

УДК 371.32:378

О ТОЧНЫХ И ПРИБЛИЖЕННЫХ МОДЕЛЯХ В ВУЗОВСКОМ КУРСЕ ТЕОРИИ ВЕРОЯТНОСТИ

¹Краснощеков В.В., ¹Семенова Н.В., ^{1,2}Алсалама А.М., ^{1,3}Михолитсис А.Г.
¹ФГАОВ ВО «Санкт-Петербургский политехнический университет Петра Великого»,
 Санкт-Петербург, e-mail: krasno_vv@spbstu.ru;
²Университет Дамаска, Дамаск, Сирия, e-mail: alialsalameh12345@gmail.com;
³Университет имени Аристотеля, Салоники, Греция, e-mail: a_micholitsis@yahoo.gr

Авторы продолжают исследование проблем формирования вероятностных компонентов когнитивного стиля студентов вузов. В работе анализируются причины противопоставления детерминистского и вероятностного подходов к познанию, а также последствия такой оппозиции в дидактике преподавания вероятностных и статистических разделов курса математики бакалаврских программ. В качестве одного из механизмов формирования вероятностного когнитивного стиля авторы предлагают использовать задания, связанные с анализом точности вероятностных моделей. Авторы проводят вычислительные эксперименты для анализа погрешностей вычисления вероятностей редких событий с помощью точной формулы Бернулли и приближенной формулы Пуассона. Авторы варьируют как вероятности появления редкого события в одном опыте, так и число опытов. Общими для всех экспериментов остаются вариации интенсивности потока событий, что дает возможность анализа сопоставимых результатов. Авторы находят абсолютные и относительные погрешности вычислений для всех случаев, представляя информацию в графическом виде. Авторы определяют параметры приближенных вычислений с высокой и неудовлетворительной точностью, обозначая таким способом границы применимости приближенных моделей. При проведении занятий авторы рекомендуют делать акцент на влиянии эффекта проявления ошибок округления, усиливающегося при операциях с малыми величинами, которые неизбежно возникают при анализе редких событий.

Ключевые слова: вероятностный подход к научному познанию, преподавание математики, вероятностные модели, точность моделирования, редкие события

ON EXACT AND APPROXIMATE MODELS IN UNIVERSITY COURSE OF PROBABILITIES

¹Krasnoshchekov V.V., ¹Semenova N.V., ^{1,2}Alsalamah A.M., ^{1,3}Micholitsis A.G.
¹Peter the Great St. Petersburg Polytechnic University, St. Petersburg, e-mail: krasno_vv@spbstu.ru;
²Damascus University, Damascus, Syria, e-mail: alialsalameh12345@gmail.com;
³Aristotle University of Thessaloniki, Thessaloniki, Greece, e-mail: a_micholitsis@yahoo.gr

The authors continue to study the problems of the formation of probabilistic components of the cognitive style of university students. The authors analyze the reasons for the opposition of deterministic and probabilistic approaches to cognition, as well as the consequences of such opposition in the didactics of teaching probabilistic and statistical parts of the mathematics course for bachelor's programs. As one of the mechanisms for the formation of a probabilistic cognitive style, the authors propose to use tasks related to the analysis of the accuracy of probabilistic models. The authors carry out computational experiments to analyze the errors in calculating the probabilities of rare events using the exact Bernoulli formula and the approximate Poisson formula. The authors vary both the probability of occurrence of a rare event in one experiment and the number of experiments. Variations in the intensity of the flow of events remain common to all experiments, which makes it possible to analyze comparable results. The authors find absolute and relative computational errors for all cases, presenting information in a graphical form. The authors define the parameters of approximate calculations with high and unsatisfactory accuracy, thus marking the limits of applicability of approximate models. When conducting classes, the authors recommend focusing on the influence of the effect of the manifestation of round-off errors, which intensifies during operations with small values, which inevitably arise when analyzing rare events.

