

УДК 681.518.3.08

DOI: <http://dx.doi.org/10.20535/0203-3771342017130262>Н. А. Яремчук¹, професор, к.т.н., А. А. Кравченко², асистент**АНАЛІЗ СКЛАДОВОЇ НЕВИЗНАЧЕНОСТІ ВІД НЕСТАБІЛЬНОСТІ
ЛАТЕНТНОГО ПАРАМЕТРУ ПІД ЧАС ТЕСТУВАННЯ****En**

This article presents the results of the research component of the uncertainty of instability latent parameter - the level of knowledge that is determined by the test results according to the Item Response Theory (IRT). Earlier studies of the latent parameter instability have already been done by Neiman and Khlebnikov. They used linearization functions. But, these results are valid only if we have a large number of questions and when the measurement nonlinearity can be neglected. The authors of this study analyzed error results measuring the level of knowledge. The research allowed the evaluating of the latent instability parameter influence considering the measurement nonlinearity. In the resulting ratios we used the number of correct answers and the number of questions in the test. These ratios allow calculating the characteristics of uncertainty at different elevations scale of the knowledge level.

The authors demonstrat that the instability of latent parameter leads to a shift of the knowledge level assessing, and this shift is significant and commensurate with the random component of uncertainty that is characterized by standard deviation error.

To improve the accuracy of the knowledge level measuring the measurement result correcting is offered. Thus, the authors calculated the adjustments for each mark range. Standard deviation uncertainty component from the latent parameter instability obtained by the formula is summarized with other components of uncertainty.

Ru

В работе проведен анализ неопределенности от нестабильности латентного параметра при тестировании в условиях ограниченности количества вопросов в тесте. Было показано, что полученные в литературе соотношения справедливы только для большого количества вопросов в тесте. В других случаях несправедлива линеаризация уравнения погрешности. В работе по-

¹ НТУУ «Київський політехнічний інститут імені Ігоря Сікорського», кафедра інформаційно вимірною вальної техніки

² НТУУ «Київський політехнічний інститут імені Ігоря Сікорського», кафедра інформаційно вимірною вальної техніки

лучены составляющие погрешности и смещение оценки уровня обученности и среднее квадратическое отклонение случайной составляющей погрешности. Смещение оценки может быть скорректировано, а случайная составляющая учитывается при вычислении суммарной неопределенности уровня обученности.

Вступ

В роботі наведено результати дослідження складової невизначеності від нестабільності латентного параметру – рівня навченості, який визначають за результатами тестування. У відповідності з теорією *IRT (Item Response Theory)* рівень навченості θ в прийнятих одиницях – логітах визначається за формулою:

$$\theta_i = \ln \frac{b_i}{k - b_i},$$

де b_i – кількість правильних відповідей, отриманих за k запитаннями, заданими під час тестування.

За однакової кількості запитань значення b_i може змінюватися у залежності від психоемоційного стану особи, що проходить тестування, і від умов тестування. Ці зміни, що впливають на результат тестування, називають нестабільністю латентного параметру. Щоб зменшити цю складову невизначеності та підвищити надійність тестування, процедуру тестування проводять багаторазово в однакових умовах, а результати усереднюють. Але далеко не завжди можна проводити такі експерименти, і в реальних системах тестування результати вимірювання рівня навченості отримують за одним сеансом тестування. Тому виникає необхідність у дослідженні цієї складової невизначеності у разі вимірювання рівня навченості за результатами тестування.

У роботі [1] проведено дослідження впливу нестабільності латентного параметру за умов великої кількості запитань k і використання лінеаризації функції $\theta = f(b_i)$. Математичною моделлю змін b_i від нестабільності латентного параметру вважають біноміальний розподіл, за яким ймовірність правильної відповіді $p(\xi = b_i)$ визначають за формулою:

$$p(\xi = b_i) = C_k^{b_i} \cdot p^{b_i} (1 - p)^{k - b_i},$$

де p – ймовірність відповіді.

