

2. *Шубин, А. В.* Разработка многофункционального устройства контроля жизненных показателей человека. Фотоплетизмография / А. В. Шубин, К. О. Хохлов. Текст: электронный // Физика. Технологии. Инновации: тезисы докладов V Международной молодежной научной конференции, посвященной памяти Почетного профессора УрФУ В. С. Кортова (Екатеринбург, 14–18 мая 2018 г.). Екатеринбург: Урал. федерал. ун-т, 2018. С. 35–36. URL: <https://elar.urfu.ru/handle/10995/110062?ysclid=lhs332kdpo110855367>.

УДК 378.146.012.4

Братищенко В. В.

СРАВНЕНИЕ ОЦЕНОК УЧАЩИХСЯ НА ОСНОВЕ ВЕРОЯТНОСТНЫХ МОДЕЛЕЙ

Владимир Владимирович Братищенко

Кандидат физико-математических наук, доцент

e-mail: vvb@bgu.ru

ФГБОУ ВО «Байкальский государственный университет», Россия, Иркутск

COMPARISON OF STUDENT GRADES BASED ON PROBABILITY MODELS

Vladimir Vladimirovich Bratischenko

Baikal State University, Russia, Irkutsk

Аннотация. В работе предлагается вероятностная модель оценки, основанная на пошаговом выполнении экзаменационного задания. По аналогии с моделью «*Partial Credit*» в модели используются латентные параметры для подготовленности студентов и трудности шагов задания, которые определяют вероятности успешного выполнения шага задания. Методом максимального правдоподобия получены уравнения, связывающие наблюдения оценок и латентные параметры. Решения уравнений численными методами позволяют получить

оценки латентных параметров. Параметры определяют распределения вероятностей оценок. Полученные функции распределения оценок используются для построения отображения одного набора оценок в другой.

Abstract. *The paper proposes a probabilistic model for student grades based on the step-by-step execution of an examination task. By analogy with the “Partial Credit” model, the model uses latent parameters for student ability and the difficulty of task steps, which determine the probabilities of successful completion of a task step. The maximum likelihood method is used to obtain equations relating the observations of estimates and latent parameters. Solving the equations by numerical methods makes it possible to obtain estimates of latent parameters. The parameters define the probability distributions of the grades. The resulting score distribution functions are used to determine the mapping from one set of grades to another.*

Ключевые слова: *стохастическая модель оценок, Partial Credit Model, Item Response Theory, латентные параметры, сравнение шкал измерений.*

Keywords: *stochastic model of student grades, Partial Credit Model, Item Response Theory, latent parameters, comparison of grade scales.*

Задача адекватного педагогического измерения знаний и навыков учащихся является сложной и актуальной задачей [1]. Любая область обучения нуждается в оценках для измерения достижений и соответствия полученных знаний образовательным стандартам. Оценка измеряет уровень знаний и навыков по дисциплине. Невысокие оценки демонстрируют недостаточность знаний и стимулируют учащихся. Перекосы в распределении оценок позволяют понять степень соответствия объема и уровня сложности предмета выделенному времени на изучение предмета и подготовленности учащихся. На величину оценки влияют много факторов: это и подготовленность студента, и сложность экзамена, и технология оценки знаний, и межличностные отношения студента и преподавателя, и множество случайных событий, так или иначе влияющих на процесс оценивания. Все это приводит к вероятностному описанию педагогических измерений. Очевидно, что стандартная шкала

оценок применяется экзаменаторами по-разному — каждой аттестации соответствует свое распределение вероятностей. В таких условиях необходимо иметь инструмент сопоставления оценок, полученных в соответствии с разными законами распределения вероятностей.

Исчерпывающий анализ всех факторов затруднен по причинам отсутствия инструментов измерения их влияния. В Item Response Theory (IRT) все факторы сведены к двум латентным параметрам: сложность тестового задания и подготовленность испытуемого [2]. Указанных двух параметров достаточно для описания распределения оценки с двумя значениями: «сдано» и «не сдано». Если результаты задания оцениваются по шкале с более чем двумя градациями, то применяется модель "Partial Credit" [3], в которой вводятся несколько латентных параметров для описания вероятностей получения каждого значения по шкале оценивания.

