Математическая Теория Игр и её Приложения, т.1, в.4, с. 63-77

УДК 519.234+519.832+378
ББК 22.172+22.18+ Ч421,6(2)284я8/2

ТЕОРЕТИКО ИГРОВОЙ ПОДХОД К ОЦЕНКЕ ТОЧНОСТИ ТЕСТИРОВАНИЯ

Михаил М. Луценко
Петербургский государственный университет путей сообщения
Санкт-Петербург
e-mail: ML4116@mail.ru

В настоящей работе мы построим несколько игровых моделей тестирования, и укажем их надежность (вероятность правильной оценки тестируемого), оптимальные решающие правила, наихудшее априорное распределение. Задача оценивания формулируется как задача принятия решений и сводится к статистической игре между Статистиком и Природой с пороговой функцией выигрыша. Предлагаемый подход не предполагает нормальности распределения уровней знаний испытуемых и прост в реализации. При построении моделей используются результаты работ автора [1], [2].

Ключевые слова: тестирование, антагонистическая игра, статистическая игра, рандомизированная решающая функция, наихудшее априорное распределение.

1. Введение

Важнейшая задача теории тестирования – оценка уровня знаний испытуемого по результатам теста. Задача становится особенно актуальной, если по результатам теста принимают важные административные решения: выдача аттестата об образовании, зачисление в ВУЗ и др. Анализу тестов и оценкам их точности посвящена обширная литература (см. например [3]). При этом прямо или косвенно используются интервальные оценки параметра биномиального (гипергеометрического) распределений. Кроме того, все оценки точности (надежности) тестирования проводят в предположении о нормальности распределения уровней знаний испытуемых. Но с этим предположением трудно согласиться. При тестовом контроле уровня знаний учащихся существенная часть процесса обучения посвящена подготовке к прохождению теста. А, зная форму теста, темы и типы задач, учащиеся способны так представить свои знания, что их объективная оценка будет затруднена. Таким образом, мы имеем конфликтную ситуацию (игру), участники которой: испытуемый, стремящийся меньше времени затратить на подготовку к тесту, но при этом получить наивысший балл и лицо, принимающее решение (Статистик), стремящийся наиболее точно оценить уровень знаний испытуемого.

Заметим, что предлагаемая здесь модель оценки уровня знаний испытуемого по результатам тестирования не предполагает каких-либо ограничений на априорное распределение уровней знаний испытуемых.

2. Задачи классификации и оценивания

Сформулируем задачу оценки уровня знаний испытуемого по результатам тестирования более точно. Предположим, что по результатам теста группа учащихся разбита на N классов, и оценка уровня знаний учащегося осуществляется по номеру того класса, в который он попал. Выбор номера класса может осуществляться разными методами, например, по числу правильных ответов на вопросы теста или по числу набранных баллов за решенные задачи, если веса задач были различны. Обозначим через $\mathbf{X}_\theta$ номер класса, в который попал учащийся, имеющий уровень знаний θ (тип тестируемого). Таким образом, Статистик, наблюдая значение случайной величины $\mathbf{X}_\theta$, должен оценить тип тестируемого θ.

Сведем задачу классификации к задаче статистического оценивания. Для этого введем следующие обозначения. Конечное множество возможных уровней знаний учащегося обозначим через $\Theta = \{\theta_1, \theta_2, \ldots, \theta_m\}$ (множество параметров); множество возможных значений случайной величины $\mathbf{X}_\theta$ обозначим через $X = \{x_1, x_2, \ldots, x_N\}$

(множество наблюдений), а через $\{P_\theta(x)\}_{\theta\in\Theta}$ семейство распределений этих величин на множестве X. Множество допустимых оценок знаний учащегося обозначим через $D = \{d_1, d_2, \ldots, d_n\}$ (множество решений). Обозначим через $\delta(x)$ решение Статистика в том случае, когда значение случайной величины $\mathbf{X}_\theta$ равно x. Функцию $\delta : X \to D$ называют решающей функцией. Множество решающих функций мы обозначим через $\mathbf{D} = D^X$, то есть $\mathbf{D}$ есть прямое произведение $N = |X|$ экземпляров пространств D. Таким образом, каждую решающую функцию можно представить в виде вектора $\delta = (\delta_1, \delta_2, \ldots, \delta_N)$, у которого $\delta_k = \delta(x_k) \in D$.

