УДК 378.14:[51:004]

І.В. АЛЄКСЄЄВА, В.О. ГАЙДЕЙ, О.О. ДИХОВИЧНИЙ, Н.Р. КОНОВАЛОВА, Л.Б. ФЕДОРОВА

ЗАСТОСУВАННЯ МАТЕМАТИЧНИХ МОДЕЛЕЙ ТЕСТІВ У КОМПЛЕКТІ ДИСТАНЦІЙНОЇ ОСВІТИ «ВИЩА МАТЕМАТИКА»

Анотація. *У статті проінформовано про застосування дихотомічної і політомічної IRT-моделей в аналізі тестів комплекту електронних курсів з вищої математики, створених у Національному технічному університеті України «Київський політехнічний інститут».*
Ключові слова: *IRT, математичні методи параметризації тестових завдань, дихотомічна і політомічна моделі Раша.*

Аннотация. *Статья информирует о применении дихотомической и политомической IRT-моделей в анализе тестов комплекта электронных курсов по высшей математике, разработанных в Национальном техническом университете Украины «Киевский политехнический институт».*
Ключевые слова: *IRT, математические методы параметризации тестовых заданий, дихотомическая и политомическая модели Раша.*

Abstract. *The report informs on the application dichotomous and polytomous IRT-models for the Item Analysis of the set of e-courses on higher mathematics designed by National Technical University of Ukraine "Kyiv Polytechnic Institute".*
Key words: *IRT, dichotomous and polytomous Rasch models.*

1. Вступ

Сучасний період розвитку освіти характеризується всебічним поширенням застосування тестового підходу до перевірки знань, значна увага якому приділяється у більшості розвинених країн, серед яких належне місце намагається посісти й Україна. З іншого боку, спостерігається інтенсивний розвиток різноманітних форм дистанційної освіти. Провідне місце в засобах дистанційної освіти посідає система електронного тестування. Це, перш за все, зумовлено високою ефективністю та об'єктивністю форм контролю, розвиненням електронних курсів, удосконаленням технічного забезпечення вишів. Водночас відносна простота організації відповідних контрольних заходів та висока швидкість оброблення їхніх результатів, висока інформативність цих результатів, можливість визначення рейтингу студентів саме за допомогою електронного тестування поєднуються із зацікавленістю самих студентів у проведенні такого контролю і позитивним сприйняттям його як студентами, так і викладачами.

Кафедрою математичного аналізу та теорії ймовірностей НТУУ «Київський політехнічний інститут» був створений комплект дистанційних курсів «Вища математика». Центральною компонентою комплекту є система електронних текстів, розроблена з використанням платформи MOODLE, і яка містить більше 1000 тестів. Продуктивне використання тестової системи ставить проблему аналізу її якості та ефективності, який можна проводити у двох напрямках: спроможності всієї системи в цілому до адекватного оцінювання знань студентів, тобто її змістовній валідності, та в «калібруванні» окремих тестових завдань.

У першому напрямку дослідження проводились методами класичного статистичного аналізу, які довели статистично значущий збіг результатів традиційних та електронних засобів контролю [1].

Аналіз тестових завдань було здійснено із застосуванням сучасних математичних методів параметризації тестових завдань, які мають назву Item Response Theory (IRT).

Item Response Theory останнім часом привертає найбільшу увагу фахівців у галузі педагогічних і психологічних вимірювань. Одна з можливих причин популярності цієї теорії – це її застосування в широко відомих тестових службах. Серед них: Національна Рада розвитку освіти (National Assessment of Educational Progress, NAEP), Служба з розроблення тестів для оцінки здатності до оволодіння освітніми програмами різного рівня складності (Scholastic Aptitude Tests, SAT), а також Американська атестаційна служба the Graduate Record Examination (GRE).

Крім цих служб, IRT використовується в таких масштабних міжнародних порівняльно-оціночних дослідженнях, як Третє міжнародне дослідження рівня підготовленості з математики і природничих наук (Third International Math and Science Survey, TIMSS), а також у Міжнародній програмі оцінки якості підготовленості студентів (the Programme of International Student Assessment, PISA).

Інші можливі причини підвищеної уваги до IRT – це безліч нових статистичних розробок даної теорії, що спираються на можливості їх активного застосування в обчислювальних і освітніх технологіях.

