MSC 62P10, 62P35 **DOI: 10.14529/jcem170405**

CONDITIONS OF OBTAINING THE DISCRETE KURTOSIS SPECTRUM OF STATISTICAL DISTRIBUTIONS OF BIOMETRIC DATA FOR SMALL SAMPLES

V. I. Volchikhin[1], president@pnzgu.ru,
A. I. Ivanov[2], ivan@pniei.penza.ru,
A. I. Gazin[3], yearn@bk.ru,
A. G. Bannykh[1], ibst@pgzgu.ru.
[1] Penza State University, Penza, Russian Federation,
[2] Penza Scientific Research Electrotechnical Institute, Penza, Russian Federation,
[3] Lipetsk State Pedagogical University named after P.P. Semenov-Tyan-Shan, Lipetsk, Russian Federation.

The aim of the paper is to amplify the statistic criterions in small test samples. We propose to use the simulation tools and numerically get the density of distribution of statistical excess criterion values in small samples. The spectrum of excess criterion states becomes discrete, when the histogram intervals are synchronized with the mathematical expectation of the sample. The chi-square Pearson's molecule constructed before was created with the use of the second-order statistical moment. In this paper, we prove that such constructions are also efficient for forth-order statistical moments. The chi-square mathematical Pearson's molecule and mathematical excess molecule are analogous. We surmise that there are infinitely many mathematical molecules, which are similar to the actual physical molecules in their properties. The Schrödinger equations are not unique; their analogues can be constructed for each mathematical molecule. We can expect a synthesis of the mathematical molecules with inner multidimensional continuum states of "electrons" and their displays in the form of discrete output spectrums of states for sixth-, eighth-order and higher even statistical moments.

Keywords: quantum superposition, chi-square Pearson's criterion, discrete spectrum of states, statistical analysis of small samples.

1. Two Branches of Classical Statistics Dealing with the Description of Continuous and Discrete Distribution of the Values of a Random Variable

As is known, the modern classical statistics has two independent branches. The first one deals with discrete distributions [1]:

- hypergeometric distribution;
- polynomial distribution;
- binomial distribution;
- Poisson distribution;
- geometric distribution;
- Pascal distribution;
- Polya distribution.

The second branch deals with continuous distribution:

- Gaussian distribution;
- Student's t-distribution;
- uniform distribution;
- Cauchy distribution;
- Laplace distribution;
- beta distribution;
- gamma distribution;
- Fisher–Snedecor distribution \ the F-distribution;
- chi-square Pearson's distribution.

At the beginning, two above-mentioned groups of distribution were considered as not intersect. Nevertheless, the conditions [2, 3], under which the chi-square Pearson's distribution becomes discrete, were published in 2015. The fact that the same statistical distribution under certain conditions can be refer to different classes raised and still raises doubts. Nevertheless, subsequent publications [4, 5] answered some of the expert society's questions.

The important theoretical question is how many distributions previously considered as continuous can be made discrete. We show that the mathematical chi-square molecule with three [6] allowed orbitals can be transformed into the mathematical excess molecule with discrete spectrum of the output states.

This is important to evaluate the number of continuous-quantum equations, which are analogous to the Schrödinger equations. The quantum mathematics is based on the Schrödinger equations [7, 8]. It is obvious that for the chi-square Pearson's equations, the quantum mathematics of the Schrödinger's wave equations should be changed. Preliminary examination shows that the changes are not numerous. Nevertheless, there may exist other equations, which essentially correct the quantum mathematics based on the Manin–Schrödinger paradigm [7, 8].

2. Continuous Distribution of the Excess Coefficient Values

Mathematical statistics uses the excess coefficient of the sample vector from $-n$ values of a random variable:

$$\gamma(\bar{x}) = \frac{M_4(\bar{x})}{(\sigma(\bar{x}))^4} - 3, \tag{1}$$

where $M_4(\cdot)$ is the fourth statistical moment, $\sigma(\cdot)$ is the standard deviation.

It is obvious that for high value of sample, i.e. $n \to \infty$, we obtain a parameter, which is equal to zero for the normal distribution. If the sample has "heavy" tails and "blunt" peak, then the excess coefficient is negative. On the contrary, the data with the "sharp" peak and "thin" tails have the positive excess coefficient.

The distribution of the excess coefficient (1) essentially depends on the volume of sample. An example for biometric data samples having actual volumes is given in Fig. 1.

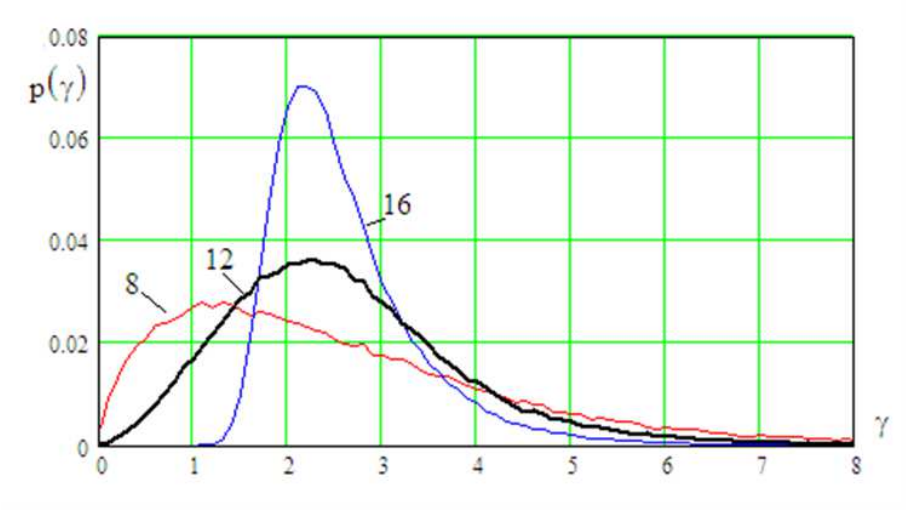

Fig. 1. The distribution density of the excess coefficients for samples of biometric data consisting of 8, 12, 16 tests with the normal distribution of values

Fig. 1 shows that the distribution density functions of the excess coefficient essentially depend on the sample volume. However, the functions also essentially depend on the distribution of the initial data. If the initial biometric data have an even distribution, then the distribution density functions of the excess coefficient are others, see Fig. 2.

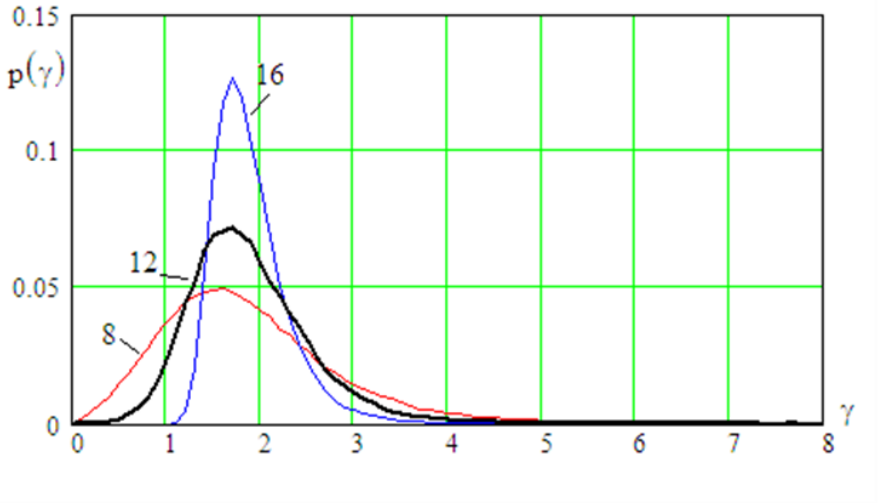

Fig. 2. The distribution density of the excess coefficient for biometric data samples consisting of 8, 12, 16 tests with even distribution of the values

It is easy to see that Fig. 1 and Fig. 2 are different. The distribution in Fig. 2 is narrower than in Fig. 1, i.e. the excess coefficient can be considered as a criterion of the normal distribution of initial data in the test sample. The chi-square Pearson's criterion [2, 3, 4, 5] and excess criterion in the continuous case of their implementation have a comparable power of decision making.

3. The Mathematical Chi-square Molecule with Three Levels of Allowed States

Classical quantum calculations [8] and quantum calculations of neural network emulation of quantum states [7] have many common features. Therefore, by analogy with the planetary model of the hydrogen molecule, we consider the mathematical chi-square Pearson's molecule. Both structures are illustrated in Fig. 3.

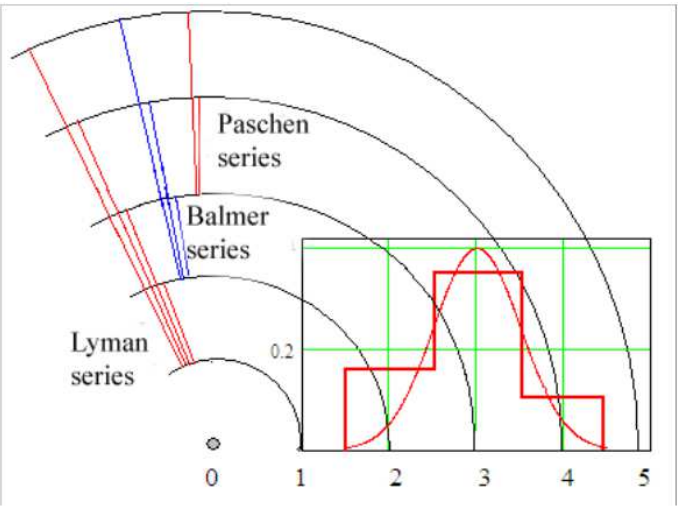

Fig. 3. The planetary model of hydrogen molecule based on the hypothesis of the normal distribution of electron state continuum with quantization of data on 3 orbitals (three bars of the histogram)

In the planetary model of the hydrogen atom, the electron moves from one orbit to another and emits a light photon. As a result, the spectrum of hydrogen absorption (or radiation) contains lines (series of lines). The left part of Fig. 1 gives the corresponding names of series of hydrogen spectrum lines.

The papers [5, 6, 7] describe the spectrums of mathematical chi-square molecule for 4 and 6 orbitals. The spectrums of states for three orbitals can be constructed by analogy. An example of the spectrum is given in Fig. 4.

Fig. 4 shows that the spectrum of states of two compared Pearson's molecules is not monotonous relatively power of its components. The neighboring components of the spectrum of the same molecule can be described by a complex correlation.

4. Discrete Character of the Spectrum of States of the Mathematical Excess Molecule

The idea of the excess coefficient (1) is to compensate the correlation of the fourth statistical moment to the fourth power of standard deviation for the normal distribution. The excess molecule should be based on the similar idea. In order to synthesize the molecule, we can use the normal distribution and calculate the limits of quantization leading to similar probabilities of random data appearance in the central bar of the histogram (Fig. 3) and in two outside bars. The following condition that the probabilities

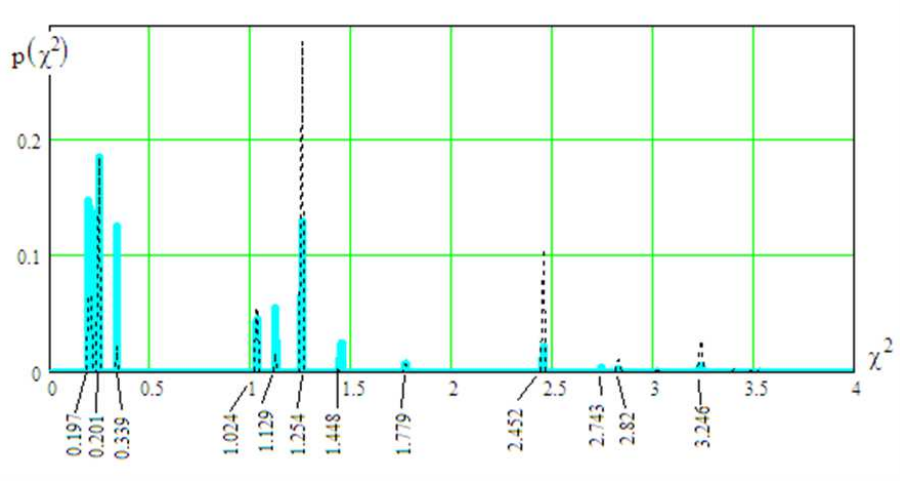

Fig. 4. An example of the correlation of 11 the most powerful components of output spectrum of the chi-square Pearson's molecule with three orbitals for the normal distribution (pale thick lines) and for the even distribution (dotted line) in the case of 16 tests

are equal

$$E(P_1) + E(P_3) = E(P_2), \tag{2}$$

is fulfilled, if the data in each sample are quantized by comparing them using the following two limits.

$$\begin{cases} k_1 = E(x) - 0.675 \cdot \sigma(x), \\ k_2 = E(x) + 0.675 \cdot \sigma(x), \end{cases} \tag{3}$$

where $E(\cdot)$ is the operator to calculate the mathematical expectation, $\sigma(\cdot)$ is the operator to calculate the standard deviation.

For such limits of data quantization, a histogram for each sample has the different number of tests n1, n2, n3 in the first, second and third bars of the histogram, respectively. However, for samples of the same size, these numbers have a discrete spectrum of states. Therefore, we can normalize the number of tests and calculate the spectrum of states of the excess molecule in the following way:

$$\tilde{\gamma} = 2\left\{\frac{n_1}{N} + \frac{n_3}{N}\right\} - 1, \tag{4}$$

where N is the number of tests in the sample under examination.

The spectrum of states (4) is obviously discrete, because the calculations are based on counting the number of tests, which are in the two outside intervals of the histogram. The type of the distribution of discrete spectrum values (4) can be easily obtained by the numerical experiment. The data of the numerical experiment for the sample of 16 tests are given in Fig. 5.

Fig. 5 shows that the most probable value of the spectrum is $\tilde{\gamma} = 0$ for the data with the normal distribution. For the even distribution of the sample data, the value $\tilde{\gamma} = 0.25$ m is the most probable. Spectrum lines for the normal law of continuum inside the molecule coincide with those for the continuum with the even law, and a step between the lines is the same.

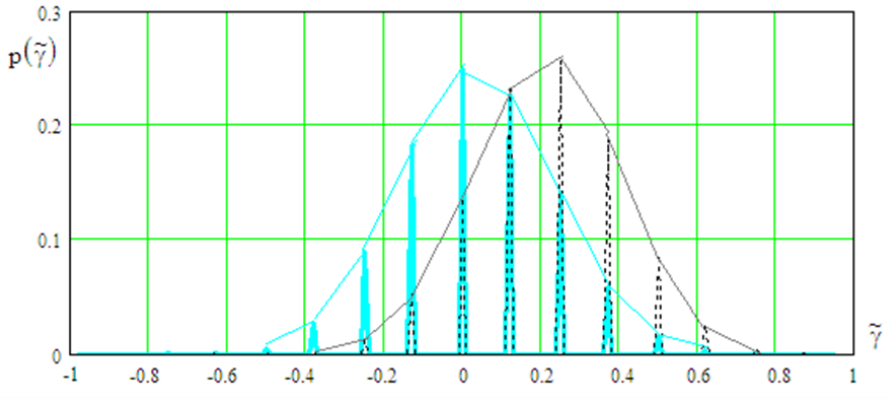

Fig. 5. An example of output spectrums of the excess molecule with 3 orbitals for the normal distribution (pale thick lines) and for the even distribution (dotted lines)

The spectrum of the output states of the excess molecule for the inner continuum with the normal and even distribution of values are similar and can be obtained by linear shift with the mirror reflection with respect to the equiprobable component.

Therefore, the descriptions of the spectrums of the excess molecule and the chi-square molecule are significantly different. That is, the quantum superposition and quantum entanglement of these mathematical structures are described by different equations. Moreover, the continuum-quantum equation for the excess molecule should be much easier than for the chi-square molecule. The simplification is caused by the symmetrization of the molecule by alignment of the states in outside bars and the central one of the histogram (2). The method is similar to the method of symmetrization of the quantum entanglement (i.e., the correlations between the discharges of the quantum superposition) used previously [9].

5. Operations on the Quantum Superposition

After quantization of a random state of 16 continuous data into three intervals of the histogram (Fig. 3), each interval includes different number of tests. For each probable state of the histogram there exists its own value of the position of chi-square spectrum lines. The numerical simulation with rounding off the data to three decimals after the decimal point gives the following results, see Fig. 6.

If the quantum superposition is arranged in ascending order of the value of spectral lines, then the lines corresponding to the spectrum $\chi^2 = 0.197$ and $\chi^2 = 0.201$ are the first and appear with the probability $P(0.197)$ and $P(0.201)$. Other lines are on the right on the first two lines according to Fig. 6.

In order to describe binary states of the quantum superposition in the Dirac bracket it is necessary to use three binary digits for the outside bars of the histogram and four binary digits for the central bar of the histogram.

Therefore, the first two elements of the quantum superposition can be given in the following way:

$$|\psi\rangle = \sqrt{P(0.197)} \cdot \left| \begin{pmatrix} 2 \\ 11 \\ 3 \end{pmatrix} \right\rangle + \sqrt{P(0.201)} \cdot \left| \begin{pmatrix} 3 \\ 11 \\ 2 \end{pmatrix} \right\rangle + \dots . \tag{5}$$

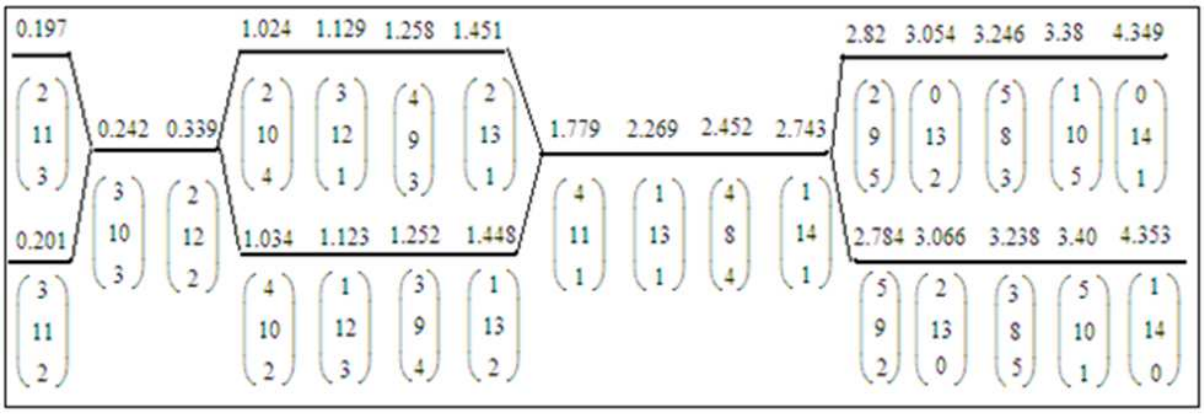

Fig. 6. The states of three bars of the histogram corresponding to different values of the χ^2 molecule spectrum lines in the interval from 0 to 4.353

For the binary form, the same notation is the following:

$$|\psi\rangle = \sqrt{P(0.197)} \cdot \left| \begin{pmatrix} 010 \\ 1011 \\ 011 \end{pmatrix} \right\rangle + \sqrt{P(0.201)} \cdot \left| \begin{pmatrix} 011 \\ 1011 \\ 010 \end{pmatrix} \right\rangle + \dots . \tag{6}$$

Therefore, the quantum superposition of chi-square molecule with 3 orbitals and 16 electrons allows to model the quantum superposition having length 10 qubits.

Note that in order to convert the chi-square molecule to the excess molecule, we reject the data in the central interval of the histogram. That is, we create an intermediate molecule with the masked central part of codes in the Dirac brackets keeping the quantum superposition of 6 qubits.

Furthermore, in order to synthesize the excess molecule, we sum the states of the outside right and left bars of the histogram (4). As a result, we obtain the quantum superposition of the excess molecule having the lengh 4 qubits (13 spectral lines are given in Fig. 5).

Therefore, the quantum superposition of the mathematical molecule is susceptible to masking the data, which are in the intervals of the quantize and it is possible to perform addition, subtraction and other arithmetic operations with the data of different quantization intervals. In this case, we obtain a mathematical molecule in order to get and maintain the quantum superposition and quantum entanglement, which are characteristic of the molecule.

Conclusion

The Manin–Schrödinger paradigm was developed in 1980s. Some mathematicians predicted the possibility of creating very effective computational algorithms within the framework of the Manin–Schrödinger paradigm. Nevertheless, the attempt to realize the algorithms in practice faced the problem of synchronizing hardware-based "Schrödinger's cats". The Schrödinger's equation effectively describes the hydrogen molecule states, though it is hard to use for creating macro objects and their synchronization. The software package to solve the Schrödinger's equation has about 250000 lines in high-level machine

code. Moreover, the increase in the number of electrons in the equation up to 30 and more leads to the necessity of using super computers. The increase in the number of electrons in the Schrödinger's equation leads to the the exponential increase in the difficulty of its solution. The case of the chi-square molecule and excess molecule is different. The software for realization of these molecules consists of several code lines, and the increase in the number of tests (electrons) gives a linear growth of the computational complexity.

We show that the mathematical chi-square Pearson's molecule and the mathematical excess molecule can be software implemented by analogy with the hydrogen molecule. Therefore, the Schrödinger equation is not the unique basis for the software quantum calculators. Along with the Manin–Schrödinger paradigm, many other paradigms, which use the equation of the chi-square molecule or the excess molecule or other mathematical molecules created on the same principle, should be considered.

It is of fundamental importance that the equations of any mathematical molecules can be symmetrized, that simplifies their formal description.

Moreover, the discrete character of the spectrum of excess molecules gives hope of the increase in the power of the statistical analyses of excesses in small samples of biometrical data by analogy with the increase in the power of the chi-square criterion described previously [5, 6].

References

1. Korn G.A., Korn T.M. *Mathematical Handbook for Scientists and Engineers*. New York, San Francisco, Toronto, London, Sydney, McGraw-Hill Book Company, 1968. doi: 10.1002/zamm.19690490921.

2. Akhmetov B.B., Ivanov A.I., Serikova N.I., Funtikova Yu. V. Discrete Character of the Law of Chi-square Distribution Criterion for small Test Selections of Values. *Bulletin of the National Academy of Sciences of the Republic of Kazakhstan*, 2015, no. 1, pp. 17–25. (in Russian)

3. Akhmetov B., Ivanov A., Gilmutdinov A., Bezyayev A., Funtikova Y. The Family of Chi-Square Molecules Pearson: Software-Continuum Quantum Accelerators of High-Dimensional Calculations. *15th International Conference on Control, Automation and Systems (ICCAS 2015), October 13-16, 2015*. Busan, Korea, 2015, pp. 1337–1341.

4. Kulagin V., Ivanov A., Gazin A., Akhmetov B. Cyclic-continuum Quantum Computing: Power Amplification chi-square Test on Small Sample. *Analytics*, 2016, issue 5, pp. 22–29. (in Russian)

5. Volchikhin V.I., Ivanov A.I., Pashchenko D.V., Akhmetov B.B., Vyatchanin S.Ye. The Prospect of Creation of a Cyclic Continual-Cuantum chi-square Machine for Checking Statistical Hypetheses on Small Test Samples of Biometric and Other Types of Data. *University proceedings. Volga region. Technical sciences*, 2017, no. 1(41), pp. 5–15. (in Russian) doi: 10.21685/2072-3059–2017-1-1.

6. Volchikhin V.I., Ivanov A.I., Serikov A.V., Serikova U.I., Using the Effects of Quantum Superposition of the Regularization of the Standard Deviation Calculation on Small Samples of Biometric Data. *Measuring. Monitoring. Management. Control*, 2017, no. 1, pp. 51–63. (in Russian)

7. Ivanov A.I. *Multifunctional Neural Network Processing of Biometric Data with Software Playback Effects of Quantum Superposition*. Penza, Publisher "PNIAI", 2016. (in Russian)

8. Nielsen M., Chuang I. *Quantum Computation and Quantum Information*. New York, Cambridge University Press, 2010.

9. Volchikhin V. I., Akhmetov B. S., Ivanov A. I. A Fast Symmetrization Algorithm for Correlations of Biometric Data of High Dimension. *University proceedings. Volga region. Technical sciences*, 2016, no. 1(37), pp. 5–15. (in Russian)

Vladimir I. Volchikhin, DSc (Techn), Professor, President of Penza State University (Penza, Russian Federation), president@pnzgu.ru.

Alexander I. Ivanov, DSc (Techn), Associate Professor, Laboratory of Biometric and Neural Network Technologies, Penza Scientific Research Electrotechnical Institute (Penza, Russian Federation), ivan@pniei.penza.ru.

Alexei I. Gazin, PhD (Techn), