

ТЕХНОЛОГИЧЕСКИЕ ОСНОВЫ ПОВЫШЕНИЯ НАДЕЖНОСТИ И КАЧЕСТВА ИЗДЕЛИЙ

TECHNOLOGICAL BASES OF IMPROVING THE RELIABILITY AND QUALITY OF PRODUCTS

УДК 519.718.2

DOI 10.21685/2307-4205-2019-4-7

А. И. Иванов, Е. Н. Куприянов, К. Н. Савинов, А. Г. Банньих, А. В. Безяев

СОКРАЩЕНИЕ ТРЕБОВАНИЙ К ОБЪЕМУ ВЫБОРКИ ПРИ ПРОВЕРКЕ ГИПОТЕЗЫ НОРМАЛЬНОСТИ БИОМЕТРИЧЕСКИХ ДАННЫХ ЧЕРЕЗ НЕЙРОСЕТЕВОЕ ОБЪЕДИНЕНИЕ СЕМИ СТАТИСТИЧЕСКИХ КРИТЕРИЕВ

A. I. Ivanov, E. N. Kupriyanov, K. N. Savinov, A. G. Bannykh, A. V. Bezyaev

REDUCING SAMPLE SIZE REQUIREMENTS WHEN TESTING THE NORMALITY HYPOTHESIS BIOMETRIC DATA THROUGH A NEURAL NETWORK ASSOCIATION OF SEVEN STATISTICAL CRITERIA

Аннотация. Актуальность и цели. Целью работы является описание основных теоретических положений, позволяющих оценить доверительные вероятности совместных решений семи нейронов, воспроизводящих работу хи-квадрат критерия, двух критериев Крамера – Мизеса, двух вариантов критерия Андерсона – Дарлинга, критерия Шапиро – Уилка, критерия среднего геометрического. **Материалы и методы.** Каждому из существующих статистических критериев может быть поставлен в соответствие свой искусственный нейрон. Нейроны настраиваются так, чтобы ошибки первого и второго рода при разделении нормальных данных и данных с равномерным распределением оказывались равновероятными. **Результаты.** Заранее обученная сеть из семи искусственных нейронов дает фиктивный рост размеров тестовой выборки с 21 до 29 примеров, в сравнении с применением одного хи-квадрат нейрона. Если оставаться в рамках линейного прогнозирования, то увеличение числа искусственных нейронов до 70 должно приво-

Abstract. *Background.* The aim of the work is to describe the main theoretical provisions that allow us to estimate the confidence probabilities of joint solutions of 7 neurons that reproduce the work of the chi-square criterion, two Cramer-von Mises criteria, two versions of the Anderson-Darling criterion, the Shapiro-Wilk criterion, the geometric mean criterion. *Materials and methods.* Each of the existing statistical criteria can be assigned its own artificial neuron. Neurons are tuned so that errors of the first and second kind when separating normal data and data with a uniform distribution are equally probable. *Results.* A pre-trained network of 7 artificial neurons gives a fictitious increase in the size of the test sample from 21 to 29 examples, in comparison with the use of a single chi-square neuron. If we remain within the framework of linear forecasting, then an increase in the number of artificial neurons to 70 should lead to an equivalent increase in the initial sample to 101 experiments or an almost 5-fold decrease in the error probabilities from 0.32 to 0.06 for the chi-squared neuron. *Conclusions.* A neural network

доть к эквивалентному росту исходной выборки до 101 опыта или почти 5-кратному снижению вероятностей ошибок с 0,32 до значения 0,06 для хи-квадрат нейрона. **Выводы.** Нейросеть, обученная распознавать малые выборки нормальных данных, можно рассматривать как некоторую математическую молекулу, имеющую 128 выходных спектральных линий. Выигрыш от нейросетевого объединения нескольких статистических критериев обусловлен выполнением вычислений на пограничном переходе между непрерывным спектром вероятности и его аналогом – дискретным линейчатым спектром амплитуд вероятности малых выборок.

Ключевые слова: биометрические данные, надежность, статистическая выборка, моделирование, критерий, нейросеть.

trained to recognize small samples of normal data can be considered as some mathematical molecule with 128 output spectral lines. The gain from the neural network of several statistical criteria is due to the calculations at the boundary transition between the continuous probability spectrum and its analogue, the discrete line spectrum of the probability amplitudes of small samples.

Keywords: biometric data, reliability, statistical sampling, modeling, criterion, neural network.