У Росії математичні моделі IRT використовуються для аналізу результатів «ЕГЭ», в Україні такі дослідження майже відсутні.

У роботі [2] повідомлялось про такий аналіз для комплекту «Вища математика» з використанням моделей Раша і Бірнбаума для дихотомічних таблиць відповідей. Тобто, коли правильну відповідь позначають одиницею, а неправильну – нулем.

У даній роботі мова йтиме про політомічну модель Раша, в якій відповідь на тестове питання має декілька градацій у певній шкалі.

2. Основи IRT

Базова ідея IRT, автором якої є датський математик Г. Раш [3], полягає у впровадженні так званих латентних параметрів:

- підготовленість іспитника θ_i, $i=\overline{1,N}$, де N – кількість іспитників;
- складність завдання тесту $\beta_j, j=\overline{1,K}$, де K – кількість завдань у тесті.

Зв’язок між цими параметрами встановлює ймовірність правильної відповіді i -го іспитника на j -те завдання тесту:

$$P_{ij} = \frac{1}{1+\exp\left(-\left(\theta_i - \beta_j\right)\right)}, i=\overline{1,N}, j=\overline{1,K} . \quad (1)$$

Залежність імовірності правильної відповіді на j -те завдання від підготовленості іспитника задається функцією

$$P_i(\theta) = \frac{1}{1+\exp\left(-\left(\theta_i - \beta_j\right)\right)}, j=\overline{1,K} .$$

Відповідний графік називається характеристичною кривою j -го завдання тесту (рис. 1).

Залежність імовірності правильної відповіді i -го іспитника від складності завдання задається функцією

$$P_i(\theta) = \frac{1}{1+\exp(-(\theta_i - \beta))}, i=\overline{1,N}.$$

Відповідний графік називається характеристичною кривою i -го іспитника (рис. 2).

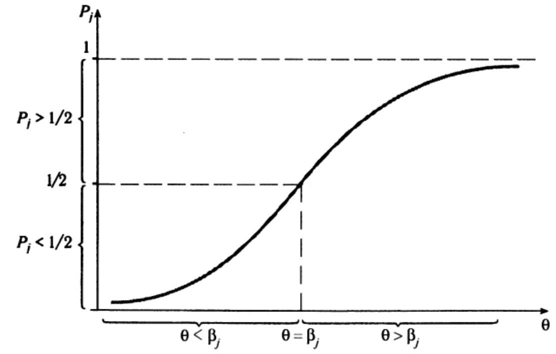


Рис. 1. Характеристична крива j-го завдання

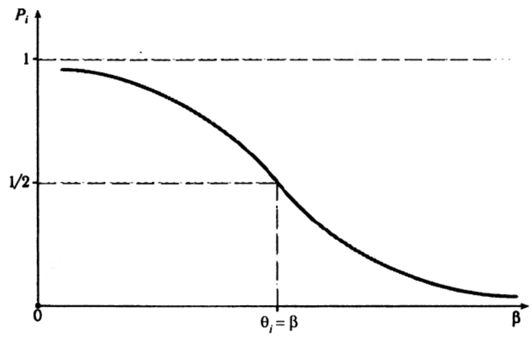


Рис. 2. Характеристична крива i-го іспитника

Модель Раша була узагальнена А. Бірнбаумом шляхом впровадження додаткового параметра – диференціювальної спроможності завдань:

$$\alpha_j = \frac{r_{bis}(j)}{\sqrt{1 - r_{bis}^2(j)}}, \; j = \overline{1, K},$$

де $r_{bis}(j)$ – коефіцієнт кореляції між балами, отриманими за кожне тестове завдання і балами за тест в цілому.

Тоді відповідна ймовірність правильної відповіді має вигляд:

$$P_j(\theta) = \frac{1}{1 + \exp(-\alpha_j(\theta - \beta_j))}, \; j = \overline{1, K},$$

а характеристичні криві для різних рівнів складності $(\alpha_3 < \alpha_2 < \alpha_1)$ наведено на рис. 3.

