasch measurements. Chicago: MESA Press, 1982.
4. Lord F. M., Novic M. Statistical Theories of Mental Test Scores. Mass.: Addison-Wesley Publisher Co. Reading, 1968.
5. Попов А.П. Критический анализ параметрических моделей Раша и Бирнбаума // Инновационные методы и средства оценки качества образования: материалы 6-й Междунар. науч. конф., М.: Изд-во МГУП, 2006. С. 231—235.
6. Попов А.П. О параметрических моделях тестирования // Инновационные методы и средства оценки качества образования: материалы 6-й науч. Междунар. конф. М.: Изд-во МГУП, 2008. С. 122-128.
7. Попов А. П., Богомолов А. А., Попова Л. А. Новая математическая модель тестирования. Наука и образование. 2005. № 3. С. 221 — 228.
8. Попов А.П., Богомолов А.А., Попова Л.А. Поиск решения как однородный во времени случайный процесс. Новая математическая модель тестирования // Инновационные методы и средства оценки качества образования: материалы 6-й Междунар. науч. конф. М.: Изд-во МГУП, 2006. С. 236—247.
9. Попов А.П. Новое направление в теории тестирования // Изв. ЮФУ. Сер.: Педагогические науки. 2008 № 1—2. С.24—31.
10. Гнеденко Б. В. Курс теории вероятностей. М.: Наука, 1988.
11. Справочник по теории вероятностей и математической статистике / В.С. Королюк, Н.И. Портенко, А.В. Скороход [и др.]. М.: Наука, 1985.