З використанням моментів біноміального розподілу та емпіричної ймовірності p :

$$M(\xi) = kp; \quad D(\xi) = kp \cdot (1 - p); \quad p = \frac{b_i}{k},$$

отримуємо наступні співвідношення:

$$M(\xi) = b_i; \quad D(\xi) = b_i \cdot \left(1 - \frac{b_i}{k}\right).$$

Тоді, якщо використати лінеаризацію, середнє квадратичне відхилення (СКВ) похибки від нестабільності латентного параметру може бути визначене за формулою:

$$\sigma(\theta) = \frac{\partial \theta}{\partial b_i} \sigma(b_i) = \frac{\partial \theta}{\partial b_i} \sqrt{D(\xi)} = \sqrt{\frac{k}{b_i(k-b_i)}}. \quad (1)$$

Якщо результат вимірювання знаходиться в середині діапазону, тобто

$$b_i = k/2.$$

Тоді

$$\sigma(\theta)|_{b_i=k/2} = \frac{2}{\sqrt{k}}.$$

Так як на кінцях діапазону $\sigma(\theta)$ збільшується, у [1] прийнята узагальнена формула:

$$\sigma(\theta) \approx \frac{2,6}{\sqrt{k}}.$$

Але, як показали дослідження, проведені авторами, за малу кількість запитань k лінеаризація рівняння похибки неможлива, і є необхідність в урахуванні зміщення лінеаризації у разі отримання результату вимірювання θ , а також, у отриманні співвідношення для оцінювання дисперсії оцінки за різну кількість правильних відповідей, тобто на різних відмітках шкали рівня навченості.

Постановка задачі

В роботі проведено аналіз складової невизначеності від нестабільності латентного параметру – рівня навченості за умов малої кількості запитань у тесті. Визначено зміщення оцінки рівня навченості, значення якого можна використати для корекції оцінки рівня навченості. Отримано співвідношення для обчислення СКВ оцінки рівня навченості від нестабільності латентного параметру, значення якого треба враховувати при обчисленні сумарної невизначеності.

Визначення зміщення оцінки рівня навченості від нестабільності латентного параметру

Для функції $y = f(x)$ із суттєвою нелінійністю треба враховувати зміщення оцінки за формулою [2]:

$$\bar{y} = f(\bar{x}) \cdot 1 + \frac{1}{2} \cdot f''(\bar{x}) \cdot D(x).$$

Тому для зміщення оцінки θ отримуємо вираз:

$$\Delta\theta = \frac{1}{2} \cdot \frac{\partial^2 \theta}{\partial b_i^2} \cdot D(\xi);$$

$$\frac{\partial^2 \theta}{\partial b_i^2} = \frac{-(k - 2b_i) \cdot k}{b_i^2 (k - b_i)^2};$$

$$D(\xi) = b_i \left(1 - \frac{b_i}{k}\right) = \frac{b_i (k - b_i)}{k};$$

$$\Delta\theta = \frac{1}{2} \cdot \frac{-(k - 2b_i) \cdot k \cdot b_i \cdot (k - b_i)}{b_i^2 (k - b_i)^2 \cdot k} = -\frac{k - 2b_i}{2b_i (k - b_i)}.$$

В середині діапазону ($b_i = k/2$) зміщення оцінки дорівнює нулю. На крайніх відмітках діапазону вимірювання при $b_i = k - 1$ і $b_i = 1$ отримуємо:

$$\Delta\theta|_{b_i=(k-1)} = \frac{k - 2}{2(k - 1)};$$

$$\Delta\theta|_{b_i=1} = -\frac{k - 2}{2(k - 1)}.$$

Якщо для прикладу $k = 8$; $\theta|_{b_i=7} = 1,95$ логіта; $\Delta\theta|_{b_i=7} = 0,43$ логіта, тобто відносно зміщення оцінки складає 22%.

