

УДК 378.14

## О ВЫБОРЕ ИЗ КОНЕЧНОГО И БЕСКОНЕЧНОГО ПОЛЕЙ В КУРСЕ ТЕОРИИ ВЕРОЯТНОСТИ

<sup>1</sup>Краснощеков В.В., <sup>1</sup>Семенова Н.В., <sup>2</sup>Мухамед Б.М.М., <sup>3</sup>Баккар М.М.А.

<sup>1</sup>ФГАОУ ВО «Санкт-Петербургский политехнический университет Петра Великого»,  
Санкт-Петербург, e-mail: krasno\_vv@spbstu.ru;

<sup>2</sup>Каирский университет, Эль-Гиза, Египет, e-mail: boss\_3ds@yahoo.com;

<sup>3</sup>Университет Аль-Баас, Дамаск, Сирия, e-mail: mamadyan1997@gmail.com

Авторы продолжают исследования точности и границ применимости вероятностных моделей. Представление о точности является важным компонентом компетенций выпускников университетов в области математического моделирования. В настоящей работе авторы проводят сравнение значений вероятностей, вычисленных по точным и приближенным формулам. Авторы находят вероятности выбора из бесконечного поля вариантов по точной формуле Бернулли, которая в данном случае имеет основанием статистическое определение вероятности. Очевидно, что в практических задачах возможен только выбор из конечного поля вариантов, тогда вычисления проводятся по классической формуле выбора. Авторы проводят все исследования на материале одной задачи, которая является нейтральной в смысле содержания текста и в то же время допускает простую интерпретацию полученных результатов. Найденные значения абсолютной и относительной ошибки вычисления вероятностей демонстрируют достаточно быструю сходимость приближенных результатов к точным. Таким образом, авторы эмпирически установили пограничные значения объема банка вариантов, при которых точные и приближенные результаты различаются не более чем на 1%. Подобранные аппроксимации линий сходимости дают формулы для минимального требуемого объема банка вариантов. Эти формулы могут использоваться при средних уровнях риска рассматриваемых процессов.

**Ключевые слова:** преподавание теории вероятностей, формула Бернулли, выбор на основе классического определения, абсолютная ошибка, относительная ошибка

## ON CHOICE FROM FINITE VS INFINITE FIELDS IN COURSE OF PROBABILITY THEORY

<sup>1</sup>Krasnoschekov V.V., <sup>1</sup>Semenova N.V., <sup>2</sup>Mukhamed B.M.M., <sup>3</sup>Bakkar M.M.A.

<sup>1</sup>Peter the Great Saint Petersburg Polytechnic University, Saint Petersburg, e-mail: krasno\_vv@spbstu.ru;

<sup>2</sup>Cairo University, Giza, Egypt, e-mail: boss\_3ds@yahoo.com;

<sup>3</sup>University Al-Baath, Damascus, Syria, e-mail: mamadyan1997@gmail.com

The authors continue to study the accuracy and limitations of applicability of probabilistic models. The concept of accuracy is an important component of the competencies of university graduates in the field of mathematical modeling. In this paper, the authors compare the probabilities calculated using exact and approximate formulas. The authors find the probabilities of selection from an infinite field of options using the exact Bernoulli formula, which, in this case, bases on the statistical definition of probability. Obviously, in practical problems, only a choice from a finite field of options is possible, then they make calculations according to the classical selection formula. The authors conduct all research on the material of the same task, which is neutral in terms of the content of the text, and, at the same time, allows for a simple interpretation of the results. The found values of the absolute and relative errors of calculating the probabilities demonstrate a fairly fast convergence of the approximate results to the exact ones. Thus, the authors empirically established the limit values of the size of the variants bank, at which the exact and approximate results differ by no more than 1%. The selected approximations of the lines of convergence give formulas for the minimum required size of the bank of variants. It is possible to use these formulas at medium risk levels of the processes under consideration.

