

ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫЕ И УПРАВЛЯЮЩИЕ СИСТЕМЫ

УДК 519.24; 53; 57.017

В. И. Волчихин, А. И. Иванов, А. В. Сериков, Ю. И. Серикова

ИСПОЛЬЗОВАНИЕ ЭФФЕКТОВ КВАНТОВОЙ СУПЕРПОЗИЦИИ ПРИ РЕГУЛЯРИЗАЦИИ ВЫЧИСЛЕНИЙ СТАНДАРТНОГО ОТКЛОНЕНИЯ НА МАЛЫХ ВЫБОРКАХ БИОМЕТРИЧЕСКИХ ДАННЫХ

V. I. Volchikhin, A. I. Ivanov, A. V. Serikov, Y. I. Serikov

USING THE EFFECTS OF QUANTUM SUPERPOSITION OF THE REGULARIZATION OF THE STANDARD DEVIATION CALCULATION ON SMALL SAMPLES OF BIOMETRIC DATA

А н н о т а ц и я. *Актуальность и цели.* Целью работы является снижение погрешности вычисления стандартного отклонения на малых выборках биометрических данных в рамках гипотезы нормального распределения значений. *Материалы и методы.* Рассматривается метод регуляризации оценки стандартного отклонения. Показано, что для этой цели нужно использовать еще один (второй) метод вычисления стандартного отклонения. Если ошибка второго метода слабо коррелирована с ошибкой классического метода вычисления корреляции, то появляется возможность их взаимного уточнения. При реализации второго метода использован один из эффектов квантовой суперпозиции, возникающий при квантовании континуума с переходом к симметричным гистограммам с регулируемой шириной интервалов. *Результаты.* Предложено усреднять результат вычисления стандартного отклонения, полученный двумя разными методами для повышения точности оценок на малых выборках. *Выводы.* Результирующая ошибка при вычислении стандартного отклонения двумя методами снижается примерно на 30 %, что эквивалентно повышению размеров тестовой выборки с 16 до 30 опытов. Это эквивалентно повышению числа обусловленности программ для вычисления стандартного отклонения примерно в 2 раза при примерном удвоении числа строк программной реализации для языков высокого уровня.

A b s t r a c t. *Background.* The aim is to decrease the error calculating the standard deviation on small samples of biometric data as part of the normal distribution of values hypothesis. *Materials and methods.* The method of estimation of the standard deviation regularization. It is shown that for this purpose need to use another (second) method of calculating the standard deviation. If the error of the second method is weakly correlated with the error of the classical

method of calculating the correlation, there is a possibility of their mutual clarification. When implementing the second method used by one of the effects of quantum superposition occurring in the quantization of the continuum with the transition to the balanced histogram with adjustable width intervals. **Results.** Proposed average result of the calculation of the standard deviation obtained by two different methods to improve the accuracy of estimates on small samples. **Conclusions.** The resulting error in the calculation of the standard deviation of the two methods is reduced by about 30 %, equivalent to increasing the size of the test sample with 16 to 30 experiments. This is equivalent to improve the condition number of programs to calculate the standard deviation of about 2 times at roughly double the number of lines of software implementation for high-level languages.

К л ю ч е в ы е с л о в а: малые тестовые выборки биометрических данных, регуляризация вычислений, гипотеза нормального закона распределения значений, стандартное отклонение.

К e y w o r d s: small test sample of biometric data, regularization calculations, the hypothesis of normal distribution of values of the law, the standard deviation.

**Проблема вычисления стандартного отклонения
на малых тестовых выборках биометрических данных**

При обработке биометрических данных, например при обучении больших искусственных нейронных сетей, используется от 8 до 16 примеров образа «Свой». Стандартный алгоритм обучения [1] опирается на вычисление нескольких сотен математических ожиданий биометрических параметров $E(v_i)$ и столько же стандартных отклонений $\sigma(v_i)$. При этом ошибка вычисления математических ожиданий на малых выборках, как правило, оказывается меньше, чем ошибка вычисления стандартных отклонений. Так, из рис. 1 видно, что для выборки, состоящей из 16 примеров, относительная ошибка вычисления стандартного отклонения может достигать 50 %. По этой причине одной из основных проблем применения стандарта [1] на малых выборках является повышение точности вычисления стандартных отклонений. Ранее было показано, что систематическая составляющая погрешности вычисления стандартного отклонения [2] устранима. В данной статье мы попытаемся показать, что случайная составляющая погрешности также может быть значительно уменьшена.