Для порівняння за $k = 20$; $\theta|_{b_i=19} = 2,94$ логіта; $\Delta\theta|_{b_i=99} = 0,47$ логіта, що складає 16%. Тобто зміщення оцінки суттєве навіть за умов великої кількості запитань у тесті.

Тому мають бути враховані зміщення для кожного значення θ .

На рис. 1 зображено графік стандартної невизначеності вимірювання рівня навченості $\sigma(\theta)$ та зміщення оцінки рівня навченості для різної кількості запитань.

Визначення СКВ або стандартної невизначеності від нестабільності латентного параметру

Дисперсію оцінки θ із урахуванням нелінійної залежності від b_i , можна отримати за формулою [2]:

$$D(y) = [f'(\bar{x})]^2 D(x) + \frac{1}{4} [f''(\bar{x})]^2 [\mu_4(x) - D^2(x)] + f'(\bar{x}) \cdot f''(\bar{x}) \cdot \mu_3(x).$$

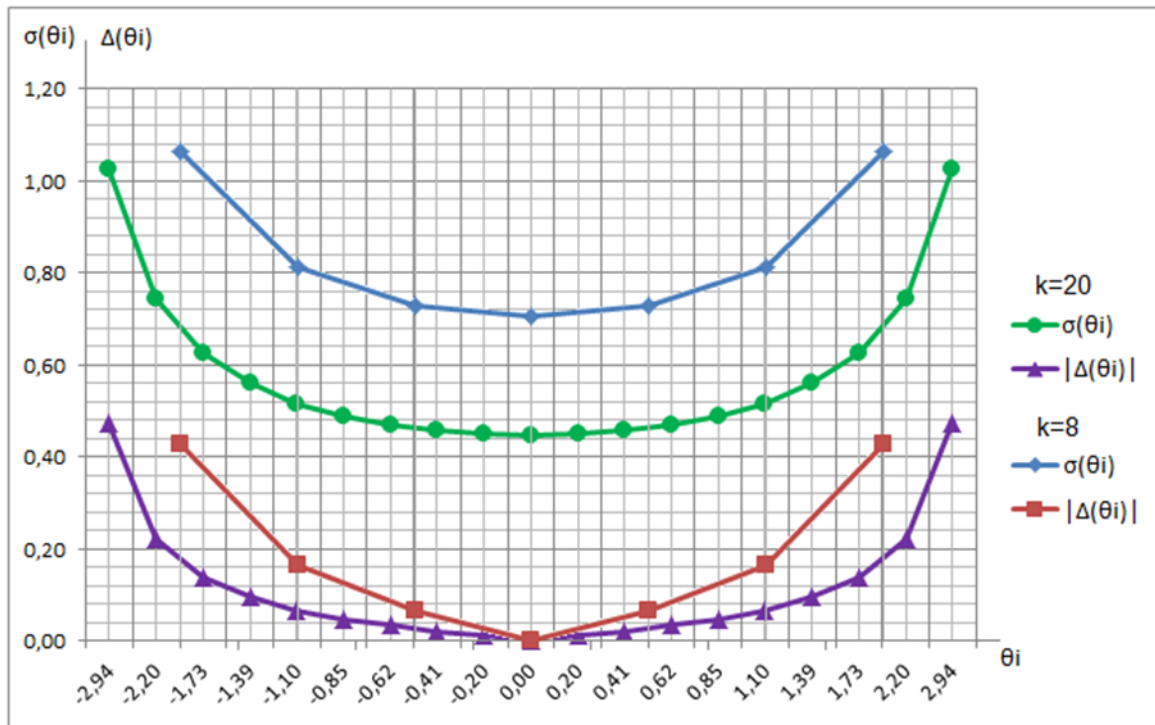


Рис. 1. Графік стандартної невизначеності вимірювання рівня навченості $\sigma(\theta)$ та зміщення оцінки рівня навченості $\Delta\theta$ для різної кількості запитань ($k = 20$ та $k = 8$)

