ИНСТИТУТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ
И МАТЕМАТИЧЕСКОЙ ГЕОФИЗИКИ
СИБИРСКОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК

МАРЧУКОВСКИЕ НАУЧНЫЕ ЧТЕНИЯ – 2019

Труды

Международной конференции "Актуальные проблемы вычислительной и прикладной математики"

1–5 июля 2019 г.
Академгородок, Новосибирск, Россия

УДК 519.6
ББК 22.19
М30

М30 Марчуковские научные чтения - 2019 : Труды Международной конференции "Актуальные проблемы вычислительной и прикладной математики" / Ин-т вычислительной математики и матем. геофизики СО РАН. Новосибирск, 1–5 июля 2019 г. – Новосибирск : ИПЦ НГУ, 2019. - 587 с.

ISBN 978-5-901548-42-4

Целью Международной конференции "Актуальные проблемы вычислительной и прикладной математики" является привлечение специалистов по численному анализу, прикладной математике и вычислительным технологиям к обсуждению актуальных вопросов математики и математического моделирования, а также вопросов практического применения современных численных методов. Основные темы конференции: вычислительная алгебра и методы аппроксимации, численное решение дифференциальных уравнений, методы Монте-Карло и численное статистическое моделирование, математическое моделирование в задачах физики атмосферы, океана, климата и охраны окружающей среды, обратные задачи, математическое моделирование в задачах геофизики и электрофизики, математические модели и методы в науках о Земле, математическое моделирование в информационных технологиях, компьютерная биология.

Конференция проводится при поддержке

Новосибирского государственного университета
Министерства науки и высшего образования Российской Федерации
Сибирского отделения Российской академии наук
Института вычислительной математики им. Г. И. Марчука РАН
ФИЦ Институт цитологии и генетики СО РАН

Спонсор
ЗАО РСК Технологии

Информационная поддержка
Пресс-служба СО РАН

Сайт конференции: http://conf.nsc.ru/amca2019/ru

ISBN 978-5-901548-42-4

КРИТЕРИИ ПРОВЕРКИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ ПРИ АНАЛИЗЕ БОЛЬШИХ ВЫБОРОК: ПРОБЛЕМЫ И ИХ РЕШЕНИЕ

Б. Ю. Лемешко, С. Б. Лемешко, И. В. Веретельникова, П. Ю. Блинов[1]

[1] *Новосибирский государственный технический университет, 630073, Новосибирск*

УДК 519.24
DOI: 10.24411/9999-016A-2019-10044

В работе рассмотрены методы построения оценок при анализе больших данных (Big Data). Демонстрируется влияние на результаты выводов по критерию χ^2 Пирсона выбора числа интервалов и способа группирования. Показывается, как влияет на распределения статистик непараметрических критериев согласия ограниченная точность представления данных в больших выборках. Даются рекомендации по применению критериев для анализа больших выборок. Показано, что на распределения статистик критериев однородности влияет неравноточность представления данных в сравниваемых выборках.

Ключевые слова: Big Data, оценивание параметров, проверка гипотез, критерии согласия, критерии однородности, статистическое моделирование.

Введение

В связи с увеличением объёмов информации в последние годы резко возрос интерес к вопросам применения статистических методов для анализа больших массивов данных (Big Data). При попытках применения для анализа больших выборок классического аппарата прикладной математической статистики, как правило, сталкиваются со специфическими проблемами, ограничивающими возможности корректного применения этого аппарата.

В настоящей работе мы будем касаться только методов и критериев, связанных с анализом одномерных случайных величин, реальные проблемы которых нам наиболее знакомы. Можно рассмотреть, по крайней мере, три ситуации, где рост размерности выборок вызывает проблемы в применении методов или критериев.

Во-первых, вследствие "проклятия размерности" хорошо зарекомендовавшие себя методы и алгоритмы становятся неэффективными. При оценивании параметров моделей с ростом размерности анализируемых выборок кардинально растут вычислительные затраты, ухудшается сходимость итерационных алгоритмов, используемых при нахождении оценок. Естественным способом разрешения данной ситуации видится применение методов, предусматривающих группирование данных. В таком случае возникают вопросы, как использование оценок по группированным данным отразится на свойствах критериев проверки гипотез, в которых будут использоваться эти оценки [1]?