Проблема малых выборок при статистической проверке гипотезы нормальности

При работе с реальными данными биометрии, биологии, медицины, экономики, социологии зачастую выборки оказываются малыми. Так, стандарт России ГОСТ Р 52633.5–2011 ориентирует при автоматическом обучении искусственных нейронных сетей на выборке от 16 до 20 примеров биометрического образа «Свой» [1]. Пользователи не хотят при обучении нейронной сети утруждать себя предъявлением 200 примеров рисунка отпечатка пальца, они считают систему биометрической защиты эргономичной, если ее сеть искусственных нейронов обучается на 20 примерах.

Совершенно такая же ситуация возникает, когда на подопытных кроликах проверяют действие нового лекарства. Потратить деньги на 20 подопытных кроликов и на 200 подопытных кроликов не одно и то же. Более того, очень часто речь идет о значительных затратах времени. Так, если биолог вывел новую породу кроликов и на текущий момент имеет 20 кроликов новой породы, то получить 200 кроликов он сможет, вырастив их только через 5–6 месяцев.

Как правило, ускоренная статистическая обработка данных опирается на гипотезу их нормальности. Стандартные рекомендации [2, 3] при проверке статистических гипотез дают приемлемый результат с доверительной вероятностью 0,94 на выборках в 200 и более примеров.

Ярким примером рассматриваемой ситуации является использование при оценках хи-квадрат критерия [2]. Для того, чтобы показать возможности этого критерия, на малых выборках сформируем некоторый квадратичный нейрон, эквивалентный хи-квадрат критерию на выборке в 21 опыт. Функциональные связи его элементов приведены в табл. 1.

Таблица 1

Хи-квадрат нейрон для выборки в 21 опыт

$\left\{ \begin{array}{l} x \leftarrow \text{sort}(x), \\ \Delta \leftarrow \frac{x_{20} - x_0}{5}, \\ \tilde{x}_i \leftarrow x_0 + \Delta \cdot i, \quad i = 0, 1, \dots, 5, \\ \chi^2 \leftarrow 21 \cdot \sum_{i=0}^4 \frac{\left(\frac{n_i}{21} - (P(\tilde{x}_{i+1}) - P(\tilde{x}_i)) \right)^2}{P(\tilde{x}_{i+1}) - P(\tilde{x}_i)}, \\ z(\chi^2) \leftarrow "0" \text{ if } \chi^2 \leq 4,2, \\ z(\chi^2) \leftarrow "1" \text{ if } \chi^2 > 4,2, \\ P_1 \approx P_2 \approx P_{EE} \approx 0,309 \end{array} \right.$	где Δ – ширина интервала гистограммы; n_i – число отсчетов, попавших в i-й интервал гистограммы; $P(\tilde{x}_i)$ – теоретическая вероятность для нормального распределения данных; P_1 – вероятность ошибки первого рода; P_2 – вероятность ошибки второго рода
---	--

Очевидно, что нетрудно написать программу, моделирующую поведение хи-квадрат нейрона на малых выборках в 21 опыт. Результаты работы такой программы представлены на рис. 1.



Рис. 1. Плотности распределения значений хи-квадрат критерия для нормальных и равномерных входных данных объемом выборки в 21 опыт (проверка нормальности, нейроном табл. 1)

Из рис. 1 видно, что распределения хи-квадрат отклика для нормальных данных и данных с равномерным распределением существенно отличаются. Настройка порогового элемента нейрона (выходного квантователя нейрона) выполняется таким образом, чтобы вероятности ошибок первого и второго рода оказывались одинаковыми $P_1 = P_2 = P_{EE} = 0,309$. Площади, соответствующие ошибкам первого и второго рода хи-квадрат нейрона, на рис. 1 помечены более темной заливкой.

Вполне очевидно, что доверительная вероятность верных решений – 0,691 недостаточно. Попробуем повысить убедительность оценок, увеличив объем выборки до 29 опытов. При этом программная реализация хи-квадрат нейрона изменяется незначительно, функциональные связи, его описывающие, приведены в табл. 2.