В этих обозначениях Статистик, наблюдая случайную величину $\mathbf{X}_\theta$ с неизвестным значением параметра θ, должен принять решение $\delta(\mathbf{X}_\theta) \in D$, которое наиболее точное оценивает значение параметра θ или же он должен найти такую решающую функцию δ, значения которой наиболее близки к θ.

В статистике рассматривают две группы оценок: точечные и интервальные. Для построения первых необходимо знать потери Статистика от ошибки при оценке им неизвестного параметра θ, то есть функцию потерь Статистика $L(\theta, d)$. К сожалению, из условий задачи такую функцию построить трудно.

Для построения интервальной оценки мы с каждым решением $d \in D$ свяжем подмножество уровней знаний $\Theta(d) \subseteq \Theta$, приемлемых (допустимых) при этом решении. По заданному семейству подмножеств $\{\Theta(d)\}_{d\in D}$ построим функцию выигрыша Статистика:

$$h(d,\theta) = \mathbf{1}_{\Theta(d)}(\theta) = \begin{cases} 1, & \text{при} \quad \theta \in \Theta(d), \\ 0, & \text{при} \quad \theta \notin \Theta(d). \end{cases}$$

Таким образом, выигрыш Статистика равен единице лишь в том случае, когда он правильно оценил испытуемого, или тип тестируемого θ оказался во множестве типов $\Theta(d)$ допустимых при данном решении d.

Зафиксируем семейство приемлемых интервалов $\{\Theta(d)\}_{d\in D}$. Каждая решающая функция δ порождает набор $\{\Theta\left(\delta(x)\right)\}_{x\in X}$ доверительных интервалов.

Для каждого параметра θ и решающей функции $\delta(x)$ найдем вероятность того, что набор доверительных интервалов $\{\Theta\left(\delta(x)\right)\}_{x\in X}$

накроет неизвестное значение параметра. Для этого воспользуемся формулой полной вероятности:

$$\mathbf{P}\left(\theta \in \Theta\left(\delta(\mathbf{X}_\theta)\right)\right) = \sum_{x \in X} \mathbf{P}\left(\mathbf{X}_\theta = x\right) \mathbf{P}\left(\theta \in \Theta(\delta(x)) | \mathbf{X}_\theta = x\right).$$

А, используя введенные выше обозначения для функции h и семейства распределений, мы получим:

$$\mathbf{P}\left(\theta \in \Theta\left(\delta(\mathbf{X}_\theta)\right)\right) = \sum_{x \in X} P_\theta(x) h(\delta(x), \theta) = H(\delta, \theta).$$

Функцию $H(\delta, \theta)$ мы будем называть функцией успеха по аналогии с Вальдовской функцией риска.

Наименьшая вероятность того, что набор $\left\{\Theta\left(\delta(x)\right)\right\}_{x \in X}$, порожденный решающей функцией δ, накроет неизвестный параметр θ, называется доверительной вероятностью для этого набора (для решающей функции δ), то есть

$$\gamma = \gamma(\delta) = \min_{\theta \in \Theta} \mathbf{P}\left(\theta \in \Theta\left(\delta(\mathbf{X}_\theta)\right)\right).$$

Целью Статистика становится определение такой решающей функции δ (такого набора доверительных интервалов), для которой доверительная вероятность будет наибольшей.

С другой стороны предположим, что параметр θ сам является случайной величиной с известным распределением ν, то есть Статистик наблюдает случайную величину $\mathbf{X}_\nu$ с априорным распределением ν:

$$P(\mathbf{X}_\nu = x) = \int\limits_{\Theta} P_\theta(x) d\nu(\theta).$$

Средне взвешенная функция успеха при заданных значениях δ и ν равна:

$$\mathbf{H}(\nu, \delta) = \int\limits_{\Theta} H(\theta, \delta) d\nu(\theta) = \sum_{x} \int\limits_{\Theta} P_\theta(x) h(\delta(x), \theta) d\nu(\theta).$$

И она равна вероятности того, что неизвестный параметр θ попадет в доверительный интервал, порожденный решающей функцией δ, если параметр θ имеет известное распределение ν.

