П.А. Ротаєнко,
Інститут педагогіки АПН України

Дисперсія розподілу частки знань умовного тестованого

Зауважимо спочатку, що поняття «умовний тестований» в нашому тлумаченні узгоджується зі знаннями практично необмеженої кількості тестованих, опитування яких дало однаковий результат за одним і тим самим тестом і тією самою методикою. Розподіл рівнів знань таких тестованих може служити об'єктивною характеристикою процесу тестування і для конкретного тестованого, оскільки для стороннього спостерігача, зокрема екзаменатора, його знання апріорі є такими ж імовірними.

Нехай є гомогенний тест з вивіреним предметним наповненням. Будемо користуватися тими ж позначеннями, що і в [2], а саме:

1) x – величина, що виражає відношення кількості запитань тесту, на які тестований може правильно відповісти, до кількості всіх запитань;

2) β – ймовірність чистого вгадування (ймовірність чистого вгадування – це ймовірність дати правильну відповідь при повному незнанні матеріалу), яку називатимемо ще коефіцієнтом тесту;

3) $n-$ кількість запитань одержаної тестованим вибірки;

4) $m-$ кількість запитань із цієї вибірки, на які він відповів правильно.

Якщо, наприклад, до кожного тестового запитання наводиться 5 відповідей, з яких тільки одна правильна, то $\beta = 1/5$, для тесту «2 із 4-х» $\beta = 1/C_4^2 = 1/6$, якщо ж кількість правильних відповідей (включаючи відсутність останніх) невідома, то $\beta = 1/2^r$, де $r-$ кількість усіх наведених відповідей. У випадку відкритої форми тестових завдань значення β можна задати наближено чи взяти нульовим.

Величину x можна вважати випадковою характеристикою. Крім того, якщо тест практично необмежений, то її можна вважати неперервною випадковою величиною (принаймні у граничному тлумаченні).

Математичне сподівання і дисперсію для величини x будемо позначати відповідно через M і D.

У [2] доведено, що при заданих (фіксованих) значеннях величин β, n, m функція

$$f(x) = (1-\beta)(n+1)C_n^m (x+\beta(1-x))^m (1-x-\beta(1-x))^{n-m} \qquad (1)$$

виражає щільність розподілу імовірностей на множині значень випадкової величини x на відрізку $[\beta/(\beta-1);1]$, а функцією розподілу для неї є функція

$$F(x) = \sum_{k=m+1}^{n+1} C_{n+1}^k (x+\beta(1-x))^k (1-x-\beta(1-x))^{n+1-k}. \qquad (2)$$

Математичне сподівання величини x на вказаному відрізку, яке ми назвали імовірнісним коефіцієнтом знань і позначали через α, виведено у вигляді

$$\alpha = M = \frac{m+1-\beta(n+2)}{(1-\beta)(n+2)}. \qquad (3)$$

Крім того, оскільки

$$\int_{\beta/(\beta-1)}^{1} f(x)dx = F(1) - F(\beta/(\beta-1)) = 1 - 0 = 1,$$

то $$\int_{\beta/(\beta-1)}^{1} C_n^m (x+\beta(1-x))^m (1-x-\beta(1-x))^{n-m} dx = \frac{1}{(1-\beta)(n+1)}, \qquad (4)$$

які б не були значення m $(0 <= m <= n)$.

На основі (1), (2), (4) можна записати:

$$\int_{\beta/(\beta-1)}^{1} (\sum_{s=t}^{T} C_T^s (x+\beta(1-x))^s (1-x-\beta(1-x))^{T-s})dx =$$

$$= \sum_{s=t}^{T} \int_{\beta/(\beta-1)}^{1} C_T^s (x+\beta(1-x))^s (1-x-\beta(1-x))^{T-s} dx =$$

$$= \sum_{s=t}^{T} \frac{1}{(1-\beta)(T+1)} = \frac{(T-t+1)}{(1-\beta)(T+1)},$$

де t, s, T – натуральні числа, $t \le T$.

