

## Модели и процедуры интервальной оценки результатов контроля знаний в компьютерных системах тестирования ВМФ

Денис А. Печников<sup>1</sup> 19pda72@bk.ru

<sup>1</sup> Военно-морская академия им. Адмирала Флота Советского Союза Н.Г. Кузнецова, Ушаковская набережная, 17/1, г. Санкт-Петербург, 197045, Россия

**Реферат.** Для осуществления эффективной военно-профессиональной подготовки специалистов ВМФ необходима соответствующая учебно-материальная база. В результате сокращения в 1990 годы отраслей военно-промышленного комплекса, разрабатывающих технику и вооружения для ВМФ, новейшие образцы этой техники теперь изготавливаются не сериями, а отдельными экземплярами. Вопрос о производстве учебных и тренировочных образцов не стоит вообще. В этих условиях в качестве единственного средства, способного обеспечить военно-профессиональную подготовку, могут рассматриваться только виртуальные аналоги образцов военной техники и вооружения, разрабатываемые средствами информационных технологий, т. е. тренажерно-обучающие системы (ТОС). На современном уровне развития информационных технологий тестирование является единственным универсальным техническим средством текущего контроля знаний обучаемых. Процедуры контроля знаний в современных компьютерных системах тестирования не соответствуют предъявляемым к ним требованиям по следующим характеристикам: 1) отсутствию возможности оценки погрешности результатов тестирования; 2) отсутствию возможности прекращения тестирования при достижении заданной достоверности его результатов. Для эффективного внедрения средств оперативного критериально-ориентированного педагогического контроля знаний в процесс подготовки специалистов ВМФ и обеспечения возможности совместного анализа и обработки оценок результатов обучения, целесообразно реализовать следующие практические рекомендации: 1. Формулировку системы предпочтений преподавателя в отношении качества подготовки обучаемых и системы предпочтений преподавателя в отношении значимости единичных тестовых заданий в тесте следует рассматривать в качестве наиболее ответственных этапов подготовки теста к практическому использованию. 2. Преподавателю впервые производящему ввод своих систем предпочтений следует проверить их фактическое соответствие на выборке из 5–10 таких результатов выполнения теста, которые покрывают весь спектр возможных результатов (от полностью верных до полностью неверных). Наличие разработанных и представленных моделей обеспечивает возможность постановки и решения задачи разработки моделей и процедур управления процессом критериально-ориентированного тестирования, обеспечивающих оперативность этого процесса.

**Ключевые слова:** интервальная оценка, доверительный интервал, биномиальный закон распределения

## Models and procedures for interval evaluating the results of control of knowledge in computer systems testing of Navy

Denis A. Pechnikov<sup>1</sup> 19pda72@bk.ru

<sup>1</sup> Naval Academy named after Admiral of the Fleet of the Soviet Union N.G. Kuznetsov, Ushakov embankment, 17/1, St. Petersburg, 197045, Russia

**Summary.** To implement effective military and professional training of Navy specialists, a corresponding educational and material base is needed. As a result of the reduction in the 1990s in the branches of the military-industrial complex developing weapons and equipment for the Navy, the latest models of this technology are now produced not in batches, but in individual copies. The question of the production of training and training samples is not worth it at all. Under these conditions, only virtual analogues of military equipment and weapons, developed by means of information technology, i.e., training and training systems (TOS), can be considered as the only means capable of providing military-professional training. At the modern level of the development of information technologies, testing is the only universal technical means of monitoring the knowledge of students. Procedures for knowledge control in modern computer testing systems do not meet the requirements for them according to the following characteristics: 1) the absence of the possibility of evaluating the error of the test results; 2) the absence of the possibility of stopping testing when the specified reliability of its results is achieved. In order to effectively implement the means of operational criteria-based pedagogical control of knowledge in the process of training specialists of the Navy and to enable joint analysis and processing of evaluations of learning outcomes, it is advisable to implement the following practical recommendations: 1. Formulating the teacher's preferences system regarding the quality of trainee training and the teacher's preferences system in relation to The significance of single test tasks in the test should be considered as the most important The essential steps in preparing a test for practical use. 2. The teacher who first enters the input of his preference systems should check their actual compliance on a sample of 5-10 such test results that cover the full range of possible outcomes (from fully faithful to completely incorrect). The presence of the developed and presented models provides the possibility of setting and solving the task of developing models and procedures for managing the process of criterion-oriented testing, ensuring the efficiency of this process.

**Keywords:** interval estimation, confidence intervals, binomial distribution

Для цитирования

Печников Д.А. Модели и процедуры интервальной оценки результатов контроля знаний в компьютерных системах тестирования ВМФ // Вестник ВГУИТ. 2018. Т. 80. № 1. С. 97–104. doi:10.20914/2310-1202-2018-1-97-104

For citation

Pechnikov D.A. Models and procedures for interval evaluating the results of control of knowledge in computer systems testing of Navy. *Vestnik VGUIT* [Proceedings of VSUET]. 2018. vol. 80. no. 1. pp. 97–104. (in Russian). doi:10.20914/2310-1202-2018-1-97-104

### Введение

Для того чтобы осуществлять военно-профессиональную подготовку специалистов ВМФ необходима соответствующая учебно-материальная база. Основу такой базы раньше всегда составляли действующие учебные и тренировочные образцы военной техники и вооружения, которые ранее поступали в военно-морские учебные заведения в достаточном количестве и обеспечивали необходимый уровень как теоретической, так и практической подготовки будущих офицеров ВМФ. В результате резкого сокращения в 1990 годы отраслей военно-промышленного комплекса, разрабатывающих технику и вооружения для ВМФ, новейшие образцы этой техники и вооружений теперь изготавливаются не сериями, а отдельными экземплярами. Вопрос о производстве учебных и тренировочных образцов не стоит вообще. В этих условиях в качестве единственного средства, способного обеспечить военно-профессиональную подготовку, могут рассматриваться только виртуальные аналоги образцов военной техники и вооружения, разрабатываемые средствами информационных технологий, т. е. тренажерно-обучающие системы (ТОС) ВМФ.