Функция δ_ν, максимизирующая $\mathbf{H}(\nu, \delta)$, называется байесовским решением (байесовской решающей функцией) относительно распределения ν, а величина этого максимума называется байесовским успехом для априорного распределения ν, то есть байесовский успех равен:

$$\mathbf{H}(\nu, \delta_\nu) = \max_{\delta} \mathbf{H}(\nu, \delta).$$

Байесовская решающая функция δ_ν порождает такой набор интервалов, для которого средняя вероятность покрытия была бы наибольшей при данном априорном распределении параметра θ.

В нашем случае множество $\mathbf{D}$ есть прямое произведение N экземпляров множеств D, а функция $\varphi(\delta) = \varphi(\delta_1, \delta_2, \ldots, \delta_N) = \mathbf{H}(\nu, \delta)$ – сепарабельная функция относительно переменных $\delta_1, \delta_2, \ldots, \delta_N$. Следовательно, баесовский успех может быть найден по формуле:

$$\mathbf{H}(\nu, \delta_\nu) = \sum_{k=1}^{N} \max_{\delta_k \in D} \left[\int\limits_{\Theta} P_\theta(x) h(\theta,\ \delta_k)\, d\nu(\theta) \right], \delta_k = \delta(x_k).$$

Или значение байесовской решающей функции δ_ν в точке x есть максимум слагаемого, зависящего лишь от этого значения. Следовательно, значения δ_ν можно находить в каждой точке независимо от значений в других точках.

Наихудшим для Статистика будет то априорное распределение ν параметра θ, для которого байесовский успех будет наименьшим. В этом случае минимальный байесовский успех равен:

$$\mathbf{H}(\nu^*, \delta_{\nu^*}) = \min_{\nu} \max_{\delta} \mathbf{H}(\nu, \delta)$$

Априорное распределение ν^*, на котором достигается этот минимум, называется наихудшим априорным распределением.

3. Статистическая игра “Тестирование”

Для выше определенной задачи определим статистическую игру “Тестирование” $\Gamma = \langle \mathbf{D},\ \Theta,\ H \rangle$ между Статистиком и Природой. В этой игре: $\mathbf{D} = D^X$ – пространство решающих функций Статистика, Θ – множество параметров (стратегии Природы) и функция выигрыша:

$$H(\delta, \theta) = \sum_{x} P_\theta(x) h(\delta(x), \theta), \quad \delta \in \mathbf{D} = D^X\,.$$

Статистик (игрок 1) стремится к увеличению своей доверительной вероятности, а Природа (тестируемый) стремится получить наибольший балл, то есть исказить результаты оценивания.

Нижнее значение игры Γ равно максимальной доверительной вероятности, которую может обеспечить себе Статистик независимо от действий Природы

$$\underline{v} = \max_{\delta \in \mathbf{D}} \min_{\theta \in \Theta} H(\delta, \theta) = \max_{\delta \in \mathbf{D}} \gamma(\delta).$$

Решающая функция δ^*, на которой этот максимум достигается, – оптимальный набор доверительных интервалов. Заметим, что верхнее значение игры Γ равно единице:

$$\bar{v} = \min_{\theta \in \Theta} \max_{\delta \in \mathbf{D}} H(\delta, \theta) = 1.$$

Так как верхнее значение игры $\Gamma = \langle \mathbf{D},\ \Theta,\ H \rangle$ больше нижнего значения, то ее решение мы будем искать в смешанных стратегиях. Обозначим через $\overline{\mathbf{D}}$, $\overline{\Theta}$ пространства вероятностных мер (распределений), заданных на соответствующих множествах и содержащие все вырожденные меры, а функцию выигрыша смешанного расширения игры Γ через:

$$\mathbf{H}(\mu, \nu) = \int\limits_{\mathbf{D} \times \Theta} H\left(\delta, \theta\right) d\mu(\delta) d\nu(\theta).$$