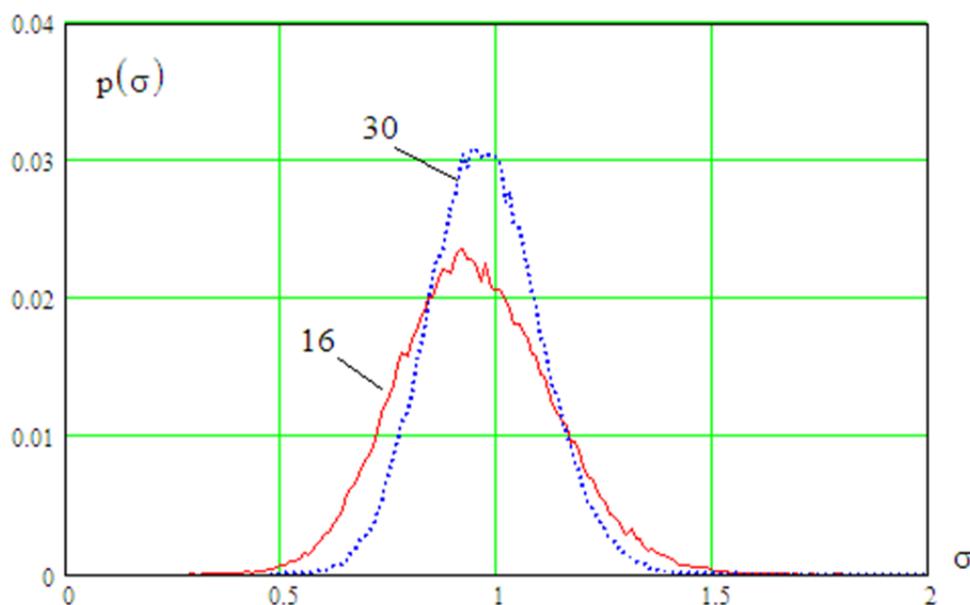


Рис. 1. Плотности распределения значений среднеквадратического отклонения при выборке из 16 и 30 опытов для нормированных данных

Та же самая проблема малых выборок возникает и при тестировании [3] нейросетевых преобразователей биометрия-код. При оценке вероятности ошибок второго рода (ошибочный пропуск «Чужого») можно заранее сформировать достаточно большую тестовую базу образов «Чужой» по ГОСТ Р 52633.1 [4]. В этом случае проблема решается за счет предварительного выполнения комплекса предварительных организационно-технических мероприятий. Если же требуется определить вероятность ошибок первого рода (отказ в доступе примерам образа «Свой»), нет иного способа, как предъявлять средству биометрической аутентификации выборку из тестовых примеров, не участвовавших при обучении.

Очевидно, что с ростом числа опытов в обучающей выборке [1] или тестовой выборке [2] $n \rightarrow \infty$ происходит монотонное снижение методической ошибки $E(\sigma) \rightarrow 1$. Также происходит монотонное снижение стандартного отклонения континуума стандартных отклонений $\sigma(\sigma) \rightarrow 0$. Эта ситуация отображена на рис. 1, из которого видно, что при выборке из 30 опытов стандартное отклонение сжимается примерно на 30 % по отношению к распределению, построенному на выборке из 16 опытов.

Математическая молекула асимметрии континуума данных с нормальным распределением значений

Классические квантовые вычисления [5] и квантовые вычисления нейросетевой эмуляции квантовых состояний [6] имеют много общего. В связи с этим (по аналогии с планетарной моделью некоторой физической молекулы) рассмотрим математическую молекулу, построенную для наблюдения асимметрии нормального распределения континуума внутренних состояний 16 электронов, размещенных на трех орбиталях. Обе эти конструкции иллюстрируются рис. 2.