В тенденциях развития ТОС ВМФ отмечается следующий ряд основных противоречий: 1) между ростом возможностей ТОС как средств моделирования управляемых объектов и неизменностью их дидактических характеристик; 2) между необходимостью индивидуализации процесса подготовки и ограниченными психофизиологическими возможностями обучающихся по управлению обучением в группах с количественным составом более 5–7 обучаемых; 3) между ростом числа и сложности задач управления обучением и неизменностью психофизиологических возможностей руководителей обучения. В качестве средства разрешения этих противоречий рассматривается автоматизация в ТОС ВМФ функций управления обучением.

Ключевой функцией, без автоматизации которой нельзя автоматизировать все остальные функции управления обучением, является функция педагогического контроля качества подготовки обучаемых.

Качество реализации в ТОС ВМФ функции контроля качества подготовки различно в отношении знаний и действий (умений и навыков) обучаемых.

В направлении автоматизации функции оценивания деятельности выполнен целый ряд исследований (Чельшкова М.Б. [31], Вадзинский Р.Н. [3], Ю.А. Ветров [4], Ю.Ф. Волынец [5], В.Н. Иванов [8], В.Д. Мосин [11], Д.А. Печников [13–15], А.Н. Платов [16], В.Ю. Пузырев [17], А.А. Скоров [19], А.М. Стручков [20], Е.В. Хекерт

[23, 24], О.Е. Чудаков [25], К.Ю. Шилов [26], Н.В. Щербак [27] Kaftandjieva, F [36]), Wim J. van derLinden [37] завершившихся созданием оригинальной автоматизированной системы оценки, которая прошла испытания и внедряется в программное обеспечение перспективных ТОС ВМФ.

В направлении автоматизации функции оценивания знаний выполнено менее значительное число исследований. Первая их группа (В.Н. Наумов [12], А.Н. Ханников [29], П.В. Филиппов [28]) носит поисковый характер и ориентирована на создание технологий искусственного интеллекта. На совершенствование моделей и процедур тестирования направлена вторая группа исследований (Р.Р. Туктаров [21], А.О. Туровская [22]), которые были направлены на совершенствование моделей и процедур оценивания отдельных видов тестовых заданий и не касались вопросов управления процессом тестирования.

На достижение целей контроля усвоения знаний в ТОС ВМФ ориентирована компьютерная система тестирования военного назначения (КСТ ВН) “Система автоматизированного контроля (САК)”, которая входит в состав комплекса программ инструментальных средств (КПИС) “Медиатор”. Эта КСТ является типичным представителем современных КСТ и полностью соответствует современному уровню их развития.

ГОСТ РВ 51540-2005 [6] и ГОСТ РВ 15.205-2004 [7] определяют КСТ по функциональному предназначению как продукцию двойного назначения и относятся КСТ к комплектующим изделиям межотраслевого применения (КИМП), представляющим собой “изделие военной техники (ВТ), предназначенное для выполнения определенных технических функций в составе изделий ВТ или их составных частей, создаваемое не для конкретного изделия ВТ по самостоятельным комплектам и не подвергаемое изменениям в процессе создания изделий ВТ, в котором его применяют”. В соответствии с классификациями, приведенными в ГОСТ РВ 51540-2005 [6] и ГОСТ Р 50-605-80-93 [7], КСТ должны быть необходимо отнесены к изделиям ВТ, которые входят в категорию учебно-тренировочных средств, принадлежащих к учебной военной техники как виду военной техники.

### Основная часть

В настоящее время в компьютерных системах тестирования наибольшее распространение получили тестовые задания закрытого типа с единственным верным ответом. Задания этого типа применяются не менее чем в 92% случаев. В рассматриваемых тестовых заданиях успешность выполнения задания оценивается

путем установления факта совпадения ответа обучаемого с заранее заданным правильным вариантом ответа. Любой другой вариант ответа является неправильным. Для последующей статистической обработки результаты (верно – не верно) выполнения таких тестовых заданий представляются представляет в виде дихотомическая переменная  $(0 - 1)$ .

Безошибочность (правильность) ответов обучаемого определяется в виде показателя частности  $B$ , который имеет вид:

$$B = \frac{\sum_{i=1}^n j_i}{n} \quad (1)$$

где  $i$  ( $i = \overline{1, n}$ ) – номер тестового задания,  $n$  – общее число тестовых заданий,  $j_i$  ( $j_i = 0, 1$ ) результат выполнения отдельного тестового задания в тесте  $j_i = 1$  – верный ответ,  $j_i = 0$  – неверный ответ.