Если μ^*, ν^* – вырожденные вероятностные меры с носителями δ^*, θ^*, соответственно, то мы будем писать:

$$\mathbf{H}(\delta^*, \nu) = \mathbf{H}(\mu^*, \nu),\ \mathbf{H}(\mu, \theta^*) = \mathbf{H}(\mu, \nu^*),\ \mathbf{H}(\delta^*, \theta^*) = \mathbf{H}(\mu^*, \nu^*) = H(\delta^*, \theta^*).$$

Смешанные стратегии (вероятностные меры) $\mu \in \overline{\mathbf{D}}$, $\nu \in \overline{\Theta}$ задают вероятности с которыми игроки выбирают те или иные чистые стратегии из своих множеств стратегий. Функция выигрыша $\mathbf{H}(\mu, \nu)$ равна математическому ожиданию выигрыша первого игрока, если игроки использовали свои смешанные стратегии μ, ν.

Решением статистической игры $\Gamma = \langle \mathbf{D},\ \Theta,\ H \rangle$ в смешанных стратегиях будет решение игры $\overline{\Gamma} = \left\langle \overline{\mathbf{D}},\ \overline{\Theta},\ \mathbf{H} \right\rangle$, то есть тройка $\langle \mu^*, \nu^*, v \rangle$, для которой выполняются неравенства

$$\mathbf{H}(\mu, \nu^*) \le v \le \mathbf{H}(\mu^*, \nu) \text{ при всех } \mu \in \overline{\mathbf{D}},\ \nu \in \overline{\Theta}.$$

Легко доказать, что для проверки оптимальности тройки $\langle \mu^*, \nu^*, v\rangle$ достаточно ограничится вырожденными мерами μ, ν, то есть достаточно выполнения следующих неравенств

$$\mathbf{H}(\delta, \nu^*) \le v \le \mathbf{H}(\mu^*, \theta) \text{ при всех } \delta \in \mathbf{D},\ \theta \in \Theta.$$

Выполнение последних неравенств равносильно равенствам, в которых внешние экстремумы достигаются на μ^* и ν^*, соответственно.

$$v = \min_{\nu} \max_{\delta} \mathbf{H}(\delta, \nu) = \max_{\mu} \min_{\theta} \mathbf{H}(\mu, \theta)$$

Таким образом, оптимальная стратегия игрока 1 есть рандомизированная решающая функция μ^*, для которой доверительная вероятность покрытия неизвестного параметра была бы наибольшей.

Оптимальная стратегия Природы ν^* (наихудшее априорное распределение) – это такое распределение, для которого байесовская решающая функция δ^* была бы наименее эффективна.

Значение статистической игры – вероятность того, что наугад вызванный студент будет правильно оценен (точно указан его тип).

4. Решение конечных статистических игр

Хорошо известно (см. например [1]), что при решении игры Γ Статистик может ограничиться смешанными стратегиями, имеющими вид $\mu = (\mu_1, \mu_2, \ldots, \mu_N)$, где $N = |X|$ – размерность пространства наблюдений (число классов), а μ_k, $k = \overline{1, N}$ – вероятностные меры на множестве решений D.

Так как множества X, D, Θ – конечны, и их мощности соответственно равны: N, n, m, то Γ – матричная игра с матрицей выигрыша размера $Nn \times m$. Обозначим через $b_{i,j} = h(\theta_i, d_j)$, $i = \overline{1, m}$, $j = \overline{1, n}$ – элементы матрицы B; через $\lambda^k_{i,i} = P_{\theta_i}(x_k)$, $i = \overline{1, m}$ – ненулевые элементы квадратной матрицы Λ^k, $k = \overline{1, N}$; через $\mu_k = (\mu^1_k, \mu^2_k, \ldots, \mu^n_k)^t$, $k = \overline{1, N}$ – компоненты рандомизированной решающей функции $\mu = (\mu_1, \mu_2, \ldots, \mu_N)$, через $\nu = (\nu_1, \nu_2, \ldots, \nu_m)^t$ вектор априорного распределения параметра θ, через $\mathbf{1}_m$ – m-мерный столбец, все элементы которого единицы.