Таблица 2

Хи-квадрат нейрон для выборки в 29 опытов

$\left\{ \begin{array}{l} x \leftarrow \text{sort}(x), \\ \Delta \leftarrow \frac{x_{20} - x_0}{5}, \\ \tilde{x}_i \leftarrow x_0 + \Delta \cdot i, \quad i = 0, 1, \dots, 5, \\ \chi^2 \leftarrow 29 \cdot \sum_{i=0}^4 \frac{\left(\frac{n_i}{29} - (P(\tilde{x}_{i+1}) - P(\tilde{x}_i)) \right)^2}{P(\tilde{x}_{i+1}) - P(\tilde{x}_i)}, \\ z(\chi^2) \leftarrow "0" \text{ if } \chi^2 \leq 4,6, \\ z(\chi^2) \leftarrow "1" \text{ if } \chi^2 > 4,6, \\ P_1 \approx P_2 \approx P_{EE} \approx 0,231 \end{array} \right.$	Коэффициенты взаимной корреляции $\begin{aligned} \text{corr}(\chi^2, \omega^2) &= 0,542, \\ \text{corr}(\chi^2, \omega_c^2) &= 0,424, \\ \text{corr}(\chi^2, a^2) &= 0,313, \\ \text{corr}(\chi^2, al) &= 0,632, \\ \text{corr}(\chi^2, ds^2) &= -0,104, \\ \text{corr}(\omega^2, a^2) &= 0,280, \\ \text{corr}(\omega^2, al) &= 0,863 \end{aligned}$
---	---

Увеличение выборки примерно на одну треть приводит к снижению на одну треть вероятностей появления ошибок первого и второго рода. Можно попытаться показать, что на выборках, увеличенных в 10–20 раз, вероятности ошибок первого и второго рода становятся приемлемыми для практики.

Снижение вероятностей ошибок первого и второго рода за счет совместного использования нескольких статистических критериев

В связи с проблемой малых выборок реальных данных возникает соблазн применить сразу несколько статистических критериев для анализа одной и той же выборки. Для этой цели построим

нейросетевой аналог статистического критерия Шапиро – Уилка [4–7]. Функциональные связи этой программной конструкции описываются табл. 3.

Таблица 3

Нейрон Шапиро – Уилка, 21 опыт

$\left\{ \begin{array}{l} x \leftarrow \text{sort}(x), \\ v^2 \leftarrow \frac{1}{\sigma(x)} \cdot \left\{ \sum_{i=0}^9 a_i \cdot (x_{20-i} - x_i) \right\}^2, \\ z(v^2) \leftarrow "0" \text{ if } v^2 \geq 20,44, \\ z(v^2) \leftarrow "1" \text{ if } v^2 < 20,44, \\ P_1 \approx P_2 \approx P_{EE} \approx 0,301 \end{array} \right.$	где a_i – коэффициенты Шапиро – Уилка ($a_0 = 0,4834, a_1 = 0,3185, a_2 = 0,2575, a_3 = 0,2119, a_4 = 0,1736, a_5 = 0,1399, a_6 = 0,1039, a_7 = 0,0804, a_8 = 0,0530, a_9 = 0,0263$); $\sigma(x)$ – стандартное отклонение малой выборки в 21 опыт
--	--

В табл. 4 даны функциональные связи нейрона Крамера – Мизеса. Для этой математической конструкции в правой части табл. 4 даны значения корреляционных связей хи-квадрат нейрона и других нейронов с данными нейрона Крамера – Мизеса. Тот факт, что эти корреляционные связи существенно меньше единицы, служит предпосылкой совместного применения рассматриваемых статистических критериев.

Таблица 4

Нейрон Крамера – Мизеса, 21 опыт

$\left\{ \begin{array}{l} x \leftarrow \text{sort}(x), \\ \omega^2 \leftarrow \sum_{i=0}^{19} (i+1-21 \cdot P(x_i))^2 \cdot \frac{21 \cdot (x_{i+1} - x_i)}{x_{20} - x_0}, \\ z(\omega^2) \leftarrow "0" \text{ if } \omega^2 \leq 31,1, \\ z(\omega^2) \leftarrow "1" \text{ if } \omega^2 > 31,1, \\ P_1 \approx P_2 \approx P_{EE} \approx 0,359 \end{array} \right.$	Примечание: коррелированность критериев составляет: $\text{corr}(\omega^2, \chi^2) = 0,556$ $\text{corr}(\omega^2, v^2) = -0,779$ $\text{corr}(\omega^2, a^2) = 0,226$ $\text{corr}(\omega^2, al) = 0,827$
--	---

Для критерия проверки статистических гипотез Смирнова – Крамера – Мизеса функциональные связи даны в табл. 5. Сравнивая данные по корреляции табл. 4 и 5, легко заметить, что оба нейрона семейства Крамера – Мизеса не только различаются по их мощности, но и имеют разные коэффициенты корреляции с другими классическими математическими конструкциями.