Keywords: probabilistic approach to scientific knowledge, teaching mathematics, probabilistic models, modeling accuracy, rare events

В настоящее время анализ точности вероятностных моделей получил широкое распространение в областях, связанных с проблемами Big Data: теории алгоритмов [1], проектировании машинного обучения [2, 3], исследовании социальных сетей [4], вычислительной механике [5], энергетике [6] и т.д. Авторы настоящей статьи показали, что представления о точности вероятностных моделей являются важным компонентом формирующегося у студентов вероят-

ностного подхода к научному познанию [7]. Также авторы доказательно утверждают, что наиболее естественным было формирование вероятностного подхода в ходе освоения студентами вероятностных и статистических курсов в рамках их математической подготовки [8]. В то же время в российской университетской среде до сих пор бытует отношение к вероятностному подходу как к чему-то второстепенному, вспомогательному. Причины такого отношения,

с одной стороны, известны, с другой – нуждаются в дальнейшем анализе, что и будет сделано в настоящей статье. Также в научно-методической литературе практически отсутствуют рекомендации по формированию представлений о точности вероятностных моделей, которые невозможно дать без вычислительных экспериментов, часть из которых будет представлена в результатах исследования.

Основной целью настоящего исследования является развитие методики формирования представлений студентов о точности вероятностных и статистических моделей.

Материалы и методы исследования

Задачи исследования:

Во-первых, авторы продолжают исследования по формированию вероятностного подхода студентов к научному познанию [9], одной из составляющих которого является представление учащихся о точности вероятностных моделей. В целом проблема остается дискуссионной. Необходим поиск новых аргументов в ее поддержку, поскольку именно сильная или слабая опора преподавателя на вероятностный подход определяет методику преподавания теории вероятностей и математической статистики в вузе.

Во-вторых, авторы ставят задачу усовершенствовать методику формирования компетенций студентов в сфере точности вероятностных моделей, которая востребована в научной и проектной практике. В работе [7] авторы исследовали следующие частные случаи сопоставления точных и приближенных моделей.

1. Проведено исследование изменения формы полигона биномиального распределения при увеличении числа опытов. Графическими средствами показано, что уже при 20 опытах полигон распределения несимметричного биномиального распределения близок к дискретизации соответствующего нормального распределения. Результаты были продемонстрированы студентам в весеннем семестре 2020/2021 учебного года, планируются к использованию в дальнейшем в курсе теории вероятности. Этот этап исследования можно считать завершенным.

2. Проведено сравнение вычисленных на интервалах значений вероятностей по точной формуле Бернулли и по приближенной интегральной теореме Лапласа. Построен график изменения относительной ошибки с ростом числа опытов, демонстрирующих ее убывание по обратной корневой зависимости. В то же время само значение ошибки нельзя признать удовлет-

ворительным. Результаты были продемонстрированы студентам в весеннем семестре 2020/2021 учебного года, планируется дальнейшее исследование вопроса.

3. Проведено сравнение вычисленных вероятностей редких событий по точной формуле Бернулли и приближенной формуле Пуассона. Визуальное сравнение полигонов не дает возможности заметить расхождение уже при 20 опытах. Таким образом, графики не являются наглядными для демонстрации в студенческой аудитории, что порождает необходимость в дополнительном исследовании, результаты которого представлены в настоящей статье.

4. Проведено сравнение вычисленных интервальных вероятностей редких событий по точной формуле Бернулли, по приближенной формуле Пуассона и по приближенной интегральной теореме Лапласа. При числе опытов, равном 100 и 1000, названные значения практически совпадают. С одной стороны, это дает студентам уверенность в возможности применения приближенных формул вместо точных при возникновении необходимости в подобных расчетах в профессиональной деятельности. С другой стороны, полученные результаты не дают представления о динамике ошибки в приближенных расчетах. Это обстоятельство и послужило одним из побудительных мотивов к настоящему исследованию.

