Methodology for calculating the probabilistic characteristics of the objectivity of the test results of students in the distance learning system Moodle

Lyzlov Sergey Sergeevich, Uvarov Sergey Sergeevich, Katina Marina Vladimirovna
Russian University of Transport (MIIT), Moscow, Russia

Contribution on the State of the Art
https://doi.org/10.7251/JIT2202111S UDC: 316.644-057.875:502/504

Abstract: The article discusses the methodology for calculating the probabilistic characteristics of the objectivity of test results in distance learning. Calculation expressions are obtained and simulation modeling of the process of forming test tasks for students is carried out. The results of calculations and simulation modeling are given, estimates of a random discrete value are obtained, defined as the number of tests at which information about the content of all questions in test tasks becomes known to all students.

Keywords: Moodle distance learning system, simulation modeling, calculation expressions and simulation results for assessing the probabilistic characteristics of the objectivity of student testing results.

To implement distance learning, there are a large number of software products that are well covered in print and on the Internet [1, 2, 3]

The Institute of Transport Engineering and Control Systems of the Russian University of Transport operates a distance learning server, on which the Moodle distance learning system, which is widely used in Russian universities, is installed.

At the Department of Information Management and Protection, the above system is used to test students in a number of disciplines.

The Moodle system has ample opportunities for testing, there are only 12 types of questions. The Moodle system is quite fully described on the Internet, so we will not describe all the testing possibilities in the Moodle system, but touch on the main parameters.

For testing, the teacher must create a database of questions in the bank of questions, a system related to the entire course. The bank of questions can be structured by creating categories of questions related to a specific topic of the discipline. This allows you to organize both intermediate testing on a specific topic, and the final exam, which includes all topics of the discipline.

When developing a testing scenario, the teacher must determine the number of questions in the bank of questions of the system, the number of categories and questions in them, the number of questions in the test task for each student, the time required to pass the test, the grading system, the number of allowed attempts to pass the test.

When analyzing the results of the test control of students from several groups who were tested non-simultaneously, it was noticed that the grades increased, and the time to answer questions decreased. This is explained by the fact that the students of the first group, using information technologies (for example, social networks), transmit information about the questions to the second group, etc. This leads to the fact that the objectivity of the test results is reduced. To increase the objectivity of the test results, it is necessary to increase the number of questions in the bank of questions of the system and change them frequently for the following groups of students. This greatly complicates the work of the teacher. Therefore, the problem arose of assessing the probabilistic characteristics of the objectivity of the test results.

Consider the algorithm for generating test items for each student when dividing the bank of questions into categories related to specific topics of the discipline. Figure 1 shows a graph illustrating the process of generating test items.

Fig. 1. *Illustration of the process of forming test items*

The bank of questions of the system is divided into m (m=1, 2, 3…) categories, in each category i (i=1, 2, 3…) questions.

For each student, the system generates a test task containing one question from each category, so the test task will contain m questions. The system selects a question from a category randomly according to a uniform distribution law.

If we have 5 categories (m=5) and there are 5 questions (i=5) in each category, the number of questions in the system's question bank should be m·i = 25. Using five questions from a category, $i^i = 5^5 = 3°125$ combinations of questions in the test item, composed of one category.

Categories can contain the same or different number of questions. Therefore, to determine the probabilistic characteristics of the objectivity of the test results, it is necessary to conduct an analysis for the category containing the maximum number of questions.

Consider the simplest example of testing (exam), if there are only two questions in a particular category (k = 2). The first question will be denoted by the number 1, and the second by the number 2. Naturally, the questions differ from each other. For the first student, the system will generate a test task from one question, in which there will be question 1 or 2 with a probability of 0.5. This is true for the second student and for all subsequent students.

Let us find the probability that after testing two students (n=2), where n is the number of students, all other students will know the content of two questions.

The analysis was carried out using the apparatus of probability theory and combinatorics [4, 5].

All possible combinations (results of outcomes) are calculated by the formula $k^k = 2^2 = 4$.

Naturally, in order to find out the content of two questions, the number of students must satisfy the condition n ≥ k.

Table 1 shows all possible combinations of outcomes Ii, where i=1,2,3,4, there are 4 such outcomes.

Table 1.

n	I_1	I_2	I_3	I_4
1	1	2	1	2
2	2	1	1	2

|←A→|←B→|

The probabilities of outcomes are equally probable, therefore, the probability of any of the outcomes Pi is equal to Pi =0.25, where i - 1, 2, 3, 4.

Table 1 can be divided into two areas:

- area A corresponds to the fact that the content of the two questions will be known to all other students taking the test;
- region B corresponds to the fact that either of the two questions will be unknown.

If the system generates a test task for the second student, taking into account the results of the formation of the question for the first student, and the result of the formation falls into area A, this corresponds to the fact that all questions are known, otherwise the system must generate a test task for the third student, etc.

Therefore, with two independent tests (two students participate in testing), the probability that the combination of test results will fall into area A is determined by the expression PA=0.25+0.25=0.5, otherwise, in order to have information about the content of two questions, it is necessary to conduct a third test. In this case, the probability that the content of two questions will be known is equal to P3=PA·0,5 = 0,25.

Similarly, we can write: P4=P3·0.5 = 0.25·0.5= 0.125; P5=P4·0.5=0.125·0.5=0.0625.

Let's write down the general recursive formula for calculating the probabilities:

P2=0.5; P3 = P2·0.5=0.25; P4 = P3·0.5= 0.125; ... Pn = P(n-1)·0.5 (1)
where n is the number of tests (number of students participating in testing).

Figure 2 shows a graph corresponding to the process of forming test items for students.

Fig. 2. *An illustration of the process of forming test items for five students*

For the first student, the system generates question 1 or 2. Similarly, for the second student, the system generates question 1 or 2. In this case, the combinations of all possible variants of questions generated by the system for two students (n=2) will correspond to the following combinations: 11, **12, 21** , 22. Combinations **12** and **21** correspond to a situation in which the content of two questions is known. Combinations 11 and 22 indicate that the content of the first or second question is unknown.

Similarly, in the third test (n=3), the system generates the following combinations: 111, **112, 221,** 222. Combinations **112** and **221** correspond to a situation in which the content of two questions is known. With combinations of 111 and 222, the content of any question is unknown. Further, the process of forming test tasks continues.

A discrete random variable is the number of tests conducted by n students, at which information about two questions will be known. The distribution law of a discrete random variable n given in tabular form will correspond to Table 2.

Table 2.

n	2	3	4	5	6	7
P(n)	0,5	0,25	0,125	0,0625	0,03125	0,01563

The distribution polygon of a discrete random variable n is shown in Figure 3, and the distribution function of a discrete random variable n is shown in Figure 4.