Для решения матричной игры Γ составим пару взаимно двойственных задач. Из первой (прямой) задачи мы найдем: наилучшую рандомизированную решающую функцию $\mu = (\mu_1, \mu_2, \ldots, \mu_N)$, из второй (двойственной) – наихудшее априорное распределение ν, а общее значение этих игр – значение игры Г.

Прямая задача:	Двойственная задача:
$v \to \max$, $\sum\limits_{k=1}^{N} \Lambda^k B \mu_k \ge v \mathbf{1}_m;\ \sum\limits_{j=1}^{n} \mu_k^j = 1;$ $\mu_k^j \ge 0;\ k = \overline{1,N};\ j = \overline{1,n};$	$v = \sum\limits_{k=1}^{N} u_k \to \min,$ $\nu^t \Lambda^k B \le u_k \mathbf{1}_n^t;\ k = \overline{1,N};\ \sum\limits_{i=1}^{m} v_i = 1.$

Известно много способов решения задач линейного программирования. Наиболее уместными здесь был бы динамический метод [1], специально разработанный автором для статистических игр с пороговыми функциями выигрыша. Однако в простейших случаях конечную статистическую игру можно решить средствами MS Excel. Хотя эти методы часто не дают точное решение, но они всегда указывают допустимые решения задач и, следовательно, верхнюю и нижнюю оценки матричной игры.

5. Примеры решения статистических игр "Тестирование"

Приведем примеры решения нескольких игр "Тестирование" и проинтерпретируем полученные результаты.

Пример 5.1. Группа учащихся четырех уровней подготовки по результатам тестирования разбита на 10 подгрупп $X = \{\Delta_0, \Delta_1, \ldots, \Delta_9\}$ (пространство наблюдений). Требуется разбить исходную группу на четыре класса так, чтобы в первый класс входили только отличники, во второй только хорошисты, в третий – успевающие и в четвертый – неуспевающие.

Обозначим через $\Theta = \{$отл, хор, уд, неуд $\}$ множество типов студентов (пространство параметров). Считаем известными вероятности, с которыми учащиеся разных типов попадут в ту или иную подгруппу по результатам тестирования. Обозначим через $P_\theta(x)$ вероятность того, что учащийся уровня θ попал в подгруппу x по результатам тестирования. Составим следующую таблицу распределений $P_\theta(x)$ (таб. 1), исходя из понимания того, что даже отличник может ошибиться, а плохо занимающийся может довольно много случайно отгадать.

Таблица 1. Распределение вероятностей $P_\theta(x)$.

$x=$	Δ_0	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6	Δ_7	Δ_8	Δ_9
$\theta=$ отл	0	0	0	0	0	0	0	0	0.1	0.9
$\theta=$ хор	0	0	0	0	0	0	0.05	0.8	0.1	0.05
$\theta=$ уд	0	0	0.05	0.1	0.7	0.1	0.05	0	0	0
$\theta=$ неуд	0.1	0.15	0.6	0.1	0.05	0	0	0	0	0

Табл. 1 составлена так, что все типы студентов довольно хорошо отделены друг от друга.

Если считать, что подгруппа Δ_i состоит из учащихся, верно решивших от $10i\%$ до $10(i+1)\%$ задач, то данные таблицы интерпретируются следующим образом. Отличники в 90% случаев решают свыше 90% задач, хорошисты в 80% случаев решают от 70% до 80% задач, успевающие в 70% случаев решают от 40% до 50% задач, неуспевающие в 95% случаев решают менее 40% задач.

Обозначим через $D=\{$отл, хор, уд, неуд $\}$ – множество решений Статистика, а функция

$$h(d,\theta)=\begin{cases} 1, & \text{при } \theta=d, \\ 0, & \text{при } \theta\neq d \end{cases}$$

задает множество параметров допустимых при данном решении. В данном случае Статистик выигрывает лишь в том случае, когда он правильно назвал тип тестируемого. Иными словами с каждым мы связываем интервал, состоящий из одного элемента.