Все существующие модели и процедуры обработки результатов выполнения систем тестовых заданий основаны на точечной оценке вида (1). Эти точечные оценки результатов тестирования не обеспечивают оперативность процедуры и не вполне обеспечивают достоверность результатов тестирования. Ниже предлагаются модели и процедуры интервальной оценки результатов выполнения систем тестовых заданий тестовые задания закрытого типа с единственным верным ответом.

Предлагаемые модели и процедуры базируются на том факте, что показатель “правильность” является дихотомической переменной, а процедура его определения в процессе решения обучаемым тестирующей выборки из  $n$  заданий соответствует схеме Бернулли.

Схемой Бернулли называется последовательность независимых в совокупности испытаний, в каждом из которых возможны лишь два исхода – “успех” и “неудача”, при этом успех в одном испытании происходит с вероятностью  $p \in (0, 1)$ , а неудача – с вероятностью  $q = 1 - p$ . “Схема Бернулли лежит в основе таких целочисленных распределений, как биномиальное, геометрическое и отрицательное биномиальное [4].

Для корректного применения схемы Бернулли должны быть выполнены следующие условия [1]: 1) каждое испытание имеет ровно два исхода, условно называемых успехом и неудачей; 2) независимость испытаний: результат очередного эксперимента не должен зависеть от результатов предыдущих экспериментов; 3) вероятность успеха должна быть постоянной (фиксированной) для всех испытаний.

Выполнение первого условия определяется характеристиками самих систем тестовых заданий закрытого типа с единственным верным ответом, в которых результаты выполнения задания обозначаются дихотомической переменной  $j_i$  ( $j_i = 0, 1$ ). Выполнение второго условия обуславливается независимостью тестовых заданий в тесте и возможностью их предъявления в случайном порядке. Выполнение третьего условия обеспечивается тем, фактом, что в процессе тестирования обучаемым никаких новых сведений об объектах тестируемой предметной области не сообщается.

Примем допущение о том, что вероятность правильного выполнения тестовых заданий обучаемым определяется только знаниями обучаемого о тестируемой предметной области и не зависит ни от каких-либо других факторов (например, от фактора угадывания). Если принятое допущение выполняется, то выполняется и третье условие соответствия процедуры тестирования схеме Бернулли. В соответствии с принятым допущением безошибочность выполнения тестовых заданий обучаемым характеризуется некоторой постоянной вероятностью  $p = const$  правильного выполнения задания и соответствующей ей вероятностью  $q = 1 - p$  ее неправильного выполнения.

Пусть в результате решения обучаемым выборки из  $n$  тестовых заданий получен вектор (примеры векторов приведены в таблице 1)

$\vec{J} = (j_1, \dots, j_i, \dots, j_n)$ , (2), в котором  $j_i = 0, 1$  есть оценки успешности выполнения соответствующего задания ( $j_i = 1$  – задание выполнено без ошибок,  $j_i = 0$  – задание выполнено с ошибками). Требуется оценить вероятность  $p$  безошибочной деятельности.

Таблица 1.  
Некоторые варианты векторов результатов тестирования для  $n = 12$

Table 1.  
Some variants of vectors of test results for  $n = 12$

Вариант Option	$j_1$	$j_2$	$j_3$	$j_4$	$j_5$	$j_6$	$j_7$	$j_8$	$j_9$	$j_{10}$	$j_{11}$	$j_{12}$
1	1	1	1	1	1	1	1	1	1	1	1	1
2	0	0	0	0	0	0	1	1	1	1	1	1
3	1	1	1	1	1	1	0	0	0	0	0	0
4	1	0	1	0	1	0	1	0	1	0	1	0
5	0	1	0	1	0	1	0	1	0	1	0	1
6	1	1	0	0	0	0	0	0	0	1	1	1
7	1	1	0	0	1	1	1	1	1	0	1	1
8	1	1	1	1	0	1	1	1	0	1	1	1
9	1	1	1	1	0	0	0	1	1	1	1	0
10	1	1	0	1	1	0	0	1	1	0	1	0

Рассмотрим те законы распределений, которые могут быть использованы для интервальной оценки вероятности  $p$  безошибочного выполнения тестовых заданий.

Как отмечалось выше, эмпирические данные, полученные по схеме Бернулли, подчиняются законам таких целочисленных распределений, как биномиальное, геометрическое и отрицательное биномиальное. “Механизм возникновения этих распределений определяется тем способом, которым обрывается последовательность испытаний Бернулли” [4].

Биномиальное распределение имеет место в тех случаях, когда последовательность испытаний Бернулли обрывается после проведения фиксированного числа  $n$  испытаний. При этом под биномиальной случайной величиной  $X$  понимается число успехов в серии из  $n$  испытаний Бернулли.

Вероятность  $P(J)$  появления вектора  $J$  при биномиальном распределении оценивается как:

$$P(J) = C_n^k p^{n-k} q^k \text{ при } C_n^k = \frac{n!}{k!(n-k)!}, \quad (3)$$

где  $J$  – рассматриваемый вариант (вектор) результатов тестирования;  $p$  – вероятность безошибочного выполнения тестового задания,  $q = 1 - p$  – вероятность выполнения задания с ошибкой,  $k$  – число заданий, выполненных с ошибкой,  $n$  – общее число заданий,  $n - k$  – число заданий, выполненных без ошибок.