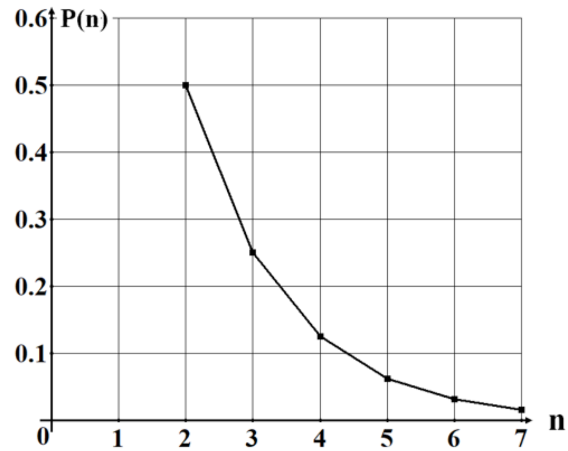


Fig. 3. *Distribution polygon*

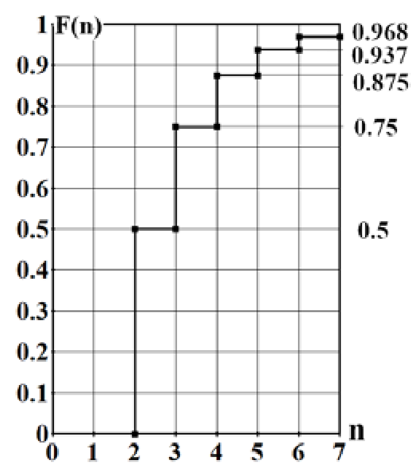


Fig. 4. *Distribution function*

Analysis of the distribution function graph (Fig. 4) allows us to draw the following conclusions:

The probability that when testing two students (n=2) information about two questions will be known is P(2) = 0.5. The probability that when testing three students (n = 3) information about two questions will be known is P(3) = 0.75, etc.

The main numerical characteristics of a discrete random variable n:

The mathematical expectation of a discrete random variable n is determined by the formula:

$$M(n) = \sum_{n=2}^{\infty} n \cdot P_n \tag{2}$$

Since the probability value Pn for each next value of the random variable n corresponds to the principle of dichotomy, i.e. Pn = Pn-1·0.5, formula (2) can be transformed into a number series:

$$M(n) = 1 + \sum_{n=2}^{\infty} \frac{n+1}{n^2} = 3 \tag{3}$$

The variance of a discrete random variable n is determined by the formula

$$D(n) = \sum_{n=2}^{\infty} (n - M(n))^2 \cdot P_n = 2 \tag{4}$$

Consider the case when k=3, the system generates three test tasks for each student on one question. The dimension of the problem increases exponentially. The total number of all combinations of test outcomes becomes $k^k = 3^3 = 27$. All possible combinations of questions in test tasks, taking into account the location of the three numbers 1, 2, 3, are shown in Table 3.

In Table 3, three areas A, B and C can be distinguished. In area A there are n! = 3! = 6 possible combinations of questions in test tasks for three students, therefore, the probability of an event falling into area A is $P(A) = \frac{6}{27} = 0.2222 \ldots$ Area A corresponds to the situation when information about the content of all three questions becomes known.

Region B contains three sub-regions B1, B2, B3. In sub-area B1, only two digits 1 and 2 are used, and one of the digits can be repeated. If the number 1 is repeated, such combinations correspond to the outcome numbers 7, 8, 9, if the number 2 is repeated, such combinations correspond to the outcome numbers 10, 11, 12, which is shown in Table 3. The number of permutations with repetitions is determined by the formula:

$$\Pi(k1, k2) = \frac{k!}{k_1! \cdot k_2!} = \frac{3!}{2! \cdot 1!} = \frac{3!}{1! \cdot 2!} = 3 \tag{5}$$

where k1=2 the first question is included in combinations twice, similarly k2=2 – the second question is included in combinations twice. Therefore, the number of combinations of numbers in area B1 is 6.

A similar calculation can be carried out for subareas B2 and B3, therefore, the total number of combinations of numbers in area B is 18.

The probabilities that the combination of numbers will belong to subdomains B1 or B2 or B3.

$$\text{B3} \;\; P(B_1) = P(B_2) = P(B_3) = \frac{6}{27} = \frac{2}{9} = 0.2222 \ldots$$

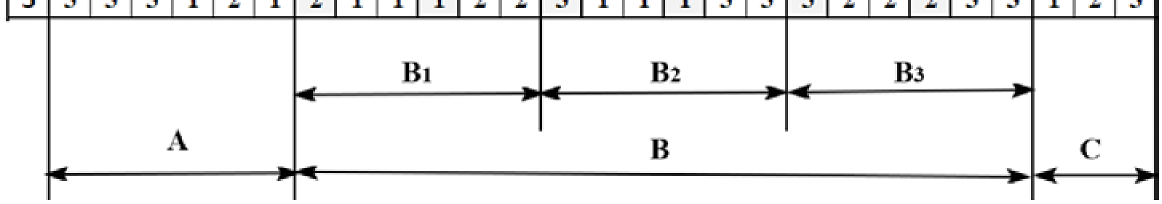
From this it follows that the probability of getting into area B is $P(B) = P(B_1) + (B_2) + P(B_3) = \frac{18}{27} = \frac{2}{3} = 0.666 \ldots$

In area C, there are 3 possible combinations, therefore, the probability of getting into area C is $P(C) = \frac{3}{27} = \frac{1}{9} = 0.1111 \ldots$ In area C, the content of two questions is unknown - either 2 and 3 or 1 and 3 or 1 and 2.

If the combination of question numbers in the test tasks did not fall into area A, therefore, the combination of question numbers fell into area B or C. But since 2 questions are unknown in area C, then when testing the fourth student, area C will not give an answer to the content of all three questions. If the event B is realized, the probability of which is P(B)=0.666…then in order for the information

Table 3.

	123						112			221			113			331			223			332			1	2	3
n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	1	1	2	2	3	3	1	1	2	2	2	1	1	1	3	3	3	1	2	2	3	3	3	2	1	2	3
2	2	2	1	3	1	2	1	2	1	2	1	2	1	3	1	3	1	3	2	3	2	3	2	3	1	2	3
3	3	3	3	1	2	1	2	1	1	1	2	2	3	1	1	1	3	3	3	2	2	2	3	3	1	2	3



about all three questions to be known, it is necessary that one of the missing question numbers (either 1, or 2, or 3) falls out. The probability of this event is $P(X) = \frac{1}{3}$. Events B and X are independent, using the formula for multiplying the probabilities of independent events B and X, you can determine the probability that when testing the fourth student, information about the content of all three questions will be known.

$$P(4) = P(B \cdot X) = \frac{18}{27} \cdot \frac{1}{3} = \frac{6}{27} = \frac{2}{9} = 0.222 \ldots \tag{6}$$

Thus, the probability that information about all three questions will be known after testing the fourth student is P(n) = P(4) = 0.222…

If, during testing of the fourth student, information about one of the three questions (1 or 2 or 3) was not known, the following conclusions can be drawn:

- when forming question number 1 for the fourth student, the combination of combinations of questions for the three previous students corresponds to sub-areas B1 and B2 in which information about the content of question number 1 is known, in sub-area B1 information about question number 3 is unknown, and in sub-area B2 information about question number 2 is unknown;
- when forming question number 2 for the fourth student, the combination of combinations of questions for the three previous students corresponds to sub-areas B1 and B3 in which information about the content of question number 2 is known, in sub-area B1 information about question number 3 is unknown, and in sub-area B3 information about question number is unknown one;
- when forming question number 3 for the fourth student, the combination of combinations of questions for the three previous students corresponds to sub-areas B2 and B3 in which information about the content of question number 3 is known, in sub-area B2 information about question number 3 is unknown, and in sub-area B3 information about question number is unknown one.