**Keywords:** teaching the theory of probability, Bernoulli's formula, choice based on the classical definition, absolute error, relative error

Проблематика развития вероятностного подхода к математическому моделированию в области научных исследований и технических разработок является весьма актуальной. Можно указать для примера только самые востребованные в аспекте разработки вероятностных и статистических моделей направления: медицина [1], энергетика [2], химические технологии, в первую очередь связанные с нефтегазовым комплексом [3], информационные технологии [4], анализ бизнес-процессов [5] и т.д. В то же время, с точки зрения работодателей, именно построение математических моделей являет-

ся слабым местом подготовки инженеров и экономистов в современных условиях. Это не вызывает удивления по двум основным причинам. Во-первых, на освоение курса теории вероятности, а иногда и математической статистики, в типовых учебных планах большинства инженерных направлений подготовки отводится один семестр, что не дает преподавателям возможности охватить весь материал, научить решать задачи по теории вероятностей и одновременно углубиться в вопросы построения вероятностных моделей явлений окружающего мира с анализом их точности и границ применимости.

Во-вторых, освоение вероятностных университетских дисциплин преимущественно на 1–2 курсах осложняет ориентацию задач на будущую профессиональную деятельность инженеров и экономистов [6]. Авторы встречались с различными подходами к формированию смыслового и содержательного компонентов заданий по теории вероятностей: от строго классического, лишённого какого-то ни было эмоционального окраса, до ёрнического, в духе «Вредных советов» Григория Остера. Авторские попытки увязать содержание задач с инженерной и экономической проблематикой привели к продуцированию громоздких и скучных для студентов заданий. Можно рекомендовать остановиться на «умеренных» текстах житейской и студенческой тематики либо опираться на общекультурные ориентиры [7]. Вопросы выбора содержания текста задач важны, поскольку связаны с поддержкой мотивации к изучению вероятностных дисциплин, однако они не могут решить проблем формирования компетенций студентов в области построения и анализа математических моделей, которым и посвящено настоящее исследование.

#### Материалы и методы исследования

Настоящая работа продолжает исследования авторов, направленных на поиск механизмов формирования системного, в частности вероятностного, подхода студентов к научному познанию [8, 9]. В качестве соавторов работ были приглашены иностранные обучающиеся программ предмагистерской подготовки Санкт-Петербургского политехнического университета Петра Великого (СПбПУ). Подготовку иностранных граждан к освоению основных образовательных программ на русском языке осуществляют несколько сотен российских вузов, но менее двух десятков из них реализуют специализированные образовательные программы подготовки в магистратуру и аспирантуру. Дело в том, что для запуска и функционирования таких программ необходима численность контингента иностранных подготовительных отделений в несколько сотен обучающихся. Кроме того, такая подготовка имеет смысл только при условии формирования проектных исследовательских компетенций [10] будущих магистрантов, что требует наличия специально подготовленных научно-педагогических кадров. Именно в предмагистратуре преподаватели могут сосредоточиться на анализе вероятностных моделей, опираясь на сформированные в зарубежных вузах компетенции обучающихся в области теории вероятностей и математической статистики.

С помощью предмагистрантов авторы выполнили в предшествующие годы два исследования в области точности вероятностных моделей. Тематика формирования компетенций студентов в этой области представляется актуальной, что подтверждается, в частности, работами других авторов [11, с. 109–116].

