2 0 1 0 г **№ 15**

Т р у д ы Ф О Р А

ВЕРОЯТНОСТНЫЙ АНАЛИЗ ДОСТОВЕРНОСТИ ОЦЕНКИ ПРИ ТЕСТИРОВАНИИ С ВЫБОРОМ ОТВЕТА

Х.М. Андрухаев

Адыгейский государственный университет, г. Майкоп

Рассматривается один из вариантов анализа достоверности полученной информации с помощью тестирования с выбором правильного ответа из нескольких вариантов, один из которых верный, а остальные не верны. При этом применяется схема независимых испытаний Я. Бернулли и интервальные оценки для случая нормального закона распределения. Возможны и другие подходы, например, байесовский. Приведены некоторые примеры, иллюстрирующие выводы, сделанные в статье.

Введение

В настоящее время метод тестирования для получения информации применяется в самых разнообразных областях науки и практики. Естественно, он имеет свою историю возникновения и развития, свои положительные и отрицательные стороны.

В системе начального, среднего и высшего образования (особенно за рубежом) тестирование с выбором правильного ответа из нескольких вариантов начали применять давно, например, в нашей стране это было модно использовать в системе программированного обучения (1950-70 гг.). Особенно было распространено «да-нет» - тестирование. Этот примитивный вариант проведения тестирования применяется и в настоящее время при получении самой разнообразной информации в социологических исследованиях, а также для оценки знаний учащихся по различным дисциплинам. Возникает вопрос: насколько можно доверять выводам, сделанным в результате тестирования?

В общем случае научный анализ проблемы достоверности выводов на основе тестирования является сложной вероятностно-статистической задачей.

В предлагаемой статье рассматривается более узкая задача анализа достоверности оценочных выводов при проведении контрольных мероприятий в форме тестирования с выбором ответа.

I. Точечные оценки

Пусть тест содержит N заданий, к каждому из которых дается S ($S \geq 2$) ответов, один из них верный, а остальные – нет. При этом неверные ответы, так же как и верный, должны быть одинаково правдоподобными.

Предположим, что тестируемый знает как выполнить m заданий и поэтому выберет правильный ответ на них наверняка, т. е. с вероятностью 1. На остальные $N-m=n$ заданий тестируемый выбирает ответы «наудачу» и тогда вероятность того, что выбранный ответ будет верным для каждого из n заданий равна $p=1/S$.

Обозначим через k (k - целое, $0 \leq k \leq n$) число правильных ответов, которое наберет тестируемый, выбирая ответы наудачу на указанные выше n заданий. Таким образом общее число правильных ответов на N заданий будет равно $m+k$ (на m заданий наверняка ответы будут верными, а k - угадает).

Назовем опытом выбор наудачу ответа на каждый из n заданий, успехом – событие – выбранный ответ верный, неудачей – выбранный ответ неверный и обозначим через A_i - событие – в $i^{м}$ опыте успех, а через $\overline{A}_i$ - событие – в $i^{м}$ опыте неудача, $i=1,2,...,n.$ Отметим, что выборы ответов на те m заданий, которые наверняка тестируемый может правильно выполнить, не считаются опытами.

В описанной ситуации возникают ряд вопросов, например:

1. Чему равна вероятность того, что в n опытах тестируемый получит k успехов?
2. Чему равно наиболее вероятное число успехов в n опытах?

Ответом на первый вопрос служит известная в теории вероятностей формула Я. Бернулли (1654-1705, Швейцария):

$$P_n(k) = \frac{n!}{k!(n-k)!} p^k q^{n-k}, \qquad (1)$$

где $p = 1/S$, $q = 1 - p = 1 - 1/S$, $P_n(k)$ - вероятность k успехов в n опытах. (2)

Чтобы ответить на второй вопрос, обозначим через k_0 наиболее вероятное число успехов в n опытах и рассмотрим систему двух неравенств

$$\begin{cases} P_n(k_0) \geq P_n(k_0+1), \\ P_n(k_0) \geq P_n(k_0-1) \end{cases} \qquad (3)$$

которые по определению k_0, очевидно, выполняются.