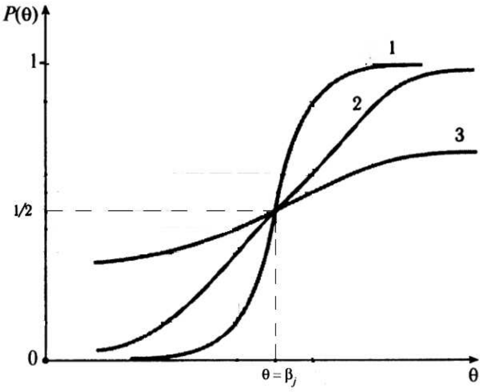


Рис. 3. Характеристичні криві завдань різних диференціювальних спроможностей $(\alpha_3 < \alpha_2 < \alpha_1)$

3. Дихотомічна модель Раша

Емпіричне оцінювання латентних параметрів проводять на підставі спеціальних статистичних процедур аналізу дихотомічної таблиці відповідей X, кожен елемент якої

$$\{x_{ij}\} = \begin{cases} 1, \textit{ якщо відповідь } i-\textit{го студента на } j-\textit{те завдання правильна}; \\ 0, \textit{ якщо відповідь } i-\textit{го студента на } j-\textit{те завдання неправильна}. \end{cases}$$

Елементи цієї матриці можна вважати спостереженнями за незалежними випадковими величинами ξ_{ij}, які набувають значення 1 з імовірністю P_{ij} і значення 0 з імовірністю $Q_{ij} = 1 - P_{ij}$, де P_{ij} задається формулою (1).

Тоді первинний бал i-го іспитника $b_i = \sum_{j=1}^{K} \xi_{ij}, \; i = \overline{1, N}$ є випадковою величиною з математичним сподіванням:

$$M\{b_i\} = \sum_{j=1}^{K} P_{ij}, \; i = \overline{1, N},$$

а первинний бал j-го завдання $c_j = \sum_{i=1}^{N} \xi_{ij}, \; j = \overline{1, K}$ – з математичним сподіванням:

$$M\{c_j\} = \sum_{i=1}^{N} P_{ij}, \; j = \overline{1, K}.$$

Емпіричні перші моменти випадкових величин b_i та c_j обчислюються за матрицею спостережень так:

$$\tilde{b}_i = \sum_{j=1}^{K} x_{ij}, \; i = \overline{1, N},$$

$$\tilde{c}_j = \sum_{i=1}^{N} x_{ij}, \quad j = \overline{1, K}.$$

Прирівнюючи за методом моментів перші теоретичні та емпіричні моменти з урахуванням вигляду P_{ij}, дістаємо систему нелінійних рівнянь:

$$\begin{cases} \tilde{b}_i = \sum_{i=1}^{N} P_{ij}, \\ \tilde{c}_j = \sum_{j=1}^{K} P_{ij}, \\ P_{ij} = \dfrac{1}{1+\exp\left(-\left(\theta_i - \beta_j\right)\right)}, \; i = \overline{1, N}, \; j = \overline{1, K}. \end{cases} \tag{2}$$

Розв'язуючи цю систему, одержуємо оцінки параметрів:

$$\tilde{\theta}_i, \; i = \overline{1, N}, \; \tilde{\beta}_j, \; j = \overline{1, K},$$

які є спроможними, асимптотично-незсуненими, ефективними та асимптотично-нормальними [4]. Для розв'язання цієї системи можна використовувати наближені чисельні методи, наприклад, метод дотичних. Ту ж саму систему можна отримати, використавши метод найбільшої вірогідності. Для розвязання був застосований спрощений алгоритм [5].

За допомогою процедур, реалізованих у пакеті Excel для дихотомічної таблиці відповідей, сформованої автоматично системою MOODLE, на підставі моделей Раша і Бірнбаума аналізувались результати електронних контрольних робіт і іспитів 2008–2009 рр. Так, аналізуючи ансамбль характеристичних кривих іспиту 2008 р. (рис. 4) за моделлю Раша, було відмічено певну нерівномірність та наявність завдань з надто низькою складністю, наприклад, завдання 13.