В работе [12], выполненной в 2020 г. с помощью предмагистранта из Сирии, авторы, прежде всего, получили графическое подтверждение приближения формы полигона биномиального распределения в существенно несимметричном случае к форме кривой Гаусса. Эти графики были включены авторами в тексты лекций по теории вероятностей для российских и иностранных студентов 2-го курса, что позволило подойти в этих лекциях к проблеме оценки точности приближенных вычислений в теории вероятностей. Далее в этой же работе авторы провели сравнение значений вероятностей, вычисленных на интервалах, по точным формулам Бернулли и приближенным формулам, следующим из интегральной теоремы Лапласа. Такие же вычисления были проделаны для случая редких событий на основе сравнения результатов, полученных по точной формуле Бернулли и приближённым формулам Пуассона и Лапласа. Выбор распределений Бернулли и Пуассона объясняется тем, что, по мнению исследователей, именно эти распределения являются наиболее важными во множестве случаев практических приложений [13]. Несмотря на результаты, подтверждающие возможность замены расчётов по формуле Бернулли расчётами по формуле Пуассона уже при 20 измерениях, сами величины относительной ошибки нельзя было признать удовлетворительными. Это, во-первых, позволило использовать полученные результаты только как качественные, тем более что было получено подтверждение положения общей теории ошибок С.Г. Михлина об обратном корневом убывании ошибки с ростом числа измерений. Во-вторых, это же обстоятельство породило необходимость дальнейшего исследования точности вероятностных моделей Пуассона, аппроксимирующих формулу Бернулли с детальным анализом поведения абсолютной и относительной ошибок вычислений. Такое исследование было проведено авторами в 2021 г. с помощью предмагистрантов из Сирии и из Греции [14].

В этой работе было показано, что при вероятности наступления редкого события  $p = 0,005$  наихудшее значение абсолютной ошибки достигается для  $P(X = 1)$  и не превышает 0,12%, в то время как при вероятности наступления редкого события  $p = 0,0002$  наи-

худшее значение абсолютной ошибки также достигается для  $P(X = 1)$ , но не превышает 0,005%. Таким образом, студентам можно продемонстрировать положение, что точность формулы Пуассона существенно повышается при уменьшении вероятности редкого события. Эти результаты можно было бы считать превосходными, если бы не характер поведения относительной ошибки, которая неуклонно возрастает. Это связано с накоплением ошибок округления при операциях с малыми величинами. Действительно, наихудшие значения относительной ошибки получаются при вычислении  $P(X = 4)$ , для нахождения которой требуется произвести наибольшее число операций. Эти результаты также пригодны для использования в учебном процессе. Студентам можно продемонстрировать конкретные ограничения, связанные с приближенными вычислениями в теории вероятностей и математической статистике.

В настоящем исследовании авторы обратились к другой стороне оценки вероятностных моделей, связанной с проблемой ограниченности выбора. А именно, предмагистрантам было предложено сравнить результаты вычисления вероятности событий, связанных с выбором из бесконечного и конечного полей (банков). Бесконечный выбор определяется формулами Бернулли и/или теоремами о сложении и умножении вероятностей событий, конечный – формулами на основе классического определения вероятности. Эта проблематика связана с современными воззрениями на определения вероятности [15]. Точнее, сопоставляются классическое и статистическое определения вероятности. Эта проблема является многозначной, что находит отражение и в курсе теории вероятностей. Можно привести характерные примеры формулировок задач на сумму и произведение событий: «вероятность того, что цель будет засечена первым локатором, составляет 85%», «вероятность выхода устройства из строя составляет 2%», «вероятность аварии на атомной станции составляет 0,05%» и т.д. Ясно, что эти примеры основаны на неких упрощенных статистических данных, однако студенты должны задуматься над тем, сколько опытов необходимо провести, чтобы быть уверенным в точности модели выбора из бесконечного банка. Например, какова была статистика аварий на атомных станциях, достаточна ли она для того, чтобы определение вероятности было статистическим, а не экспертным? С этой же проблематикой отчасти связана задача о минимальном объеме выборки в статистике. Таким образом вопросы, поднятые в настоящем исследо-

вании, актуальны и представляют дидактический интерес в аспекте формирования вероятностного подхода в научном познании студентов.

### Результаты исследования и их обсуждение

Исследование было проведено на базе следующей модельной задачи:

*Летом 75% дней солнечных. Построить ряд распределения числа солнечных дней из 3 выбранных (0, 1, 2 или 3).*

Выбор тематики определяется нейтральным содержанием, что важно в интернациональных группах студентов. Кроме того, у обучающихся создаётся представление о реальной возможности накопления статистического материала за счет повторяющихся наблюдений. Очевидно, что для решения используется формула Бернулли

$$P_B(i) = P(X = i) = C_n^i p^i q^{n-i}$$

со значениями параметров

$$p = 0,75, q = 1 - p = 0,25, n = 3.$$

Это точное решение  $P_B(0), P_B(1), P_B(2), P_B(3)$ .