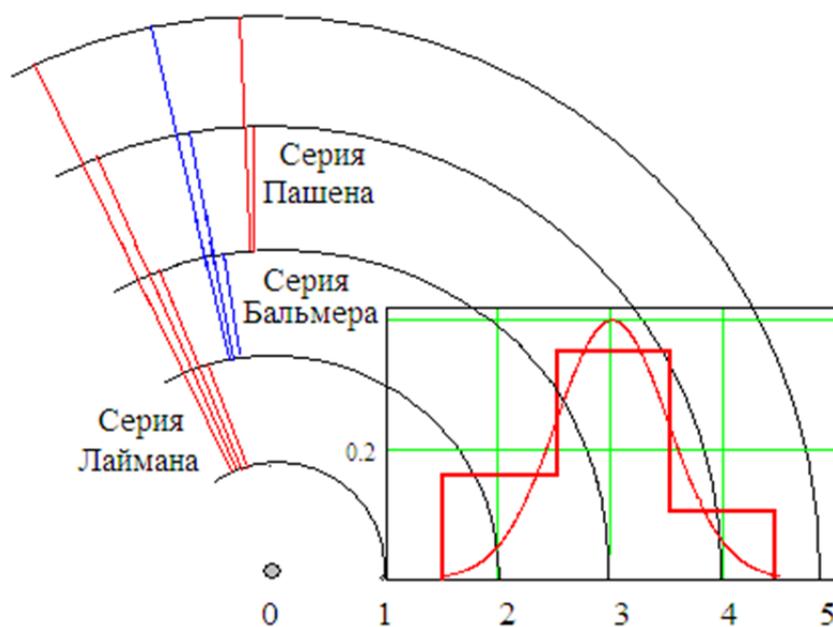


Рис. 2. Планетарная модель молекулы водорода, построенная исходя из гипотезы нормального закона распределения значений континуума состояний электронов, с квантованием данных по трем орбитам (по трем столбцам гистограммы)

Для планетарной модели некоторой физической молекулы Нильсом Бором постулировано отсутствие излучения при нахождении электрона на стационарной орбите, спектр излучения (поглощения) возникает только при переходе электрона с одной орбиты на другую.

Математическая модель синтезирована иначе. Ее внутренний континуум воспроизводится программным генератором псевдослучайных чисел, который имеет нормальное распределение данных и создает выборки по 16 опытов. Данные таких выборок накрывают 3 интервала гистограммы (рис. 2), каждый интервал гистограммы соответствует нахождению некоторого числа электронов на одной из трех, разрешенных для них орбиталей.

Принципиальным отличием физической молекулы от математической является то, что мы наблюдаем спектр возможных состояний числа электронов (опытов) на каждой из трех орбиталей. Воспроизвести на обычном компьютере молекулу водорода, точно воспроизводящую спектр излучения серий линий Лаймана, Бальмера, Пашена, очень трудно (таких программных моделей нет). Напротив, создать программную модель математической молекулы с тремя орбиталями и 16 электронами несложно. В этом корень проблемы. Создать квантовый вычислитель с поддержкой даже на 10 кубит квантовой суперпозиции, находясь в рамках парадигмы Манина – Шредингера, очень сложно из-за проблемы синхронизации состояния «котов Шредингера». При отказе от аппаратно-программного воспроизведения уравнений Шредингера положение кардинально изменяется. Поясним это на примере математической молекулы с тремя уровнями квантования состояний нормального континуума при 16 опытах.

Математическая молекула, созданная для наблюдения асимметрии хвостов нормального распределения на малых выборках

Ранее были созданы математические хи-квадрат молекулы Пирсона [6–10], использующие процедуры квантования данных, характерные для гистограмм, содержащих 4 и 6 интервалов. Для каждой математической молекулы получается свой спектр выходных состояний, зависящий от закона распределения внутреннего континуума, разметов выборки (числа электронов), параметров квантователя (числа столбцов гистограммы и способа их синхронизации с математическим ожиданием и стандартным отклонением).

В нашем случае будем использовать гистограмму с тремя столбцами. При этом центр второго столбца будем всегда совмещать с математическим ожиданием каждой выборки по 16 примеров. В этом случае компараторы первого и второго уровней квантователя должны быть синхронизированы между собой следующим образом:

$$\begin{cases} k_1 = E(x) - 0,673\sigma(x); \\ k_2 = E(x) + 0,673\sigma(x). \end{cases} \quad (1)$$

В этом случае в первый столбец гистограммы примеры тестовой выборки должны попадать с вероятностью $P_1 = 0,25$. Во второй и третий столбцы гистограммы опыты тестовой выборки будут попадать с вероятностями $P_2 = 0,5$ и $P_2 = 0,25$. Однако эта вероятность возникает только в пределе при очень большом повторении серии опытов. В реальной жизни каждая выборка по 16 примеров будет давать гистограммы с разным заполнением столбцов.