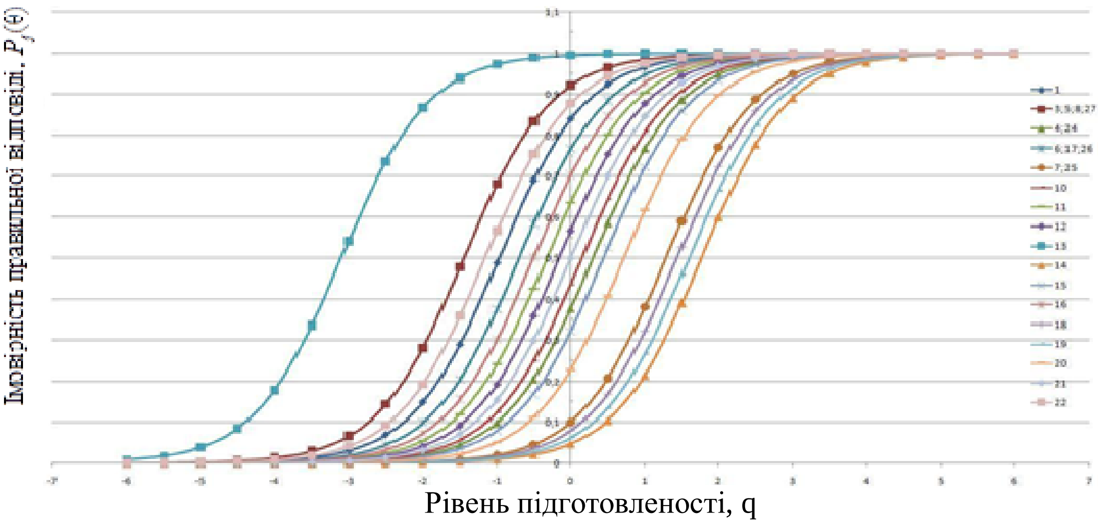


Рис. 4. Характеристичні криві завдань тесту 2008–2009 рр. Модель Раша

Також на підставі моделі Бірнбаума помічено, що завдання з номерами 3, 5, 8, 27 (рис. 5) мають однакову складність: $\beta_j = -1{,}45$, але при цьому $\alpha_3 = 0{,}025$, $\alpha_5 = 0{,}18$,

$\alpha_8 = 0{,}24$, $\alpha_{27} = 0{,}44$. Отже, завдання 3 і 5 вилучено, а завдання 8 і 27 залишено. Для деяких завдань значення α_j взагалі виявились від'ємними. Такі завдання було доречно вилучити. Наприклад, $\alpha_{16} = -0{,}06715$.