Тоді матимемо:

$$D = \int_{\beta/(\beta-1)}^{1} (x-M)^2 f(x)dx = \int_{\beta/(\beta-1)}^{1} x^2 f(x)dx - 2M \int_{\beta/(\beta-1)}^{1} xf(x)dx + M^2 \int_{\beta/(\beta-1)}^{1} f(x)dx = \left| x^2 F(x) \right|_{\beta/(\beta-1)}^{1} -$$

$$-2\int_{\beta/(\beta-1)}^{1} xF(x)dx - 2M*M + M^2 \left| F(x) \right|_{\beta/(\beta-1)}^{1} = 1 - 2\left| x\int F(x)dx \right|_{\beta/(\beta-1)}^{1} + 2\int_{\beta/(\beta-1)}^{1} dx \int F(x)dx - 2M^2 + M^2 =$$

$$= 1 - M^2 - 2\left| x\int dx \sum_{k=m+1}^{n+1} C_{n+1}^k (x+\beta(1-x))^k (1-x-\beta(1-x))^{n+1-k} \right|_{\beta/(\beta-1)}^{1} + 2\int_{\beta/(\beta-1)}^{1} dx \int dx \sum_{k=m+1}^{n+1} C_{n+1}^k (x+\beta(1-$$

$$-a)^k (1-x-\beta(1-x))^{n+1-k} = 1 - M^2 - 2\left| x\sum_{k=m+1}^{n+1} \int C_{n+1}^k (x+\beta(1-x))^k (1-x-\beta(1-x))^{n+1-k} dx \right|_{\beta/(\beta-1)}^{1} +$$

$$+2\int_{\beta/(\beta-1)}^{1} dx \sum_{k=m+1}^{n+1} \int C_{n+1}^k (x+\beta(1-x))^k (1-x-\beta(1-x))^{n+1-k} dx = 1 - M^2 - 2\frac{(n-m+1)}{(1-\beta)(n+2)} +$$

$$+2\int_{\beta/(\beta-1)}^{1} dx \sum_{k=m+1}^{n+1} \frac{1}{(1-\beta)(n+2)} \sum_{j=k+1}^{n+2} C_{n+2}^j (x+\beta(1-x))^j (1-x-\beta(1-x))^{n+2-j} = 1 - M^2 - 2\frac{(n-m+1)}{(1-\beta)(n+2)} +$$

$$+2\frac{1}{(1-\beta)(n+2)} * \frac{1}{(1-\beta)(n+3)}((n-m+1)+(n-m)+(n-m-1)+..+1) = 1 - (\frac{(m+1)-\beta(n+2)}{(1-\beta)(n+2)})^2 -$$

$$-2\frac{(n-m+1)}{(1-\beta)(n+2)} + \frac{1}{(1-\beta)(n+2)} * \frac{1}{(1-\beta)(n+3)}(n-m+2)(n-m+1) = \frac{(m+1)(n-m+1)}{(1-\beta)^2 (n+2)^2 (n+3)}.$$

Отже, $$D = \frac{(m+1)(n-m+1)}{(1-\beta)^2 (n+2)^2 (n+3)}. \qquad (5)$$

З (5) випливає, що значення дисперсії найменше на кінцях проміжку можливих значень величини m, тобто при $m = 0$ і $m = n$. Дисперсія в таких випадках дорівнює $\frac{(n+1)}{(1-\beta)^2 (n+2)^2 (n+3)}$.

Оскільки $(n+2)^2 = (n+1)(n+3) + 1 \approx (n+1)(n+3)$, то для цих випадків можна записати

$$D \approx \frac{1}{(1-\beta)^2 (n+3)^2}. \qquad (6)$$

Середньо-квадратичне відхилення відносно M, тобто середньо-квадратична похибка для цих випадків дорівнює

$$\sigma = \sqrt{D} \approx \frac{1}{(1-\beta)(n+3)} . \qquad (7)$$

З (7) випливає, що похибка при обчисленні частки знань за формулою (3) для найкращого і найгіршого результатів знаходиться приблизно в обернено пропорційній залежності від обсягу вибірки. Взагалі, з (5) випливає, що відношення середньо-квадратичних похибок для різних тестів і однакових кількостей правильних і неправильних відповідей є таким

$$\frac{\sigma_1}{\sigma} = \frac{1-\beta}{1-\beta_1} , \qquad (8)$$

де σ, σ_1 – середньо-квадратичні похибки, β, β_1 – відповідні коефіцієнти тестів.

Якщо через σ позначити середньо-квадратичну похибку коефіцієнта знань, визначеного за відкритим тестом ($\beta = 0$), то на основі (8) відповідну похибку будь-якого тесту можна виразити через цю похибку:

$$\sigma_i = \sigma / (1-\beta_i) . \qquad (8')$$

Не менш важливо мати і найбільше значення диперсії чи середньо-квадратичної похибки, бо саме такий критерій визначає надійність усіх результатів тестування («найслабша ланка визначає надійність усього ланцюга»). З (5) легко одержати ці значення: для парного значення n вони досягаються при $m = \frac{n}{2}$:

$$D = \frac{1}{4(1-\beta)^2(n+3)} , \quad \sigma = \frac{1}{2(1-\beta)\sqrt{n+3}} , \qquad (9)$$

а для непарного при $m = \frac{n-1}{2}$ та при $m = \frac{n+1}{2}$:

$$D = \frac{n+1}{4(1-\beta)^2(n+2)^2} , \quad \sigma = \frac{\sqrt{n+1}}{2(1-\beta)(n+2)} . \qquad (9')$$

Коефіцієнт знань α (3) при $n = 2m$ дорівнює

$$\alpha = \frac{m+1-\beta(2m+2)}{(1-\beta)(2m+2)} = \frac{1-2\beta}{2(1-\beta)} .$$

Оскільки величини (9) дещо більші за відповідні величини (9'), останньому значенню коефіцієнта знань відповідає найбільша середньо-квадратична похибка порівняно з випадками $m \neq \frac{n}{2}$, а отже, такі знання можна вважати «найменш визначальними». Назвемо це значення коефіцієнта знань (частки знань) «критичним» і позначатимемо через α_{kr},

$$\alpha_{kr} = \frac{1-2\beta}{2(1-\beta)} . \qquad (10)$$

Як бачимо, для кожного тесту є критичне значення коефіцієнта знань, яке найважче визначити за цим тестом. Для $\beta = 1/5$, наприклад, $\alpha_{kr} = \frac{1-0.4}{2(1-0.2)} = 0.375$, для відкритого тесту $\alpha_{kr} = \frac{1-2*0}{2(1-0)} = 0.5$.

Відповідне значення критичної частки знань у балах називатимемо критичним балом даного тесту у даній системі оцінювання і позначатимемо через O_{kr}. Наприклад, у 12-бальній системі для тесту «1 з 5-и» $O_{kr} = 12*0.375 = 4.5(б.)$, для відкритого тесту $O_{kr} = 12*0.5 = 6(б.)$.

Оскільки критичним значенням коефіцієнта знань для різних тестів завжди відповідає відношення $m = n/2$, то для похибок цих значень завжди мають місце рівності (8), (8').

З'ясуємо тепер, як можна оцінити звичайну (абсолютну) похибку значення α. Як уже зазначалося, таку оцінку можна дати через оцінку її імовірності, тобто оцінивши ймовірність нерівності $|x-\alpha| \le \delta$, де δ – задане додатне число, з точністю до якого шукається абсолютна похибка:

$$P(|x-\alpha| \le \delta) = ?$$

Оскільки розподіл імовірностей на множині значень величини x є унімодальним і асиметричним (крім випадку $n = 2m$), можна скористатися нерівностями Б'єнеме-Чебишева та Гауса з коефіцієнтом асиметрії Пірсона [1, с. 206-207]. Однак одразу ж зауважимо, що одержана таким чином оцінка, незважаючи на свою надійність і строгість, матиме, перш за все, теоретичне значення, тому що практично вона, як правило, є завищеною, тобто може служити достатнім критерієм, але не завжди може служити необхідним критерієм.

Нерівність Чебишева має вигляд

$$P(|x-\alpha| \le K*\sigma) > 1 - 1/K^2, \quad (11)$$

де σ – середньо-квадратичне наближення, $K > 0$.

Нерівність Гауса має вигляд

$$P(|x-\alpha| \ge K*\sigma) \le \frac{4(1+|s|^2)}{9(K-|s|)^2}, \ |s| < K, \quad (12)$$

де s – коефіцієнт асиметрії Пірсона,

$$s = \frac{\alpha - \mu}{\sigma}, \quad (12')$$

α – математичне сподівання, μ – мода розподілу.