Результаты исследования и их обсуждение

Прежде всего рассмотрим дискурс вероятностного подхода в российской науке и практике высшей школы. Существующее неоднозначное отношение к вероятностному подходу объясняется как объективными, внешними, так и субъективными, внутривузовскими причинами. Во-первых, в отечественной науке в 1930–1950-е гг. сложилось противопоставление детерминистского, одобряемого и вероятностного, «сомнительного» подхода к научному познанию [10]. Для преодоления этой антитезы, синтеза подходов в рамках единого системного потребовались годы, но устойчивые представления отчасти сохраняются («вероятностный детерминизм» [11]). Проявляется это противопоставление в вузовских курсах с учетом признанной инерционности образовательной системы. Во-вторых, в ряде отраслей, связанных с обороной и безопасностью, к вероятностным моделям относятся настороженно, поскольку цена даже одной человеческой жизни велика [12]. В то же время к вероятностным моделям в экономике и бизне-

се, где риски угрожают преимущественно финансовым активам, отношение всегда было более лояльным [13]. В-третьих, значительная часть вузовских преподавателей математики не уделяют должного внимания формированию вероятностного подхода у студентов, поскольку «не одобряют» эвристических оснований теории вероятностей и математической статистики по сравнению со строгими основаниями математического анализа и линейной алгебры. Люди, проникшие с детства красотой и гармонией логических построений, стараются «не замечать» хаотичности окружающего мира. Вероятность явно противоречит архетипическому представлению о математике как о точной науке. Такие преподаватели предпочитают «задерживаться» на изложении абстрактной теории вероятностей (комбинаторики и геометрических вероятностей), «не успевая» добраться до статистики, в том числе такого важнейшего стохастического инструмента, как регрессионный анализ. Разумеется, таких взглядов придерживаются далеко не все математики. В качестве наиболее ярких примеров, созвучных авторскому подходу к теории вероятности, можно назвать работы Н.П. Пучкова [14] и Г.Д. Гефана [15]. Ссылки на исследования других сторонников приближения вероятностных курсов к практико-ориентированным задачам можно найти в цитированных выше работах авторов.

Таким образом, формирование вероятностного подхода к научному познанию студентов не является авторской инновацией и находится в поле определенной полемики

со сторонниками современного и завуалированного детерминистского подхода. Не отвергая важности детерминистского подхода, авторы настаивают на его равноправии с вероятностным в рамках единого системного подхода к научному познанию.

Для поиска эффективных методических средств формирования у студентов представлений о точности вероятностных моделей был проведен ряд вычислительных экспериментов, для выполнения и анализа результатов которых в качестве соавторов были привлечены предмагистранты, закончившие бакалавриат в зарубежных университетах. Целью экспериментов был анализ точности приближенной формулы Пуассона для оценки вероятности наступления редких событий в зависимости от числа опытов.

Задача 1. Событие A может наступить в одном опыте с вероятностью $p = 0,005$. Сравнить вероятности неоявления события A : $P(X = 0)$, появления одного $P(X = 1)$, двух $P(X = 2)$, трех $P(X = 3)$ или четырех $P(X = 4)$ событий A в серии из n опытов, вычисленные по точной формуле Бернулли $P_B = P(X = i) = C_n^i p^i q^{n-i}$ и приближенной формуле Пуассона $P_p = P(X = i) = \frac{a^i}{i!} e^{-a}$, где $q = 1 - p = 0,995$, для n от 625 до 10000 с соответствующим изменением $a = np$ от 0,125 до 2,00.

Для оценки точности приближения Пуассона были вычислены абсолютная ошибка $\Delta = P_B - P_p$ и относительная ошибка $\varepsilon = \Delta / P_B$, графики изменения которых с ростом числа опытов приведены на рис. 1 и 2.