Во-вторых, многие критерии проверки статистических гипотез не приспособлены даже для анализа выборок порядка тысячи наблюдений, так как информация о распределениях статистик этих критериев при справедливости проверяемой гипотезы представлена лишь краткими таблицами критических значений. Возможность применения таких критериев при "разумных" величинах объёмов выборок n легко разрешается статистическим моделированием распределений статистик в интерактивном режиме в ходе статистического анализа [2].

В-третьих, замечено, что применение, например, непараметрических критериев согласия, для которых известны предельные (асимптотические) распределения статистик, с ростом объёмов выборок всегда приводит к отклонению даже справедливой проверяемой гипотезы. В [3] показано, что корни этой проблемы

Работа выполнена при поддержке Министерства образования и науки РФ в рамках государственной работы «Обеспечение проведения научных исследований» (№ 1.4574.2017/6.7) и проектной части государственного задания (№ 1.1009.2017/4.6).

ISBN 978-5-901548-42-4

связаны с ограниченной точностью представления анализируемых данных. В данном случае будет показано, что причина некорректности применения к большим выборкам критериев проверки гипотез об однородности кроется в неравноточности измерений в анализируемых выборках.

1 Оценивание параметров законов распределения

Оценки параметров законов могут находиться различными методами. Наилучшими асимптотическим свойствами обладают оценки максимального правдоподобия (ОМП), вычисляемые в результате максимизации функции правдоподобия

$$\hat{\theta} = arg\max_{\theta} \prod_{j=1}^{n} f(x_j, \theta), \tag{1}$$

или её логарифма, где θ — неизвестный вектор параметров, $f(x,\theta)$ — функция плотности закона распределения, $x_1, x_2, \ldots, x_n$ — выборка, по которой вычисляется оценка $\hat{\theta}$. ОМП вектора параметров в большинстве случаев находятся в результате использования некоторого итерационного метода.

При вычислении MD-оценок (оценок минимального расстояния) по θ минимизируется некоторая мера близости (расстояние) $\rho(F(x,\theta), F_n(x))$ между теоретическим $F(x,\theta)$ и эмпирическим $F_n(x)$ распределениями. MD-оценки находятся в процессе решения задачи

$$\hat{\theta} = arg\min_{\theta} \rho(F(x,\theta), F_n(x)), \tag{2}$$

В качестве мер близости можно использовать, например, статистики непараметрических критериев согласия (Колмогорова, Крамера–Мизеса–Смирнова, Андерсона–Дарлинга, Купера, Ватсона и других).

ОМП параметров законов распределения, как правило, не являются робастными. Наличие аномальных наблюдений или ошибочность предположения о виде закона приводят к построению моделей с функциями распределения, неприемлемо отклоняющимися от эмпирических распределений. MD-оценки обладают большей устойчивостью.

Очевидно, что при очень больших выборках вычисление оценок (1) и (2) связано с серьёзными вычислительными трудностями.

В случае группированной выборки имеющаяся в нашем распоряжении информация связана с множеством непересекающихся интервалов, которые делят область определения случайной величины на k непересекающихся интервалов граничными точками

$$x_{(0)} < x_{(1)} < \cdots < x_{(k-1)} < x_{(k)},$$

где $x_{(0)}$ — нижняя грань области определения случайной величины X; $x_{(k)}$ — верхняя грань области определения случайной величины X.

ОМП по группированной выборке вычисляется в результате максимизации функции правдоподобия

$$\hat{\theta} = arg\max_{\theta} \prod_{i=1}^{k} P_i^{n_i}(\theta), \tag{3}$$

где $P_i(\theta) = \int\limits_{x_{(i-1)}}^{x_{(i)}} f(x,\theta)dx$ — вероятность попадания наблюдения в i-й интервал значений, n_i — количество наблюдений, попавших в i-й интервал, $\sum\limits_{i=1}^{k} n_i = n$.