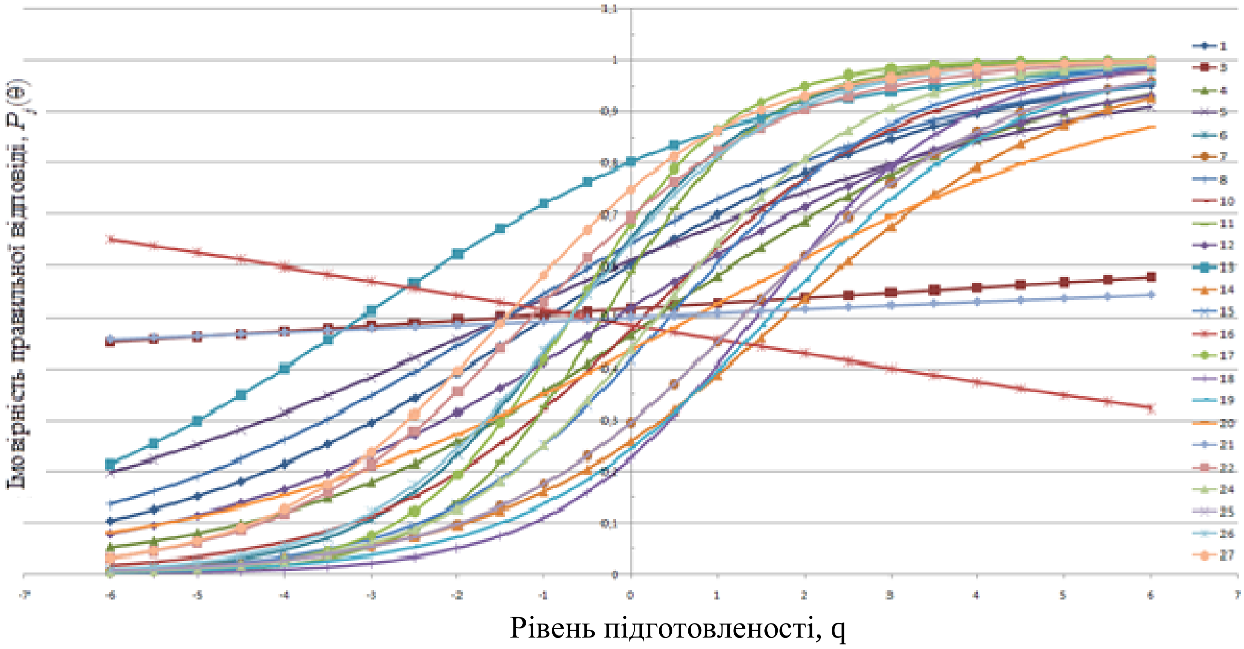


Рис. 5. Характеристичні криві завдань тесту 2008–2009 рр. Модель Бірнбаума

Після послідовного вилучення певних завдань та включення нових було сформовано нові контрольні роботи та іспит, які були запропоновані студентам протягом 2009–2010 рр. навчального року для уточнення відповідних параметрів.

Результати відповідної корекції тестів видно на рис. 6, 7. Як бачимо, відбулось покращання рівномірності ансамблю характеристичних кривих і відсутні завдання з від'ємною диференціювальною спроможністю.

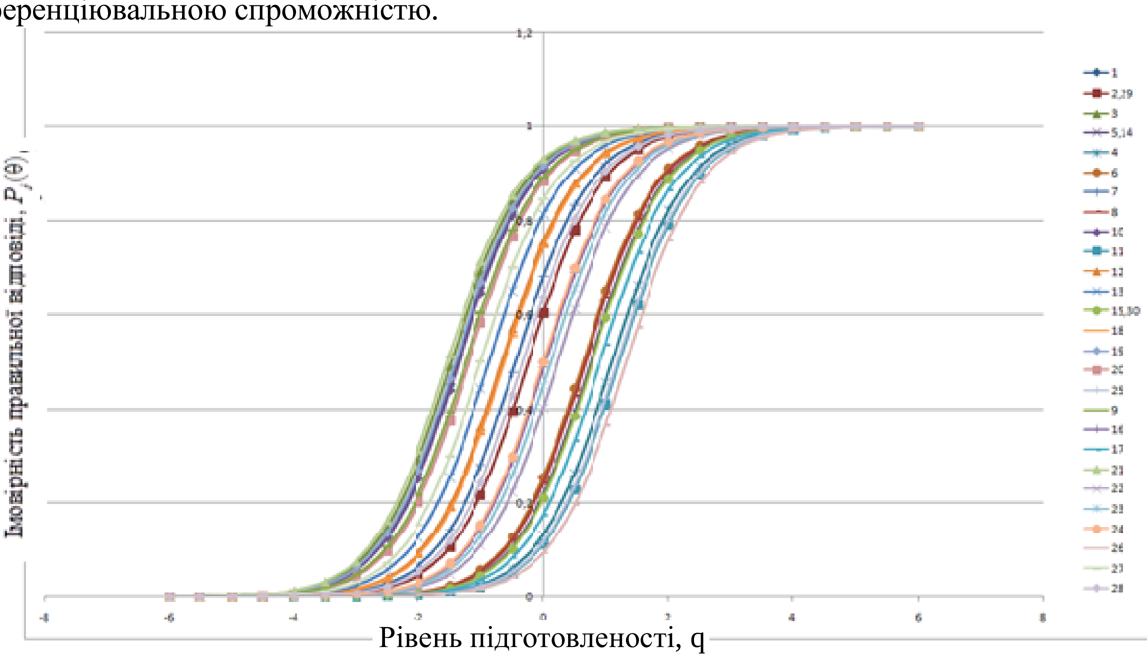
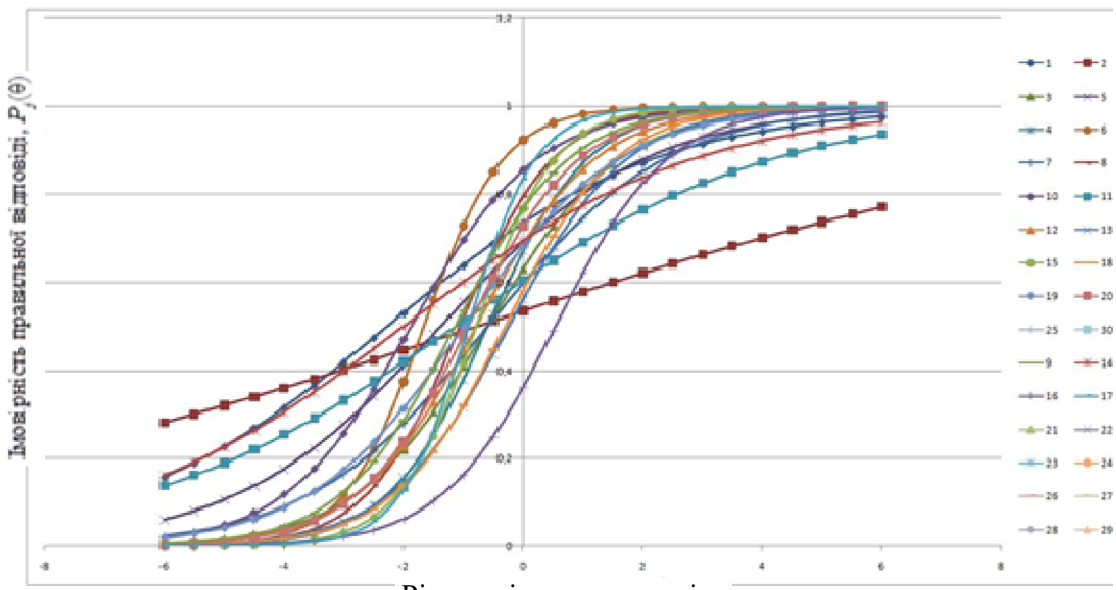