It should be taken into account that after testing the fourth student in area C, combinations of question numbers appear in which information about the content of two questions is known (see Table 4 of combinations of question numbers in area C), this circumstance must be taken into account when calculating probabilities.

In Table 4, combinations of question numbers in area C correspond to 111, 222, 333. When testing the fourth student, depending on the question number (1 or 2 or 3), combinations of question numbers appear in area C in which information about the content of one of the three questions is unknown.

Table 4.

	1	2	3
n	25	26	27
1	1	2	3
2	1	2	3
3	1	2	3
4	1	1	1
	2	2	2
	3	3	3

From what has been said, it follows that in order for information about all three questions to be known, it is necessary that the system for the fifth student form questions from areas, the probability of falling into which is equal to:

$$P = \left(\frac{2}{27} + 2 \cdot \frac{6}{27}\right) = \frac{14}{27} = 0.518518 \ldots \tag{7}$$

To calculate the probability that information about all three questions will be known after testing the fifth student, it is necessary to take into account the probability of one of the numbers falling out either 1 or 2 or 3, this probability is equal to $\frac{1}{3}$, therefore, we can write:

$$P(5) = \frac{14}{27} \cdot \frac{1}{3} = \frac{14}{81} = 0.17283\ldots \tag{8}$$

Having carried out a similar analysis for n = 6, 7, 8… you can get the calculated expressions for P(n).

When k=4 (the number of questions in the category of the bank of questions of the system), the dimension of the problem increases significantly k^k = 44 = 256, which significantly complicates the receipt of calculation expressions, therefore, a simulation model was created for the process of forming test tasks for students by the Moodle system. The simulation results are shown below.

k=2 *Table 5*

n	2	3	4	5	6	7	8	9	10
P(n)	0,5	0,25	0,125	0,0625	0,03125	0,01563	0,00781	0,0039	0,00195

k=3 *Table 6*

n	3	4	5	6	7	8	9	10	11
P(n)			0,17283	0,1233	0,0848	0,05760	0,0384	0,0264	0,01722

k=4 *Table 7*

n	4	5	6	7	8	9	10	11		12
P(n)	0,0936	0,1407	0,1465	0,1319	0,1101	0,0882	0,0675	0,0535		0,0407

k=5 *Table 8*

n	5	6	7	8	9	10	11	12	13
P(n)	0,0382	0,0768	0,1002	0,1076	0.1043	0,0956	0,0837	0,0716	0,0601

k=6 *Table 9*

n	6	7	8	9	10	11	12	13	14
P(n)	0,0155	0,0385	0,0599	0,0750	0,0826	0,08444	0,0817	0,0760	0,06141

Tables 5, 6, 7, 8, 9 show the laws of distribution of a discrete random variable n - the number of students tested, given in tabular form, for various k - the number of questions in the category of the bank of questions of the system. The condition n ≥ k must be satisfied.

The polygons of the distributions of a discrete random variable n for various k are shown in Figure 5 and the distribution functions in Figure 6.