Элементами множества решающих функций $\mathbf{D}=D^X$ будут вектора $\mathbf{d}=(d_0,d_1,\ldots,d_9)$, координаты которых будут решениями Статистика при наблюдении им ответа с соответствующим номером. Математическое ожидание выигрыша Статистика при использовании им решающего правила $\delta=(\delta_0,\delta_1,\ldots,\delta_9)$ будет иметь следующий вид:

$$H(\mathbf{d},\theta)=\sum_{i=0}^{9}P_\theta(\Delta_i)h(d_i,\theta),$$

где через θ обозначен оцениваемый тип тестируемого (уровень его подготовки).

Рассмотрим статистическую игру $\Gamma = \langle \mathbf{D},\ \Theta,\ H\rangle$ с описанными выше компонентами. У нас получилась матричная игра с матрицей размера 40×4, которая может быть решена средствами MS Excel.

Запишем теперь решение этой задачи тестирования. Значение этой игры равно 0.900. То есть Статистик будет правильно оценивать испытуемого лишь в 90% случаев. Рандомизированная решающая функция Статистика примет следующий вид. Ему рекомендуется называть "отличником" только испытуемого, правильно решившего свыше 90% задач ($\delta_9 =$ отл); хорошистом испытуемого, решившим от 70% до 90% задач ($\delta_8 = \delta_7 =$ хор); успевающим, решившего от 40% до 70% задач и с вероятностью 0.5 решившего от 30% до 40% задач ($\delta_6 = \delta_5 = \delta_4 =$ уд, $\delta_3 =$ уд с вероятностью 0.5 и неуд с вероятностью 0.5). В остальных случаях Статистику рекомендуется называть испытуемого неуспевающим ($\delta_2 = \delta_1 = \delta_0 =$ неуд).

В заключении мы приведем оптимальную стратегию Природы (табл. 2).

Таблица 2. Наихудшее априорное распределение параметра θ примера 5.1.

θ_i	отл	хор	уд	неуд
ν_i	0.175	0.275	0.275	0.275

Пример 5.2. Предположим, что тест состоит из 10 вопросов, и Статистик принимает решение по результатам этого теста. Пространство наблюдений X состоит из 11 чисел: от нуля до 10. Вероятность правильного ответа на один вопрос теста равна уровню знаний тестируемого, возможные значения которого составляют множество $\Theta = \{\,0.95;\ 0.85;\ 0.75;\ 0.65;\ 0.55;\ 0.45;\ 0.35;\ 0.25;\ 0.15;\ 0.05\,\}$. Тогда вероятность правильно ответить на x вопросов находится по формуле Бернулли: $P_\theta(x) = C_{10}^x \theta^x (1-\theta)^{10-x}$, $x = \overline{0, 10}$.

Статистик, оценивая уровень знания испытуемого, ставит одну из следующих четырех оценок: {отл, хор, уд, неуд} $= D$. Причем отличная оценка ставится испытуемым, имеющим 95% и 85% знаний, хорошая от 55% до 75%, удовлетворительная от 35% до 45% и неудовлетворительная остальным.

Составим статистическую игру $\Gamma = \langle \mathbf{D},\ \Theta,\ H \rangle$ и решим ее в смешанных стратегиях. Матрица выигрыша в рассматриваемой игре имеет размер 44×10. К сожалению, средства MS Excel не позволяют точно решить две взаимно-двойственные задачи. Но мы получаем верхнюю и нижнюю оценки значения игры, рандомизированную решающую функцию и наихудшее априорное распределение параметра θ.

В результате расчетов мы для значения игры получим: нижнюю (0.519) и верхнюю (0.562) оценки значения игры.

Таблица 3. Компоненты рандомизированной решающей функции μ примера 5.2.