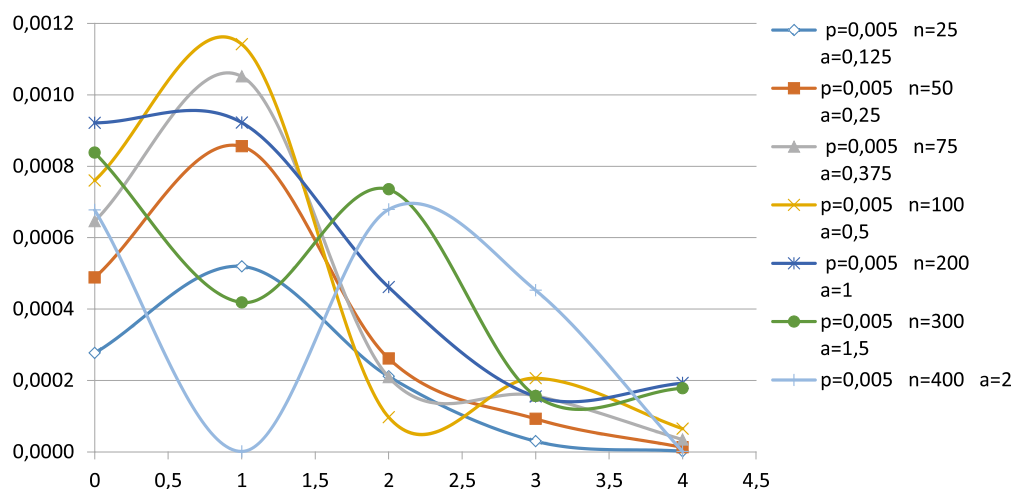


Рис. 1. Зависимость абсолютной ошибки Δ от числа опытов n в задаче 1

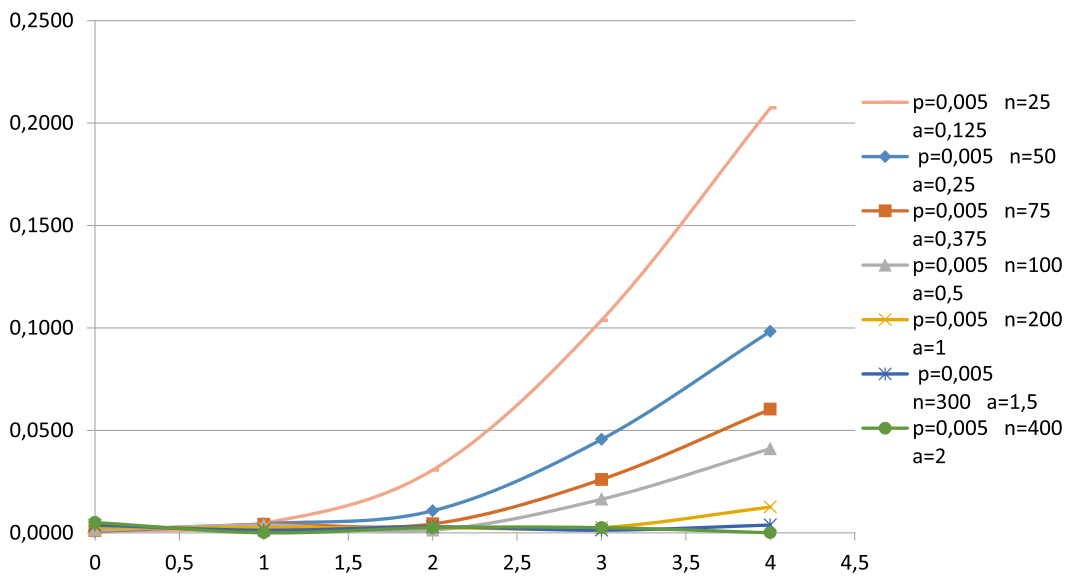


Рис. 2. Зависимость относительной ошибки ε от числа опытов n в задаче 1

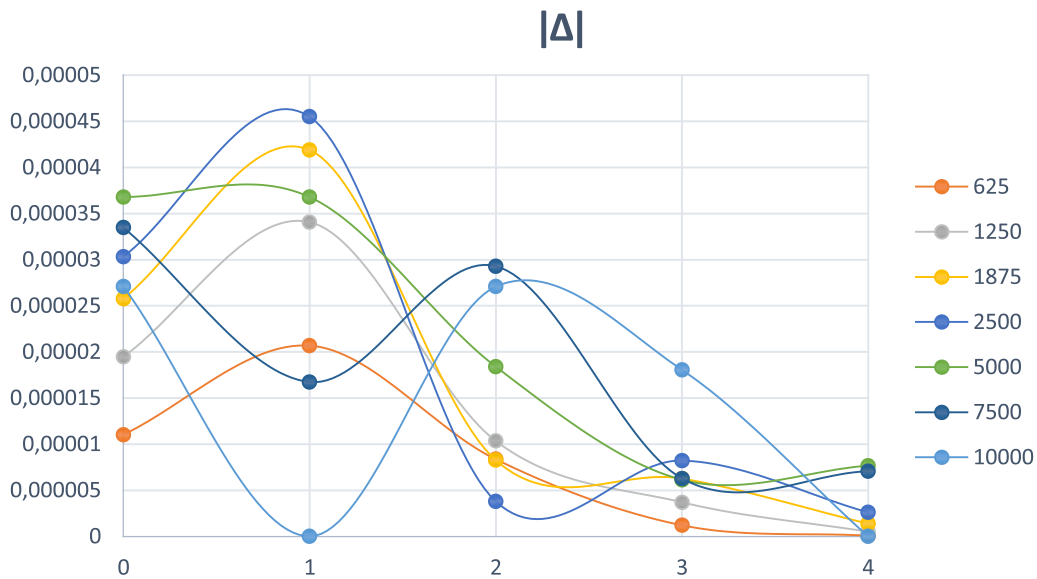


Рис. 3. Зависимость абсолютной ошибки Δ от числа опытов n в задаче 2

Задача 2. В условиях задачи 1 событие A может наступить в одном опыте с вероятностью $p = 0,0002$. Сравнить вероятности неоявления события A : $P(X = 0)$, появления одного $P(X = 1)$, двух $P(X = 2)$, трех $P(X = 3)$ или четырех $P(X = 4)$ событий A в серии из n опытов, вычисленных по точной формуле Бернулли и приближенной формуле Пуассона, с $q = 1 - p = 0,9998$, для n от 625 до 10000 с соответствующим

изменением $a = np$ от 0,125 до 2,00. Таким образом, интенсивности потока события, связанные со значениями параметра a , одинаковы для обеих задач.

Как и для задачи 1, для оценки точности приближения Пуассона были вычислены абсолютная ошибка $\Delta = P_B - P_P$ и относительная ошибка $\varepsilon = \Delta / P_B$, графики изменения которых с ростом числа опытов приведены на рис. 3 и 4.