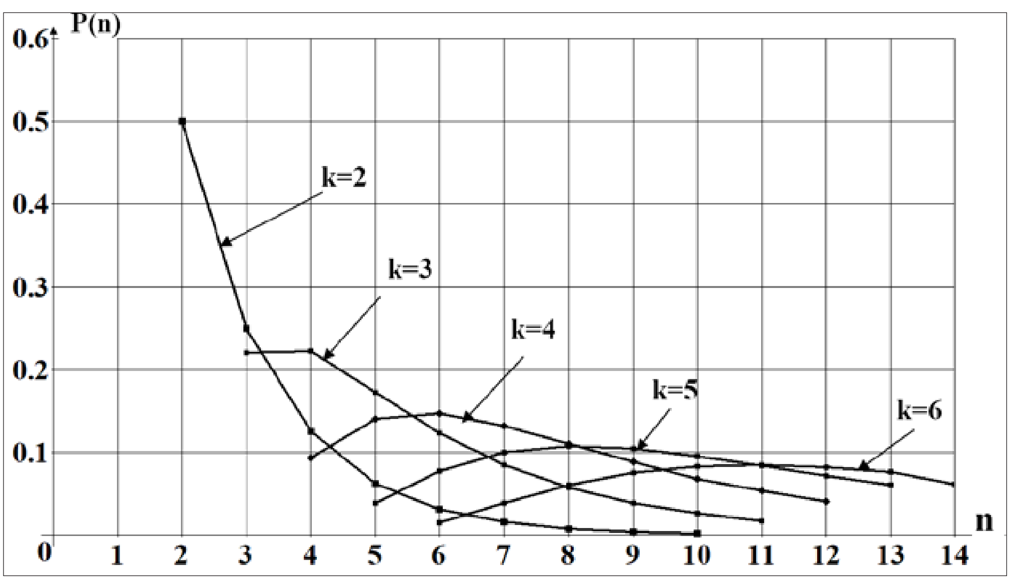


Fig. 5. *Polygons of distributions of a discrete random variable n*

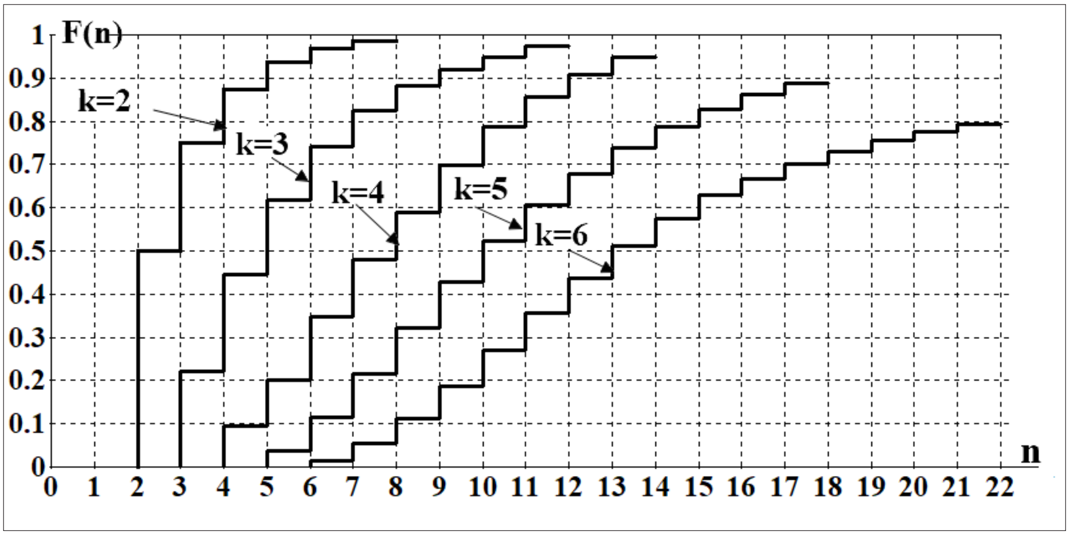


***Fig.6. Distribution** functions of a discrete random variable n*

Table 10

k	2	3	4	5	6
M(n)	3	5,5	8,33	11,42	14,69
D(n)	2	6,82	14,4	25,25	38,94
δ(n)	1,41	2,57	3,79	5,02	6,24

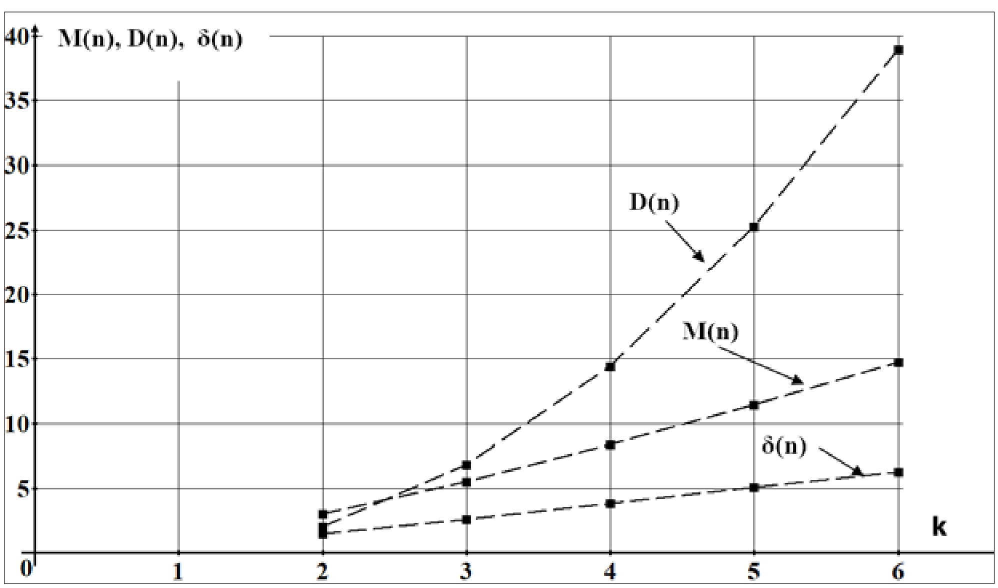


***Fig. 7.** Estimates of mathematical expectations M(n), variances D(n) and standard deviations δ(n)*

Analysis of the graphs in Figure 6 shows that with six questions in a category (k=6) when testing 21 students (n=21), the probability that information about the content of all six questions with a probability of 0.8 will be known to all students of the group.

Table 10 shows for various k estimates of mathematical expectations M(n), variances D(n) and standard deviations δ(n) for a discrete random variable n.

Figure 7 shows the values of estimates of mathematical expectations M(n), variances D(n) and standard deviations δ(n) for a discrete random variable n for various k, where k is the number of questions in the category of the Moodle question bank.

The sample size in simulation modeling for each value of k is more than one million realizations.

After analyzing the calculated expressions and simulation results, we can conclude that the simulation model adequately reflects the process of forming test tasks for students.

The results obtained will allow teachers to rationally develop scenarios for testing students in the Moodle system, taking into account the number of students in groups and the number of questions in the categories of the system's question bank.

INFORMATION ABOUT AUTHORS

Lyzlov Sergey Sergeevich - Associate Professor of the Department of Information Management and Security, Russian University of Transport (MIIT)
academic degree: candidate of technical sciences
academic title: senior researcher
e-mail: postmain@mail.ru

Uvarov Sergey Sergeevich - Associate Professor, Department of Information Management and Security, Russian University of Transport (MIIT)
academic degree: candidate of technical sciences
academic title: senior researcher

Katina Marina Vladimirovna - Senior Lecturer, Department of Information Management and Security, Russian University of Transport (MIIT)

Методика расчета вероятностных характеристик объективностирезультатов тестирования студентов в системе дистанционного обучения Moodle

Лызлов Сергей Сергеевич, Уваров Сергей Сергеевич, Катина Марина Владимировна

Российский университет транспорта (МИИТ), Москва, postmain@mail.ru

Оригинальная научная статья

Аннотация: В статье рассматривается методика расчета вероятностных характеристик объективности результатов тестирования при дистанционном обучении. Получены расчетные выражения и проведено имитационное моделирования процесса формирования тестовых заданий для студентов. Приведены результаты расчетов и имитационного моделирования, получены оценки случайной дискретной величины, определяемое как количество тестирований при котором информация о содержании всех вопросов в тестовых задания становится известна всем студентам.

Ключевые слова: Система дистанционного обучения Moodle, имитационное моделирование, расчетные выражения и результаты имитационного моделирования оценки вероятностных характеристик объективности результатов тестирования студентов.

Для реализации дистанционной формы обучения существует большое количество программных продуктов, хорошо освещенных в печати и в сети Интернет [1, 2, 3]

В институте транспортной техники и систем управления Российского университета транспорта функционирует сервер дистанционного обучения, на котором установлена система дистанционного обучения Moodle, широко используемая в университетах РФ.

На кафедре «Управление и защита информации» для тестирования студентов по ряду дисциплин используется указанная выше система.

Система Moodle обладает широкими возможностями для проведения тестирования, одних только типов вопросов существует 12. Система Moodle достаточно полно описана в Интернете, поэтому не будем описывать все возможности тестирования в системе Moodle, коснемся главных параметров.

Для тестирования преподаватель должен создать базу вопросов в банке вопросов, системы, относящихся ко всему курсу. Банк вопросов возможно структурировать, создав категории вопросов, относящиеся к определенной теме дисциплины. Это позволяет организовывать как промежуточное тестирование по определенной теме, так и проведение итогового экзамена, который включает все темы дисциплины.

При разработки сценария тестирования преподаватель должен определить количество вопросов в банке вопросов системы, количество категорий и вопросы в них, количество вопросов в тестовом задании для каждого студента, время, необходимое для прохождения теста, систему оценивания, количество разрешенных попыток пройти тест.