Отрицательное биномиальное распределение имеет место в тех случаях, когда последовательность испытаний обрывается сразу же после  $m$ -ого успеха. При этом рассматриваются две случайные величины: случайная величина  $Z$  – число неудач, предшествовавших  $m$ -ому успеху, и случайная величина  $W$  – общее число испытаний до  $m$ -ого успеха (включая  $m$ -ый успех).

Вероятность  $P(J)$  появления вектора  $J$  при отрицательном биномиальном распределении оценивается как:

$$P(J) = C_{z+m-1}^z p^m q^z \text{ при } C_{z+m-1}^z = \frac{(z+m-1)!}{z!(m-1)!}, \quad (4)$$

где  $J$  – рассматриваемый вариант (вектор) результатов тестирования;  $p$  – вероятность безошибочного выполнения тестового задания,  $q = 1 - p$  – вероятность выполнения задания с ошибкой,  $z$  – число заданий, которые выполнены с ошибкой и предшествуют  $m$ -ому успеху.

Геометрическое распределение является частным случаем отрицательного биномиального распределения и возникают при обрыве серии испытаний сразу же после первого успеха или неудачи. При этом рассматриваются две

случайные величины: случайная величина  $X$  – число неудач (успехов), предшествовавших первому успеху (неудаче), и случайная величина  $Y$  – число испытаний до первого успеха (неудачи). Сам успех (неудача) может включаться и не включаться в число  $Y$ .

Геометрическое распределение – это единственное дискретное распределение со свойством отсутствия памяти (последствия): знание об отсутствии успеха (неудачи) в предыдущих опытах никак не влияет на распределение оставшегося числа опытов до появления успеха (неудачи).

При геометрическом распределении вероятность  $P(J; j_i = 1)$  появления вектора  $J(j_i = 1; i = \overline{1, n})$  оценивается как:

$$P(J; j_i = 1) = p^h q, \quad (5)$$

а вектора  $J(j_i = 0; i = \overline{1, n})$  как

$$P(J; j_i = 0) = q^h p, \quad (6)$$

где  $p$  – вероятность безошибочного выполнения задания,  $q = 1 - p$  – вероятность выполнения задания с ошибкой,  $h (h = \overline{1, n})$  – номер  $i$  последнего задания, выполненного без ошибки или, наоборот, с ошибкой (число “неудач (ошибок)” до первого “успеха” или число “успехов” до первой “неудачи (ошибки)”.

В целях интервальной оценки вероятности  $p$  безошибочного выполнения тестовых заданий целесообразно использовать те законы распределения, которые принципиально обеспечивают обработку любого числа ( $i, i = \overline{1, +\infty}$ ) тестовых заданий. Такому требованию соответствуют только законы биномиального и геометрического распределения.

Закон отрицательного биномиального распределения может быть использован для оценивания результатов тестирования только после появления  $m$ -ого успеха. При его использовании минимальная длина оцениваемого вектора  $J$  лежит в пределах  $z + m \leq J \leq +\infty$ . Другими словами, при самом благоприятном стечении обстоятельств отрицательное биномиальное распределение обеспечит возможность оценки результатов тестирования только при числе  $i = z + m$  выполненных заданий, а при самом неблагоприятном – не обеспечит никогда.

Введем понятие вектора  $\vec{n}$  непрерывного результата (успеха или неудачи), под которым будем понимать любую часть вектора (1), каждый из элементов которой равен 1 или 0: если все элементы вектора равны 1 ( $j_i = 1$ ) –

вектор  $\vec{n}$  непрерывного успеха, если все элементы вектора равны 0 ( $j_i = 0$ ) – вектор  $\vec{n}$  непрерывной неудачи. Например, в таблице 1 для 6 варианта вектора  $\vec{j}$  это вектора  $\vec{n}_1 = (j_1, j_2)$ ,  $\vec{n}_2 = (j_3, j_4, j_5, j_6, j_7, j_8, j_9)$ ,  $\vec{n}_3 = (j_{10}, j_{11}, j_{12})$ .

Вероятность появления вектора (1) можно оценивать с двух позиций:

1) как вероятность события  $L$ , заключающегося в том, что при выполнении  $n$  заданий обучаемым было достигнуто  $(n-k)$  успехов (допущено  $k$  ошибок);

2) как вероятность события  $H$ , заключающегося в появлении в векторе (1)  $m$  векторов непрерывного результата.

Вероятность  $P(L)$  наступления события  $L$  оценивается по биномиальному закону (2). Соответственно при определении доверительного интервала (ДИ) для оценок вероятности  $p$  успешного выполнения тестового задания корректно использование следующих формул оценки доверительных интервалов биномиального распределения:

1) оценка точного ДИ [2, 3]:

$$\left( \frac{(n-k)}{(n-k) + (k+1)F_{2(k+1), 2(n-k), 1-\varepsilon/2}}, \frac{(n-k+1)F_{2(n-k+1), 2k, 1-\varepsilon/2}}{k + (n-k+1)F_{2(n-k+1), 2k, 1-\varepsilon/2}} \right), \quad (7)$$

где  $n$  – число испытаний,  $k$  – число ошибок, а  $F_{f, g, \alpha}$  – квантиль порядка  $\alpha$  распределения  $F$  с  $f, g$  степенями свободы;