По первым двум столбцам гистограммы осуществим нумерацию возможных состояний гистограмм по формуле $n = 10n_1 + n_2$. На рис. 3 даны линии спектра состояний математической молекулы, высота которых пропорциональна вероятности их появления.

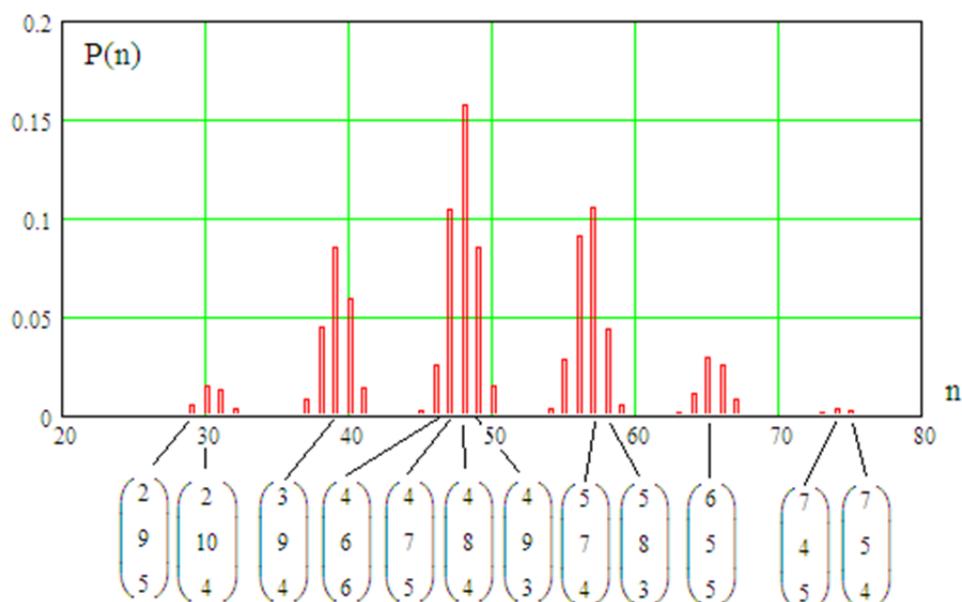


Рис. 3 Спектр состояний гистограмм трехуровневой молекулы с нормальным внутренним континуумом

Описанная выше математическая конструкция хорошо описывает асимметрию правого и левого хвостов распределения значений в малой выборке. В левой части рисунка оказываются линии спектра, соответствующие слишком коротким левым хвостам распределений. В правой части рис. 3 спектральные линии соответствуют ситуации, когда укороченными оказываются правые хвосты распределений. В этом смысле эту математическую конструкцию следует рассматривать как молекулу, удобную для контроля асимметрии распределений малых выборок.

Использование математической молекулы для контроля среднеквадратического отклонения в малой выборке

Заметим, что появление линии в левой части рис. 3 и/или в правой части всегда свидетельствует о высокой вероятности ошибок вычисления стандартного отклонения. Если же линия спектра окажется в центре рисунка, то стандартное отклонение с высокой вероятностью будет вычислено правильно.

Пользуясь этим, осуществим оценку верного значения стандартного отклонения путем регулирования ширины центрального столбца гистограммы. Для этой цели введем коэффициент масштаба M в формулу вычисления положения порогов квантования данных:

$$\begin{cases} k_1 = E(x) - M \cdot 0,673\sigma(x); \\ k_2 = E(x) + M \cdot 0,673\sigma(x). \end{cases} \quad (2)$$

Если теперь плавно менять масштаб в интервале от 0,3 до 1,3, то соотношения между столбцами гистограммы будет меняться. В качестве критерия процесса оптимизации будем использовать равенство вероятности попадания примеров в двух крайних столбцах и в центральный столбец:

$$\begin{cases} P_1 + P_3 - P_2 = 0, \\ n_1 + n_3 - n_2 = 0. \end{cases} \quad (3)$$

На рис. 4 дается функция изменения масштаба для двух разных выборок, приводящая изначально асимметричную гистограмму к условию частичной симметрии (3).