При анализе результатов тестового контроля студентов нескольких групп, проходивших тестирование неодновременно, было замечено, что оценки повышались, а время ответа на вопросы уменьшалось. Это объясняется тем, что студены первой группы, используя информационные технологии (например, социальные сети), передают информацию о вопросах второй группе и. т. д. Это приводит к тому, что объективность результатов тестирования снижается. Для повышения объективности результатов тестирования необходимо увеличивать количество вопросов в банке вопросов системы и часто менять их для следующих групп студентов. Это значительно усложняет работу преподавателя. Поэтому возникла задача оценки вероятностных характеристик объективности результатов тестирования.

Рассмотрим алгоритм формирования тестовых заданий для каждого студента при разбиении банка вопросов на категории, относящиеся к конкретным темам дисциплины. На рисунке 1 приведен граф, иллюстрирующий процесс формирования тестовых заданий.

Рис. 1. *Иллюстрация процесса формирования тестовых заданий*

Банк вопросов системы разбит на m (m=1, 2, 3…) категорий, в каждой категории i (i=1, 2, 3…) вопросов.

Для каждого студента система формирует тестовое задание, содержащее по одному вопросу из каждой категорий, таким образом в тестовом задании окажется m вопросов. Вопрос из категории система выбирает случайным образом по равномерному закону распределения.

Если имеем 5 категорий (m=5) и в каждой категории 5 вопросов (i=5), количество вопросов в банке вопросов системы должно быть $m \cdot i = 25$. Используя пять вопросов из категории можно сформировать $i^i = 5^5 = 3°125$сочетаний вопросов в тестовом задании, составленных из одной категории.

Категории могут содержаться как одинаковое, так и разное количество вопросов. Поэтому для определения вероятностных характеристик объективности результатов тестировании необходимо провести анализ для категории, содержащей максимальное количество вопросов.

Рассмотрим простейший пример тестирования (экзамена), если в конкретной категории есть только два вопроса (k = 2). Первый вопрос обозначим цифрой 1, а второй цифрой 2. Естественно, вопросы отличаются друг от друга. Первому студенту система сформирует тестовое задание из одного вопроса, в котором окажется вопрос 1 или 2 с вероятностью 0,5. Это справедливо и для второго студента и для всех следующих студентов.

Найдем вероятность того, что после тестирования двух студентов (n=2), где n – количество студентов, всем остальным студентам будет известно содержание двух вопросов.

Анализ проведен с использованием аппарата теории вероятностей и комбинаторики [4, 5].

Все возможные сочетания (результаты исходов) вычисляем по формуле $k^k = 2^2 = 4$.

Естественно, для того что бы узнать содержание двух вопросов, количество студентов должно удовлетворять условию $n \geq k$.

В таблице 1 приведены все возможные сочетания исходов Ii, где i=1,2,3,4, таких исходов 4.

Таблица 1

n	I_1	I_2	I_3	I_4
1	1	2	1	2
2	2	1	1	2

|←A→|←B→|

Вероятности исходов равновероятны, следовательно, вероятность любого из исходов Pi равна Pi =0.25, где i - 1, 2, 3, 4.

В таблице 1 можно выделить две области:

- область А соответствует тому что содержание двух вопросов будет известно всем остальным студентам, проходящих тестирование;
- область В соответствует тому что любой из двух вопросов будет неизвестен.

Если система сформирует тестовое задание для второго студента с учетом результатов формирования вопроса для первого студента и результат формирования попадает в область А, это соответствует тому, что все вопросы известны, в противном случае система должна сформировать тестовое задание для третьего студента. и. т. д.

Следовательно, при двух независимых испытаниях (в тестировании участвуют два студента) вероятность того что сочетание результатов испытаний попадет в область А определяются выражением $P_A=0.25+0.25=0.5$, в противном случае для того чтобы иметь информацию о содержании двух вопросов необходимо провести третье испытание. В этом случае вероятность того, что будет известно содержание двух вопросов равна $P_3=P_A \cdot 0{,}5 = 0{,}25$

Аналогично можно записать: $P_4=P_3 \cdot 0.5 = 0.25 \cdot 0.5= 0.125$; $P_5=P_4 \cdot 0{,}5=0{,}125 \cdot 0{,}5=0{,}0625$.

Запишем общую рекуррентную формулу для вычисления вероятностей:

$$P_2 =0.5;\ P_3 = P_2 \cdot 0.5=0.25;\ P_4 = P_3 \cdot 0.5= 0.125;\ \ldots\ P_n = P_{(n-1)} \cdot 0.5 \quad (1)$$

где n – количество испытаний (количество студентов, участвующих в тестировании).

На рисунке 2 представлен граф, соответствующий процессу формирования тестовых заданий для студентов.

Рис. 2. *Иллюстрация процесса формирования тестовых заданий для пяти студентов.*

Для первого студента система формирует вопрос 1 или 2. Аналогично, для второго студента система формирует вопрос 1 или 2. В этом случае сочетания всех возможных вариантов сформированных системой вопросов для двух студентов (n=2) будут соответствовать следующим комбинациям: 11, **12**, **21**, 22. Сочетания **12** и **21** соответствуют ситуации, при которой содержание двух вопросов известно. Сочетания 11 и 22 говорят о том, что неизвестно содержание первого или второго вопроса.

Аналогично при третьем испытании (n=3) система формирует следующие комбинации: 111, **112**, **221**, 222. Сочетания **112** и **221** соответствуют ситуации при который содержание двух вопросов известно. При сочетаниях 111 и 222 – содержание какого-либо вопроса неизвестно. Далее процесс формирования тестовых заданий продолжается.

Дискретной случайной величиной является количество проведенных тестирований n студентов, при котором будет известна информация о двух вопросах. Закон распределения дискретной случайной величины n, заданный в табличной форме будет соответствовать таблице 2.

Таблица 2

n	2	3	4	5	6	7
P(n)	0,5	0,25	0,125	0,0625	0,03125	0,01563

Многоугольник распределения дискретной случайной величины n представлен на рисунке 3, а функция распределения дискретной случайной величины n на рисунке 4.

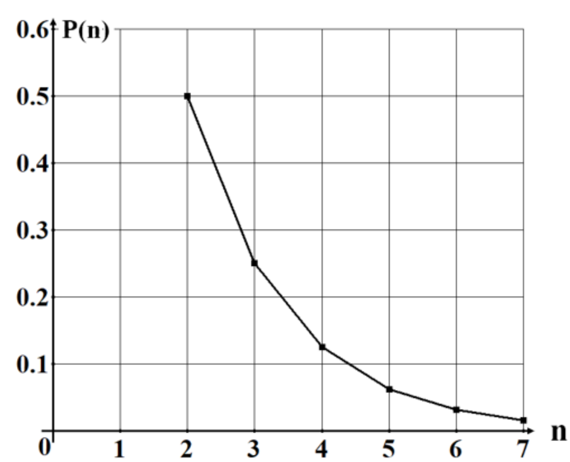


Рис. 3. *Многоугольник распределения.*

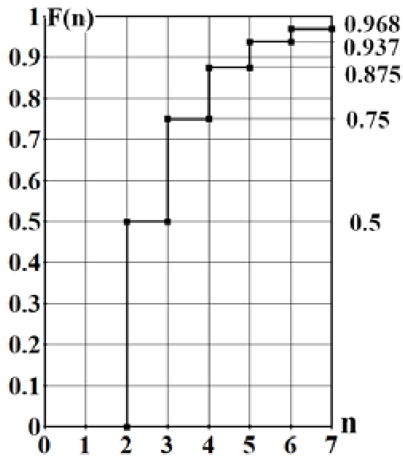


Рис. 4. *Функция распределения.*

Анализ графика функции распределения (рис. 4) позволяет сделать следующие выводы:

Вероятность того что при тестирование двух студентов (n=2) будет известна информация о двух вопросов равна P(2) = 0.5. Вероятность того что при тестировании трех студентов (n = 3) будет известна информация о двух вопросах равна P(3) = 0.75 и т. д.

Основные числовые характеристики дискретной случайной величины n:

Математическое ожидание дискретной случайной величины n определяется по формуле:

$$M(n) = \sum_{n=2}^{\infty} n \cdot P_n \tag{2}$$

Поскольку значение вероятности Pn для каждого следующего значения случайной величины n соответствует принципу дихотомии, т. е. Pn = Pn-1·0.5 формулу (2) можно преобразовать в числовой ряд:

$$M(n) = 1 + \sum_{n=2}^{\infty} \frac{n+1}{n^2} = 3 \tag{3}$$

Дисперсия дискретной случайной величины n определяется по формуле

$$D(n) = \sum_{n=2}^{\infty} (n - M(n))^2 \cdot P_n = 2 \tag{4}$$

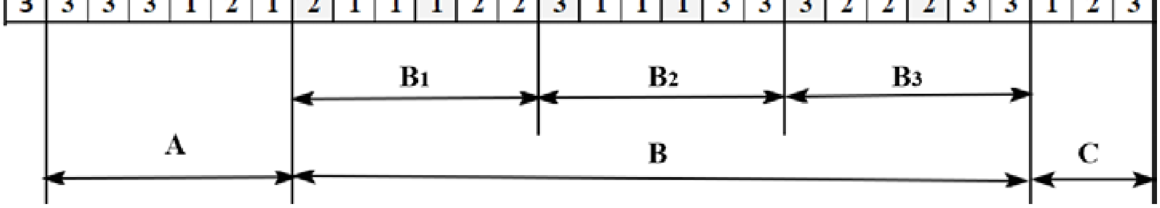
Рассмотрим случай при k=3, система формирует три тестовых задания для каждого студента по одному вопросу. Размерность задачи возрастает в геометрической прогрессии. Общее число всех комбинаций исходов тестирования становится $k^k = 3^3 = 27$. Все возможные сочетания вопросов в тестовых заданий с учетом места расположения трех цифр 1, 2, 3 приведены в таблице 3.

В таблице 3 можно выделить три области A, B и C. В области A находится n! = 3! = 6 возможных сочетаний вопросов в тестовых заданиях для трех студентов, следовательно, вероятность события попасть в область A равна $P(A) = \frac{6}{27} = 0.2222\ldots$ Область A соответствует ситуации, когда информация о содержании всех трех вопросов становится известна.

Область B содержит три подобласти B1, B2, B3. В подобласти B1 используются только две цифры 1 и 2, причем одна из цифр может повторяться. Если повторяется цифра 1, такие сочетания соответствуют номерам исходов 7, 8, 9, если повторяется цифра 2, такие сочетания соответствуют номерам исходов 10, 11, 12, что отображено в таблице 3. Число перестановок с повторениями определяется по формуле:

Таблица 3

	123						112			221			113			331			223			332			1	2	3
n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	1	1	2	2	3	3	1	1	2	2	2	1	1	1	3	3	3	1	2	2	3	3	3	2	1	2	3
2	2	2	1	3	1	2	1	2	1	2	1	2	1	3	1	3	1	3	2	3	2	3	2	3	1	2	3
3	3	3	3	1	2	1	2	1	1	1	2	2	3	1	1	1	3	3	3	2	2	2	3	3	1	2	3



$$\Pi(k1,k2) = \frac{k!}{k_1!\cdot k_2!} = \frac{3!}{2!\cdot 1!} = \frac{3!}{1!\cdot 2!} = 3 \qquad (5)$$

где $k_1=2$ первый вопрос входит в сочетания дважды, аналогично $k_2=2$ – второй вопрос входит в сочетания дважды. Следовательно, число сочетаний цифр в области В1 равно 6.

Аналогичный расчет можно провести для подобластей В2 и В3, следовательно, общее количество сочетаний цифр в области В равно 18.

Вероятности того что сочетание цифр будет принадлежать подобластям В1 или В2 или В3

В3 $P(B_1) = P(B_2) = P(B_3) = \frac{6}{27} = \frac{2}{9} = 0.2222\ldots$

Из этого следует, что вероятность попасть в область В равна $P(B) = P(B_1) + (B_2) + P(B_3) = \frac{18}{27} = \frac{2}{3} = 0.666\ldots$

В области С возможных сочетаний 3, следовательно, вероятность попасть в область С равна $P(C) = \frac{3}{27} = \frac{1}{9} = 0.1111\ldots$В области С неизвестно содержание двух вопросов - или 2 и 3 или 1 и 3 или 1 и 2.

Если сочетание номеров вопросов в тестовых заданиях не попало в область А, следовательно, сочетание номеров вопросов попало в область В или С. Но поскольку в области С неизвестно 2 вопроса, то при тестировании четвертого студента область С не даст ответ на содержание всех трех вопросов. Если реализовалось событие В вероятность которого Р(В)=0.666…то для того чтобы информация о всех трех вопросов была известна необходимо чтобы выпал один из недостающих номеров вопроса (либо 1, либо 2, либо 3). Вероятность этого события равна $P(X) = \frac{1}{3}$.События В и Х являются независимыми, используя формулу для умножения вероятностей независимых событий В и Х можно определить вероятность того что при тестировании четвертого студента информация о содержании всех трех вопросах будут известна.

$$P(4) = P(B\cdot X) = \frac{18}{27}\cdot\frac{1}{3} = \frac{6}{27} = \frac{2}{9} = 0.222\ldots \qquad (6)$$

Таким образом, вероятность того что информация о всех трех вопросах будет известна после тестирования четвертого студента равна P(n) = P(4) = 0.222…

Если при тестировании четвертого студента не стало известна информации об одном из трех вопросах (1или 2 или 3) можно сделать следующие выводы:

- при формировании вопроса номера 1 для четвертого студента комбинация сочетаний вопросов для трех предыдущих студентов соответствует подобластям В1 и В2 в которых информация о содержании вопроса номер 1известна, в подобласти В1 неизвестна информация о вопросе номер 3, а в подобласти В2 неизвестна информация о вопросе номер 2;
- при формировании вопроса номер 2 для четвертого студента комбинация сочетаний вопросов для трех предыдущих студентов соответствует подобластям В1 и В3 в которых информация о содержании вопроса номер 2 известна, в подобласти В1 неизвестна информация о вопросе номер 3, а в подобласти В3 неизвестна информация о вопросе номер 1;
- при формировании вопроса номер 3 для четвертого студента комбинация сочетаний вопросов для трех предыдущих студентов соответствует подобластям В2 и В3 в которых информация о содержании вопроса номер 3 известна, в подобласти В2 неизвестна информация о вопросе номер 3, а в подобласти В3 неизвестна информация о вопросе номер 1;

Необходимо учитывать, что после тестирования четвертого студента в области С появляются сочетания номеров вопросов, в которых известна информация о содержании двух вопросах (см. таблицу 4 сочетаний номеров вопроса в области С) это обстоятельство необходимо учитывать при вычислении вероятностей.

В таблице 4 сочетания номеров вопросов в области С соответствуют 111, 222, 333. При тестировании четвертого студена, в зависимости от номера вопроса (1 или 2 или 3) в области С появляются сочетания номеров вопросов, в которых неизвестна информация о содержании одного из трех вопросов.

Таблица 4

	1	2	3
n	25	26	27
1	1	2	3
2	1	2	3
3	1	2	3
4	1	1	1
	2	2	2
	3	3	3

Из сказанного следует, то чтобы информация о всех трех вопросах была известна, необходимо, чтобы система для пятого студента сформировала вопросы из областей, вероятность попадания в которые равна:

$$P = \left(\frac{2}{27} + 2 \cdot \frac{6}{27}\right) = \frac{14}{27} = 0.518518 \ldots \qquad (7)$$

Для вычисления вероятности, того, что информация о всех трех вопросах будет известна после тестирования пятого студента, необходимо учитывать и вероятность выпадения одной из цифр либо 1 либо 2 либо 3, эта вероятность равна $\frac{1}{3}$, следовательно можно записать:

$$P(5) = \frac{14}{27} \cdot \frac{1}{3} = \frac{14}{81} = 0.17283\ldots \qquad (8)$$

Проведя аналогичный анализ для n = 6, 7, 8… можно получить расчетные выражения для P(n).

При k=4 (количество вопросов в категории банка вопросов системы) размерность задачи значительно увеличивается $k^k = 4^4 = 256$, что существенно усложняет получение расчетных выражений, поэтому была создана имитационная модель процесса формирования системой Moodle тестовых заданий для студентов. Результаты имитационного моделирования приведены ниже.

В таблицах 5, 6, 7, 8, 9 приведены законы распределения дискретной случайной величины n – количества тестируемых студентов, заданные в табличной форме, для различных k – количество вопросов в категории банка вопросов системы. Необходимо выполнение условия n ≥ k.

Многоугольники распределений дискретной случайной величины n для различных k представлены на рисунке 5 и функции распределения на рисунке 6.

k=2 *Таблица 5*

n	2	3	4	5	6	7	8	9	10
P(n)	0,5	0,25	0,125	0,0625	0,03125	0,01563	0,00781	0,0039	0,00195

k=3 *Таблица 6*

n	3	4	5	6	7	8	9	10	11
P(n)			0,17283	0,1233	0,0848	0,05760	0,0384	0,0264	0,01722

k=4 *Таблица 7*

n	4	5	6	7	8	9	10	11		12
P(n)	0,0936	0,1407	0,1465	0,1319	0,1101	0,0882	0,0675	0,0535		0,0407

k=5 *Таблица 8*

n	5	6	7	8	9	10	11	12	13
P(n)	0,0382	0,0768	0,1002	0,1076	0.1043	0,0956	0,0837	0,0716	0,0601

k=6 *Таблица 9*

n	6	7	8	9	10	11	12	13	14
P(n)	0,0155	0,0385	0,0599	0,0750	0,0826	0,08444	0,0817	0,0760	0,06141