Для исследования приближенных решений были использованы задачи с выбором на основе классического определения вероятности:

*Из  $N = 12$  дней  $m = 9$  солнечных. Найти вероятности того, что среди этих 12 дней будет 0 солнечных, 1 солнечный, 2 солнечных, 3 солнечных.*

Это будет приближенное решение  $P_K(0), P_K(1), P_K(2), P_K(3)$ , которое находится по формуле

$$P_K(i) = P(X = i) = \frac{C_m^i C_{N-m}^{3-i}}{C_N^3}$$

Для повышения точности приближенного решения находятся также значения обозначенных выше вероятностей для последовательности начальных условий:  $N = 16, m = 12; N = 20, m = 15; N = 40, m = 30; N = 60, m = 45; N = 80, m = 60; N = 100, m = 75$ . Для иллюстрации выбраны результаты крайних случаев грубого ( $N = 12, m = 9$ ) и точного ( $N = 100, m = 75$ ) приближенных решений по классической формуле. Соответствующие полигоны  $P_K$  представлены на рис. 1 и 2 вместе с полигоном точного решения, полученного по формуле Бернулли  $P_B$ .

Очевидна сходимость приближенного решения к точному, причём ее характер можно наблюдать, рассмотрев последовательно все построенные полигоны. Наибольшее расхождение наблюдается для вероятности  $P(2)$ , но уже для  $N = 80, m = 60$  оно не превышает 1%.

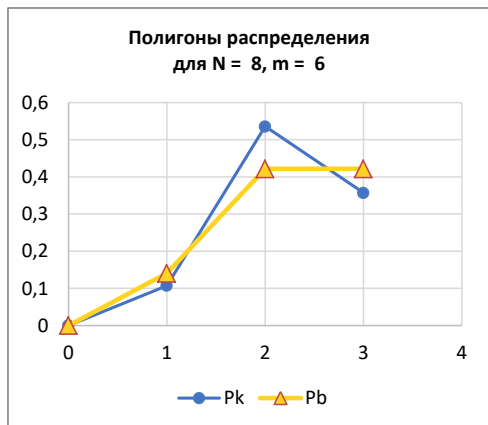


Рис. 1. Полигоны распределения по формуле Бернулли  $P_B$  и по классической формуле  $P_K$  в грубом приближении

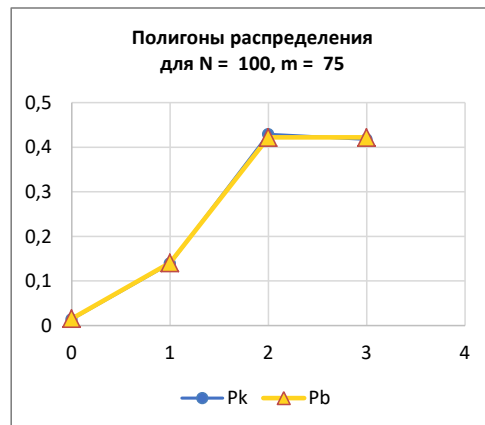


Рис. 2. Полигоны распределения по формуле Бернулли  $P_B$  и по классической формуле  $P_K$  в точном приближении