При наличии негруппированных данных к оценкам по группированным данным обращаются редко. Связано это с большей трудоёмкостью вычислительного процесса и необходимостью многократного использования численного интегрирования при вычислении $P_i(\theta)$ и требует соответствующей программной поддержки. В случае больших объёмов выборок ситуация меняется. При фиксированном числе интервалов группирования с ростом объёмов выборок вычислительные затраты не меняются, а возрастают только с увеличением количества интервалов k. Это свидетельствует о целесообразности в условиях Big Data использовать ОМП по группированным выборкам. Это асимптотически эффективные и, к тому же, робастные оценки. При малом k качество оценок можно улучшать, используя асимптотически оптимальное группирование (АОГ) [4], при котором минимизируются потери в информации Фишера, связанные с группированием.

2 Применение критерия χ^2 Пирсона к большим выборкам

Статистику критерия согласия χ^2 Пирсона вычисляют по формуле

$$X_n^2 = n \sum_{i=1}^{k} \frac{(n_i/n - P_i(\theta))^2}{P_i(\theta)} \tag{4}$$

В случае проверки простой гипотезы при $n \to \infty$ эта статистика подчиняется χ_r^2-распределению с $r = k-1$ степенями свободы, если верна нулевая гипотеза.

При проверке сложной гипотезы и оценивании по выборке m параметров закона статистика (4) в случае справедливости H_0 подчиняется χ_r^2-распределению с $r = k - m - 1$ степенями свободы, если оценки получаются минимизацией (4) этой статистики, или используются ОМП (3) (или другие асимптотически эффективные оценки по группированным данным).

При оценивании параметров по негруппированным данным распределение статистики (4) не подчиняется χ_{k-m-1}^2-распределению. При использовании ОМП по негруппированным данным рекомендуется применять критерий Никулина–Рао–Робсона [5,6]. Хотя, опираясь на статистическое моделирование и интерактивный режим [2] моделирования распределений статистики, не исключается использование критерия χ^2 Пирсона [7].

Принципиальные проблемы, препятствующие применению критерия χ^2 Пирсона для анализа Big Data, отсутствуют: возможны только вычислительные трудности.

Проиллюстрируем результаты применения критерия χ^2 Пирсона на примере достаточно большой выборки объёмом $n = 10^7$, смоделированной по стандартному нормальному закону $N(0,1)$.

В таблице 1 представлены результаты применения критерия при проверке простой гипотезы о принадлежности выборки закону $N(0,1)$ при различном числе интервалов в случае равночастотного группирования (РЧГ) и при $k = 15$ в случае (АОГ). При АОГ максимизируется мощность критерия χ^2 Пирсона относительно близких конкурирующих законов [8]. В таблице приведены значения X_n^{2*} статистики (4), вычисленные по выборке, и соответствующие значения достигнутого уровня значимости $p_{value} = P\{X_n^2 \geq X_n^{2*} | H_0\}$. Как можно видеть, результаты зависят как от способа разбиения, так и от числа интервалов. От этого же зависит и мощность критерия [9].

Таблица 1: Результаты проверки простой гипотезы

	АОГ	РЧГ						
	$k = 15$	$k = 15$	$k = 50$	$k = 75$	$k = 100$	$k = 500$	$k = 1000$	$k = 2000$
X_n^{2*}	7.75162	9.18380	56.8942	79.4904	96.5701	493.995	1044.57	2099.91
p_{value}	0.90186	0.81910	0.20475	0.31026	0.55038	0.55482	0.15403	0.05702

Таблица 2: Результаты проверки сложной гипотезы

	АОГ	РЧГ						
k	15	15	50	75	100	500	1000	2000
$\hat{\theta}_0$	0.000276	0.000301	0.0002440	0.000270	0.000268	0.000277	0.000273	0.000274
$\hat{\theta}_1$	1.007150	1.002629	1.001730	1.001338	1.001123	1.000399	1.000305	1.000236
X_n^{2*}	927.9202	99.99627	101.7669	104.5111	112.1514	493.7161	1043.471	2098.605
p_{value}	0.0	5.58e-16	6.50e-06	0.007396	0.139377	0.533166	0.149218	0.055723

В таблице 2 представлены результаты проверки сложных гипотез. Там приведены ОМП $\hat{\theta}_0$ параметра сдвига и $\hat{\theta}_1$ параметра масштаба нормального закона по группированным данным, полученные при соответствующем числе интервалов k, значения статистик X_n^{2*} и p_{value}.