Предлагается следующая модель, описывающая получение оценки в виде последовательности шагов. В случае «неудачи» на l -м шаге ($l = 1, \dots, k$) с вероятностью $1 - p_l$ оценка принимается равной $l - 1$, в случае «успеха» с вероятностью p_l испытание продолжается на следующем шаге. Предполагается независимость шагов. Успешно выполнив все шаги учащийся получает максимальный балл k . Случайная величина X — итоговая оценка будет иметь следующее распределение вероятностей

$$P\{X = x\} = (1 - p_{x+1}) \prod_{l=0}^x p_l, \quad x = 0, \dots, k,$$

где для удобства полагается $p_0 = 1, p_{k+1} = 0$.

Математическое ожидание X можно вычислить в виде следующей суммы

$$M[X] = \sum_{h=1}^k \prod_{l=1}^h p_l.$$

По аналогии с моделью "Partial Credit" [3], применяемой для анализа результатов тестирования, подготовленность i -го студента будем измерять пара-

метром θ_i , а l -й шаг оценивания j -й аттестации будем характеризовать параметром δ_{jl} . Параметры определяют для i -го студента вероятности успешного выполнения шагов j -й аттестации

$$p_l(i, j) = \frac{e^{\theta_i}}{e^{\theta_i} + e^{\delta_{jl}}}, l = 1, \dots, k$$

и вероятность оценки

$$P\{X_{ij} = x_{ij}\} = \left(1 - p_{x_{ij}+1}(i, j)\right) \prod_{l=0}^{x_{ij}} p_l(i, j)$$

Очевидно, что чем больше θ_i , тем сильнее распределение оценки X_{ij} будет смещаться в сторону более высоких оценок. Параметры $\delta_{j1}, \dots, \delta_{jk}$ определяют распределение оценки j -й аттестации при некотором фиксированном θ и могут рассматриваться как трудности шагов аттестации. Предполагается независимость всех шагов.

Для определения латентных параметров можно применить метод максимального правдоподобия. Полученные этим методом уравнения решают численными методами. По этой методике был обработан набор оценок 23 студентов для 72 дисциплин. Шкала оценивания: «неявка», «неудовлетворительно», «удовлетворительно», «хорошо», «отлично» — была переведена в числовую: 0, 1, 2, 3, 4. Дисперсионный анализ подтвердил гипотезы об отсутствии влияния на отклонения оценок от математических ожиданий со стороны студентов или дисциплин.

Обработка набора оценок подтвердила гипотезу, что на разных экзаменах применяются шкалы с разными распределениями вероятностей. Для корректной статистической обработки таких данных необходимо установить соответствие разных шкал. Традиционно сопоставление шкал связывают с сопоставлением интервалов [4] интегральных характеристик, по которым определяются рейтинги. Соответствие можно определять на основе статистических распределений оценок [5]. Для данных, изучаемых в данной работе, такой подход оправдан статистической однородностью рассматриваемых данных.

В соответствии с предложенной моделью для каждого студента и дисциплины оценка будет иметь свое распределение вероятностей. Для определения соответствия значений оценок с разными распределениями вероятностей можно использовать функцию распределения $F_X(x) = P(X < x)$. Для непрерывных величин X и Y можно использовать преобразование

$$y = F_Y^{-1}(F_X(x)).$$

Для дискретных случайных величин такой подход не является точным.

Идея следующего метода связана с сопоставлением значений двух вариационных рядов с одинаковыми номерами. Для дискретных случайных величин в вариационном ряду одинаковые значения образуют интервал. Сопоставление компонентов вариационных рядов приводит к ситуации, когда интервалу одинаковых значений x будет соответствовать часть вариационного ряда для Y , возможно с неодинаковыми значениями y . Значению x можно поставить в соответствие среднее значение соответствующей части вариационного ряда y . Если вместо вариационных рядов использовать вероятности, то получается следующая схема. Значение x определяет интервал $(F_X(x), F_X(x + 1)]$. Этому интервалу могут соответствовать несколько значений Y : $y_1, y_1 + 1, \dots, y_2$

$$y_1: F_Y(y_1 - 1) < F_X(x) \leq F_Y(y_1),$$

$$y_2: F_Y(y_2) < F_X(x + 1) \leq F_Y(y_2 + 1).$$

Для пересчета значения вычисляется сумма по перекрывающимся интервалам

$$y = \sum_{i=y_1}^{y_2} i \frac{\min(F_Y(i + 1), F_X(x + 1)) - \max(F_Y(i), F_X(x))}{F_X(x + 1) - F_X(x)}$$

Преобразование оценки на основе распределений следует проводить для исключения тенденции преподавателя ставить завышенные или заниженные оценки. Такая тенденция определяется латентными параметрами $\delta_{j1}, \dots, \delta_{jk}$. Чем больше математическое ожидание оценки экзамена для студента со средним параметром подготовленности, тем больше вероятность завышения оценки. На рис. 2 представлены функции преобразования оценок для разных

распределений экзаменов к усредненному распределению с математическим ожиданием 2,25.