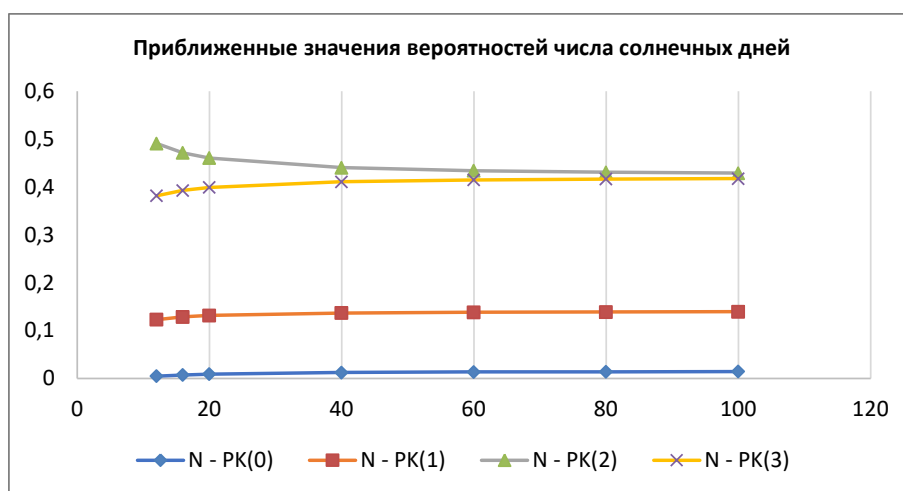


Рис. 3. Зависимость приближенных значений вероятностей  $P_K(i)$  от объема банка для выбора  $N$

Таким образом, вероятно именно эти параметры можно считать приемлемыми для замены в условиях данной задачи выбора из бесконечного банка выбором из конечного банка. Разумеется, в случае расчетов параметра безопасной работы атомной станции необходима большая точность, а соответственно, и большие объемы данных наблюдений.

Графики сходимости компонентов полигонов  $P_K(i)$  представлены на рис. 3.

Видно, что все компоненты полигонов приближенных решений выходят на свои предельные (точные) значения снизу, и только наиболее грубые оценки  $P(2)$  – сверху.

Для детального анализа сходимости приближенных значений вероятностей к точным были вычислены величины абсолютных ошибок  $\Delta = P_K - P_B$  (рис. 4) и относительных ошибок  $\varepsilon = \Delta / P_B$  (рис. 5) всех изученных случаев. Как и в предшеству-

ющих исследованиях авторов [12] и [14], величины абсолютных ошибок достигают долей процента с гарантией при  $N = 100$ , чего нельзя сказать о величинах относительных ошибок, остающихся в окрестности 10%. Это объясняется, во-первых, выполнением деления малых величин, во-вторых, накоплением ошибок округления в связи с ростом числа операций. Ошибки округления становятся ведущим источником погрешности по сравнению с ошибками моделирования.

Были подобраны аппроксимации для графиков абсолютных и относительных ошибок. Оказалось, что

$$\varepsilon[\Delta] \approx \frac{\beta}{N^\alpha}, \Rightarrow N \approx \left(\frac{\beta}{\varepsilon}\right)^{1/\alpha},$$

где константа  $\beta$  варьируется от 1,2 до 8,1 для разных компонентов полигона  $P_K(i)$  в слу-

чае относительной ошибки, или от 0,13 до 0,95 в случае абсолютной ошибки, а показатель  $\alpha$  близок к 1, причём превышает 1 для трех компонентов из четырёх, что указывает на относительно быструю сходимость приближенного решения к точному.

**Заключение**

Проведенный анализ дает возможность сделать вывод о правомерности замены выбора из бесконечного банка вариантов выбором из конечного банка. Авторы провели анализ значений абсо-

лютных и относительных ошибок приближенного вычисления вероятностей. Авторы показали, что уже при 100 вариантах выбора приближенные значения достаточно хорошо приближают точные. Это значит, что такое количество случаев оказывается достаточным для генерирования значений вероятности в соответствии со статистическим подходом к их определению. Эти данные могут быть использованы при прогнозировании бизнес-процессов и в теории надёжности при умеренных уровнях рисков.