Таблица 5

Нейрон Смирнова – Крамера – Мизеса, 21 опыт

$\left\{ \begin{array}{l} x \leftarrow \text{sort}(x), \\ \omega_c^2 \leftarrow \sum_{i=0}^{20} (i+1-21 \cdot P(x_i))^2, \\ z(\omega_c^2) \leftarrow "0" \text{ if } \omega_c^2 \leq 38,1, \\ z(\omega_c^2) \leftarrow "1" \text{ if } \omega_c^2 > 38,1, \\ P_1 \approx P_2 \approx P_{EE} \approx 0,337 \end{array} \right.$	Примечание: коррелированность критериев составляет: $\text{corr}(\omega_c^2, \chi^2) = 0,401,$ $\text{corr}(\omega_c^2, \omega^2) = 0,885,$ $\text{corr}(\omega_c^2, a^2) = 0,393$
--	---

Аналогичная ситуация возникает и для нейронов Андерсона – Дарлинга (табл. 6, 7). И в этом случае статистические критерии, принадлежащие к одному семейству, имеют разные корреляционные связи с другими классическими статистическими критериями.

Таблица 6

Нейрон Андерсона – Дарлингa, 21 опыт

$\left\{ \begin{array}{l} x \leftarrow \text{sort}(x), \\ a^2 \leftarrow \sum_{i=0}^{20} \frac{(i+1-21 \cdot P(x_i))^2}{P(x_i) \cdot (1-P(x_i))}, \\ z(a^2) \leftarrow "0" \text{ if } a^2 \leq 241, \\ z(a^2) \leftarrow "1" \text{ if } a^2 > 241, \\ P_1 \approx P_2 \approx P_{EE} \approx 0,336 \end{array} \right.$	Примечание: $\text{corr}(a^2, \chi^2) = 0,424,$ $\text{corr}(a^2, \omega^2) = 0,226,$ $\text{corr}(a^2, \omega_c^2) = 0,393,$ $\text{corr}(a^2, v^2) = -0,614$
---	--

Таблица 7

Логарифмический нейрон Андерсона – Дарлингa, 21 опыт

$\left\{ \begin{array}{l} x \leftarrow \text{sort}(x), \\ al \leftarrow -\sum_{i=0}^{20} \left[\frac{i+1}{21} \cdot \ln(P(x_i)) - \left(1 - \frac{i+1}{21} \right) \cdot \ln(1-P(x_i)) \right], \\ z(al^2) \leftarrow "0" \text{ if } al \leq 10,72, \\ z(al^2) \leftarrow "1" \text{ if } al > 10,72, \\ P_1 \approx P_2 \approx P_{EE} \approx 0,321 \end{array} \right.$	Примечание: $\text{corr}(a^2, al) = 0,644,$ $\text{corr}(\omega^2, al) = 0,827,$ $\text{corr}(\omega_c^2, al) = 0,832,$ $\text{corr}(\chi^2, al) = 0,672,$ $\text{corr}(a^2, al) = 0,628,$ $\text{corr}(v^2, al) = -0,678$
---	--

Мы знаем, что в XX в. математиками были созданы десятки статистических критериев. При этом математики, занимающиеся этой веткой исследований, стремились синтезировать статистические критерии, которые сами по себе имели высокую мощность (низкое значение вероятностей ошибок первого и второго рода). Предположительно, множество статистических критериев было отброшено из-за того, что они имели низкий показатель мощности по отношению к уже найденным (опубликованным) статистическим критериям.

Из приведенных выше таблиц следует, что математические исследования XXI в. должны вестись иначе. Необходимо искать новые статистические критерии, руководствуясь не только их мощностью, при совместном использовании нескольких статистических критериев на первый план выходит уровень коррелированности вновь синтезированного статистического критерия и его давно созданных аналогов.

В этом отношении значительный интерес представляет семейство статистических критериев среднего геометрического [8–11]. Нейрон, воспроизводящий один из критериев этого семейства, описывается связями, приведенными в табл. 8.

Таблица 8

Нейрон интегро-дифференциального квадрата среднего геометрического выборка в 21 опыт

$\left\{ \begin{array}{l} x \leftarrow \text{sort}(x), \\ ds^2 \leftarrow \sum_{i=0}^{20} \left(\sqrt{\frac{(i+1) \cdot p(x_i)}{21}} \right)^2, \\ z(ds^2) \leftarrow "0" \text{ if } ds^2 \geq 7,42, \\ z(ds^2) \leftarrow "1" \text{ if } ds^2 < 7,42, \\ P_1 \approx P_2 \approx P_{EE} \approx 0,265 \end{array} \right.$	где $p(x) = \frac{\partial P(x)}{\partial x}$ – плотность распределения значений нормального закона; $\text{corr}(\chi^2, ds^2) = -0,124,$ $\text{corr}(\omega^2, ds^2) = -0,144,$ $\text{corr}(\omega_c^2, ds^2) = -0,176,$ $\text{corr}(v^2, ds^2) = -0,074$
--	--

Из табл. 8 видно, что для нового статистического критерия корреляционные связи с другими известными статистическими критериями оказываются отрицательными и рекордно низкими. Это крайне важно для нейросетевых обобщений статистических критериев.