2) оценка приближенного ДИ в соответствии с центральной предельной теоремой, когда при больших  $n$  распределение биномиальной случайной величины будет близко к нормальному [2, 3]

$$\left( p^* - u_{1-\varepsilon/2} \sqrt{\frac{p^*(1-p^*)}{n}}, p^* + u_{1-\varepsilon/2} \sqrt{\frac{p^*(1-p^*)}{n}} \right) = \left( \frac{n-k}{n} - u_{1-\varepsilon/2} \sqrt{\frac{(n-k)k}{n^2}}, \frac{n-k}{n} + u_{1-\varepsilon/2} \sqrt{\frac{(n-k)k}{n^2}} \right), \quad (8)$$

где  $p^*$  – оценка эмпирической частоты правильного выполнения тестовых заданий;  $u_{1-\varepsilon/2}$  – квантили стандартного нормального распределения порядка  $1 - \varepsilon / 2$ .

Оценим вероятность события  $H$ , которое заключается в появлении в векторе (2)  $m$  векторов непрерывного результата.

В соответствии с введенным выше понятием вектора  $\vec{n}_s$  непрерывного результата события  $\vec{n}_s \in \pi$  появления таких векторов относятся к категории попарно несовместных событий. Поэтому вероятность  $P(H)$  любого из вариантов их реализаций (см. табл. 1) оценивается по формуле:

$$P(H) = P(h_1 \cup \dots \cup h_s \cup \dots \cup h_m) = P(h_1) + \dots + P(h_s) + \dots + P(h_m), \quad (9)$$

где  $s$  ( $s = \overline{1, m}; m \leq n$ ) – номер вектора непрерывного результата;  $n$  – число выполненных тестовых заданий.

В (8) вероятности  $P(h_s)$  оцениваются по геометрическому закону в соответствии с (5, 6)

$$P(h_s) = \begin{cases} p^h q & \text{при } j_i = 1; \\ q^h p & \text{при } j_i = 0, \end{cases} \quad (10)$$

где  $p$  – вероятность безошибочного выполнения задания,  $q = 1 - p$  – вероятность выполнения задания с ошибкой,  $h$  ( $h = \overline{1, (n-k)}$ ) – номер  $i$  последнего задания, выполненного без ошибки или, наоборот, с ошибкой (число “неудач (ошибок)” до первого “успеха” или число “успехов” до первой “неудачи (ошибки)”).

Для оценки в (9, 10) неизвестного параметра  $p$  предлагается использовать метод максимального правдоподобия, при котором в качестве оценки вероятности  $p$  принимается ее значение

$$p = \arg \max_{p \in (0,1)} f(\vec{j} | p) = \arg \max_{p \in (0,1)} P(H), \quad (11)$$

максимизирующее вероятность  $P(H)$  появления при выполнении  $n$  тестовых заданий вектора  $\vec{j} = (j_1, \dots, j_i, \dots, j_n)$ , включающего  $m$  ( $m = \overline{1, n}$ ) векторов  $\vec{n}_s$  непрерывного результата.

Каждый из полученных векторов  $\vec{n}_s$  непрерывного результата позволяет по (10) получить оценку  $p_s$  с ошибкой  $\sigma_s$ :  $p_s \pm \sigma_s$ . Соответственно функция  $W(p_s, p)$  правдоподобия (совместной плотностью вероятности) для любой из реализаций события  $H = \cup h_s$  может быть представлена в виде

$$W(p_s, p) = \prod_{s=1}^{s=m} P(h_s) = \prod_{s=1}^{s=l} P(h_s^+) \prod_{s=l}^{s=m} P(h_s^-) = p^l q^k \quad (12)$$

при  $t > 0, k > 0,$

где  $p$  – вероятность безошибочного выполнения тестового задания,  $q = 1 - p$  – вероятность выполнения задания с ошибкой,  $k$  – число заданий, выполненных с ошибкой,  $t$  – число заданий, выполненных правильно,  $n = t + k$  – общее число выполненных заданий,  $s (s = \overline{1, m})$  – номера векторов  $\vec{n}_s$  непрерывного результата,  $s (s = \overline{1, l})$  – номера векторов  $\vec{n}_s$  непрерывного успеха,  $s (s = \overline{l+1, m})$  – номера вектор  $\vec{n}_s$  непрерывной неудачи.

Например, для представленного в таблице 1 варианта 6 вектора  $\vec{j}$ , включающего вектора  $\vec{n}_1 = (j_1, j_2)$ ,  $\vec{n}_2 = (j_3, j_4, j_5, j_6, j_7, j_8, j_9)$  и  $\vec{n}_3 = (j_{10}, j_{11}, j_{12})$ . функция (11) имеет вид

$$W(p_s, p) = \prod_{s=1}^{s=m} P(h_s) = P(h_1)P(h_2)P(h_3) = p^2 q^7 p^3 = p^5 q^7$$

Если рассматриваемый вектор является единичным вектором непрерывного успеха  $\vec{n}_s$  (в таблице 1 вариант 1) или непрерывной

неудачи  $\vec{n}_s$  то такие вектора оцениваются по формуле (10).

Поиск  $p$  целесообразно реализовать путем решения следующей задачи нелинейного программирования

$$\left. \begin{aligned} p^{n-k+1} q^{k+1} &\rightarrow \max, \\ p &\in (0, 1), k = \overline{1, n} \end{aligned} \right\} \quad (13)$$

Для решения (12) можно использовать стандартную функцию “Solve (Поиск решения)” Microsoft Excel.

ДИ оценки  $p$  может быть определен как

$$\left( p - u_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}, p + u_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right) \quad (14)$$

Сравним оценки обучаемых, полученные обычным путем на основе предлагаемых моделей.

Пусть обучаемыми выполнено 13 заданий и получены результаты, приведенные в таблице 1. Оценки этих результатов тестирования приведены в таблице 2.

Таблица 2.

Оценки результатов тестирования, приведенных в таблице 1

Table 2.

Evaluation of test results given in table 1

№ варианта option	n	n-k	k	l	m-l	Точечная оценка по модели (2.1) Point estimate by the model (2.1)	Интервальные оценки Interval estimates					
							на основе биномиального распределения based on binomial distribution			на основе геометрического распределения on the basis of geometric distribution		
							Вероятнейшее значение по (2.3) The most probable value according to (2.3)	Границы ДИ по (2.7) The boundaries of CI by (2.7)		Вероятнейшее значение по (2.13) The most probable value according to (2.13)	Границы ДИ по (2.14) The boundaries of the CI by (2.14)	
								Нижняя Lower	Верхняя Upper		Нижняя Lower	Верхняя Upper
1	12	12	0	1	0	1,000	1,000	0,735	1,000	0,923	0,754	1,000
2	12	6	6	1	1	0,500	0,500	0,211	0,789	0,500	0,182	0,818
3	12	6	6	1	1	0,500	0,500	0,211	0,789	0,500	0,182	0,818
4	12	6	6	6	6	0,500	0,500	0,211	0,789	0,500	0,182	0,818
5	12	6	6	6	6	0,500	0,500	0,211	0,789	0,500	0,182	0,818
6	12	5	7	2	1	0,417	0,417	0,152	0,723	0,400	0,089	0,711
7	12	9	3	3	2	0,750	0,750	0,428	0,945	0,647	0,343	0,951
8	12	10	2	4	2	0,833	0,833	0,516	0,979	0,667	0,367	0,966
9	12	8	4	2	2	0,667	0,667	0,349	0,901	0,625	0,317	0,933
10	12	7	5	4	4	0,583	0,583	0,277	0,848	0,550	0,234	0,866

Оценки, приведенные в таблице 2, свидетельствуют, что применение предлагаемых моделей оценки результатов тестирования обеспечивает переход от точечных оценок эмпирической частоты к адекватным теоретически обоснованным оценкам вероятности правильного действия, которые имеют количественно оцениваемый ДИ.

Поскольку задачи текущего контроля решаются непосредственно в процессе обучения и имеют целью выработку адресных обучающих воздействий, то процесс критериально-ориентированного тестирования должен занимать минимум времени. Иначе говоря, к самому процессу критериально-ориентированного тестирования, а также процедурам обработки и презентации его результатов должно быть предъявлено требование оперативности.

Оперативность процедур обработки и отображения результатов действий обучаемых проблем не вызывает, поскольку полностью обеспечивается за счет быстродействия аппаратно-программных средств КСТ. А вот оперативность самой процедуры тестирования соответствующего обеспечения не имеет. Поэтому условия и методы ее реализации требуют более подробного рассмотрения.

Наличие моделей (3, 7, 13, 14) интервальной оценки результатов тестирования обеспечивает возможность постановки и решения задачи разработки моделей и процедур управления процессом критериально-ориентированного тестирования, обеспечивающих оперативность этого процесса.

#### ЛИТЕРАТУРА

- 1 Аванесов В.С. Проблема формы тестовых заданий. URL: <http://testolog.narod.ru/Theory32.html>
- 2 Ван дер Варден Б.Л. Математическая статистика. М.: Изд-во Иностранной литературы, 1960. 436 с. URL: <http://www.twirpx.com/file/275786/>
- 3 Вадзинский Р.Н. Справочник по вероятностным распределениям. СПб.: Наука, 2001. 295 с. URL: <http://налекцию.рф/content/786>
- 4 Буравлев А.И., Переверзев В.Ю. Выбор оптимальной длины педагогического теста и оценка надежности его результатов // Открытое образование. 1999. № 2. С. 23–26. URL: [http://www.e-joe.ru/sod/99/2\\_99/st160.html](http://www.e-joe.ru/sod/99/2_99/st160.html)
- 5 Ветров Ю.А. Повышение качества тактической подготовки офицеров запаса ВМФ на основе рационального использования компьютерных технологий обучения. СПб: БГТУ «Военмех», 1999. 193 с.
- 6 Вольнец Ю.Ф. Концепция, принципы построения и функционирования инфологической автоматизированной среды подготовки специалистов ВМФ. Петродворец: ВМИРЭ, 1999. 199 с.
- 7 ГОСТ РВ 51540–2005. Военная техника. Термины и определения. М.: Стандартинформ, 2011. -12 с. URL: <http://www.twirpx.com/file/1699773/>
- 8 ГОСТ РВ 15.205–2004. Система разработки и постановки продукции на производство. Военная техника. Порядок выполнения опытно-конструкторских работ по созданию комплектующих изделий межотраслевого

#### Выводы

На современном уровне развития информационных технологий тестирование является единственным универсальным техническим средством текущего контроля знаний обучаемых. Процедуры контроля знаний в современных компьютерных системах тестирования не соответствуют предъявляемым к ним требованиям по следующим характеристикам: 1) отсутствию возможности оценки погрешности результатов тестирования; 2) отсутствию возможности прекращения тестирования при достижении заданной достоверности его результатов.

Для эффективного внедрения средств оперативного критериально-ориентированного педагогического контроля знаний в процесс подготовки специалистов ВМФ и обеспечения возможности совместного анализа и обработки оценок результатов обучения, полученных компьютерных систем тестирования и преподавателем, целесообразно реализовать следующие практические рекомендации:

1. Формулировку системы предпочтений преподавателя в отношении качества подготовки обучаемых и системы предпочтений преподавателя в отношении значимости единичных тестовых заданий в тесте следует рассматривать в качестве наиболее ответственных этапов подготовки теста к практическому использованию.

2. Преподавателю впервые производящему ввод своих систем предпочтений следует проверить их фактическое соответствие на выборке из 5–10 таких результатов выполнения теста, которые покрывают весь спектр возможных результатов (от полностью верных до полностью неверных).

применения. М.: Стандартинформ, 2005. 65 с. URL: <http://www.twirpx.com/file/520927/>

9 Иванов В.Н. Метод автоматизированной оценки качества профессиональной деятельности в корабельных автоматизированных системах. Петродворец: ВМИРЭ, 1999. 187 с.

10 Кремер Н.Ш. Теория вероятностей и математическая статистика. М.: Юнити-Дана, 2004. 573 с. URL: [http://www.e-joe.ru/sod/99/2\\_99/st160.html](http://www.e-joe.ru/sod/99/2_99/st160.html)

11 Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов. М.: Логос, 2010. 668 с. URL: <http://www.twirpx.com/file/1346290/>

12 Мосин В.Д. Моделирование процесса адаптивного управления деятельностью курсантов в ходе практического обучения на автоматизированных тренажерах. Петродворец: ВМИРЭ, 1999. 164 с.

13 Наумов В.Н. Теоретические основы представления содержания профессиональной подготовки специалистов ВМФ в интеллектуальных обучающих системах. Петродворец: ВМУРЭ им. А.С. Попова, 1994. 296 с.

14 Печников А.Н. Комплекс методов и моделей оценки качества и эффективности деятельности операторов вооружения и военной техники в тренажерно-обучающих системах ВМФ. СПб.: СПбГЭТУ, 2011. 343 с.

15 Печников А.Н., Ветров Ю.А. Проектирование и применение компьютерных технологий обучения. Ч. 1. Концепция систем автоматизированного обучения и моделирование процессов деятельности. СПб: БГТУ, 2002. 195 с.

16 Печников А.Н., Стручков А.М., Хекерт Е.В. и др. Генератор оценки: Свидетельство о государственной регистрации программ для ЭВМ № 2009613362 от 26.06.2009.

17 Платов А.Н. Автоматизация управления контролем подготовки на тренажерах ВМФ. СПб: ЦНИИ МО РФ, 2000. 125 с.

18 Пузырев В.Ю. Диагностика деятельности операторов корабельных электромеханических систем в специализированных тренажерах ВМФ. Новороссийск: НГМА, 2005. 129 с.

19 ГОСТ Р 5060580-93. Система разработки и постановки продукции на производство. Термины и определения. М.: ВНИИстандарт, URL: <http://gostrf.com/normadata/1/4293827/4293827526.htm>

20 Скоров А.А. Методологические основы использования компьютерных технологий обучения в процессе профессиональной подготовки специалистов ВМФ. Петродворец: ВМИРЭ, 2001. 318 с.

21 Стручков А.М. Проектирование компьютерных систем обучения и интеллектуального тренажа. Петродворец: ЗАО "Фирма "Пассат", 2002. 166 с.

#### REFERENCES

1 Avanesov V.S. Problema formy [The problem of the form of test tasks] Available at: <http://testolog.narod.ru/ThErou32.html> (in Russian)

2 Van der Waerden B.L. Matematicheskaya statistika [Math statistics] Moscow, Publishing House of Foreign Literature, 1960. 436 p. Available at: <http://www.twirpx.com/file/275786> (in Russian)

3 Vadzinsky R.N. Spravochnik [Handbook of probability distributions] Saint-Petersburg, Science, 2001. 295 p. Available at: <http://налекцию.рф/content/786> (in Russian)

4 Buravlev A.I., Pereverzev V.Yu. Choice of the optimal length of the pedagogical test and evaluation of the reliability of its results. *Otkrytoe obrazovanie* [Open Education] 1999. no. 2. pp. 23 - 26. Available at: [http://www.e-joe.ru/sod/99/2\\_99/st160.html](http://www.e-joe.ru/sod/99/2_99/st160.html) (in Russian)

5 Vetrov Yu.A. Povyshenie kachestva takticheskoi podgotovki [Improving the quality of tactical training of Navy reserve officers based on the rational use of computer learning technologies] Saint-Petersburg, BSTU "Voenmeh", 1999. 193 p. (in Russian)

6 Volynets Yu.F. Kontseptsiya, printsipy postroeniya i funktsionirovaniya [The concept, principles of construction and functioning of the informational automated environment for training specialists of the Navy] Petrodvorets, VMIRE, 1999. 199 p. (in Russian)