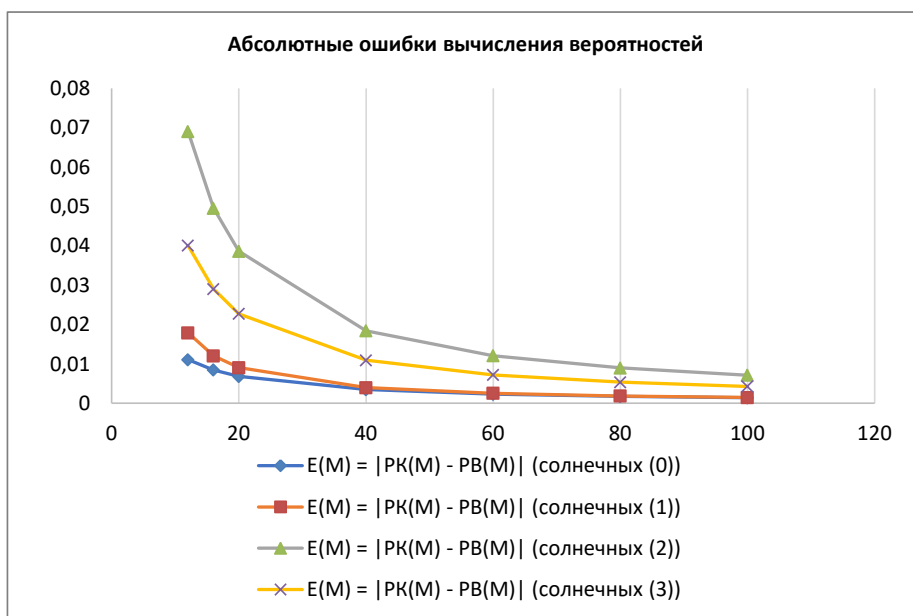


Рис. 4. Зависимость абсолютной ошибки вычисления вероятностей выбора от объёма банка для выбора N

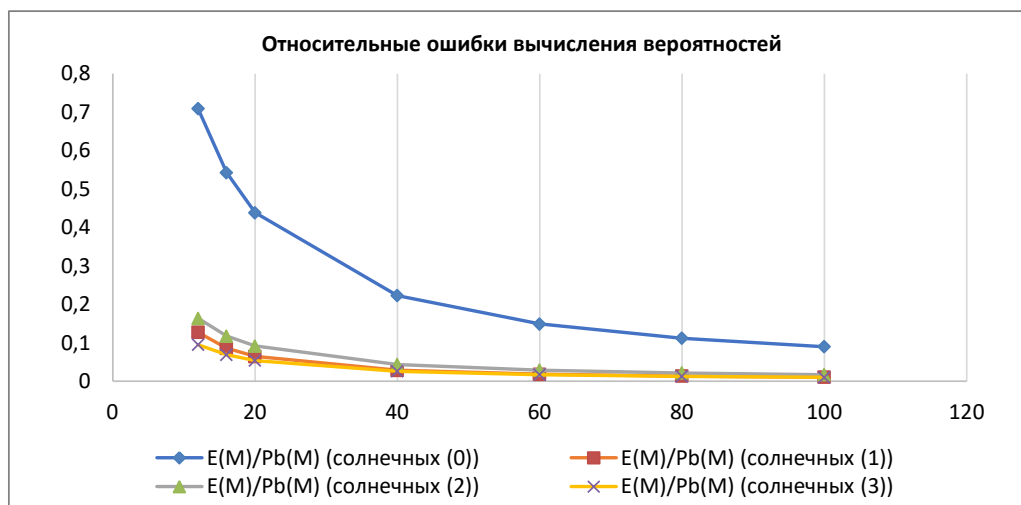


Рис. 5. Зависимость относительной ошибки вычисления вероятностей выбора от объёма банка для выбора N