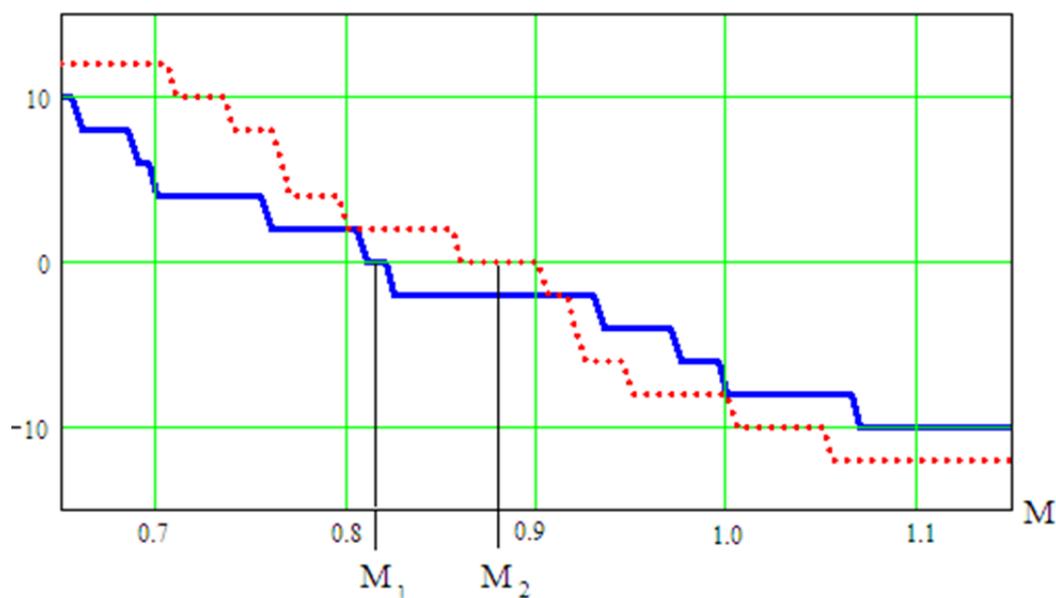


Рис. 4. Примеры поиска оптимального масштаба стандартного отклонения, приводящего к частично симметризованной гистограмме

Из рис. 4 видно, что первая выборка (ее след отображен сплошной линией) свертывается до частично симметричной гистограммы при масштабе $M_1 = 0,82$. Вторая выборка свертывает-

ся до частично симметричной гистограммы при масштабе $M_2 = 0,87$. Естественно предположить, что разные масштабы M_1 и M_2 вполне могли оказаться следствием ошибки вычисления стандартного отклонения, обусловленной конечной выборкой из 16 примеров.

Проверка статистических свойств второго метода определения стандартного отклонения показала, что его ошибка не коррелирована с ошибкой классического метода вычисления. Стандартные отклонения случайных ошибок первого и второго методов сопоставимы. Это позволяет корректировать случайную составляющую ошибки путем простого усреднения данных, полученных двумя разными методами:

$$\sigma(x) = \sigma(x) \frac{1+M}{2}. \quad (4)$$

Такая коррекция позволяет снизить стандартное отклонение примерно на 30 %, что эквивалентно повышению выборки с 16 опытов до 30. При этом методическая составляющая ошибки увеличивается, однако она поддается компенсации известным методом [2].

Заключение

Если оставаться в рамках классического статистического анализа, то задача повышения точности оценки стандартного отклонения на малых выборках кажется невыполнимой. Положение меняется, когда мы выполняем параллельные вычисления при континуальном и квантовом представлении данных. Результаты вычислений в непрерывном и дискретном пространствах оказываются независимыми, что и дает возможность для их взаимного дополнения. Примерно так же, как непрерывная физика Ньютона дополняется квантовой физикой микромира, классический статистический анализ непрерывных данных может быть дополнен квантовой статистикой тех же самых данных.