7 GOST RV 51540-2005. Voennaya tekhnika [State standard 51540-2005 Military equipment. Terms and Definitions] Moscow, StandardInform, 2011. 12 p. Available at: <http://www.twirpx.com/file/1699773> (in Russian)

8 GOST RV 15.205-2004. Sistema azrabotki i postanovki [State standard 15.205-2004 System for the development and production of products for production. Military equipment. The order of performance of developmental works on creation of completing products of interbranch application] Moscow, Standartinform, 2005. 65 p. Available at: <http://www.twirpx.com/file/520927> (in Russian)

#### СВЕДЕНИЯ ОБ АВТОРАХ

Денис А. Печников к.т.н., доцент, докторант, Военно-морская академия им. Адмирала Флота Советского Союза Н.Г. Кузнецова, Ушаковская набережная, 17/1, г. Санкт-Петербург, 197045, Россия, 19pda72@bk.ru

#### КРИТЕРИЙ АВТОРСТВА

Денис А. Печников полностью подготовил рукопись и несет ответственность за плагиат

#### КОНФЛИКТ ИНТЕРЕСОВ

Автор заявляет об отсутствии конфликта интересов.

ПОСТУПИЛА 22.01.2018

ПРИНЯТА В ПЕЧАТЬ 19.02.2018

9 Ivanov V.N. Metod avtomatizirovannoi otsenki kachestva [The method of automated assessment of the quality of professional activity in shipborne automated systems] Petrodvorets, VMIRE, 1999. 187 p. (in Russian)

10 Kremer N.Sh. Teoriya veroyatnosti i matematicheskaya statistika [Theory of Probability and Mathematical Statistics] Moscow, Unity-Dana, 2004. 573 p. Available at: [http://www.e-joe.ru/sod/99/2\\_99/st160.html](http://www.e-joe.ru/sod/99/2_99/st160.html) (in Russian)

11 Crocker L., Algina J. Vvedenie v klassicheskuyu i sovremennuyu teoriyu testov [Introduction to the classical and modern theory of tests] Moscow, Logos, 2010. 668 p. Available at: <http://www.twirpx.com/file/1346290> (in Russian)

12 Mosin V.D. Modelirovanie protsessa adaptivnogo upravleniya [Modeling the process of adaptive management of cadets' activities in the course of practical training on automated simulator] Petrodvorets, VMIRE, 1999. 164 p. (in Russian)

13 Naumov V.N. Teoreticheskie osnovy predstavleniya sodержaniya professioal'noi podgotovki [Theoretical bases of representation of the content of professional training of specialists of the Navy in intellectual training systems] Petrodvorets, VVMURE, 1994. 296 p. (in Russian)

14 Pechnikov A.N. Kompleks metodov i modelei otsenki [A set of methods and models for assessing the quality and efficiency of the activities of weapons and military equipment operators in the training and training systems of the Navy] Saint-Petersburg, SPbGETU, 2011. 343 p. (in Russian)

15 Pechnikov A.N., Vetrov Yu.A. Proektirovanie i primeneniye komp'yuternykh tekhnologii [Design and application of computer learning technologies. Part 1. The concept of automated learning systems and the modeling of activity processes] Saint-Petersburg, BSTU, 2002. 195 p. (in Russian)

16 Pechnikov A.N., Struchkov A.M., Heckert E.V. et al. Generator otsenki [Evaluation generator: Certificate of state registration of computer programs No. 2009613362 of 26.06.2009] (in Russian)

17 Platov A.N. Avtomatizatsiya upravleniya kontrol'em [Automation of control of training control on simulators of the Navy] Saint-Petersburg, Central Research Institute of the Ministry of Defense of the Russian Federation, 2000. 125 p. (in Russian)

18 Puzyrev V.Yu. Diagnostika deyatel'osti operatorov [Diagnostics of the activity of operators of ship electromechanical systems in specialized simulators of the Navy] Novorossiysk, NGMA, 2005. 129 p. (in Russian)

19 GOST R 5060580-93. Sistema razrabotki i postanovki [System for the development and production of products for production. Terms and Definitions] Moscow, VNIИstandart. Available at: <http://gostrf.com/normadata/1/4293827/4293827526.htm> (in Russian)

20 Skorov A.A. Metodologicheskie osnovy ispol'zovaniya komp'yuternykh [Methodological bases of using computer training technologies in the process of professional training of Navy specialists] Petrodvorets, VMIRE, 2001. 318 p. (in Russian)

21 Struchkov A.M. Proektirovanie kompyuternykh sistem obucheniya [Designing of computer training systems and intellectual training] Petrodvorets, Passat, 2002. 166 p. (in Russian)

#### INFORMATION ABOUT AUTHORS

Denis A. Pechnikov Cand. Sci. (Engin.), associate professor, doctoral, Naval Academy named after Admiral of the Fleet of the Soviet Union N.G. Kuznetsov, Ushakov embankment, 17/1, St. Petersburg, 197045, Russia, 19pda72@bk.ru

#### CONTRIBUTION

Denis A. Pechnikov completely prepared the manuscript and is responsible for plagiarism

#### CONFLICT OF INTEREST

The author declare no conflict of interest.

RECEIVED 1.22.2018

ACCEPTED 2.19.2018