ОМП параметров по полной негруппированной выборке $\hat{\theta}_0 = 0.000274$, $\hat{\theta}_1 = 1.000177$. В [7] построены модели распределений статистики (4) для случая проверки сложной гипотезы относительно нормального закона с использованием ОМП по негруппированным данным и применением АОГ. Вычисленное по выборке значение статистики $X_n^{2*} = 6.600521$ при $k = 15$, а полученная в соответствии с приведенной в [7] моделью предельного распределения оценка $p_{value} = 0.886707$, что свидетельствует о хорошем согласии полной выборки с нормальным законом $N(0.000274, 1.000177)$.

В случае анализа больших выборок [3] критерий χ^2 Пирсона демонстрирует как свои положительные качества [8], так и недостатки: зависимость результатов анализа (и мощности критерия) как от способа разбиения, так и от числа интервалов [9].

3 Непараметрические критерии согласия при больших выборках

Как правило, в Big Data объёмы выборок практически неограничены, но сами данные представлены с ограниченной точностью (округлены с некоторым $\triangle$). По сути, "нарушается предположение" о том, что наблюдается непрерывная случайная величина. Именно этот факт, как подробно показано в [3], является основной причиной возможной некорректности выводов при анализе больших данных с использованием непараметрических критериев согласия.

В данном случае мы ограничимся демонстрацией зависимости от $\triangle$ распределения статистики критерия согласия Крамера–Мизеса–Смирнова. Статистика критерия имеет вид

$$S_\omega = \frac{1}{12n} + \sum_{i=1}^{n} \left\{ F(x_i, \theta) - \frac{2i-1}{2n} \right\}^2, \tag{5}$$

и при проверке простой гипотезы в пределе подчиняется закону с функцией распределения $a1(s)$ [10].

Покажем зависимость распределения статистики от $\triangle$ для случая принадлежности выборок стандартному нормальному закону. При округлении с точностью до 1 в выборках, принадлежащих $N(0, 1)$, может появляться 9 уникальных значений, при округлении с точностью до $\triangle = 0.1$ — порядка 86 уникальных значений, с точностью $\triangle = 0.01$ — порядка 956, с точностью до $\triangle = 0.001$ — порядка 9830.

На рис. 1 представлена зависимость распределений статистики (5) критерия Крамера–Мизеса–Смирнова от степени округления $\triangle$ при объёмах выборок $n = 1000$ для случая проверки простой гипотезы о принадлежности выборки стандартному нормальному закону. На рисунке приведено предельное распределение $a1(S)$, а также реальные распределения $G(S_{1000}|H_0)$ статистики при степени округления $\triangle$ =0.01, 0.05, 0.1, 0.2, 0.3. Как можно видеть, при $\triangle = 0.01$ распределение $G(S_{1000}|H_0)$ практически не отличается от $a1(S)$, но с ростом $\triangle$ отклонение $G(S_{1000}|H_0)$ от $a1(S)$ быстро увеличивается.

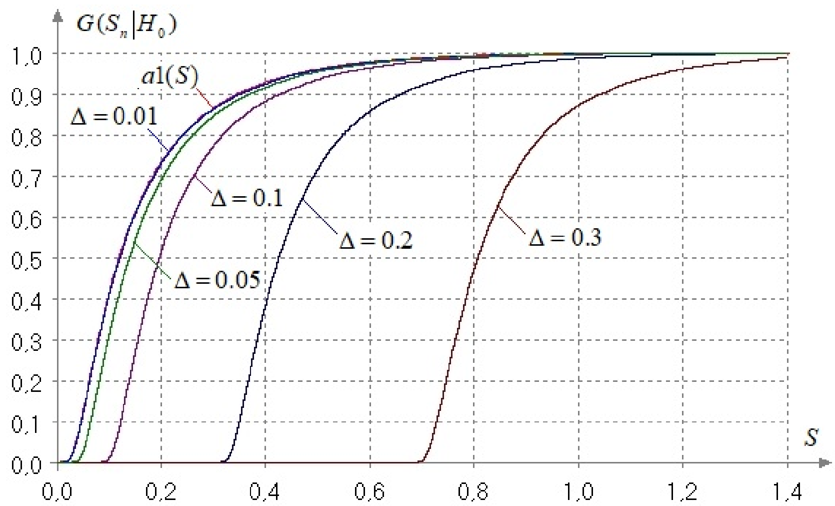

Рис. 1: Распределения статистики $G(S_n|H_0)$ критерия Крамера–Мизеса–Смирнова в зависимости от $\triangle$ при $n = 1000$

Таким же образом меняются распределения статистик непараметрических критериев согласия при проверке сложных гипотез вида H_0: $F(x) \in \Big\{F(x, \theta), \theta \in \Theta\Big\}$, где Θ — область определения параметра θ. Следует напомнить, что при проверке сложных гипотез уже нельзя использовать классические результаты [11], так как предельные распределения зависят от ряда факторов, связанных с оцениванием параметров [12]. И на всё это ещё накладывается влияние степени округления данных.

В работе [3] на основании результатов исследований сформулированы следующие рекомендации. Для того чтобы при анализе больших выборок с применением соответствующего непараметрического критерия согласия можно было использовать классические результаты, статистика должна вычисляться не по всему большому массиву, а по выборкам, извлекаемым по равномерному закону из той "генеральной совокупности", роль которой играет анализируемый большой массив данных. Объём извлекаемой выборки должен учитывать точность $\triangle$ и не превышать некоторой величины n_{max}, при которой (при данном $\triangle$) распределение статистики $G(S_{n_{max}}|H_0)$ критерия при справедливости H_0 ещё реально не отличается от предельного распределения $G(S|H_0)$ статистики этого критерия.

Вышесказанное в полной мере относится к применению к большим выборкам любых непараметрических критериев согласия.

Распределения статистик 3-х критериев согласия Жанга [13], представляющих собой развитие критериев Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга, зависят от объёмов выборок n. В этом случае не может идти речи об использовании предельных распределениях статистик. Но распределения статистик $G(S_n|H_0)$ этих критериев таким же образом зависят от от степени округления $\triangle$. Проблема может разрешаться статистическим моделированием (в том числе, в интерактивном режиме [2]) распределений статистик при заданных n и $\triangle$ при справедливости проверяемой гипотезы H_0. Именно так в подобной ситуации эта проблема разрешается в развиваемой программной системе ISW [17].

Отметим, что подобным же образом степень округления регистрируемых данных влияет на свойства множества других критериев, в частности, специальных критериев, ориентированных на проверку гипотезы о принадлежности выборок нормальному закону, о принадлежности выборок равномерному закону, показательному закону и др.

4 Критерии однородности при больших выборках

В критериях однородности, где сравнивается 2 и более выборок, на распределения статистик влияет неравноточность данных, представленных в выборках.

Двухвыборочный критерий однородности Лемана–Розенблатта предложен в работе [15] и исследован в [16]. Статистика, построенная по двум выборкам $x_{11}, x_{12}, \ldots, x_{1,n_1}$ и $x_{21}, x_{22}, \ldots, x_{2,n_2}$, используется в форме

$$S_{LR} = \frac{1}{n_1 n_2 (n_1 + n_2)} \left[n_1 \sum_{i=1}^{n_1} (r_i - i)^2 + n_2 \sum_{j=1}^{n_2} (s_j - j)^2 \right] - \frac{4 n_1 n_2 - 1}{6(n_1 + n_2)}, \tag{6}$$

где r_i — порядковый номер (ранг) x_{1i}; s_j — порядковый номер (ранг) x_{2j} в объединенном вариационном ряде. Предельным распределением статистики (6) при справедливости проверяемой гипотезы H_0 является то же самое распределение $a1(s)$, которое является предельным для статистики критерия согласия Крамера–Мизеса–Смирнова.