		μ_{10}	μ_9	μ_8	μ_7	μ_6	μ_5	μ_4	μ_3	μ_2	μ_1	μ_0
Решения	отл	1.00	0.49	0.75	0	0	0	0	0	0	0	0
	хор	0	0.51	0.24	0.95	0.75	0.70	0	0	0	0	0
	уд	0	0	0	0.05	0.25	0.30	1.00	1.00	0	0	0
	неуд	0	0	0	0	0	0	0	0	1.00	1.00	1.00

В столбцах табл. 3 указаны вероятности, с которыми Статистик указывает то или иное решение в зависимости от наблюдения.

Итак, вероятность правильного решения Статистика об уровне знаний студента по результатам тестирования лежит в промежутке от 0.52 до 0.56. Таким образом, **примерно в 50% случаев Статистик примет неверное решение об уровне знаний испытуемых**.

Таблица 4. Наихудшее априорное распределение параметра θ примера 5.2.

θ_i	0.95	0.85	0.75	0.65	0.55	0.45	0.35	0.25	0.15	0.05
ν_i	0.00	0.11	0.01	0.04	0.25	0.19	0.15	0.26	0.00	0.00

Пример 5.3. Испытуемые, как и в предыдущем примере, отвечают на 10 вопросов теста. При этом каждый испытуемый имеет один из

10 возможных уровней знаний, значения которых изменяются от 5% до 95%. Но в этом примере мы расширим множество возможных решений, и оно будет совпадать с множеством уровней знаний (множеством параметров):

$$\Theta = D = \{\, 0.95;\ 0.85;\ 0.75;\ 0.65;\ 0.55;\ 0.45;\ 0.35;\ 0.25;\ 0.15;\ 0.05 \,\}\,.$$

В допустимый интервал $\Theta(d)$ войдут лишь те значения параметра θ, которые отклонятся от d на расстояние не большее чем 10%, то есть

$$h(d,\theta) = \mathbf{1}_{\Theta(d)}(\theta) = \begin{cases} 1, & \text{при} \quad |i-j| \le 0,1; \\ 0, & \text{при} \quad |i-j| > 0,1. \end{cases}$$

Составим статистическую игру $\Gamma = \langle \mathbf{D},\ \Theta,\ H \rangle$ и решим ее в смешанных стратегиях. Матрица выигрыша в рассматриваемой игре имеет размер 110×10. Здесь, как и в предыдущем примере, мы получаем верхнюю и нижнюю оценки значения игры, рандомизированную решающую функцию и наихудшее априорное распределение параметра θ.

В результате расчетов мы для значения игры получим: нижнюю (0.771) и верхнюю (0.788) оценки значения игры.

Таблица 5. Компоненты рандомизированной решающей функции μ примера 5.3.

		μ_{10}	μ_9	μ_8	μ_7	μ_6	μ_5	μ_4	μ_3	μ_2	μ_1	μ_0
Решения	0.95	0	0	0	0	0	0	0	0	0	0	0
	0.85	1.00	0.56	0	0	0	0	0	0	0	0	0
	0.75	0	0.41	0.89	0	0	0	0	0	0	0	0
	0.65	0	0.03	0.11	0.95	0.05	0	0	0	0	0	0
	0.55	0	0	0	0.05	0.92	0.52	0	0	0	0	0
	0.45	0	0	0	0	0.03	0.48	0.99	0.03	0	0	0
	0.35	0	0	0	0	0	0	0.01	0.96	0.10	0	0
	0.25	0	0	0	0	0	0	0	0.01	0.90	0.40	0
	0.15	0	0	0	0	0	0	0	0	0	0.60	1.00
	0.05	0	0	0	0	0	0	0	0	0	0	0

Таблица 6. Наихудшее априорное распределение параметра θ примера 5.3.

θ_i	0.95	0.85	0.75	0.65	0.55	0.45	0.35	0.25	0.15	0.05
ν_i	0.03	0.05	0.12	0.14	0.18	0.11	0.14	0.10	0.08	0.03

Заметим, что единственная наблюдаемая Статистиком случайная величина $\mathbf{X}_\nu$ – число правильных ответов на тест одним испытуемым. Распределение этой величины можно построить, по наихудшему априорному распределению ν.

$$\mathbf{P}(\mathbf{X}_\nu = x) = \sum_{i=1}^{m} P_{\theta_i}(x)\nu_i$$

Как видно из гистограммы распределения случайной величины $\mathbf{X}_\nu$ (рис.1) она мало отличается от графика плотности нормального распределения. А при наблюдении Статистиком за величиной $\mathbf{X}_\nu$ за нулевую гипотезу, вполне вероятно, будет принята гипотеза о нормальности ее распределения.