Корректировка ошибок выходного кода нейросетевого объединения в пространстве расстояний Хэмминга

Если анализируемую малую выборку в 21 пример биометрических данных подать на семь описанных выше нейронов, то каждая выборка будет давать 7 бит выходного кода. В случае, когда все семь нейронов будут давать состояния «0», следует считать, что с высокой достоверностью входная выборка имеет нормальный закон распределения. Напротив, если выходной код будет состоять из всех единиц, то с высокой вероятностью анализируемые данные будут иметь равномерный закон распределения.

Очевидным является также то, что появление в выходном коде одной единицы следует рассматривать как обнаружение нормального распределения входных данных. Единичное состояние одного разряда следует рассматривать как ошибку кода, которая должна быть исправлена.

Чисто формально для исправления ошибок может быть использован любой классический код, способный обнаруживать и корректировать ошибки [12]. В данной статье мы не будем касаться проблем применения достаточно сложных кодов с обнаружением и исправлением ошибок. В первом приближении вполне достаточно применения самых простых процедур коррекции, построенных на вычислении расстояния Хэмминга от идеального кода «0000000». Все коды, имеющие 1, 2, 3 разряда с состоянием «1», рассматриваются как факт обнаружения нормального закона. Коды, имеющие 4 и более состояний «1», рассматриваются как обнаружение равномерного распределения значений малой выборки.

Этот столь примитивный корректор позволяет рассматриваемой математической конструкции снизить значение равновероятной ошибки до величины $P_1 = P_2 = P_{\text{EE}} = 0,234$, что примерно на 30 % меньше среднего значения вероятностей ошибок первого и второго рода всех семи использованных нейронов. Последнее эквивалентно тому, что в случае применения одного хи-квадрат нейрона объем анализируемой выборки должен был вырасти с 21 примера до 29 примеров (отметим, что дополнительно появились восемь примеров в малой выборке). Именно для иллюстрации этого эффекта выше была приведена табл. 2 с данными хи-квадрат нейрона для выборки в 29 опытов. Получается, что алгоритмическое усложнение статистической обработки данных примерно в 7 раз приводит к появлению в исходной выборке дополнительно восьми примеров.

Если оставаться в рамках простейшей линейной модели прогнозирования, то усложнение статистической обработки в 70 раз (применение сети из 70 нейронов) должно приводить к появлению ниоткуда в тестовой выборке 80 дополнительных примеров, т.е. эквивалентная выборка должна составить 101 опыт вместо 21. Доверительная вероятность для хи-квадрат нейрона с пятью равными интервалами гистограммы при выборке 101 опыт составляет величину 0,94 (см. табл. 2).

Перспективы перехода от статистического анализа непрерывных спектров к статистическому анализу дискретных выходных спектров малых выборок

Все описанное выше кажется обманом и «фокусами», однако все это имеет вполне разумное обоснование, которым уже давно пользуются квантовые химики, квантовые физики, квантовые математики и просто криминалисты. Одним из самых ярких примеров является именно практика криминалистов. Для того, чтобы обнаружить факт взвешивания золота на чаше весов, необходимо протереть чашу весов ватой, смоченной спиртом. Потом вату сжигают, а пламя фотографируют. Просто фотография пламени ничего не дает (концентрация золота ничтожна), однако если свет пламени разложить призмой и сфотографировать спектральные линии, то золото обнаруживается. Происходит это потому, что спектральные линии горящих ваты и спирта не совпадают со спектральными линиями золота. Спектральные линии золота попадают в темные полосы между спектральными линиями горячей органики и становятся видны. Анализ спектральных линий излучения пламени оказывается примерно в миллиард раз чувствительнее к обнаружению примесей, чем анализ непрерывного спектра пламени. При этом обязательно нужно знать, что мы хотим обнаружить (спектр какого вещества нас интересует).

Примерно такая же ситуация возникает и в статистике, например, мы можем от обычного непрерывного спектра хи-квадрат критерия (см. рис. 1) с помощью очень простых преобразований [13–15] перейти к наблюдению дискретного выходного спектра хи-квадрат молекулы для нормальной малой выборки. Фактически преобразования, описанные в работах [13–15], есть не что иное, как реализация эквивалента оптической призмы для статистики, превращающей непрерывные спектры рис. 1 в эквивалентные дискретные спектральные линии.