## Список литературы

1. Soldatova L., Rzhetsky A., De Grave K., King R., Representation of probabilistic scientific knowledge. *Journal of Biomedical Semantics*. 2013. № 4. S7. DOI: 10.1186/2041-1480-4-S1-S7.
2. Cho Young-Been, Cho Yun-Sung, Lee Jae-Gul, Oh Seung-Chan. Design and Implementation of Probabilistic Transient Stability Approach to Assess the High Penetration of Renewable Energy in Korea. *Sustainability*. 2021. 4205. DOI: 10.3390/su13084205.
3. Taleb-Berrouane M., Khan F., Kamil M. Dynamic RAMS Analysis Using Advanced Probabilistic Approach. *Chemical Engineering Transactions*. 2019. No. 77. P. 241–246. DOI: 10.3303/CET1977041.
4. Boulaares S., Omri A., Sassi S., Benslimane D. A Probabilistic Approach: A Model for the Uncertain Representation and Navigation of Uncertain Web Resources. *Proc. 14th Int. Conf. Signal-Image Technology & Internet-Based Systems (SITIS)*, 26–29 Nov. 2018. Las Palmas de Gran Canaria, Spain. 24–31. DOI: 10.1109/SITIS.2018.00015.
5. Bayomie D., Di Ciccio C., La Rosa M., Mendling J. A Probabilistic Approach to Event-Case Correlation for Process Mining. *Proc. 38th Int. Conf. Conceptual Modeling, ER 2019*, 4–7 Nov. 2019, Salvador, Brazil. P. 136–152. DOI: 10.1007/978-3-030-33223-5\_12.
6. Старовойтова Е.Л. Методические особенности изучения вопросов теории вероятностей и математической статистики студентами-первокурсниками // Преподавание математики в высшей школе и работа с одаренными студентами в современных условиях: материалы Международного научно-практического семинара (Могилев, 20 февраля 2020 г.). Могилев, РБУ, 2020. С. 81–84.
7. Гефан Г.Д. Разъяснение положений теории вероятностей с помощью примеров из жизни и художественной литературы // Тенденции развития науки и образования. 2022. № 81–4. С. 38–40. DOI: 10.18411/tmio-01-2022-133.
8. Краснощеков В.В., Семенова Н.В. Формирование вероятностного подхода как методологии научного познания студентов вузов // Современные наукоемкие технологии. 2016. № 9 (3). С. 515–519. DOI: 10.17513/snt.36261.
9. Краснощеков В.В., Семенова Н.В. Forming of probabilistic approach to cognition as component of students professional culture. In: Shipunova O.D., Byleva D. (eds.) *European Proceedings of Social and Behavioural Sciences EpSBS. Proceedings of the joint conferences: 20th Professional Culture of the Specialist of the Future (PCSF 2020) & 12th Communicative Strategies of Information Society (CSIS 2020)*. St. Petersburg, Nov. 16–17, 2020. P. 139–149. DOI: 10.15405/epsbs.2020.12.03.14.
10. Callingham R., Watson J., Oates G. *Statistics and Probability: From Research to the Classroom. Researching and Using Progressions (Trajectories) in Mathematics Education*. Boston, Singapore, Brill, 2019. Chapter 8. P. 181–204. DOI: 10.1163/9789004396449\_008.
11. Гефан Г.Д. Проектно-исследовательская деятельность студентов транспортного вуза при обучении вероятностно-статистическим дисциплинам. Иркутск: ИрГУПС, 2021. 188 с. URL: [https://www.elibrary.ru/download/elibrary\\_47309682\\_36476724.pdf](https://www.elibrary.ru/download/elibrary_47309682_36476724.pdf) (дата обращения: 15.07.2022).
12. Краснощеков В.В., Семенова Н.В., Алдармини С.С. Методы формирования компетенций студентов в области точности вероятностных моделей // Современные проблемы науки и образования. 2020. № 5. URL: <https://science-education.ru/ru/article/view?id=30142> (дата обращения: 16.07.2022). DOI: 10.17513/spno.30142.
13. Viti A., Terzi A., Bertolaccini L. A practical overview on probability distributions. *J Thorac Dis*. 2015. № 7 (3). E7–E10. DOI: 10.3978/j.issn.2072-1439.2015.01.37.
14. Краснощеков В.В., Семенова Н.В., Алсалама А.М., Михолитсис А.Г. О точных и приближенных моделях в вузовском курсе теории вероятности // Современные наукоемкие технологии. 2021. № 10. С. 149–154. DOI: 10.17513/snt.38869.
15. Dhritikesh Chakrabarty. Latest Definition of Probability: Link with Its Earlier Definitions. Preprint. 2021, July. DOI: 10.13140/RG.2.2.32880.58886.