Розглянемо деякі випадки для відкритого тесту ($\beta = 0$). Для інших тестів, як уже зазначено, похибки σ можуть бути виражені через похибку цього тесту за формулою (8'), а отже, таке саме відношення залишається і для абсолютних похибок у (11) і (12).

Якщо взяти випадок результатів тестування, коли $m = n$, то для того, щоб у правій частині нерівності (11) мати, наприклад, не менше 90%, достатньо покласти $K = 19/6$. Тоді точність, з якою обчислюється абсолютна похибка, визначатиметься величиною $K*\sigma = 19\sigma/6$. Але ж у цьому випадку σ визначається за формулою (7), тому

$$P(|x-\alpha| \le 19/(n+3)/6) > 0,9 = 90\%,$$

звідки завжди можна знайти значення обсягу вибірки n за даною точністю δ.

При $\delta = 1/12$, наприклад, достатньо взяти $n = 35$. Це означає, що при 35 правильних відповідях на 35 запитань відкритого тесту у більше ніж 90% випадків за формулою (3) отримується оцінка з точністю до одного бала. Для оцінки з точністю до половини бала треба взяти $n \ge 73$. Оцінки, як побачимо нижче, дещо занижені (мінімальні критеріальні значення n завищені).

Якщо взяти інший крайній випадок, коли середньо-квадратична похибка найбільша, тобто якщо $n = m/2$, то нерівність Б'єнеме-Чебишева дає значно гіршу оцінку. У цьому випадку, або ж якщо значення коефіцієнта знань близьке до критичного, краще користуватися нерівністю (12). Справді, допустимо, що $2m - n = 1$. Тоді

$$\sigma = \frac{\sqrt{n+1}}{2(n+2)}, \ \mu - \alpha = \frac{2m-n}{n(n+2)} = \frac{1}{n(n+2)}, \ s = \frac{2}{n\sqrt{n+1}}.$$

При $n = 80$, наприклад,

$$\sigma = 9/164 \approx 1/18, s = 1/360.$$

Поклавши $K = 1{,}5$, матимемо:

$$P(|x-\alpha| \ge 1/12) \le 0{,}203 \approx 20\%.$$

Якщо $2m - n = 0$, то

$$\sigma = \frac{1}{2\sqrt{n+3}}, \ \mu = 0, \ s = 0.$$

При $n = 78$ $\sigma = 1/18, s = 0$, то поклавши знову $K = 1{,}5$, матимемо

$$P(|x-\alpha| \ge 1/12) \le 16/81 < 20\%.$$

Остання нерівність означає, що відповівши на 39 запитань із 80 відкритого тесту правильно, тестований заслуговує 6-и балів у 12-бальній системі з точністю до одного бала, яка гарантована для 80% опитуваних.

З (5) випливає, що при $n \to \infty$ $\sigma \to 0$, тобто при необмеженому збільшенні обсягу вибірки середньо-квадратична похибка прямує до нуля, а отже і абсолютна похибка прямує до нуля (точніше ймовірність її малості прямує до одиниці, що випливає з (11)), тому можна говорити про мінімальні обсяги вибірок для того чи іншого тесту, маючи на увазі певну точність оцінки та її імовірність, тобто говорити про репрезентативність вибірки. Але оскільки на значення дисперсії впливає і рівень знань (кількість правильних відповідей m), то при мінімальних обсягах вибірок для будь-яких знань будуть завищуватися такі обсяги для деяких знань. Очевидно, що можна було б згідно теорії користуватися такими характеристиками для критичного значення частки знань (у цьому випадку ці характеристики найгірші).

Для того, щоб мінімальні обсяги вибірок не завищувати, можна було б раз і назавжди обчислити прирости функції розподілу ймовірностей для будь-яких знань у балах і для певної імовірності відповідні співвідношення значень m і n.

ЛІТЕРАТУРА

1. *Крамер Г.* Математические методы статистики: Пер. с англ.–М.: Мир, 1975.– 648 с.

2. Ротаєнко П.А. Формула ймовірнісної оцінки знань за результатами комп'ютерного тестування // Інформатика та інформаційні технології в навчальних закладах. – 2006. – №2. – С. 52–55.