Нам надо найти k_0 из (3). Преобразуем первое неравенство, используя формулу Бернулли:

$$C_n^{k_0} p^{k_0} q^{n-k_0} \geq C_n^{k_0+1} p^{k_0+1} q^{n-k_0-1},$$

$$\frac{n!}{k_0!(n-k_0)!} p^{k_0} q^{n-k_0} \geq \frac{n!}{(k_0+1)!(n-k_0-1)!} p^{k_0+1} q^{n-k_0-1};$$

Откуда следует:

$$\frac{q}{n-k_0} \geq \frac{p}{k_0+1}$$

$$qk_0 + q \geq np - k_0 p,$$

$$(q+p)k_0 \geq np - q;$$

$$k_0 \geq np - q, \text{ т. к. } q + p = 1. \qquad (4)$$

Аналогично, преобразуя второе неравенство системы (3), получаем

$$k_0 \leq np + p \qquad (5)$$

Из (4) и (5) имеем двойное неравенство

$$np - q \leq k_0 \leq np + p \qquad (6)$$

Промежуток $[np - q, np + p]$ имеет длину 1, поэтому при не целом $np + p$, $np - q$ тоже будет не целым и $k_0 = [np + p]$ (квадратные скобки обозначают целую часть числа, заключенного в них.)

Если же $np + p$ - целое, то и $np - q$ так же будет целым. Следовательно k_0 имеет два значения $k_0' = np - q$, $k_0'' = np + p.$

Сформулируем теперь правило нахождения k_0 (наиболее вероятного числа успехов в n опытах):

находим $np + p$. Если $np + p$ не целое число, то $k_0 = [np + p]$; если же $np + p$ - целое число, то $k_0' = np - q$ или $k_0'' = np + p.$

Рассмотрим два примера

Контрольная работа состоит из N заданий (вопросов), предусматривающих два возможных варианта ответа («да» или «нет»). На вопросы, ответы на которых учащийся знет, он отвечает наверняка, а на остальные ответ выбирает наудачу «да» или «нет». Таким образом вероятность успеха $p = 1/2$ и вероятность неудачи $q = 1 - 1/2 = 1/2$. Пусть n по прежнему число вопросов, ответы на которых ученик выбирает наудачу (число опытов). Тогда для любого k $(0 \leq k \leq n)$

$$P_n(k)=\frac{n!}{k!(n-k)!}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{n-k}=\frac{n!}{k!(n-k)!2^n}.$$

К примеру, при $n=5$ и $k=2$

$$P_5(2)=\frac{5!}{2!3!\cdot 2^5}=\frac{5\cdot 4}{2\cdot 2^5}=\frac{5}{16}.$$

Следовательно вероятность двух успехов в пяти опытах равна $\frac{5}{16}$. Та же самая будет вероятность трех неудач в пяти опытах. Наиболее вероятное число успешных (верных) ответов в пяти заданиях находим по правилу: $np+p=5\cdot\frac{1}{2}+\frac{1}{2}=3$.

Значит $k_0'=2;\quad k_0''=3.$

Рассмотрим реальную ситуацию.

Контрольная работа состоит из 10 вопросов, предусматривающих два варианта ответа («да» или «нет»). Найти вероятность того, что:

а) ученик незнающий ни одного вопроса, выбирая наудачу, ответит верно на 4 вопроса правильно. Каково наиболее вероятное число k_0 правильных ответов может дать ученик?

Здесь: $n=10;\ S=2,\ p=\frac{1}{2},\ q=\frac{1}{2},\ k=4.$

По формуле Бернулли

$$P_{10}(4)=\frac{10!}{4!6!\cdot 2^{10}}=\frac{10\cdot 9\cdot 8\cdot 7}{4\cdot 3\cdot 2\cdot 2^{10}}=\frac{5\cdot 3\cdot 7}{2^9}=\frac{105}{512}\approx 0{,}2.$$

Далее находим $np+p=10\cdot\frac{1}{2}+\frac{1}{2}=5\frac{1}{2}$ - не целое, $k_0=\left[5\frac{1}{2}\right]=5.$

Наиболее вероятное число правильных ответов ученика равно 5. если, например, за правильный ответ начисляется 1 балл, за неправильный – 0 баллов, то ученик, заслуживающий 0 баллов, вероятнее заработает 5 баллов. К примеру, если учитель решает ставить: оценку «5» за 10 или 9 баллов,

оценку «4» за 8, 7, 6 баллов,

оценку «3» за 5, 4 балла,

оценку «2» за 3, 2, 1, 0 баллов,

то $P_{10}(4)+P_{10}(5)=0{,}4...;\ P_{10}(0)+P_{10}(1)+P_{10}(2)+P_{10}(3)=0{,}17...$

Учащийся, заслуживающий объективно «2», скорее всего получит оценку «3». Если же ему еще немного повезет, то может и «четверку» получить и даже не исключено, что получит «5». Вывод: такое тестирование с $S=2$ не следует проводить.