Рис. 6. Характеристичні криві завдань тесту 2009–2010 рр. Модель Раша



Рівень підготовленості, q

Рис. 7. Характеристичні криві завдань тесту 2009–2010 р. Модель Бірнбаума

Також про покращання якості тесту свідчить порівняння інформаційних функцій.

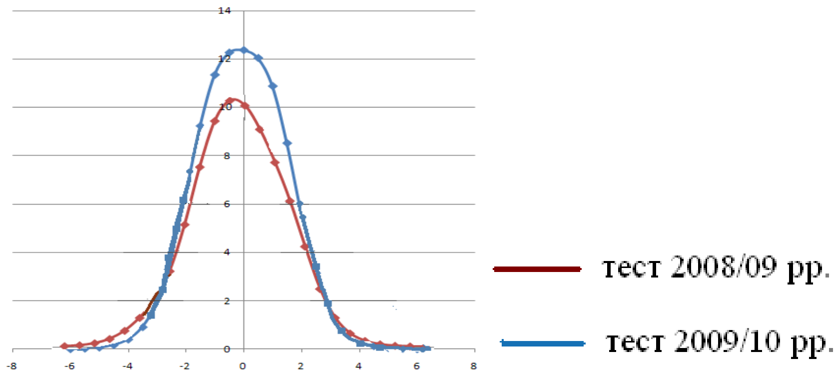


Рис. 8. Порівняння інформаційних функцій тестів 2008–2009 та 2009–2010 рр.

4. Політомічна модель Раша

Розглянемо політомічну модель Раша [6]. Нехай за виконання j-го, $j=\overline{1,K}$ завдання тесту учень може одержати від 0 до m_j балів. Щоб дістати найвищу категорію m_j, учень повинен послідовно подолати m_j кроків: на першому кроці спочатку треба досягти першого рівня в один бал, потім на другому кроці досягнути другої категорії в два бали и т.д. Рівень складності виконання g-го кроку j-го завдання ми будемо позначати через $\beta_{jg}\left(g=\overline{1,m_j}\right)$.

Зрозуміло, що складність виконання кожного кроку, в загальному випадку, різна. Умовна ймовірність $P_{ijg}^{\text{ум}}$ вірного виконання i-им учнем в j-му завданні кроку g за умови, що крок $g-1$ виконаний правильно, описується основною логістичною функцією успіху Раша:

$$P_{ijg}^{УМ}=\frac{1}{1+\exp\left(-\left(\theta_i-\beta_{jg}\right)\right)},\ j=\overline{1,K},\ \ i=\overline{1,N},\ \ g\in\overline{1,m_j}.$$