Тоді для оцінки θ отримуємо вираз:

$$D(\theta) = [f'(b_i)]^2 D(\xi) + \frac{1}{4} [f''(b_i)]^2 [\mu_4(\xi) - D^2(\xi)] + f'(b_i) \cdot f''(b_i) \cdot \mu_3(\xi), \quad (2)$$

де $\mu_3(\xi) = np(1-p)(1-2p)$ і

$\mu_4(\xi) = np(1-p)[1 + 3(n-2)p(1-p)]$ - центральні третій та четвертий моменти біноміального розподілу [3], що за прийнятою емпіричною ймовірністю дорівнюють:

$$\mu_3(\xi) = \frac{(b_i(k-b_i)(k-2b_i))}{k^2}; \quad (3)$$

$$\mu_4(\xi) = \frac{(b_i(k-b_i))}{k} \left[1 + 3(k-2) \cdot \frac{(b_i(k-b_i))}{k^2} \right]. \quad (4)$$

Тоді за (2) із урахуванням (3) та (4) дисперсія рівня навченості визначається як:

$$D(\theta) = \frac{k}{b_i(k-b_i)} + \frac{(k-2b_i)^2}{b_i^2(k-b_i)^2} \left\{ \frac{1}{4} \frac{k}{b_i(k-b_i)} \left[1 + 2 \frac{b_i(k-b_i)}{k} - \left[6 \frac{b_i(k-b_i)}{k^2} \right] \right] - 1 \right\}. \quad (5)$$

Перша складова у формулі (5) є дисперсія, яку отримують із використанням лінеаризації, друга складова додається на нелінійній частині шкали рівня навченості. Тому у середині діапазону за $b_i = \frac{k}{2}$ формули для СКВ за виразами (1) та (5) співпадають.

Якщо взяти крайні точки діапазону $b_i = k-1$ та $b_i = 1$, то отримуємо однакові вирази для обчислення дисперсії, а саме:

$$D(\theta) = \frac{k}{k-1} + \frac{(k-2)^2}{(k-1)^2} \left\{ \frac{k}{4(k-1)} \left[1 + 2 \frac{k-1}{k} - 6 \frac{k-1}{k^2} \right] - 1 \right\}.$$

У разі $k = 8$; $D(\theta) = 0,848$; $\sigma(\theta) = 0,921$ логіт. Для порівняння усереднена оцінка, що наведена у [1], $\sigma_y(\theta) \approx 1,03$ логіт.

У разі $k = 20$; $D(\theta) = 0,770$; $\sigma(\theta) = 0,878$ логіт. Для порівняння усереднена оцінка, що наведена у [1], становить $\sigma_y(\theta) \approx 0,58$ логіт. Тобто у крайніх відмітках діапазону вимірювання СКВ похибки більше за усереднену оцінку.

Отримане співвідношення для дисперсії (5) може бути використане для оцінювання стандартної невизначеності від нестабільності латентного параметру на всіх відмітках діапазону вимірювання рівня навченості (рис. 2).

Висновки

Відомо, що нестабільність латентного параметру, що викликана нестабільністю психоемоційного стану осіб, що проходять тестування, і впливом умов тестування, спотворює результати тестування і відповідно результати вимірювання рівня навченості. Для оцінки її впливу використовують наближені формули СКВ складової невизначеності від нестабільності латентного параметру, що отримані із використанням лінеаризації рівняння вимірювання. За цими формулами можна оцінити тільки усереднене значення СКВ за діапазоном вимірювання у залежності від кількості запитань у тесті.

У роботі проведено аналіз похибки результатів вимірювання рівня навченості, який дозволив оцінити вплив нестабільності латентного параметру із урахуванням нелінійності рівняння вимірювання.

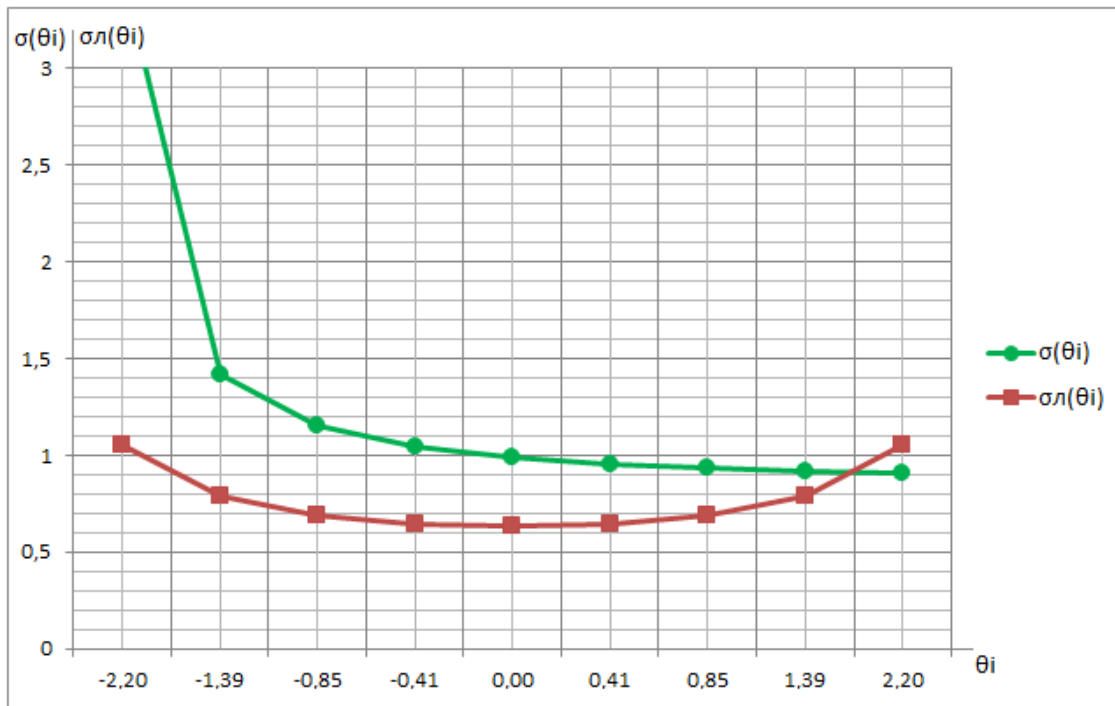


Рис. 2. Графік СКВ стандартної невизначеності вимірювання рівня навченості із лінеаризацією $\sigma_l(\theta)$ та без лінеаризації $\sigma(\theta)$ за $k = 8$

У отриманих співвідношеннях використовується кількість правильних відповідей і кількість запитань у тесті, тобто вони дозволяють обчислити характеристики невизначеності на різних відмітках шкали рівня навченості.

Показано, що нестабільність латентного параметру приводить до зміщення оцінки рівня навченості, причому це зміщення суттєве та сумірне із випадковою складовою невизначеності, яка характеризується СКВ (рис. 1).

Для підвищення точності вимірювання рівня навченості пропонується корегувати результат вимірювання, причому коригуюча поправка обчислюється для кожної відмітки діапазону. А отримане за наведеними формулами значення СКВ складової невизначеності від нестабільності латентного параметру підсумовується з іншими складовими невизначеності.

Список використаної літератури

1. Нейман Ю. В. Введение в теорию моделирования и параметризации педагогических тестов/ Ю. В. Нейман, В. А. Хлебников. – М.: Прометей, 2000. – 168 с. – ISBN 5-7042-1068-6.
2. Сборник задач по теории вероятностей, математической статистике и теории случайных функций, под редакцией А. А. Свешникова, – М: Наука, 1970. – 656 с.

3. *Корн Г.* Справочник по математике (для научных работников и инженеров)/ Г. Корн, Т. Корн, – Перевод с англ. под общей редакцией Арамановича И. Г.– М.:Наука, 1979. – 831 с.