Рис. 2 демонстрирует зависимость распределения статистики $G(S_{LR}|H_0)$ критерия однородности Лемана–Розенблатта от степени округления $\triangle_2$ наблюдений во второй выборке при округлении в первой выборке $\triangle_1 = 0.01$ при объёмах выборок $n_i = 1000$. Уже при $\triangle_2 = 0.05$ отклонение $G(S_{LR}|H_0)$ от $a1(S)$ оказывается существенным. При фиксированном $\triangle_2$ с ростом объёмов выборок отклонение $G(S_{LR}|H_0)$ от $a1(S)$ быстро увеличивается. Отклонение увеличивается с ростом $\triangle_2$ и фиксированном объёме выборки. Распределения статистики $G(S_{LR}|H_0)$ критерия однородности Лемана–Розенблатта зависят также от разности $\triangle_1$ и $\triangle_2$.

Аналогичным образом от различия в точности регистрируемых данных в выборках зависят распределения статистик других критериев однородности, рассмотренных в [17].

Заключение

При построении вероятностных моделей по большим выборкам целесообразно использование методов оценивания параметров, предусматривающих группирование данных. В отличие от оценок по негруппированным данным такие оценки робастны, а вычислительные затраты не зависят от объёмов выборок.

Критерий χ^2 Пирсона при анализе больших выборок сохраняет как свои положительные качества, так и свойственные ему недостатки.

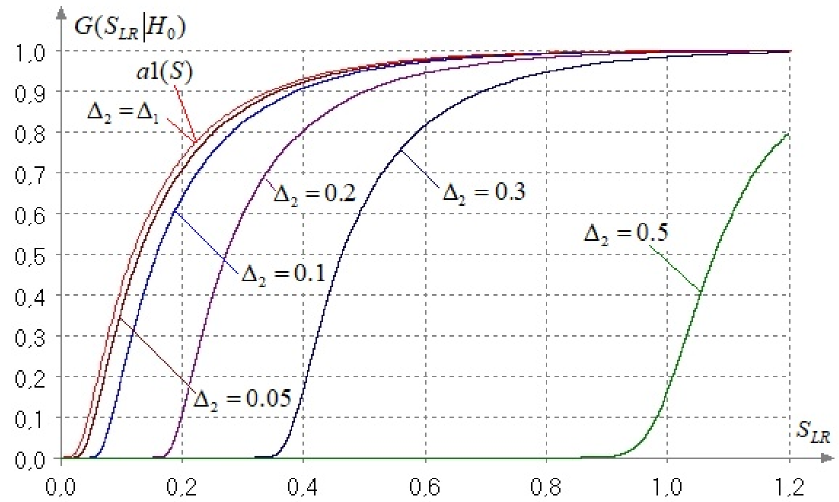

Рис. 2: Распределения статистики $G(S_{LR}|H_0)$ критерия однородности Лемана–Розенблатта при $n_i = 1000$ в зависимости от $\triangle_2$ при $\triangle_1 = 0.01$

Корректность применения непараметрических критериев согласия для анализа больших выборок можно обеспечить ограничением объёмов выборок, извлекаемых из больших совокупностей. Можно также использовать методы статистического моделирования для нахождения реальных распределений статистик (при соответствующих $\triangle$).

Возможную некорректность выводов при использовании критериев однородности, связанную с неравноточностью измерений в выборках, можно также устранить применением статистического моделирования для нахождения реальных распределений статистик.

Подобная стратегия действий при анализе больших выборок реализуется в программной системе ISW [14].

Список литературы

[1] Lemeshko B. Yu., Lemeshko S. B., Postovalov S. N. Statistic Distribution Models for Some Non-parametric Goodness-of-Fit Tests in Testing Composite Hypotheses // Communications in Statistics — Theory and Methods. 2010. Vol. 39, № 3. — P. 460-471.

