

Юрий Николаевич Каргин, преподаватель Ноябрьского колледжа профессиональных и информационных технологий, кандидат физико-математических наук
kargin04@yandex.ru

ЭНТРОПИЯ ЗНАНИЙ. КОНСТРУИРОВАНИЕ МОДЕЛИ ПЕДАГОГИЧЕСКОГО ТЕСТА

Введение

Ранее¹ была высказана идея существования взаимосвязи между теорией информации и теорией педагогических измерений. Выявление этой связи позволило по-новому:

- описать содержательный смысл основных показателей качества педагогических измерений через информационные характеристики;
- выразить традиционные единицы измеряемых величин — логит — через меры информации бит;
- рассмотреть вопрос о классах моделей педагогических измерений и способах их конструирования;
- по-новому взглянуть на вопросы эффективности педагогического теста и способы её оценивания.

В данной работе высказанная идея и её следствия находят своё дальнейшее развитие.

¹ Каргин Ю. Исследование взаимосвязи теории информации и теории педагогических измерений. ПИ №2. 2013. С. 3–22.

Проблемы МТИ

Математическая теория педагогических измерений (МТИ) начинается с достаточно очевидного высказывания — испытуемый с лучшей подготовкой к содержанию тестового задания имеет больше шансов на успех, чем испытуемый с худшей подготовкой. Это высказывание можно привести и в более строгой форме, через понятие вероятности. Вероятность правильного ответа на тестовое задание испытуемым с лучшей подготовкой больше, чем испытуемого с худшей подготовкой.

Дальнейшее построение МТИ проходит через поиск, выбор и обоснование строгих математических правил, реализующих это высказывание, правил вычисления вероятности правильного ответа на тестовое задание, в зависимости от подготовленности испытуемого и трудности тестового задания. Этот этап развития МТИ завершается формулировкой моделей педагогических измерений, анализом их свойств, по которым затем и разрабатываются технологии педагогического тестирования.

Следует отметить, что на некоторых этапах построения МТИ проявляются слабые места и недостаточно убедительные суждения. В частности, переход от исходного высказывания о монотонном росте вероятности правильного ответа с ростом подготовленности испытуемого к построению той или иной математической модели, строго говоря, не обоснован. Обычно это обоснование представлено несколько размытыми рассуждениями о необходимом виде функции, о её свойствах, включая асимптотическое поведение. Иногда математический вид функции обосновывается подбором под эмпирические данные или даже просто представляется как удачная догадка Г. Раша.

Другим «слабым местом» МТИ является отсутствие внятного ответа на вопрос, что, собственно, мы измеряем, «накладывая» ту или иную модель на наблюдаемые данные педагогического тестирования? Как понять связь между подготовленностью испытуемого к тестовому заданию и значениями измеряемых показателей? Тем более, когда шкала измерения знаний в логитах неравномерна и, вообще говоря, зависит от подготовленности испытуемых и трудности тестовых заданий. И понять эту связь не так просто. Но необходимо.

Возможно, и в этих трудностях, в том числе, кроются препятствия более широкого распространения МТИ. Отсюда возникает недоверие к результатам педагогических измерений методами МТИ или игнорирование их отличительных качеств. А иногда и примитивное толкование результатов МТИ как более сложной формы ранжирования испытуемых.

Таким образом, можно выделить две взаимосвязанные проблемы МТИ — убедительное обоснование модели педагогических измерений и внятная интерпретация её параметров. Новое понимание и толкование этих проблем в этой статье предлагается через применение основных идей и положений теории информации к педагогическим измерениям.

Частная и средняя энтропия

К числу исходных и основных понятий теории информации относят понятие энтропии как меры неопределённости системы и понятие информации сообщения как меры снятой этим сообщением неопределённости. Иногда эту логическую связь представляют иначе: если некоторое сообщение приводит к уменьшению неопределённости системы, то такое сообщение несёт информацию о познаваемой системе в количестве уменьшения неопределённости. В качестве единичной меры неопределённости (энтропии) выбрана неопределённость системы, находящейся в одном из двух равновероятных состояний. Энтропия такой системы равна 1 бит (bit, binary unit — двоичная единица).

Полагая, что не каждый читатель знаком с понятиями теории информации и основными правилами расчёта количества информации, рассмотрим их на небольшом модельном примере.

Пример

В школе стало известно, что один из учеников 10а класса победил в престижном международном шахматном турнире. Но кто этот ученик, пока неизвестно, и перед педагога-

ми школы стоит задача определить имя победителя. Рассмотрим возможности правильного решения этой задачи различными педагогами.

Сначала рассмотрим, как директор школы пытается определить имя победителя. Директор не готов точно ответить на этот вопрос, ему известно только то, что в 10а классе 16 учащихся. Для него ученики класса неразличимы, любой ученик этого класса мог стать победителем шахматного турнира. Тогда неопределённость рассмотренной системы можно охарактеризовать количеством равно-возможных вариантов — 16.

Завуч больше осведомлён о классе, ему известно, что в этом классе только 8 юношей. Т.к. один из 8 юношей мог стать победителем турнира, неопределённость системы доходит до 8 возможных вариантов.

Помощник по воспитательной работе по анкетным данным определил, что в этом классе только четверо юношей систематически занимаются спортом, но информации о видах спорта у него нет. Неопределённость системы для помощника по воспитательной работе уменьшилась до 4.

Классный руководитель ближе знаком с увлечениями своих учеников и знает, что

только два ученика класса занимаются шахматами, и только кто-то из них мог стать победителем шахматного турнира. Неопределённость системы доходит до 2.

И, наконец, руководитель шахматного кружка лучше всех подготовлен к ответу на этот вопрос. Он не просто знает этих двух учеников как шахматистов, но и имеет веские основания полагать, что шансы на победу первого из них примерно в три раза выше второго.

Таким образом, единое для всех педагогов задание воспринимается педагогами по-разному. Каждый из них имеет свои знания, которые и помогают правильно решить поставленную задачу. Неопределённость систем «отдельный педагог — задача определить победителя» различна, отсюда и различия в шансах правильно «вычислить» победителя турнира.

Дальнейшие обсуждения информационных свойств рассматриваемых систем удобно сопровождать заполнением таблицы (табл. 1). Первые два столбца таблицы заполняются по приведённым выше рассуждениям.

Пример подобран так, что неопределённости рассматриваемых систем не просто различны, а ещё и для первых систем эти различия одинаковы — количество возмож-

Таблица 1

Информационные характеристики систем «педагог-задача»

Педагог	Количество вариантов N	Вероятность выбора P	Энтропии			
			h_1 , бит	h_0 , бит	$h_1 - h_0$, бит	H , бит
директор	16	1/16	4	0,093	3,907	0,339
завуч	8	1/8	3	0,193	2,807	0,544
пом. по восп. раб.	4	1/4	2	0,415	1,585	0,811
класн. руковод.	2	1/2	1	1	0	1
руковод. кружка	—	3/4	0,415	2	—1,585	0,811

ных вариантов N в каждой последующей системе в 2 раза меньше. Действительно, анализируя второй столбец таблицы, несложно выделить отношения — $16:8=8:4=4:2=2$. Это факт на языке информационного подхода можно сформулировать так — неопределённость каждой последующей системы на 1 бит меньше предыдущей.

Мы не заполнили последний элемент таблицы второго столбца с количеством возможных вариантов для руководителя шахматного кружка. В данном примере эта мера теряет смысл. Однако дальнейшие рассуждения справедливы и для этой строки. Продолжая информационный анализ рассматриваемых систем, запишем выражения для основных величин теории информации — для частной и средней энтропии.

Общее выражение для вычисления частной энтропии некоторого единичного состояния было предложено Хартли: $h = \log_{0,5} P = -\log_2 P$, где P — вероятность наблюдаемого состояния системы. Рассчитаем по этой формуле энтропию состояния «первый ученик — победитель» с позиции руководителя кружка. По его мнению, вероятность выиграть турнир у первого ученика выше второго в три раза, т.е. равна $3/4$ (против $1/4$ второго ученика), тогда энтропия этого состояния равна $h = \log_{0,5} 0,75 = 0,415$.

В более простом, частном случае, когда неопределённость системы задаётся N равнозначными вариантами и вероятность каждого из таких состояний равна $P = 1/N$, формула Хартли принимает вид:

$$h = \log_{0,5} \frac{1}{N} = \log_2 N.$$

Именно по этим формулам рассчитаны вероятности правильного решения задачи педагогами и соответствующие значения частной энтропии h_1 (третий и четвёртый столбцы таблицы 1).

Содержательный смысл значения частной энтропии можно передать как «мера неожиданности» события. Чем меньше вероятность события, тем больше значение частной энтропии и тем неожиданней его появление. Такие толкования становятся более наглядным, если дополнительно рассчитать частные энтропии состояний «победитель турнира определён неправильно» (пятый столбец таблицы). В шестом столбце приведены разности частных энтропий, отражающие на сколько «неожиданней» правильный выбор победителя по сравнению с ошибкой.

Если показатель частной энтропии отражает неопределённость отдельного состояния системы, то средняя энтропия (или просто энтропия) отражает неопределённость всей системы и рассчитывается как среднее значение частных энтропий всех её состояний. Для расчёта этого показателя следует применять формулу Шеннона:

$$H = \sum_i P_i \times \log_{0,5} P_i = -\sum_i P_i \times \log_2 P_i.$$

Когда система может находиться только в одном из двух возможных состояний с вероятностью благоприятного исхода p , то формула Шеннона имеет вид: $H(P) = P \cdot \log_{0,5} P + (1 - P) \cdot \log_{0,5} (1 - P)$.

В последнем столбце таблицы 1 приведены средние значения энтропий системы H для педагогов школы.

В данном примере мы начинали рассматривать понятие энтропии через понимание неопределённости системы по количеству равновероятных вариантов выбора, и даже смоделировали системы с различными мерами неопределённости и рассчитали эти меры. Таким детализированным рассмотрением задачи мы пытались лишь развернуть содержание терминов неопределённость, частная и средняя энтропия, информация. На самом деле, рассматривая вопросы подготовки учебных заданий и анализа ответов на них учащихся, возможности в такой детализации вряд ли представятся. Строгий теоретический расчёт энтропий и других информационных показателей возможен только для «понятных» систем, для систем с заданными вероятностями состояний. Система «испытываемый — тест» к таким не относится. Однако применять положения теории информации можно и без такой детализации.

Информационный подход к вопросам педагогических измерений, как и другие направления МТИ, опирается не на теоретический расчёт, а на экспериментальную оценку вероятности правильного ответа испытуемым. Но в отличие от других направлений МТИ, в основе информационного подхода лежат и анализируются новые, именно информационные показатели.

Таким образом, тестовые испытания как инструмент проведения педагогических измерений, можно рассматривать и с позиций теории информации. В этом случае под системой с количественно оцениваемой неопределённостью (энтропией) следует понимать ис-

пытываемого, выполняющего тестовое задание, а под сообщением, несущим информацию, следует понимать результат тестового испытания. Тогда количество неопределённости, снятой сообщением, и есть то количество информации, которое поступает педагогу для анализа результата тестирования.

Информационные показатели МТИ

На некоторое время отвлечёмся от известных в МТИ показателей трудности тестовых заданий и подготовленности испытуемых, и попытаемся предложить свои варианты решения этой задачи, но уже на основе информационных показателей.

Первое предположение лежит на поверхности. Рассмотрим вариант описания неопределённости системы «испытываемый — тестовое задание» непосредственно показателем частной энтропии состояния с вероятностью P правильного ответа: $h = \log_{0,5} P$. В рассмотренном примере значения этого показателя приведены в четвёртом столбце таблицы 1.

Приведём основные свойства этого показателя:

- при значениях вероятности $0 \leq P \leq 1$ значения частной энтропии h изменяются в интервале от 0 до ∞ ;
- с ростом вероятности появления состояния частная энтропия этого состояния монотонно убывает;
- $h(1) = 0$ — частная энтропия (неопределённость) достоверного состояния равна нулю, $h(0) = \infty$ — частная энтропия невозможного

состояния равна бесконечности, $h(0,5) = 1$ бит — частная энтропия равновероятного дихотомического состояния равна 1. Такое состояние называют состоянием единичной неопределённости;

- свойство аддитивности $h(P_1 \cdot P_2) = h(P_1) + h(P_2)$ для независимых событий.

Если формулу Хартли переписать в виде $P = 0,5^h = \underbrace{0,5 \times \dots \times 0,5}_{h \text{ раз}}$, то состояние со

значением частной энтропии h эквивалентно h -раз совместному состоянию единичной неопределённости. Таким образом, значение показателя частной энтропии h показывает, во сколько раз неопределённость данного состояния системы отличается от состояния системы единичной неопределённости.

Приведённые свойства и интерпретация значений показателя частной энтропии позволяют ввести пару основных показателей педагогических измерений β , θ и связать их с информационным показателем отношением $h = \beta / \theta$. Тогда, из формулы Хартли следует модель педагогических измерений $P = 0,5^{\beta/\theta}$, где β — относительный показатель трудности тестового задания, θ — относительный показатель подготовленности испытуемого.

Проведём трактовку результатов рассмотренного выше примера с позиций этой модели. Собственная трудность задачи определения победителя для всех педагогов одинакова. Для упрощения рассуждений положим $\beta = 1$. Тогда относительный показатель подготовленности педагогов θ обратно пропорционален значению частной энтропии $\theta = 1/h$. Отсюда, в частности, заключаем —

классный руководитель имеет единичный уровень подготовленности (равный уровню трудности задания) к ответу на поставленный вопрос. Для директора школы $\theta = 1/4$, т.е. он в 4 раза хуже классного руководителя подготовлен к ответу на этот вопрос, а руководитель кружка — в $1/0,415 \approx 2,4$ раза лучше подготовлен к ответу на этот вопрос. Такие интерпретации подготовленности педагогов не только качественно верно отражают рассмотренную в примере ситуацию, но и количественно выверены на основе теории педагогических измерений.

В наших работах 2, 3 модель $P = 0,5^{\beta/\theta}$ представлена как альтернативная, где достаточно подробно приведены свойства модели, свойства и интерпретация показателей модели. Здесь отметим лишь то, что по этой модели измерения проводятся в шкале отношений (пропорций).

Если обратиться к данным таблицы 1, то несложно заметить, что не только значение частной энтропии отражает различия систем с педагогами при решении поставленной задачи, но и разность частных энтропий состояний правильный и неправильный выбор победителя $\Delta h = h_1 - h_0$ отражает эти различия. Например, для директора эта разность имеет значение примерно 3,9 бит (на столько бит неожиданней правильный выбор, чем ошибка), для завуча около 2,8 бит, а для классного руководителя неопределённости состояний правильный и неправильный выбор вообще одинаковы, и разность частных энтропий равна нулю. По характеру поведения разности частных энтропий Δh несложно заметить, что чем хуже подготов-

лен педагог к правильному выполнению задание, тем она больше. Отсюда можно предложить следующую связь показателей подготовленности b и трудности задания d с информационной разностью: $d - b = \ln 2 \cdot \Delta h$.

Коэффициент пропорциональности $\ln 2$ введён для приведения этих показателей к стандартному для системы Раша виду. Если в это равенство подставить выражения для частных энтропий, то после элементарных преобразований получим хорошо известную модель Раша для педагогических измерений:

$$P = \frac{\exp(b - d)}{1 + \exp(b - d)}.$$

Таким образом, разность показателей подготовленности испытуемого и трудности тестового задания в измерительной системе Раша, можно трактовать как разность (с точностью до постоянного множителя $\ln 2$) частных энтропий состояний верное и неверное решение заданий. Эта разность, традиционно измеряемая в логитах, принимает информационную меру измерения бит, с соответствующей информационной интерпретацией значений показателей подготовленности b и трудности d .

Мы предложили два варианта описания системы испытуемый-задание на основе информационного подхода. Каждый из них дал свою исходную модель педагогических измерений — альтернативную и Раша. Не исключается возможность предложить и другие варианты. Например, основанные на частной энтропии состояния «неправильное решение задачи» или другие комбинации информационных показателей. Поиск и обсуждение других вариантов моделей лежат за рамками данной работы.

В завершении подраздела остановимся на формуле Шеннона для расчёта среднего значения энтропии системы по всем её возможным состояниям и её месте в педагогических измерениях. Содержательный смысл значения средней энтропии состоит в получении среднего количества информации в сообщении о результате ответа на задание. В рассмотренном примере значения этого показателя приведены в последнем столбце таблицы 1.

Приведём основные свойства этого показателя:

- при значениях $0 \leq P \leq 1$ функция $H(P) = P \cdot \log_{0,5} P + (1 - P) \cdot \log_{0,5} (1 - P)$ симметрична относительно нейтрального значения вероятности $P = 0,5$;
- при $P = 0,5$ энтропия принимает максимальное значение $H(0,5) = 1$, на границах обращается в нуль — $H(0) = H(1) = 0$;
- свойство аддитивности $h(P_1 \cdot P_2) = h(P_1) + h(P_2)$ для независимых событий.

По своим свойствам и содержательному смыслу эта функция с успехом заменяет информационную функцию $I = P(1 - P)$ системы измерений Раша. Если учесть, что величина энтропии имеет фундаментальный смысл и широко принятые единицы измерения бит, то её применение более предпочтительно.

Таким образом, на основе информационного подхода достаточно легко и естественно удаётся построить исходные модели педагогических измерений и выразить основные показатели таких измерений через информационные показатели с единицей измерения *бит*.

Конструирование частных моделей педагогических измерений

В предыдущем подразделе мы вывели на основе информационного подхода две исходные модели педагогических измерений — модель Раша и альтернативную модель. Под термином «исходная» мы понимаем однопараметрическую дихотомическую модель, в которой единственный параметр представляет собой разность $g = b - d$ (аддитивная модель) или отношение $\gamma = \theta/\beta$ (мультипликативная модель) измеряемых показателей подготовленности испытуемого и трудности тестового задания. Таким образом, исходные модели имеют вид:

$$P = \frac{\exp(g)}{1 + \exp(g)} \text{ — модель Раша; } P = 0,5^{1/\gamma} \text{ —}$$

альтернативная модель.

По этим моделям наиболее информативны состояния с пороговыми значениями параметров соответственно $g = 0$ и $\gamma = 1$, при которых вероятности правильных и неправильных ответов совпадают и равны $P = 0,5$, а среднее значение энтропии принимает максимальное значение $H_{\max} = 1 \text{ бит} = \ln 2 \text{ нат}$.

Для иллюстрации основных положений теории педагогических измерений традиционно приводятся графики зависимости вероятности наблюдаемых событий от показателя подготовленности испытуемого при заданном уровне трудности тестового задания. Для исходных моделей графические образы тестовых заданий представлены на рисунках 1а и 1б. На этих рисунках и далее

на график вероятностей P наложен график значений энтропии $H = -\sum P_i \times \ln P_i$ в зависимости от уровня подготовленности испытуемого имеющих единицу измерения нат (1 бит $\approx 0,693$ нат). График энтропии H показывает информативные свойства тестового задания — чем выше график, тем информативней задание для данного испытуемого.

На рис. 1а приведены графики зависимости вероятности неправильного ответа P_0 и правильного ответа P_1 на тестовое задание нейтрального уровня трудности $d = 0$ и значения энтропии H , в зависимости от уровня подготовленности испытуемого b . Точка пересечения графиков вероятности $P_0 = P_1 = 0,5$ соответствует абсциссе с уровнем подготовленности испытуемого, равным уровню трудности тестового задания $b = d = 0$.

Для альтернативной модели подобные графики удобно приводить в зависимости не от относительного показателя подготовленности θ , а от его вероятностного эквивалента — вероятностного показателя $u = 0,5^{1/\theta}$. Тогда исходная модель имеет вид $P = u^\beta$ и графики вероятности неправильного ответа P_0 и правильного ответа P_1 на тестовое задания единичного уровня трудности и энтропии H представлены на рис. 1б.

Если принять предложенный в работе¹ критерий эффективности тестового задания

$$\frac{H}{H_{\max}} > \frac{1}{2 \ln 2} \approx 0,72, \text{ то можно сделать ут-}$$

верждения:

- нейтральное задание эффективно для испытуемых с уровнем подготовленности из диапазона $-1,4 < b < 1,4$ с шириной $\Delta b = 2,8$ логит;

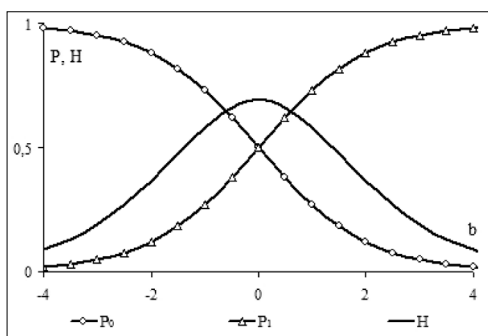


Рис. 1а. Графики вероятности P ответов на задание с нейтральным уровнем трудности ($d = 0$) в системе Раша и средней энтропии H

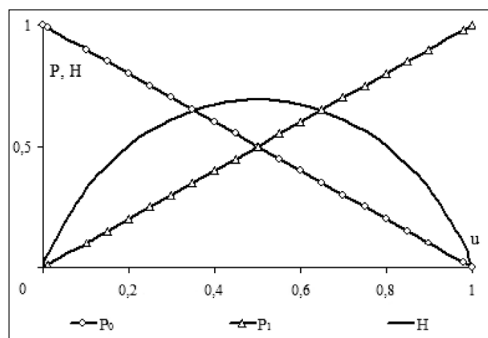


Рис. 1б. Графики вероятности P ответов на задание с единичным уровнем трудности ($\beta = 1$) в альтернативной системе и средней энтропии H

• единичное задание эффективно для испытуемых с вероятностным показателем уровня подготовленности $0,2 < u < 0,8$ с шириной диапазона $\Delta u = 0,6$ (или с относительным показателем из диапазона $0,43 < \theta < 3,1$).

Хорошо известной модификацией исходных моделей являются двухпараметрические дихотомические модели педагогических

измерений. Введение дополнительного параметра a позволяет учитывать дифференцирующие свойства показателей модели. В аддитивную модель Раша этот параметр входит в качестве множителя, в мультипликативную альтернативную модель в качестве показателя степени к основному параметру модели:

$$P = \frac{\exp(a \times g)}{1 + \exp(a \times g)} \text{ — модель Бирнбаума;}$$

$$P = 0,5^{1/\gamma^a} \text{ — двухпараметрическая альтернативная модель.}$$

Введение в модель дифференцирующего параметра приводит к следующим деформациям образов тестовых заданий, в зависимости от значения параметра (рисунки 2а и 2б):

Введение дифференцирующего показателя a отражается на диапазоне эффективности тестовых заданий следующим образом — увеличение значения дифференцирующего показателя уменьшает диапазон эффективности тестовых заданий.

В системе Раша ширина диапазона обратно пропорционален значению дифференцирующего показателя; можно примерно записать $\Delta b \approx 2,8/a$. В частности, при $a = 2$ имеем диапазон эффективности тестовых заданий $-0,07 < b < 0,07$, а при $a = 0,5$ имеем $-2,8 < b < 2,8$.

Для альтернативной модели при значении $a = 2$ имеем диапазон эффективности тестовых заданий $0,35 < u < 0,68$, а при $a = 0,5$ имеем $0,02 < u < 0,93$.

Другим направлением модификации модели является переход от дихотомического к полиномическим типам, допускающим

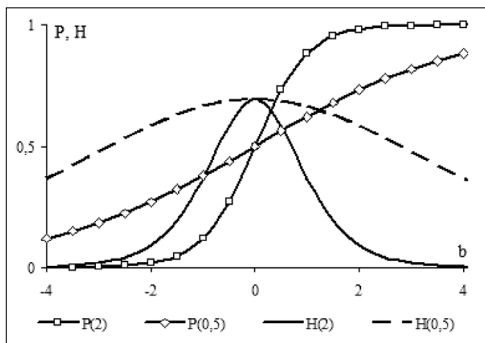


Рис. 2а. Графики вероятности P правильных ответов на задания с нейтральным уровнем трудности, значениями параметра $a = 2$ и $a = 0,5$ в системе Раша и средней энтропии H

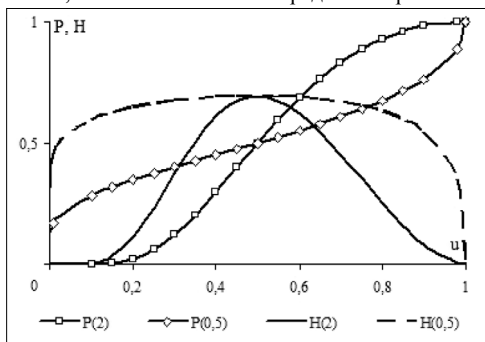


Рис. 2б. Графики вероятности P правильных ответов на задания с единичным уровнем трудности, значениями параметра $a = 2$ и $a = 0,5$ в альтернативной модели и средней энтропии H

вариативность ответа на тестовое задание. Известным примером политомической модели класса Раша является модель РСМ (Partial Credit Model). Аналогичные расширения допустимы и для класса альтернативных моделей.

Получим алгоритм построения этой модели на основе информационного подхода.

Допустим, тестовое задание содержит n градаций ответа, тогда система испытуемый-задание может находиться в $n + 1$ состояниях (включая нулевое состояние $n = 0$) с вероятностями $P_0, P_1, \dots, P_k, P_{k+1}, \dots, P_n$. Переход системы из одного состояния в другое $P_k \rightarrow P_{k+1}$ сопровождается изменением неопределённости состояний, с разностью частных энтропий $\Delta h_k = \ln P_{k+1} / P_k$, значение которой определяется только пороговым параметром g_{k+1} , отделяющим эти состояния. Т.е. отношение вероятностей последовательных состояний есть функция только от параметра g : $P_{k+1} / P_k = f(g_{k+1})$.

Для начальных состояний последнее соотношение, задающее отношение вероятности правильного ответа к вероятности ошибки можно записать для двух исходных моделей в виде: $P_1 / P_0 = \exp(g_1)$ и $P_1 / P_0 = 0,5^{1/\gamma_1} / (1 - 0,5^{1/\gamma_1})$. Отсюда несложно выразить вероятность первого состояния через вероятность нулевого состояния

$$P_1 = P_0 \exp(g_1) \text{ и } P_1 = \frac{P_0 \times 0,5^{1/\gamma_1}}{(1 - 0,5^{1/\gamma_1})} \text{ или}$$

вероятности последующего состояния через предыдущее $P_{k+1} = P_k \exp(g_1)$ и

$$P_{k+1} = \frac{P_k \cdot 0,5^{1/\gamma_{k+1}}}{1 - 0,5^{1/\gamma_{k+1}}}$$

Таким образом вероятности любого последующего состояния можно выразить через вероятность P_0 нулевого состояния. Дописывание к этим соотношениям условия нормировки $P_0 + P_1 + \dots + P_n = 1$ позволяет однозначно записать выражения для вероятностей возможных событий. Примеры кон-

струирования частных политомических моделей рассмотрим в следующем подразделе

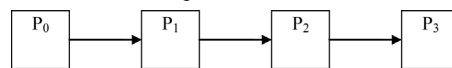
Возможны и составные схемы тестовых заданий. В этом случае само задание можно рассматривать как совокупность непосредственных подзаданий со своими вариантами ответов и системой оценивания. Такие составные задания лучше разбивать на отдельные. Отдельные задания удобнее и выполнять и оценивать, но у составителей заданий на этот счёт может быть и своя точка зрения. Примеры таких заданий можно увидеть в задачах типа С материалов ЕГЭ, олимпиадных задачах.

Допустим, задание содержит n независимых дихотомических подзаданий, с вероятностями правильного ответа на каждое из них P_1, P_2, \dots, P_n . Для расчёта вероятности правильного ответа на отдельное подзадание можно использовать любую из исходных моделей. Тогда существует n^2 различных вариантов ответа, начиная от состояния «все подзадания выполнены неправильно» с совместной вероятностью $(1 - P_1) \cdot (1 - P_2) \cdot \dots \cdot (1 - P_n)$, «только первое подзадание выполнено правильно» с вероятностью $P_1 \cdot (1 - P_2) \cdot \dots \cdot (1 - P_n)$ и завершая правильным решением всех подзаданий с совместной вероятностью $P_1 \cdot P_2 \cdot \dots \cdot P_n$. Примеры конструирования составных заданий рассмотрим ниже.

Пример конструирования линейной политомической модели

Линейная политомическая модель предполагает строгое логически последовательное

расположение всех вариантов ответа на тестовое задание. Например, задание с категориями вариантов ответа: 0 — совершенно не согласен; 1 — не согласен; 2 — согласен; 3 — полностью согласен. Или с категориями: 0 — совершенно неправильный ответ; 1 — ответ содержит правильные элементы; 2 — правильный, но не полный ответ; 3 — совершенно правильный и полный ответ. Если вероятности выбора испытуемым одного из четырёх предложенных вариантов ответа обозначить соответственно через $P_0, P_1, P_2, \dots, P_n$, то структуру такого задания можно представить схемой:



Обозначим три пороговых значения разграничивающих смежные категории в интервальной шкале логитов через d_1, d_2, d_3 , причём, $d_1 \leq d_2 \leq d_3$. Тогда для испытуемого с подготовленностью b и разностью $g_k = b - d_k$ на основе рекуррентного правила $P_{k+1} = P_k \exp(g_{k+1})$ можно записать следующие соотношения между вероятностями появления одного из вариантов ответа: $P_1 = P_0 \exp(g_1)$, $P_2 = P_1 \exp(g_2) = P_0 \exp(g_1 + g_2)$. Подставляя эти соотношения в условие нормировки $P_0 + P_1 + P_2 + P_3 = 1$, находим искомые выражения для вероятностей (линейная модель РСМ):

$$P_0 = \frac{1}{1 + \exp(g_1) + \exp(g_1 + g_2) + \exp(g_1 + g_2 + g_3)},$$

$$P_1 = \frac{\exp(g_1)}{1 + \exp(g_1) + \exp(g_1 + g_2) + \exp(g_1 + g_2 + g_3)},$$

$$P_2 = \frac{\exp(g_1 + g_2)}{1 + \exp(g_1) + \exp(g_1 + g_2) + \exp(g_1 + g_2 + g_3)},$$

$$P_3 = \frac{\exp(g_1 + g_2 + g_3)}{1 + \exp(g_1) + \exp(g_1 + g_2) + \exp(g_1 + g_2 + g_3)}.$$

В качестве иллюстрации приведём графики этих функций для пороговых значений категорий трудности задания в интервальной шкале логитов $d_1 = -1$, $d_2 = 0$, $d_3 = 1$ (рисунок 3а).

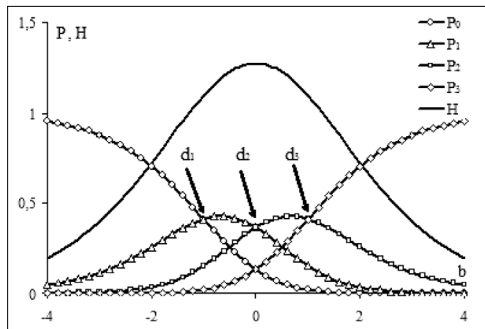


Рис. 3а. Графики вероятности P ответов испытуемым на линейное задание с уровнями трудности $d_1 = -1$, $d_2 = 0$, $d_3 = 1$ в системе Раша и средней энтропии H

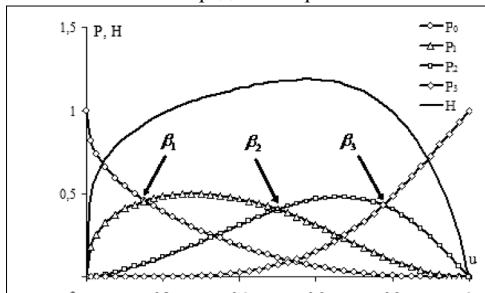


Рис. 3б. Графики вероятности P ответов испытуемым на линейное задание с уровнями трудности $\beta_1 = 1/e$, $\beta_2 = 1$, $\beta_3 = e$ в альтернативной системе и средней энтропии H

Аналогичные результаты несложно получить и для линейной альтернативной модели РСМ на основе рекуррентного соотношения $P_{k+1} = P_k \times 0,5^{1/\gamma_{k+1}} / (1 - 0,5^{1/\gamma_{k+1}})$ с пороговыми показателями в пропорциональной шкале измерений $\beta_1 \leq \beta_2 \leq \beta_3$, уровнем подготовленности испытуемого θ и отношением $\gamma_k = \theta/\beta_k$. Если ввести вероятностный показатель подготовленности испытуемого $u = 0,5^{1/\theta}$, то эти выражения примут более компактный вид:

$$P_0 = \frac{(1 - u^{\beta_1})}{(1 - u^{\beta_1})(1 - u^{\beta_2})(1 - u^{\beta_3}) + u^{\beta_1}},$$

$$P_1 = \frac{(1 - u^{\beta_2})(1 - u^{\beta_3})}{(1 - u^{\beta_1})(1 - u^{\beta_2})(1 - u^{\beta_3}) + u^{\beta_1}},$$

$$P_2 = \frac{(1 - u^{\beta_2})(1 - u^{\beta_3}) + u^{\beta_1 + \beta_2}}{(1 - u^{\beta_1})(1 - u^{\beta_2})(1 - u^{\beta_3}) + u^{\beta_1 + \beta_2 + \beta_3}},$$

$$P_3 = \frac{(1 - u^{\beta_2})(1 - u^{\beta_3}) + u^{\beta_1 + \beta_2 + \beta_3}}{(1 - u^{\beta_1})(1 - u^{\beta_2})(1 - u^{\beta_3}) + u^{\beta_1 + \beta_2 + \beta_3}},$$

$$P_0 = \frac{u^{\beta_1}}{(1 - u^{\beta_1})(1 - u^{\beta_2})(1 - u^{\beta_3}) + u^{\beta_1}},$$

$$P_1 = \frac{u^{\beta_1}}{(1 - u^{\beta_1})(1 - u^{\beta_2})(1 - u^{\beta_3}) + u^{\beta_1}},$$

$$P_2 = \frac{u^{\beta_1 + \beta_2}}{(1 - u^{\beta_1})(1 - u^{\beta_2})(1 - u^{\beta_3}) + u^{\beta_1 + \beta_2 + \beta_3}},$$

$$P_3 = \frac{u^{\beta_1 + \beta_2 + \beta_3}}{(1 - u^{\beta_1})(1 - u^{\beta_2})(1 - u^{\beta_3}) + u^{\beta_1 + \beta_2 + \beta_3}}.$$

Графики этих функций для пороговых значений трудности задания в пропорциональной шкале битов $\beta_1 = 1/e$, $\beta_2 = 1$, $\beta_3 = e$ приведены на рисунке 3б.

Точки пересечения графиков вероятности выбора того или иного варианта ответа соответствуют абсциссе с уровнем подготовленности испытуемого равным пороговому значению уровня трудности между соответствующими категориями. Например, абсцисса точки пересечения графиков $P_1(u)$ и $P_2(u)$ равна уровню подготовленности испытуемого $u = 0,5^{1/\beta_1} = 0,5^e \approx 0,152$ соответствующего пороговому значению трудности b_1 . Этот вывод несложно получить и аналитически из решения уравнения $P_1(u) = P_2(u)$.

Обратимся к анализу эффективности такого задания. Для системы, которая может находиться в одном из четырёх состояний, энтропия достигает максимально возможного значения $H_{\max} = \ln 4$ нат $\approx 1,386$ нат в точке равенства вероятностей каждого из

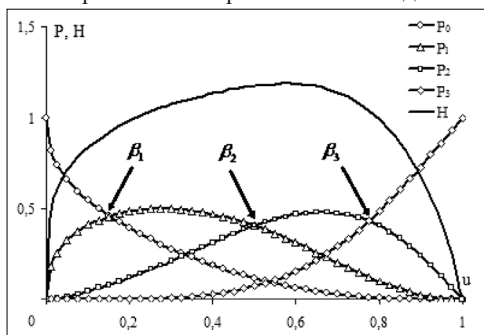


Рис. 3б. Графики вероятности P ответов испытуемым на линейное задание с уровнями трудности $\beta_1 = 1/e$, $\beta_2 = 1$, $\beta_3 = e$ в альтернативной системе и средней энтропии H

состояний. Для тестовых заданий с множественным выбором этому условию удовлетворяют состояния с одинаковыми пороговыми значениями равными уровню подготовленности испытуемого, т.е. когда все $g_k = 0$, $\gamma_k = 1$.

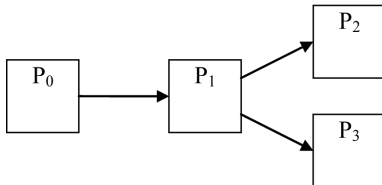
Тогда условие эффективности $\frac{H}{H_{\max}} > \frac{1}{2 \ln 2}$

для тестового задания с четырьмя возможными состояниями преобразуется в неравенство $H > 1$ нат.

Применяя это условие, имеем области уровня подготовленности испытуемых для которых задание эффективно: $-1,2 < b < 1,2$ с наибольшим значением $H_{\text{наиб}} \approx 1,28 < 1,39 \approx H_{\text{max}}$ в системе Раша; $0,22 < u < 0,79$ с наибольшим значением $H_{\text{наиб}} \approx 1,19 < 1,39 \approx H_{\text{max}}$ в альтернативной системе. Т.е. информативность линейных политомических заданий с различающимися пороговыми значениями уровней трудности несколько снижается.

Пример конструирования разветвлённой политомической модели

Разветвлённая политомическая модель допускает существование параллельных (независимых) траекторий из вариантов ответа. Например, в задании требуется не только правильно ответить на вопрос, но и обосновать его одним из методов. Тогда результаты выполнения задания могут содержать варианты: 0 — ответ неправильный; 1 — ответ правильный; 2 или 3 — правильный ответ обоснован тем или иным способом. Структуру такого задания можно представить схемой:



Обозначим три пороговых значения разграничивающих смежные категории в интервальной шкале логитов через d_1, d_2, d_3 . Тогда для испытуемого с подготовленностью b и разностью $g_k = b - d_k$ на основе рекуррентного правила $P_{k+1} = P_k \exp(g_{k+1})$ можно записать следующие соотношения между вероятностями появления одного из вариантов ответа: $P_1 = P_0 \exp(g_1)$, $P_2 = P_1 \exp(g_2) = P_0 \exp(g_1 + g_2)$, $P_3 = P_2 \exp(g_3) = P_0 \exp(g_1 + g_2 + g_3)$. Подставляя эти выражения в условие нормировки $P_1 = P_1 + P_2 + P_3 + 1$, находим искомые выражения для вероятностей (разветвлённая модель РСМ):

$$P_0 = \frac{1}{1 + \exp(g_1) + \exp(g_1 + g_2) + \exp(g_1 + g_2 + g_3)}$$

$$P_1 = \frac{\exp(g_1)}{1 + \exp(g_1) + \exp(g_1 + g_2) + \exp(g_1 + g_2 + g_3)}$$

$$P_2 = \frac{\exp(g_1 + g_2)}{1 + \exp(g_1) + \exp(g_1 + g_2) + \exp(g_1 + g_2 + g_3)}$$

$$P_3 = \frac{\exp(g_1 + g_2 + g_3)}{1 + \exp(g_1) + \exp(g_1 + g_2) + \exp(g_1 + g_2 + g_3)}$$

В качестве иллюстрации приведём графики этих функций для пороговых значений категорий трудности задания в интер-

вальной шкале логитов $d_1 = -1, d_2 = 0, d_3 = 1$ (рисунок 4а).

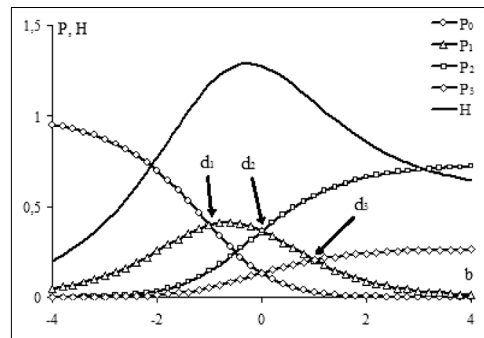


Рис. 4а. Графики вероятности P ответов испытуемым на разветвлённое задание с уровнями $d_1 = -1, d_2 = 0, d_3 = 1$ трудности в системе Раша и средней энтропии H

Аналогичные результаты несложно получить и для разветвлённой альтернативной модели РСМ с пороговыми показателями в пропорциональной шкале измерений $\beta_1, \beta_2, \beta_3$, выраженными через вероятностный уровень подготовленности испытуемого $u = 0,5^{1/\theta}$, и на основе рекуррентного соотношения $P_{k+1} = P_k \cdot u^{\beta_{k+1}} / (1 - u^{\beta_{k+1}})$:

$$P_0 = \frac{1}{(1 - u^{\beta_1})(1 - u^{\beta_2})(1 - u^{\beta_3}) + u^{\beta_1}(1 - u^{\beta_2})}$$

$$P_1 = \frac{u^{\beta_1}(1 - u^{\beta_2})}{(1 - u^{\beta_1})(1 - u^{\beta_2})(1 - u^{\beta_3}) + u^{\beta_1}(1 - u^{\beta_2})}$$

$$P_2 = \frac{\frac{(1-u^{\beta_3})}{(1-u^{\beta_3})+u^{\beta_1+\beta_2}(1-u^{\beta_3})+u^{\beta_1+\beta_3}(1-u^{\beta_2})} \cdot u^{\beta_1+\beta_2}}{(1-u^{\beta_1})(1-u^{\beta_2})(1-u^{\beta_3})+u^{\beta_1}(1-u^{\beta_2})} \cdot \frac{(1-u^{\beta_3})}{(1-u^{\beta_3})+u^{\beta_1+\beta_2}(1-u^{\beta_3})+u^{\beta_1+\beta_3}(1-u^{\beta_2})}}$$

$$P_3 = \frac{\frac{(1-u^{\beta_2})}{(1-u^{\beta_3})+u^{\beta_1+\beta_2}(1-u^{\beta_3})+u^{\beta_1+\beta_3}(1-u^{\beta_2})} \cdot u^{\beta_1+\beta_3}}{(1-u^{\beta_1})(1-u^{\beta_2})(1-u^{\beta_3})+u^{\beta_1}(1-u^{\beta_2})} \cdot \frac{(1-u^{\beta_2})}{(1-u^{\beta_2})+u^{\beta_1+\beta_2}(1-u^{\beta_3})+u^{\beta_1+\beta_3}(1-u^{\beta_2})}}$$

Графики этих функций для пороговых значений трудности задания в пропорциональной шкале битов $\beta_1 = 1/e$, $\beta_2 = 1$, $\beta_3 = e$ приведены на рисунке 4б.

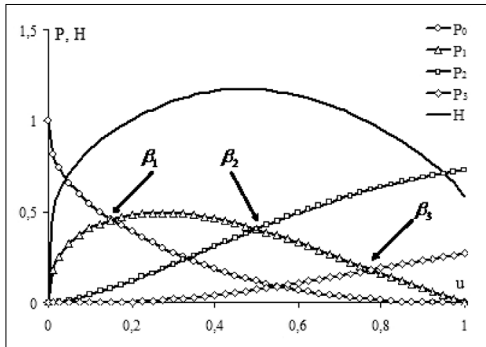


Рис. 4б. Графики вероятности P ответов испытуемым на разветвлённое задание с уровнями трудности $\beta_1 = 1/e$, $\beta_2 = 1$, $\beta_3 = e$ в альтернативной системе и средней энтропии H

Области уровня подготовленности испытуемых, для которых задание с разветвлением эффективно определяются диапазонами: $-1,4 < b < 1,3$, с наибольшим значением $H_{\text{наиб}} \approx 1,28 < 1,39 \approx H_{\text{max}}$ в системе Раша; $0,20 < u < 0,76$ с наибольшим значением $H_{\text{наиб}} \approx 1,28 < 1,39 \approx H_{\text{max}}$ в альтернативной системе. Т.е. информативность и разветвлённых политомических заданий с различающимися пороговыми значениями уровней трудности несколько снижается.

Рассматриваемые исходные модели никак не противоречат, а лишь описывают исследуемые явления с разных позиций, в разных метрических шкалах. Подтверждением этого могут служить не только единые качественные выводы и опорные количественные значения $P^{\text{Раш}}(g=0) = P^{\text{Альм}}(g=1) = 0,5$, но и количественные асимптотические оценки.

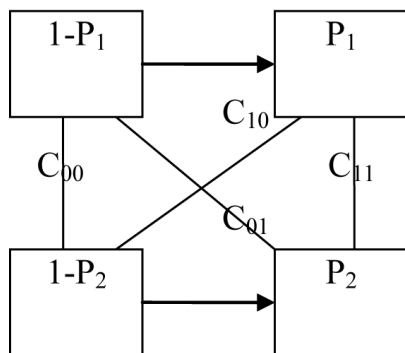
Дополнительное подтверждение находится в рассмотренном примере. При очень хорошей подготовке испытуемого из моделей следует единый, очевидный при заданных параметрах, вывод:

$$\frac{P_2^{\text{Раш}}(d \rightarrow \infty)}{P_3^{\text{Раш}}(d \rightarrow \infty)} = \frac{P_2^{\text{Альм}}(\theta \rightarrow \infty \text{ или } u = 1)}{P_3^{\text{Альм}}(\theta \rightarrow \infty \text{ или } u = 1)} = e \approx 2,718.$$

Если первый метод обоснования решения задачи в e раз (или на 1 логит) легче второго, то и вероятность правильного ответа на него в e раз выше. В системе измерений Раша подобные результаты анализа не приводятся.

Пример конструирования составной модели

Конструирование составного тестового задания покажем на примере задания, состоящего из двух относительно независимых подзаданий. Например, задание — выбрать наиболее верное описание свойств движения тела, свободно падающего на Землю из состояния покоя. Варианты ответов: прямолинейное равномерное; прямолинейное ускоренное; криволинейное равномерное; криволинейное ускоренное. В отличие от политомического такое задание предполагает предварительное решение двух независимых задач — определить вид траектории и определить характер движения тела. Если вероятности правильных ответов на каждую из задач обозначить через P_1 и P_2 , то существует четыре варианта ответа с совместными вероятностями $C_{00} = (1 - P_1)(1 - P_2)$, $C_{10} = P_1(1 - P_2)$, $C_{01} = (1 - P_1)P_2$, $C_{11} = P_1P_2$. Структуру такого задания можно представить схемой:



Обозначим два пороговых значения различающих категории правильно — неправильно для каждого из подзаданий в интервальной шкале логитов через d_1, d_2 . Тогда для испытуемого с подготовленностью b и разностью $g_k = b - d_k$ можно записать следующие соотношения для совместной вероятности ответов на подзадания:

$$C_{00} = \frac{1}{(1 + \exp(g_1))(1 + \exp(g_2))},$$

$$C_{01} = \frac{\exp(g_2)}{(1 + \exp(g_1))(1 + \exp(g_2))},$$

$$C_{10} = \frac{\exp(g_1)}{(1 + \exp(g_1))(1 + \exp(g_2))},$$

$$C_{11} = \frac{\exp(g_1)\exp(g_2)}{(1 + \exp(g_1))(1 + \exp(g_2))}.$$

Графики этих функций для пороговых значений трудности задания в интервальной шкале логитов $d_1 = -1, d_2 = 1$ приведены на рисунке 5а.

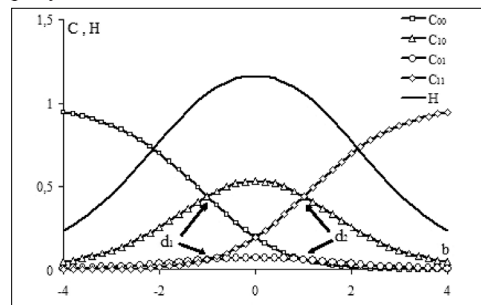


Рис. 5а. Графики совместной вероятности C ответов испытуемым на составное задание с уровнями трудности $d_1 = -1, d_2 = 1$ в системе Раша и средней энтропии H

Аналогичные результаты несложно получить и для составной альтернативной модели с пороговыми показателями двух подзаданий в пропорциональной шкале измерений β_1, β_2 , выраженными через вероятностный уровень подготовленности испытуемого $u = 0,5^{1/\theta}$:

$$C_{00} = (1 - u^{\beta_1})(1 - u^{\beta_2}),$$

$$C_{10} = u^{\beta_1}(1 - u^{\beta_2}),$$

$$C_{01} = (1 - u^{\beta_1})u^{\beta_2},$$

$$C_{11} = u^{\beta_1}u^{\beta_2}.$$

Графики этих функций для пороговых значений трудности подзаданий в пропорциональной шкале битов $\beta_1 = 1/e, \beta_2 = e$ приведены на рисунке 5б.