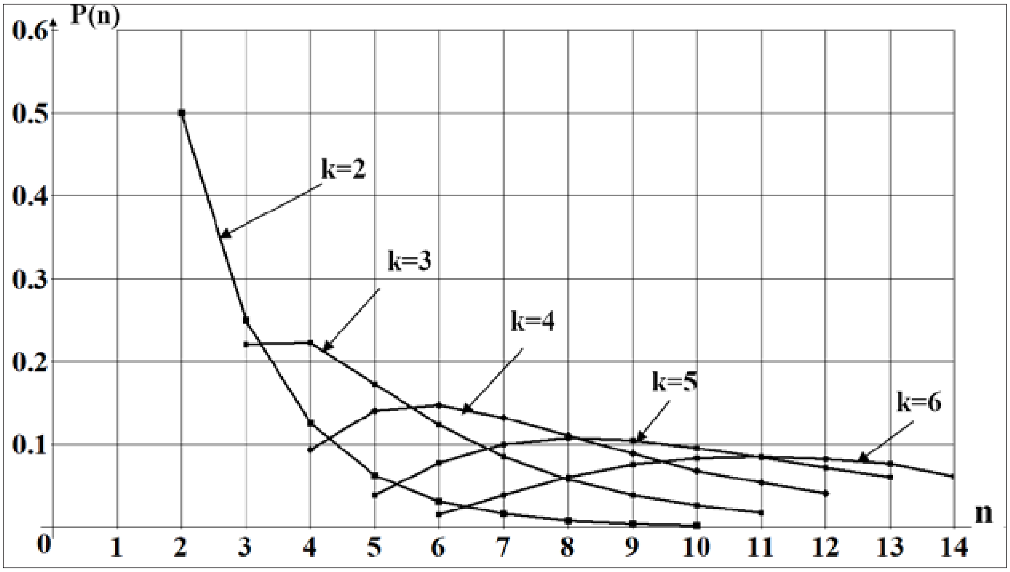


***Рис. 5.** Многоугольники распределений дискретной случайной величины n*

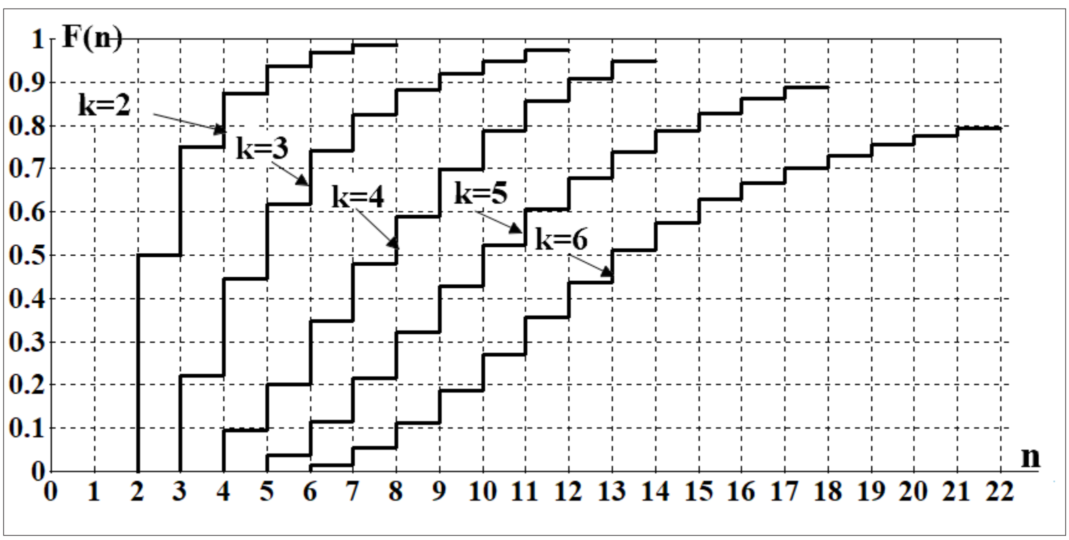


***Рис.6. Функции** распределения дискретной случайной величины n.*

Анализ графиков на рисунке 6 говорит о том, что при шести вопросах в категории (k=6) при тестировании 21 студента (n=21) вероятность того что информация о содержании всех шести вопросов с вероятностью 0,8 будет известна всем студентам группы.

В таблице 10 приведены для различных k оценки математических ожиданий M(n), дисперсий D(n) и среднеквадратических отклонений δ(n) для дискретной случайной величины n.

На рисунке 7 представлены значения оценок математических ожиданий M(n), дисперсий D(n) и среднеквадратических отклонений δ(n) для дискретной случайной величины n при различных k, где k – количество вопросов в категории банка вопросов системы Moodle.

Объем выборки при имитационном моделировании для каждого значения k составляет более одного миллиона реализаций.

Проведя анализ расчетных выражений и ре-

Таблица 10

k	2	3	4	5	6
M(n)	3	5,5	8,33	11,42	14,69
D(n)	2	6,82	14,4	25,25	38,94
δ(n)	1,41	2,57	3,79	5,02	6,24

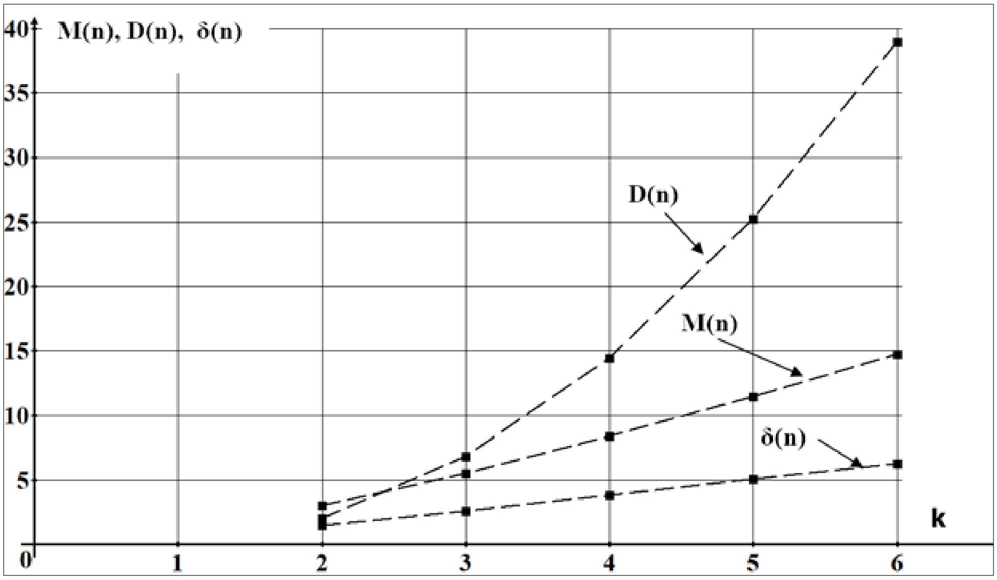


Рис. 7. *Оценки математических ожиданий M(n), дисперсий D(n) и среднеквадратических отклонений δ(n).*

зультатов имитационного моделирования можно сделать вывод о том, что имитационная модель адекватно отражает процесс формирования тестовых заданий для студентов.

Полученные результаты позволят преподавателям рационально разрабатывать сценарии проведения тестирования студентов в системе Moodle с учетом количества студентов в группах и количества вопросов в категориях банка вопросов системы.

References / Список источников литературы:

[1] Поликарпочкин М.В. Возможности использования системы дистанционного обучения Moodle в образовательном процессе // Образование, карьера, общество. – 2017. – №2 - с. 76-78.

[2] Ревунов С.В., Щербина М.М., Лубенская М.П. Инструментарно-методологические основы обеспечения дистанционного образовательного процесса средствами цифровых технологий (на примере «Microsoft Teams») // Педагогика. Вопросы теории и практики. – 2020. – Том 5. Выпуск 3. – с. 387-392.

[3] Исаева И.Э. Методические подходы и практические результаты комплексного использования ЭОР образовательного учреждения в особых условиях// Мир педагогики и психологии. – 2020. - №7. – с. 18 – 29.

[4] Вентцель Е.С. Теория вероятностей. 10-е изд., стер. - М.: Высш. Шк., 2006. – 575 с.

[5] Виленкин Н.Я., Виленкин А.Н., Виленкин П.А. Комбинаторика. – 4-е изд., исправленное. - М.: ФИМА, МЦНМО, 2013. – 400 с.

Received: August 7, 2022 / Получено: 7 августа 2022 г.
Accepted: October 23, 2022 / Принято: 23 октября 2022 г.

Сведения об авторах

Лызлов Сергей Сергеевич – доцент кафедры «Управление и защита информации» Российского университета транспорта (МИИТ)
ученая степень: кандидат технических наук
ученое звание: старший научный сотрудник
e-mail: postmain@mail.ru

Уваров Сергей Сергеевич – доцент кафедры «Управление и защита информации» Российского университета транспорта (МИИТ)
ученая степень: кандидат технических наук
ученое звание: старший научный сотрудник

Катина Марина Владимировна – старший преподаватель кафедры «Управление и защита информации» Российского университета транспорта (МИИТ)

For citation

Лызлов Сергей Сергеевич, Уваров Сергей Сергеевич, Катина Марина Владимировна, Методика расчета вероятностных характеристик объективностирезультатов тестирования студентов в системе дистанционного обучения Moodle, *JITA – Journal of Information Technology and Applications, Banja Luka*, Pan-Europien University APEIRON, Banja Luka, Republika Srpska, Bosna i Hercegovina, JITA 12(2022) 2:111-126, (UDC: 316.644-057.875:502/504), (DOI: 10.7251/JIT2202111S), Volume 12, Number 2, Banja Luka, December (65-172), ISSN 2232-9625 (print), ISSN 2233-0194 (online), UDC 004