Знайдемо формулу для знаходження безумовної ймовірності P_{ijg} того, що при виконанні j-го завдання i-ий учень виконає рівно g кроків і отримає, таким чином, рівно g балів. За припущенням політомічної моделі Раша,

$$P_{ijg}^{\text{ум}}=\frac{P_{ijg}}{P_{ijg-1}+P_{ijg}}.$$

Звідки

$$P_{ijg} = \frac{P_{ijg}^{\text{ум}}}{1 - P_{ijg}^{\text{ум}}} \cdot P_{ijg-1} = e^{\theta_i - \beta_{jg}} \cdot P_{ijg}.$$

Із отриманого рекурентного співвідношення маємо:

$$P_{ijg} = e^{\theta_i - \beta_{jg}} \cdot \ldots \cdot e^{\theta_i - \beta_{j1}} \cdot P_{ij0} = \exp\left(\sum_{h=1}^{g} \theta_i - \beta_{jh} \right) \cdot P_{ij0}, \; j = \overline{1,K}, \; i = \overline{1,N}, \; g \in \overline{1,m_j}.$$

Імовірність P_{ij0} знаходимо з умови

$$\sum_{l=0}^{m_j} p_{ijl} = 1,$$

в яку, для зручності, покладаємо $e^{\theta_i - \beta_{j0}} = 1.$ Одержуємо

$$\sum_{l=0}^{m_j} \exp\left(\sum_{h=1}^{l} \theta_i - \beta_{jh} \right) \cdot P_{ij0} = P_{ij0} \cdot \sum_{l=0}^{m_j} \exp\left(\sum_{h=1}^{l} \theta_i - \beta_{jh} \right) = 1,$$

звідки

$$P_{ijg} = \frac{\exp\left(\sum_{h=1}^{g} \theta_i - \beta_{jh} \right)}{\sum_{l=0}^{m_j} \exp\left(\sum_{h=1}^{l} \theta_i - \beta_{jh} \right)}, \quad j = \overline{1,K}, i = \overline{1,N}, g \in \overline{1,m_j}. \tag{3}$$

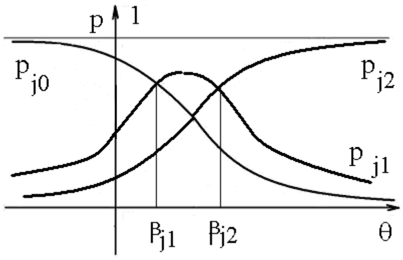


Рис. 9. Характеристичні криві двокрокового завдання

Формула (3) визначає політомічну модель Раша. На рис. 9 наведено типові графіки функцій $P = P_{j0}$, $P = P_{j1}$, $P = P_{j2}$ для двокрокового j-го завдання з $m_j = 2.$

В політомічному випадку система (2) практично не зміниться і матиме вигляд:

$$\begin{cases} \tilde{b}_i - \sum_{j=1}^{m} \sum_{k=1}^{m_j} k \cdot P_{ijk} = 0, \text{ де } \tilde{b}_i = i, \quad i = 0, \ldots, M = \sum_{j=1}^{m} m_j, \\ \tilde{c}_{jg} - \sum_{i=0}^{M} n_i \sum_{k=g}^{m_j} P_{ijk} = 0, \text{ де } \; j = 1, \ldots, m; \; g = 1, \ldots, m_j, \end{cases} \tag{4}$$

де $M = \sum_{j=1}^{K} m_j$ – максимально ймовірний первинний бал за весь тест;

n_i – кількість студентів, первинний бал яких дорівнює i;

$\tilde{c}_{jg}$ – кількість студентів, які набрали за j-те завдання не менше g балів.

Систему (4) можна наближено обчислити, наприклад, методом дотичних [7]. Відповідні ітерації мають вигляд:

$$\theta_i^{(\nu+1)} = \theta_i^{(\nu)} + \frac{b_i - \sum_{j=1}^{K} \sum_{k=1}^{m_j} k \cdot P_{ijk}^{(\nu)}}{\sum_{j=1}^{K} \left[\sum_{k=1}^{m_j} k^2 \cdot P_{ijk}^{(\nu)} - \left(\sum_{k=1}^{m_j} k \cdot P_{ijk}^{(\nu)} \right)^2 \right]}, \tag{5}$$

$$\beta_{jg}^{(\mu+1)} = \beta_{jg}^{(\mu)} + \frac{-c_{jg} + \sum_{i=0}^{M} n_i \sum_{k=g}^{m_j} P_{ijk}^{(\mu)}}{\sum_{i=0}^{M} n_i \cdot \left[\sum_{k=g}^{m_j} P_{ijk}^{(\mu)} - \left(\sum_{k=g}^{m_j} P_{ijk}^{(\mu)} \right)^2 \right]}, \qquad (6)$$