Принципиальная значимость данной статьи состоит в том, что она хорошо иллюстрирует факт дополнения друг друга непрерывной (континуальной) и квантовой статистик. Оставаясь только в непрерывной статистике или только в квантовой статистике, мы ограничиваем свои технические возможности. Из рассмотренного в статье примера видно, что действительно эффективные вычислительные алгоритмы должны рассматривать статистические объекты двусторонне, необходимо создавать континуально-квантовые алгоритмические симбиозы.

Библиографический список

1. ГОСТ Р 52633.5–2011 «Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа». – М. : Изд-во стандартов, 2011.
2. Волчихин, В. И. Компенсация методических погрешностей вычисления стандартных отклонений и коэффициентов корреляции, возникающих из-за малого объема выборок / В. И. Волчихин, А. И. Иванов, Ю. И. Серикова // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2016. – № 1 (37). – С. 103–110.
3. ГОСТ Р 52633.3–2011. Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора. – М. : Изд-во стандартов, 2011.
4. ГОСТ Р 52633.1–2009. Защита информации. Техника защиты информации. Требования к формированию баз естественных биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации. – М. : Изд-во стандартов, 2009.
5. Нильсон, М. Квантовые вычисления и квантовая информация / М. Нильсон, И. Чанг. – М. : Мир, 2006. – 821 с.
6. Иванов, А.И. Многомерная нейросетевая обработка биометрических данных с программным воспроизведением эффектов квантовой суперпозиции / А. И. Иванов. – Пенза : Изд-во ПНИЭИ, 2016. – 133 с. – URL: <http://пниэи.рф/activity/science/BOOK16.pdf>
7. The Family of Chi-Square Molecules Pearson: Software-Continuum Quantum Accelerators of High-Dimensional Calculations / В. Akhmetov, А. Ivanov, А. Gilmutdinov, А. Bezyaev, Y. Funtikova // 15th International Conference on Control, Automation and Systems (ICCAS 2015) to be held on October 13–16, 2015 in ВЕХСО(ТРО3-78). – Busan, Korea, 2015.
8. Кулагин, В. П. Циклические континуально-квантовые вычисления: усиление мощности хи-квадрат критерия на малых выборках / В. П. Кулагин, А. И. Иванов, А. И. Газин, Б. Б. Ахметов // Аналитика. – № 5. – 2016. – С. 22–29.

9. 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 (41). – С. 3–7.
10. Статистические свойства молекулы хи-квадрат Пирсона для малых тестовых выборок биометрических данных / В. И. Волчихин, А. И. Иванов, Б. Б. Ахметов, Д. В. Пашенко, С. Е. Вятчанин // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. – 2017. – № 1 (41).

Волчихин Владимир Иванович

доктор технических наук, профессор,
президент Пензенского государственного
университета
(Россия, г. Пенза, ул. Красная, 40)
E-mail: vvi@pnzgu.ru

Volchikhin Vladimir Ivanovich

doctor of technical sciences, professor,
President of Penza State University
(40 Krasnaya street, Penza, Russia)

Иванов Александр Иванович

доктор технических наук, доцент,
начальник лаборатории биометрических
и нейросетевых технологий,
Пензенский научно-исследовательский
электротехнический институт
(Россия, г. Пенза, ул. Советская, 9)
E-mail: ivan@pniei.penza.ru

Ivanov Aleksandr Ivanovich

doctor of technical sciences, associate professor,
head of biometric and neuronal nets
technology laboratory,
Penza Scientific Research Electrotechnical Institute
(9 Sovetskaya street, Penza, Russia)

Сериков Андрей Васильевич

начальник отделения,
Научно-производственное
предприятие «Рубин»
(Россия, г. Пенза, ул. Байдукова 2)
E-mail: julia-ska@yandex.ru

Serikov Andrey Vasil'evich

department director,
Scientific production company «Rubin»
(2 Baydukova street, Penza, Russia)

Серикова Юлия Игоревна

магистрант,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: julia-ska@yandex.ru

Serikova Julia Igorevna

master degree student,
Penza State University
(40 Krasnaya street, Penza, Russia)

УДК 519.24; 53; 57.017

Волчихин, В. И.

Использование эффектов квантовой суперпозиции при регуляризации вычислений стандартного отклонения на малых выборках биометрических данных / В. И. Волчихин, А. И. Иванов, А. В. Сериков, Ю. И. Серикова // Измерение. Мониторинг. Управление. Контроль. – 2017. – № 1 (19). – С. 57–63.