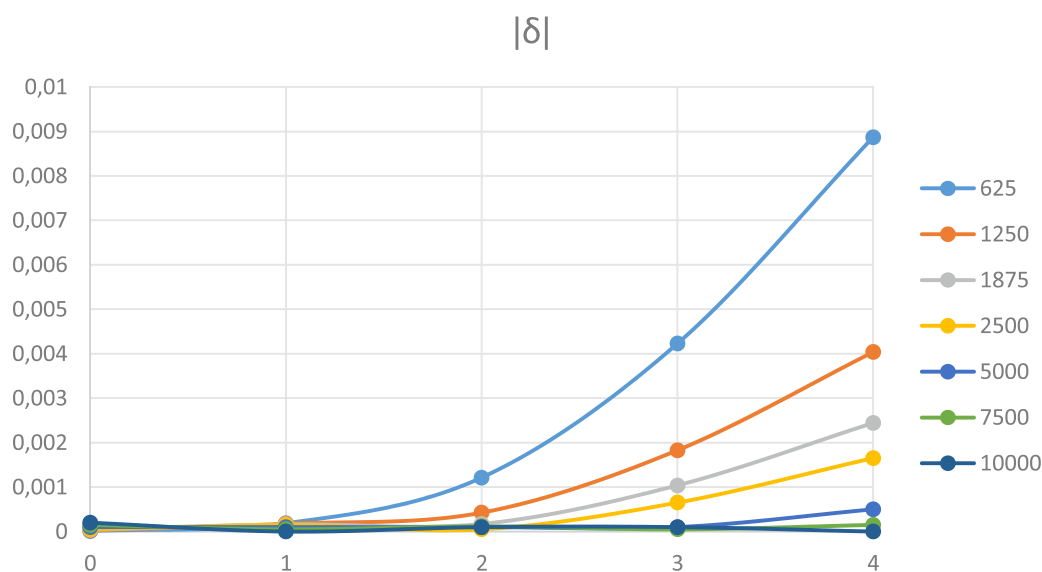


Рис. 4. Зависимость относительной ошибки ε от числа опытов n в задаче 2

Относительная ошибка вычисления вероятности появления редкого события в серии из n опытов для разных значений его вероятности p

$p = 0,005$	n	25	50	75	100	200	300	400
	$ \varepsilon $	0,0003	0,0006	0,0009	0,0013	0,0025	0,0038	0,0050
$p = 0,0002$	n	625	1250	1875	2500	5000	7500	10000
	$ \varepsilon $	0,00001	0,00002	0,00004	0,00005	0,0001	0,0002	0,0002

Для обеих задач очевидно, что с ростом числа опытов величина относительной ошибки ε снижается (рис. 2, 4). Тем не менее, сопоставляя результаты вычисления относительной ошибки для n от 25 до 400 (рис. 2) со случаем изменения n от 625 до 10 000 (рис. 4), можно отметить неудовлетворительную точность вычислений в первом случае (ошибка доходит до 20%). Разумеется, этот факт объясняется влиянием ошибок округления, которые при сравнении малых величин становятся сопоставимыми с ошибками приближения по формуле Пуассона. При вычислении же абсолютной ошибки Δ не требуется выполнять деление малых величин. Именно поэтому значения абсолютной ошибки в обоих случаях свидетельствуют о достаточно высокой точности вычисления – погрешность составляет не более 1% (рис. 1, 3). Однако осциллирующий характер абсолютной ошибки Δ (рис. 1, 3) не позволяет

сделать вывод о ее монотонном снижении с увеличением числа опытов, как в случае относительной ошибки ε (рис. 2, 4). В любом случае студентам следует учитывать возможное влияние ошибок округления при вычислении вероятностей, связанных с редкими событиями.

Следует отметить, что ведущим фактором точности вычислений является даже не общее число опытов, а величина вероятности появления редкого события. Результаты при $p = 0,0002$ (задача 2) получаются значительно более точными, чем при $p = 0,005$ (задача 1). Этот вывод подтверждается данными таблицы, представленными для случая $P(X=0)$ (таблица).

Вероятность появления редкого события $P(X=0)$ и связанная с ней вероятность появления хотя бы одного редкого события $P(X \geq 1) = 1 - P(X=0)$ являются важнейшими параметрами в теории надежности сложных систем, а также при моде-

лировании защиты от катастроф, эпидемий и других опасных и причиняющих вред явлений. Именно поэтому на значение точности вычисления вероятности неоявления разрушительных по последствиям событий следует обращать внимание студентов.