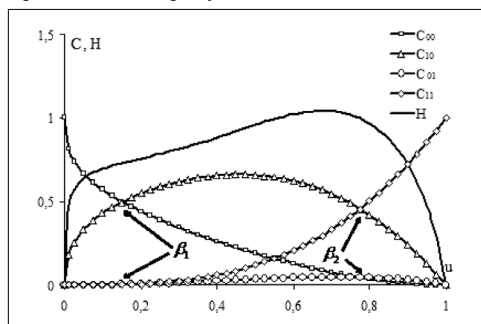


Рис. 5б. Графики совместной вероятности C ответов испытуемым на составное задание с уровнями трудности $\beta_1 = 1/e, \beta_2 = e$ в альтернативной системе и средней энтропии H

Области уровня подготовленности испытуемых для которых составное задание эффективно определяются диапазонами:

$-1,2 < b < 1,2$ с наибольшим значением $H_{\text{наиЮ}} \approx 1,16 < 1,39 \approx H_{\text{мак}}$ в системе Раша; $0,57 < u < 0,77$ с наибольшим значением $H_{\text{наиЮ}} \approx 1,04 < 1,39 \approx H_{\text{мак}}$ в альтернативной системе. Т.е. информативность составных заданий с различающимися пороговыми значениями уровня трудности заметно снижается.

В завершение рассмотренных примеров приведём одно замечание относительно выраженных различий в свойствах моделей класса Раша и альтернативных моделей. Модели класса Раша симметричны относительно нейтрального значения параметра $g = 0$. Это свойство отражено на соответствующих рисунках с литерой «а» при симметричных уровнях трудности тестовых заданий. Альтернативная система подобной симметрией относительно нейтральной единицы $\gamma = 1$ (или $u = 0,5$) не обладает, что отражено и на рисунках с литерой «б». Эти различия заложены в исходных моделях, в толкованиях её показателей, и противоречий здесь нет. Измерительные шкалы этих систем обладают различными метриками.

Информационный метод проведения измерений

Информационный подход можно распространить не только на педагогические, но и на классические физические измерения. Для примера рассмотрим процесс измерения линейкой длины небольшого бруска L . Сначала обратимся к свойствам линейки и свойствам результатов измерения линейкой. Первое. Линейка имеет метрическую шкалу

с минимальным делением 1 мм по которой собственно и проводятся измерения. Если аккуратно провести измерение, т.е. ровно наложить линейку на брусок и строго совместить левый край бруска с нулевой отметкой шкалы линейки, то положение правого края бруска указывает его длину в единицах шкалы линейки. Пример такого измерения показан на рисунке 6, по которому можно заключить, что длина бруска более 27 и менее 27,5 мм.

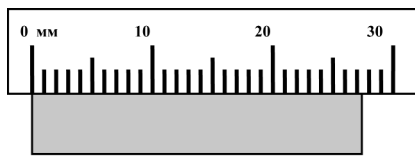


Рис. 6. Измерение бруска

Точнее провести измерение данной линейкой не удастся. Неопределённость измерения неизбежна, она заложена и в самом измерительном инструменте, и в свойствах бруска, и технологии проведения измерения. В нашем примере неопределённость измерения линейкой можно оценить значением 0,5 мм. Отсюда второе свойство линейки — присущая ей неопределённость при проведении измерений, количественно выраженная погрешностью. Таким образом, результат физического измерения характеризуется двумя основными параметрами — значение измеряемой величины и погрешность этого измерения.

Рассмотрим измерение длины бруска на основе информационной схемы. Под измерением будем понимать процесс получения наибольшего количества информации от

сравнения бруска с метрической линейкой. Будем считать, что если исследователь принимает за длину бруска значение 27 мм, то, по его мнению, вероятность события «длина бруска менее 27 мм» равна 0,5. Впрочем, как и того, что длина бруска более 27 мм. Именно в этом случае равенства вероятностей противоположных событий энтропия системы максимальная, именно в этом случае исследователь получает наибольшее количество информации от измерения.

Другой исследователь выскажет другой результат измерения, например 27,2 мм, и т.д. При проведении массовых измерений длины бруска результаты сгруппируются в диапазоне от 25 до 25,5 мм. Если результаты этих измерений выразить через показатели двухпараметрической альтернативной модели, то графический образ длины бруска будет иметь вид представленный на рисунке 7. Здесь введены следующие обозначения: $u = 0,5^{1/\theta}$ — вероятностный показатель результатов отдельного измерения по шкале линейки, $\theta = L/L_{\text{cp}}$ — относительные значения этих измерений, P — вероятность состояния «среднее значение длины бруска меньше показаний линейки», H — энтропия этого состояния.

Наиболее информативное состояние при $u_{\text{cp}} = 0,5$ с вероятностью $P(L = L_{\text{cp}}) = 0,5$ и энтропией $H_{\text{max}}(L = L_{\text{cp}}) = \ln 2$ отражает среднее значение измеряемой величины. Диапазон информационной эффективности оценивается шириной $\Delta H \approx 0,005$, что соответствует погрешности измерений $\Delta L \approx 4$ мм. Эти результаты полностью согласуются с основными показателями физического измерения,

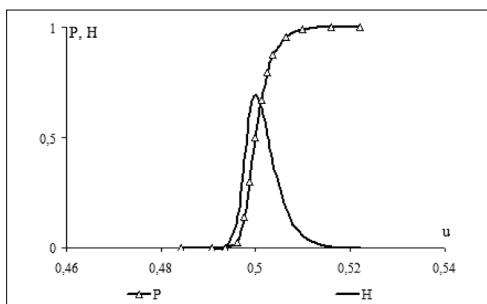


Рис. 7. График вероятности P и энтропии H измерений длины бруска

что говорит о правомерности информационного подхода к измерениям.

Однако, результаты физических и педагогических измерений имеют и выраженные различия.

Первое и принципиальное различие состоит в том, что при проведении педагогических измерений нет предварительно созданной «эталонной метрической линейки». В педагогических измерениях сама метрика формируются в процессе самих измерений, при обработке результатов тестовых испытаний.

Второе различие относится к точности измерений. При проведении педагогических измерений значение дифференцирующего показателя a незначительно отличаются от единицы, в то время как при более точных физических измерениях значения этого показателя порядка сотен. В нашем примере измерения линейкой $a = 150$.

Свойства дифференцирующего показателя таковы, что при увеличении его значе-

ния уменьшается ширина диапазона информационной эффективности измерения. Более строго — ширина диапазона информационной эффективности измерения Δ обратно пропорциональна значению дифференцирующего показателя a и по порядку величин они связаны отношением $\Delta \sim 1/a$. Если в педагогических измерениях относительная ширина диапазона информационной эффективности сопоставима со средним значением $\Delta u \sim u_{\text{ср}}$, то в физических измерениях — составляет доли процента $\Delta u \sim 0,01 \cdot u_{\text{ср}}$. Измерения, в которых погрешность измерений сопоставима с результатом, точными не назовёшь. Проблема точности педагогических измерений решается только при проведении массовых испытаний и за счёт статистической сходимости результатов измерений.

Таким образом, применение информационного подхода к измерению не ограничено отдельной областью. На наш взгляд, этот подход носит универсальный характер, суть которого можно представить выделением следующих этапов:

- поиск состояния сравниваемых величин с наибольшим значением энтропии;
- представление этого состояния в виде числовой величины с метрикой.

В этом случае результат измерения даст максимально возможное количество информации не только в смысле Шеннона, но и в смысле важнейшего метода эмпирического познания.