де p_{ijk} знаходять за формулою (3). Початкові наближення вибирають з умов:

$$\theta_i^{(0)} = \ln \frac{i}{M-i}, \ \beta_{jg}^{(0)} = \ln \frac{c_{jg-1}}{c_{jg}},$$

де c_{jg} – кількість учнів, які набрали за j - те завдання рівно g балів.

Ітераційний процес буде завершено при виконанні умови:

$$\sqrt{\sum_{i=0}^{M} \left(\theta_i^{(v+1)} - \theta_i^{(v)} \right)^2 + \sum_{j=1}^{m} \sum_{k=1}^{m_j} \left(\beta_{jk}^{(\mu+1)} - \beta_{jk}^{\mu} \right)^2} < \varepsilon.$$

Точність одержуваних при цьому оцінок характеризується середніми квадратичними похибками:

$$S_e\left(\hat{\theta}_i\right) = \left(\sum_{j=1}^{K} \left(\sum_{k=1}^{m_j} k^2 \cdot P_{ijk} - \left(\sum_{k=1}^{m_j} k \cdot P_{ijk} \right)^2 \right) \right)^{-1/2},$$

$$S_e\left(\hat{\beta}_{jg}\right) = \left(\sum_{i=0}^{M} n_i \left(\sum_{k=g}^{m_j} P_{ijk} - \left(\sum_{k=g}^{m_j} P_{ijk} \right)^2 \right) \right)^{-1/2}.$$

Одержані оцінки параметрів для нормально розподілених первинних балів є асимптотично-нормальними, асимптотично-незсуненими, асимптотично-ефективними.

5. Висновки

• Застосування дихотомічних моделей Раша і Бірнбаума для дослідження тестової системи комплекту «Вища математика» виявилось цілком обґрунтованим та доречним і дозволило суттєво підвищити ефективність формування та аналізу тестів.
• З урахуванням позитивного досвіду використання дихотомічної моделі, сформовано системи нелінійних рівнянь для визначення латентних параметрів моделі Раша у політомічному випадку. Наведено алгоритми обчислення оцінок латентних параметрів як розв’язків відповідних систем.

Наведені алгоритми обчислення латентних параметрів найближчим часом будуть застосовані до аналізу тестових завдань комплекту «Вища математика».

СПИСОК ЛІТЕРАТУРИ

1. Тестова система для комплекту дистанційних курсів «Вища математика» / І.В. Алєксеєва, В.О. Гайдей, О.О. Диховичний [та ін.] // IX международная научная конференция имени Т.А. Таран «Интеллектуальный анализ информации ИАИ-2009»: сб. трудов. – К.: Просвіта, 2009. – С. 4 – 10.
2. Застосування сучасних математичних моделей педагогічного тестування у формуванні та аналізі тестових завдань комплекту «Вища математика» / І.В. Алєксеєва, В.О. Гайдей, О.О. Диховичний [та ін.] // Дидактика математики: проблеми і дослідження: міжнародний збірник наукових робіт. – Донецьк: Вид-во ДонНТУ, 2010. – Вип. 33. – С. 50 – 56.
3. Rasch G. Probabilistic models for some intelligence and attainment tests / Rasch G. // Copenhagen: Danish Institute for Educational Research, 1960. – 126 p.

4. Нейман Ю.М. Введение в теорию моделирования параметризации педагогических тестов / Ю.М. Нейман, В.А. Хлебников. – М.: Прометей, 2000. – 168 с.
5. Челышкова М.Б. Теория и практика конструирования педагогических тестов / Челышкова М.Б. – М.: Логос, 2002. – 432 с.
6. Linden W. Handbook of Modern Item Response Theory / W. Linden, R. Hambleton R. – NY.: Springer-Verlag, 1997. – 510 p.
7. Цыганов Ш.И. Математические методы педагогических измерений / Ш.И. Цыганов // Вестник Башкирского университета. – 2009. – Т. 14, № 3. – С. 1 – 8.

Стаття надійшла до редакції 22.10.2010