Заключение

Проведенное исследование позволило найти новые аргументы для обоснования важности формирования вероятностных компонентов научного мировоззрения студентов. Анализируя мнения явных и скрытых оппонентов вероятностного подхода, авторы пришли к выводу, что основу их позиции составляют представления об опасности «неточных» вероятностных методов при моделировании вредоносных явлений. Поэтому авторы обратили особое внимание на совершенствование механизмов формирования у студентов компетенций в области точности вероятностных и статистических моделей, границ применимости некоторых из них. Результаты вычислительных экспериментов показали достоинства и ограниченность использования приближенной формулы Пуассона вместо точной формулы Бернулли при вычислении вероятностей редких событий, к разряду которых как раз и относятся природные и техногенные катастрофы. Полученные рекомендации полезны тем студентам, чья профессиональная деятельность будет связана с построением математических моделей в технико-экономической сфере.

Список литературы

1. Cusumano-Towner M.F., Mansingka V.K. AIDE: An algorithm for measuring the accuracy of probabilistic inference algorithms. *Advances in Neural Information Processing Systems*. 30 (NIPS). MIT Press. 2017. P. 3004–3014.
2. Menon A., Jiang X., Vembu S., Elkan Ch., Ohno-Machado L. Predicting accurate probabilities with a ranking loss. *Proceedings of the 29th International Conference on Machine Learning, ICML 2012*. Edinburgh, UK. 2012. P. 703–710.
3. Pita R., Mendonça E., Reis S., Barreto M., Denaxas S. A Machine Learning Trainable Model to Assess the Accuracy of Probabilistic Record Linkage. In: Bellatreche L., Chakravarthy S. (eds) *Big Data Analytics and Knowledge Discovery, DaWaK 2017. Lecture Notes in Computer Science*. 2017. V. 10440. P. 214–227. Springer, Cham. DOI: 10.1007/978-3-319-64283-3_16.
4. García Díaz V., Espada J., Gonzalez Crespo R., Pelayo García-Bustelo B., Cueva Lovelle J. An Approach to Improve the Accuracy of Probabilistic Classifiers for Decision Support Systems in Sentiment Analysis. *Applied Soft Computing*. 2018. V. 67. P. 822–833. DOI: 10.1016/j.asoc.2017.05.038.
5. Chaskalovic J., Assous F. A New Mixed Functional-probabilistic Approach for Finite Element Accuracy. In: *Computational Methods in Applied Mathematics*. 2020. V. 20 (4). P. 799–813. DOI: 10.1515/cmam-2019-0089.
6. Chihota M.J., Gaunt C.T. Impact of Input Model Accuracy on Probabilistic Load Flow Outputs. In: *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. Boise, ID, USA. 2018. P. 1–6. DOI: 10.1109/PMAPS.2018.8440208.
7. 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. [Электронный ресурс]. URL: <https://science-education.ru/ru/article/view?id=30142> (дата обращения: 06.09.2021). DOI: 10.17513/spno.30142.
8. Krasnoshchekov V.V., Semenova N.V. 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, 2020. P. 139–149. DOI: 10.15405/epsbs.2020.12.03.14.
9. 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.
10. Лёвин В.Г. Вероятность как форма научного мышления. СПб.: Литео, 2016. 168 с.
11. Карак П.С. Детерминизм и целевая причина в системе современного научного знания // *Вестник Полесского государственного университета. Серия общественных и гуманитарных наук*. 2009. № 1. С. 9–16.
12. Букринский А.М. Детерминистское нормирование и вероятностное ориентирование // *Ядерная и радиационная безопасность*. 2013. № 1 (67). С. 3–6.
13. Карасев В.В., Соложенцев Е.Д. Гибридные логико-вероятностные модели для управления социально-экономической безопасностью // *Труды СПИИРАН*. 2016. № 48. С. 125–150. DOI: 10.15622/sp.48.7.
14. Пучков Н.П. Теория вероятностей и математическая статистика в системе политехнического образования. Тамбов: Изд-во Тамб. гос. техн. ун-та, 2017. 80 с.
15. Гефан Г.Д. Концепция теоретико-эмпирического дуализма в обучении математики // *Высшее образование в России*. 2020. Т. 29. № 4. С. 85–95.