[2] Lemeshko B. Yu., Lemeshko S. B., Rogozhnikov A. P. Interactive investigation of statistical regularities in testing composite hypotheses of goodness of fit // Statistical Models and Methods for Reliability and Survival Analysis : monograph. — Wiley-ISTE , 2013. — Chap. 5. — P. 61–76.

[3] Лемешко Б. Ю., Лемешко С. Б., Семёнова М. А. К вопросу статистического анализа больших данных // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2018. № 44. — С. 40-49. DOI: 10.17223/19988605/44/5

[4] Лемешко Б. Ю. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход /Б. Ю. Лемешко, С. Б. Лемешко, С. Н. Постовалов, Е. В. Чимитова. Новосибирск : Изд-во НГТУ, 2011. — 888 с.

[5] Никулин М. С. О критерии хи-квадрат для непрерывных распределений // Теория вероятностей и ее применение. 1973. Т. XVIII. № 3. — С.75-676.

[6] Rao K. C., Robson D. S. A chi-squared statistic for goodness-of-fit tests within the exponential family // Commun. Statist. 1974. Vol. 3. — P.1139-1153.

[7] Лемешко Б. Ю. Критерии проверки отклонения распределения от нормального закона. Руководство по применению. М.: ИНФРА-М, 2015. — 160 с. DOI: 10.12737/6086

[8] Денисов В. И., Лемешко Б. Ю. Оптимальное группирование при обработке экспериментальных данных // Измерительные информационные системы. Новосибирск, 1979. — С.5-14.

[9] Лемешко Б. Ю., Чимитова Е. В. О выборе числа интервалов в критериях согласия типа χ^2 // Заводская лаборатория. Диагностика материалов. 2003. Т. 69. № 1. С. 61-67.

[10] Большев Л. Н., Смирнов Н. В Таблицы математической статистики. — М. : Наука, 1983. — 416 с.

[11] Kac M., Kiefer J., Wolfowitz J. On tests of normality and other J. tests of goodness of fit based on distance methods // Ann. Math. Stat. 1955. Vol. 26. — P.189-211.

[12] Лемешко Б. Ю. Непараметрические критерии согласия: Руководство по применению. М.: ИНФРА-М, 2014. — 163 с. DOI: 10.12737/11873

[13] Zhang J. Powerful goodness-of-fit tests based on the likelihood ratio // Journal of the Royal Statistical Society: Series B. 2002. V.64. № 2. — P.281-294.

[14] ISW–Программная система статистического анализа одномерных наблюдений. https://ami.nstu.ru/ headrd/ISW.htm. (дата обр. 30.04.2019)

[15] Lehmann E. L. Consistency and unbiasedness of certain nonparametric tests // Ann. Math. Statist. — 1951. — Vol. 22, № 1. — P. 165–179.

[16] Rosenblatt M. Limit theorems associated with variants of the von Mises statistic // Ann. Math. Statist. — 1952. — Vol. 23. — P. 617–623.

[17] Лемешко Б. Ю. Критерии проверки гипотез об однородности. Руководство по применению. М.: ИНФРА-М, 2017. — 208 с. DOI: 10.12737/22368

Лемешко Борис Юрьевич — д.т.н., профессор, г.н.с. кафедры прикладной и теоретической информатики Новосибирского государственного технического университета,
e-mail: Lemeshko@ami.nstu.ru;
Лемешко Станислав Борисович — с.н.с., к.т.н., кафедры прикладной и теоретической информатики Новосибирского государственного технического университета,
e-mail: skyer@mail.ru;
Веретельникова Ирина Викторовна — м.н.с. кафедры прикладной и теоретической информатики Новосибирского государственного технического университета,
e-mail: ira-veterok@mail.ru;
Блинов Павел Юрьевич — м.н.с. кафедры прикладной и теоретической информатики Новосибирского государственного технического университета,
e-mail: Blindizer@yandex.ru

Дата поступления — 30 апреля 2019 г.