Описанный в данной статье положительный эффект возникает из-за того, что выполняемые преобразования находятся на границе перехода от непрерывных спектральных представлений к дискретным спектральным представлениям. Конструкция из семи нейронов есть не что иное, как нейросетевая молекула, возбуждаемая данными малых выборок. При этом поле выходных кодовых состояний нейросетевой молекулы малых выборок с нормальным распределением конечно. Мы имеем ровно 128 кодовых состояний, каждое из этих 128 кодовых состояний соответствует номеру одной спектральной линии. Каждая линия имеет собственную амплитуду вероятности для непрерывного потока входных нормальных данных. Все разряды кодовых состояний сцеплены между собой (коэффициенты корреляции между разрядами или коэффициенты сцепленности разрядов приведены в таблицах). Именно операции на границе перехода от непрерывных спектров к дискретным линейчатым спектрам позволяют повышать достоверность вычислений за счет увеличения из сложности. По сути дела, это не что иное, как одна из форм реализации квантовых эффектов применительно к статистической обработке данных.

Заключение

В XX в. математиками были созданы десятки статистических критериев. При этом поиск новых критериев велся исходя из повышения их мощности для тех или иных граничных условий. Естественно, что вопрос об использовании сразу всех известных статистических критериев серьезно не рассматривался из-за отсутствия соответствующей вычислительной техники.

В XXI в. ситуация изменилась, ограничений на возможности вычислительной техники нет. Сегодня мы можем анализировать данные с использованием сотен известных и новых статистических критериев. В данной статье мы попытались показать, что каждому из известных и новых статистических критериев может быть поставлен в соответствие свой искусственный нейрон. Простое параллельное объединение нейронов может давать достаточно длинные коды. Ошибки в этих кодах могут быть скорректированы, что позволяет значительно снизить требования к размерам анализируемых выборок.

Предположительно в XXI в. акценты по методам разработки новых статистических критериев изменятся. Если в прошлом веке математики старались повышать мощность статистических критериев, то в этом веке от новых статистических критериев дополнительно требуется минимизация корреляционных связей с ранее созданными статистическими критериями.

Библиографический список

1. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа. – Москва : Госстандарт России, 2011.
2. Р 50.1.037–2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть I. Критерии типа χ^2 . – Москва : Госстандарт России, 2001. – 140 с.
3. Р 50.1.037–2002. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть II. Непараметрические критерии. – Москва : Госстандарт России, 2002. – 123 с.
4. Кобзарь, А. И. Прикладная математическая статистика. Для инженеров и научных работников / А. И. Кобзарь. – Москва : ФИЗМАТЛИТ, 2006. – 816 с.
5. Иванов, А. И. Прецизионная статистика: нейросетевое объединение хи-квадрат критерия и критерия Шапиро-Уилка при анализе малых выборок биометрических данных / А. И. Иванов, С. Е. Вятчанин, Е. А. Малыгина, В. С. Лукин // Труды Международного симпозиума Надежность и качество. – 2019. – Т. 2. – С. 131–134.
6. Волчихин, В. И. Нейросетевой анализ малых выборок биометрических данных с использованием хи-квадрат критерия и критериев Андерсона–Дарлинга / В. И. Волчихин, А. И. Иванов, А. В. Безяев, Е. Н. Куприянов // Инженерные технологии и системы. – 2019. – Т. 29, № 2. – С. 205–217. – DOI <https://doi.org/10.15507/2658-4123.029/2019.02.205-217>