б) Из 10 вопросов ученик знает ответы на 3 вопроса, а ответы на остальные 7 вопросов выбирает наудачу. Какова вероятность того, что этот ученик правильно ответит на 4 вопроса? Каково наиболее вероятное число правильных ответов этот ученик может дать?

Теперь имеем $n=7$, $s=2$, $p=q=\frac{1}{2}$, $k=1$.

По формуле Бернулли

$$P_7(1)=\frac{7!}{1!6!\cdot 2^7}=\frac{7}{2^7}=\frac{7}{128}$$

Вероятность того, что ученик заработает 4 балла равна $\frac{7}{128}$ (на три вопроса он точно ответит правильно и 1 правильный ответ из 7 угадает с вероятностью $\frac{7}{128}$).

Найдем наиболее вероятное общее число правильных ответов:

$$np+p=7\cdot\frac{1}{2}+\frac{1}{2}=4\,,\ k_0'=3\,,\ k_0''=4$$

Общее наиболее вероятное число правильных ответов $3+k_0=3+3=6$ или $3+k_0''=3+4=7$. Так как учитель решил ставить за 8, 7, 6 баллов оценку «4», то данный ученик вероятнее всего получит оценку «4», хотя объективно за знание только 3-х вопросов ему полагается «2». Еще раз подтверждается вывод: такие тестирования не следует проводить.

Напрашивается теперь вопрос: как повысить объективность оценки при тестировании?

Одно уже ясно, что надо брать $s>2$, ну, как например, в настоящее время принято, $s=4$. Другими словами, чем больше вариантов ответов на каждое задание (вопрос), вероятность правильно ответить на вопрос, выбирая ответ наудачу, меньше, т.е. оценка будет более объективнее.

Пример. Контрольное задание (тест) содержит 12 вопросов, на каждый из которых 4 варианта ответа $(s=4)$. Оказалось, что тестируемый знает ответы на 4 вопроса, а на остальные 8 ответы выбирает наудачу. Какова вероятность того, что общее число правильных ответов ученика окажется равным 8, т.е. с какой вероятностью из 8 вопросов на 4 угадает правильный ответ? Чему равно наиболее вероятное общее число правильных ответов?

В этом примере $n=8$, $p=\frac{1}{4}$, $q=1-\frac{1}{4}=\frac{3}{4}$, $k=4$.

$$P_8(4)=\frac{8!}{4!4!}\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^4=\frac{8\cdot7\cdot6\cdot5\cdot3^4}{4\cdot3\cdot2\cdot4^8}=\frac{7\cdot5\cdot81}{2\cdot4^7}\approx 0{,}1$$

Значит вероятность получить 8 баллов из 12 возможных равна приблизительно 0,1.

Далее имеем:

$$np+p=8\cdot\frac{1}{4}+\frac{1}{4}=\frac{9}{4};\quad \left[\frac{9}{4}\right]=2$$

Из 8 вопросов, ответы которых ученик не знает, вероятнее всего угадает 2. Тогда наиболее вероятное общее число правильных ответов будет 4+2=6; $P_8(2)=0{,}31...$

Если учитель решает ставить:

оценку «5» за 12 баллов,
оценку «4» за 11 или 10 баллов,
оценку «3» за 9, 8, 7, 6 баллов,
оценку «2» за 5, 4, 3, 2, 1, 0 баллов,

то вероятнее всего данный ученик получит оценку «3», т.е. более объективную оценку, чем при $s=2$.

II. Интервальные оценки

Обозначим через X - дискретную случайную величину, равную числу успехов в n опытах, в каждом из которых успех наступает с вероятностью p и неудача с вероятностью $q=1-p$.

Если X_i $(i=1,2,...,n)$ - число успехов в $i^{\text{м}}$ опыте, то эта случайная величина имеет закон распределения

X_i	0	1
p_i	q	p

Среднее значение X_i (математическое ожидание) равно $\overline{X_i} = 0 \cdot q + 1 \cdot p = p$, т.е. вероятности успеха в одном опыте. Дисперсия $D(X_i) = p - p^2 = p(1-p) = pq$ и среднее квадратическое отклонение $\sigma_{X_i} = \sqrt{pq}$.

Теперь заметим, что $X = X_1 + X_2 + ... + X_n = \sum_{i=1}^{n} X_i$ и, учитывая, что математическое ожидание суммы равно сумме математических ожиданий, а также дисперсия суммы независимых случайных величин равна сумме дисперсий, имеем:

$$M(X) = M(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} M(X_i) = \sum_{i=1}^{n} p = np,$$

$$D(X) = D(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} D(X_i) = \sum_{i=1}^{n} pq = npq,$$

$$\sigma_x = \sqrt{D(X)} = \sqrt{npq}.$$

Из центральной предельной теоремы Ляпунова (Ляпунов Александр Михайлович, СССР, 1857-1918) следует, что при $n \to \infty$ закон распределения случайной величины $X = X_1 + ... + X_n$ стремится к нормальному закону распределения с параметрами $a = np$ и $\sigma = \sqrt{npq}$, а закон распределения нормированной случайной величины $Y = \frac{X - np}{\sqrt{npq}}$ стремится к стандартному нормальному закону с параметрами $M(Y) = 0$ и $\sigma_y = 1$: как известно, уже при $n = 12$ закон распределения Y почти нормальный.

Если зададим доверительную вероятность $\gamma = 0{,}95$, то соответствующий доверительный интервал для Y можно получить из условия:

$$P(|Y|) < \varepsilon_\gamma, \text{ где } \Phi(\varepsilon_\gamma) = \frac{\gamma}{2},$$

$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ - функция Лапласа (П.С. Лаплас, 1749-1827, Франция). ε_γ находим по таблице для $\Phi(x)$.

Пример. Найти доверительный интервал для X - числа успехов (правильных ответов) при тестировании для $n = 16$, $p = 1/4$, $q = 3/4$.

Имеем: $\overline{X} = np = 16 \cdot \frac{1}{4} = 4$; $D(X) = npq = 16 \cdot \frac{1}{4} \cdot \frac{3}{4} = 3$. $\sigma = \sqrt{3}$; $\gamma = 0{,}95$

Тогда $P\left(\overline{X} - \varepsilon_\gamma \frac{\sigma_x}{\sqrt{n}} < X < \overline{X} + \varepsilon_\gamma \frac{\sigma}{\sqrt{n}}\right) \approx \gamma$.

Находим ε_γ из условия $\Phi(\varepsilon_\gamma) = \frac{\gamma}{2} = \frac{0{,}95}{2} = 0{,}475$; $\varepsilon_\gamma = 1{,}96$ (найдено по таблице из [1]) $\overline{X} - 1{,}96 \cdot \frac{\sqrt{3}}{\sqrt{16}} = 4 - 1{,}96\frac{\sqrt{3}}{4} = 4 - 0{,}49\sqrt{3} \approx 4 - 0{,}8 = 3{,}2$,

$$\overline{X} + 1{,}96 \cdot \frac{\sqrt{3}}{\sqrt{16}} = 4 + 1{,}96\frac{\sqrt{3}}{4} \approx 4 + 0{,}8 = 4{,}8.$$

Значит $P(3{,}2 < X < 4{,}8) \approx 0{,}95$. Имея в виду, что X - целое число, получаем доверительный интервал [3; 5] с доверительной вероятностью $\gamma \geq 0{,}95$. Это означает, что практически достоверно (на уровне значимости 0,05), что число правильных ответов тестируемого находится в промежутке [3; 5], если выбор ответа наудачу делается для 16 вопросов теста. Отсюда следует, что за 3 или 4 или 5 правильных ответов из 16, не говоря о меньшем числе, следует ставить оценку «2» и это будет объективно на уровне значимости 0,05, т.е. с надежностью 0,95.

Литература

1. *Солодовников А.С.* Теория вероятностей / А.С. Солодовников. – М.: Просвещение, 1983. – 207 с.
2. *Андрухаев Х.М.* Сборник задач по теории вероятностей / Х.М. Андрухаев. – М.: Высшая школа, 2005. – 180 с.

Probability analysis of reliability of an estimate at testing with an answer select

H.M. Andruhaev

In paper one of variants of the analysis of reliability of the gained information by means of testing with a select of a right answer from the several variants, one of which true is considered, and remaining are not true. The plan of independent trials of J. Bernulli and interval estimates is thus applied to a case of a normal distribution law. Other approaches, for example, байесовский are possible also. Some examples illustrating deductions, made in paper are reduced.