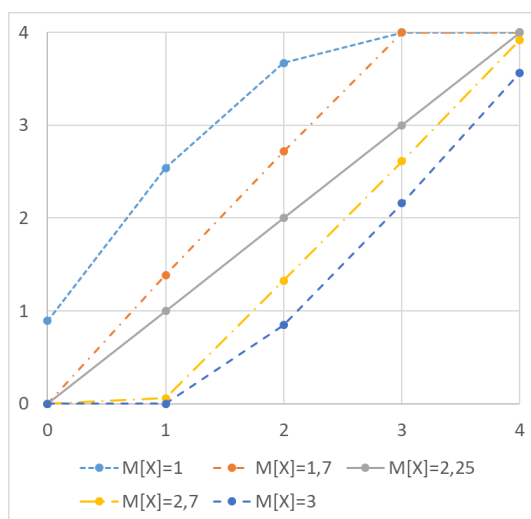


Рисунок 1 — Примеры преобразования оценок для разных распределений экзаменов к усредненному распределению

Преобразованные оценки элиминируют различие распределений вероятностей оценок на разных экзаменах. На рис. 2 представлена зависимость традиционных и преобразованных оценок от свойств распределения, характеризуемых на графике математическими ожиданиями распределений (ось X). Графики демонстрируют независимость преобразованной оценки от свойств распределения.

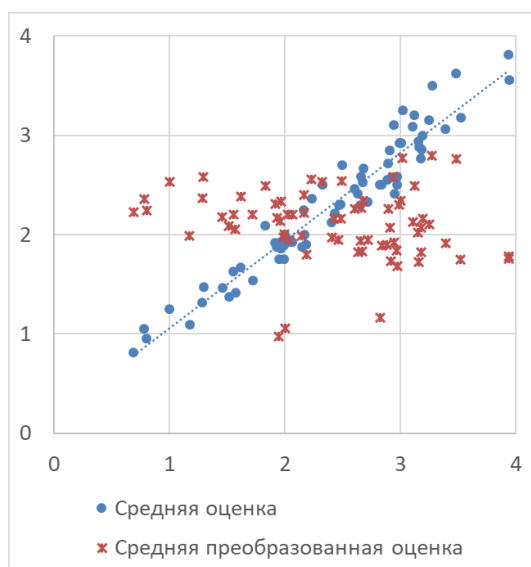


Рисунок 2 — Зависимость традиционных и преобразованных оценок от математического ожидания распределения

Предложенная методика преобразования оценок позволяет получать более точные данные успеваемости и строить более обоснованные рейтинги студентов.

Список литературы

1. *Чельшкова, М. Б.* Теория и практика конструирования педагогических тестов / М. Б. Чельшкова. Москва: Логос, 2002. 431 с. Текст: непосредственный.

2. *Нейман, Ю. М.* Введение в теорию моделирования и параметризации педагогических тестов / Ю. М. Нейман, В. А. Хлебников. Москва: Прометей, 2000. 169 с. Текст: непосредственный.

3. *Wright, B. D.* Rating Scale Analysis: Rasch Measurement / B. D. Wright, G. N. Masters. Chicago: Mesa Press, 1982. 206 p. Text: direct.

4. *О подходах к сопоставлению рейтинговых шкал* / С. А. Айвазян, С. В. Головань, А. М. Карминский, А. А. Пересецкий. Текст: электронный // Прикладная эконометрия. 2011. № 3 (23). С. 13–40. URL: <https://cyberleninka.ru/article/n/o-podhodah-k-sopostavleniyu-reytingovyh-shkal> (дата обращения: 11.09.2022).

5. *Рожков, Н. Н.* Система перезачета оценок успеваемости – инструмент поддержки академической мобильности. / Н. Н. Рожков. Текст: электронный // Университетское управление: практика и анализ. 2006. № 5. С. 104–113. URL: https://elar.urfu.ru/bitstream/10995/52912/1/UM_2006_5_017.pdf?ysclid=lhs3iuc9dc145343607 (дата обращения: 11.09.2022).