7. Иванов, А. И. Нейросетевое обобщение классических статистических критериев для обработки малых выборок биометрических данных / А. И. Иванов, Е. Н. Куприянов, С. В. Туреев // Надежность. – 2019. – № 2. – С. 22–27. – DOI 10.21683/1729-2646-2019-19-2-22-27.
8. Иванов, А. И. Оценка качества малых выборок биометрических данных с использованием дифференциального варианта статистического критерия среднего геометрического / А. И. Иванов, К. А. Перфилов, Е. А. Малыгина // Вестник Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. – 2016. – № 4 (17). – С. 864–871.
9. Перфилов, К. А. Расширение многообразия статистических критериев, используемых при проверке гипотез распределения значений биометрических данных / К. А. Перфилов, А. И. Иванов, Е. Д. Проценко // Европейский союз ученых. – 2015. – № 13, ч. 5. – С. 9–12.
10. Иванов, А. И. Оценка соотношения мощностей семейства статистических критериев «среднего геометрического» на малых выборках биометрических данных / А. И. Иванов, К. А. Перфилов // Современные охраняемые технологии и средства обеспечения комплексной безопасности объектов : XI Всерос. науч.-практ. конф. (Пенза, Заречный, 20 октября 2016 г.). – Пенза, 2016. – С. 223–229.
11. Иванов, А. И. Многомерный статистический анализ качества биометрических данных на предельно малых выборках с использованием критериев среднего геометрического, вычисленного для анализируемых функций вероятности / А. И. Иванов, К. А. Перфилов, Е. А. Малыгина // Измерение. Мониторинг. Управление. Контроль. – 2016. – № 2 (16). – С. 58–66.
12. Морелос-Сарагоса, Р. Искусство помехоустойчивого кодирования / Р. Морелос-Сарагоса. – Москва : Техносфера, 2007. – 320 с.
13. Ахметов, Б. Б. Дискретный характер закона распределения хи-квадрат критерия для малых тестовых выборок / Б. Б. Ахметов, А. И. Иванов, Н. И. Серикова, Ю. В. Фунтикова // Вестник Национальной академии наук Республики Казахстан. – 2015. – № 1. – С. 17–25.
14. Кулагин, В. Циклические континуально-квантовые вычисления: усиление мощности хи-квадрат критерия на малых выборках / В. Кулагин, А. Иванов, А. Газин, Б. Ахметов // Аналитика. – 2016. – № 5 (30). – С. 22–29.
15. Перспективы создания циклической континуально-квантовой хи-квадрат машины для проверки статистических гипотез на малых выборках биометрических данных и данных иной природы / В. И. Волчихин, А. И. Иванов, Д. В. Пашенко, Б. Б. Ахметов, С. Е. Вятчанин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2017. – № 1. – С. 5–15.

References

1. GOST R 52633.5–2011. *Zashchita informatsii. Tekhnika zashchity informatsii. Avtomaticheskoe obuchenie neyrosetevykh preobrazovateley biometriya-kod dostupa* [GOST R 52633.5-2011. Information protection. Information security techniques. Automatic training of neural network converters biometrics-access code]. Moscow: Gosstandart Rossii, 2011. [In Russian]
2. R 50.1.037–2002. *Rekomendatsii po standartizatsii. Prikladnaya statistika. Pravila proverki soglasiya opytnogo raspredeleniya s teoreticheskim. Chast' I. Kriterii tipa χ^2* [P 50.1.037-2002. Recommendations for standardization. Applied statistics. Rules for checking the agreement of the experimental distribution with the theoretical one. Part I. Criteria of the χ^2 type]. Moscow: Gosstandart Rossii, 2001, 140 p. [In Russian]
3. R 50.1.037–2002. *Prikladnaya statistika. Pravila proverki soglasiya opytnogo raspredeleniya s teoreticheskim. Chast' II. Neparametricheskie kriterii* [P 50.1.037-2002. Applied statistics. Rules for checking the agreement of the experimental distribution with the theoretical one. Part II. Nonparametric test]. Moscow: Gosstandart Rossii, 2002, 123 p. [In Russian]
4. Kobzar' A. I. *Prikladnaya matematicheskaya statistika. Dlya inzhenerov i nauchnykh rabotnikov* [Applied mathematical statistics. For engineers and researchers]. Moscow: FIZMATLIT, 2006, 816 p. [In Russian]
5. Ivanov A. I., Vyatchanin S. E., Malygina E. A., Lukin V. S. *Trudy Mezhdunarodnogo simpoziuma Nadezhnost' i kachestvo* [Proceedings of the international Symposium Reliability and Quality]. 2019, vol. 2, pp. 131–134. [In Russian]
6. Volchikhin V. I., Ivanov A. I., Bezyaev A. V., Kupriyanov E. N. *Inzhenernye tekhnologii i sistemy* [Engineering technologies and systems]. 2019, vol. 29, no. 2, pp. 205–217. DOI <https://doi.org/10.15507/2658-4123.029/2019.02.205-217> [In Russian]
7. Ivanov A. I., Kupriyanov E. N., Tureev S. V. *Nadezhnost'* [Reliability]. 2019, no. 2, pp. 22–27. DOI 10.21683/1729-2646-2019-19-2-22-27. [In Russian]
8. Ivanov A. I., Perfilov K. A., Malygina E. A. *Vestnik Sibirskogo gosudarstvennogo aerokosmicheskogo universiteta imeni akademika M. F. Reshetneva* [Bulletin of the Siberian State Aerospace University named after academician M. F. Reshetnev]. 2016, no. 4 (17), pp. 864–871. [In Russian]
9. Perfilov K. A., Ivanov A. I., Protsenko E. D. *Evropeyskiy soyuz uchenykh* [European union of scientists]. 2015, no. 13, part 5, pp. 9–12. [In Russian]