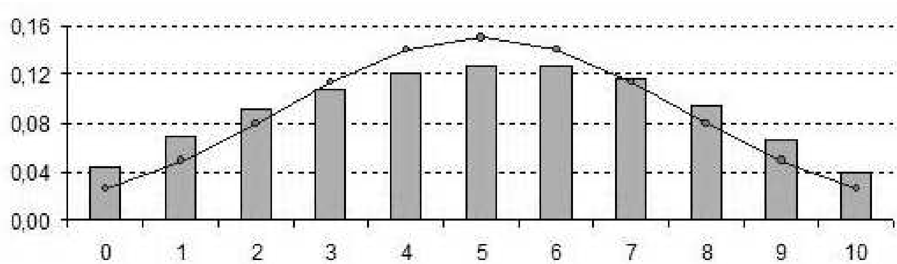

Рисунок 1. Гистограмма распределение числа правильных ответов при наихудшем априорном распределении и ее нормальная аппроксимация.

6. Заключение

Рассматриваемые здесь задачи обычно решаются статистическими методами: строятся доверительные интервалы и т.д. Предлагаемый же здесь метод решения одинаково хорошо работает для любых групп тестируемых (больших и маленьких). Однако, математическая

Таблица 7. Распределение числа правильных ответов при наихудшем априорном распределении.

Число правильных ответов:	0	1	2	3	4	5	6	7	8	9	10
Вероятность	0.04	0.07	0.09	0.11	0.12	0.13	0.13	0.12	0.09	0.07	0.04

модель (статистическая игра) жестко привязана к процедуре тестирования (принятия решений). И при ее изменении модель существенно изменяется в месте с наихудшим априорным распределением и оптимальной рандомизированной решающей функцией.

Хотя рассмотренные здесь математические модели довольно просты (малое число вопросов в тесте, искусственные семейства распределений), однако, их уточнение мало повлияет на невысокое значение игры.

Некоторые трудности, возникающие при реализации рандомизированных решающих функций, могут быть преодолены, предписывая различные решения для разных групп тестируемых. Например, в сложных случаях в городах оценки занижаются, а в небольших населенных пунктах – оценки завышаются.

К преимуществу данного подхода следует отнести то, что мы не накладываем ни каких ограничений на распределение типов тестируемых и то, что решение этих статистических игр получено стандартными методами.

Полученные значения трех игр являются нижними оценками доверительной вероятности, и могут быть улучшены при известном априорном распределении. Кроме того, байесcовское решение устойчиво при малых отклонениях априорного распределения от наихудшего.

СПИСОК ЛИТЕРАТУРЫ

1. Луценко М.М. *Теоретико игровой метод оценки параметра биномиального закона*// Теория вероятностей и ее применения. 1990. Т. 35. № 3. С. 471–481.

2. Луценко М.М., Иванов М.А. *Минимаксные доверительные интервалы для параметра гипергеометрического распределения* // Автоматика и телемеханика. 2000. № 7. С. 68–76.

3. Нейман Ю.М., Хлебников В.А. *Введение в теорию моделирования и параметризации педагогических тестов.* М: Издательство, 2000.

GAME-THEORETIC APPROACH TO ACCURACY OF TESTING

Mikhail Lutsenko, St. Petersburg Transport University,
Saint-Peterburg, Dr.Sc., prof. (ML4116@mail.ru).

Abstract: In the paper some game models of test are constructed and their reliability (probability of correct score) is shown. The matters of investigation are statistical games between Nature and Statistician. Solution of the games (the worst priori distribution, optimal decision function) are found. The results of author's work [1] are used.

Keywords: test reliability, antagonistical games, statistical game, randomize decision function, worst priori distribution.