10. Ivanov A. I., Perfilov K. A. *Sovremennye okhrannye tekhnologii i sredstva obespecheniya kompleksnoy bezopasnosti ob"ektov: XI Vseros. nauch.-prakt. konf. (Penza, Zarechnyy, 20 oktyabrya 2016 g.)* [Modern security technologies and means of ensuring complex security of objects: XI all-Russian scientific and practical conference (Penza, Zarechny, October 20, 2016)]. Penza, 2016, pp. 223–229. [In Russian]
11. Ivanov A. I., Perfilov K. A., Malygina E. A. *Izmerenie. Monitoring. Upravlenie. Kontrol'* [Measurement. Monitoring. Management. Control]. 2016, no. 2 (16), pp. 58–66. [In Russian]
12. Morelos-Saragosa R. *Iskusstvo pomekhoustoychivogo kodirovaniya* [The art of noise-tolerant coding]. Moscow: Tekhnosfera, 2007, 320 p. [In Russian]
13. Akhmetov B. B., Ivanov A. I., Serikova N. I., Funtikova Yu. V. *Vestnik Natsional'noy akademii nauk Respubliki Kazakhstan* [Bulletin of the National Academy of Sciences of the Republic of Kazakhstan]. 2015, no. 1, pp. 17–25. [In Russian]
14. Kulagin V., Ivanov A., Gazin A., Akhmetov B. *Analitika* [Analytics]. 2016, no. 5 (30), pp. 22–29. [In Russian]
15. Volchikhin V. I., Ivanov A. I., Pashchenko D. V., Akhmetov B. B., Vyatchanin S. E. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Tekhnicheskie nauki* [University proceedings. Volga region. Engineering sciences]. 2017, no. 1, pp. 5–15. [In Russian]

Иванов Александр Иванович

доктор технических наук, доцент,
ведущий научный сотрудник,
Пензенский научно-исследовательский
электротехнический институт
(440026, Россия, г. Пенза, ул. Советская, 9)
E-mail: ivan@pnici.penza.ru

Куприянов Евгений Николаевич

аспирант,
Пензенский государственный университет
(440026, Россия, г. Пенза, ул. Красная, 40)
E-mail: ibst@pnzgu.ru

Савинов Константин Николаевич

старший преподаватель,
отдел автоматизированных систем управления
и спутниковой связи,
Военно-учебный центр,
Пензенский государственный университет
(440026, Россия, г. Пенза, ул. Красная, 40)
E-mail: creativprojectcentr@gmail.com

Банных Андрей Григорьевич

аспирант,
Пензенский государственный университет
(440026, Россия, г. Пенза, ул. Красная, 40)
E-mail: ibst@pnzgy.ru

Безяев Александр Викторович

кандидат технических наук, ведущий специалист,
Пензенский филиал ФГУП НТЦ «Атлас»
(440026, Россия, г. Пенза, ул. Советская, 9)
E mail: Bezjaev_Alex@mail.ru

Ivanov Alexander Ivanovich

doctor of technical sciences, associate professor,
senior researcher,
Penza Scientific Research Electrotechnical Institute
(440026, 9 Sovetskaya street, Penza, Russia)

Kupriyanov Evgeny Nikolaevich

postgraduate student,
Penza State University
(440026, 40 Krasnaya street, Penza, Russia)

Savinov Konstantin Nikolayevich

senior lecturer,
department of automated control systems
and satellite communications,
Military Training Center,
Penza State University
(440026, 40 Krasnaya street, Penza, Russia)

Bannykh Andrey Grigoryevich

postgraduate student,
Penza State University
(440026, 40 Krasnaya street, Penza, Russia)

Bezjaev Alexander Victorovich

candidate of technical sciences, lead specialist,
STC "Atlas" Penza branch
(440026, 9 Sovetskaya street, Penza, Russia)

Образец цитирования:

Сокращение требований к объему выборки при проверке гипотезы нормальности биометрических данных через нейросетевое объединение семи статистических критериев / А. И. Иванов, Е. Н. Куприянов, К. Н. Савинов, А. Г. Банных, А. В. Безяев // Надежность и качество сложных систем. – 2019. – № 4 (28). – С. 69–77. – DOI 10.21685/2307-4205-2019-4-7.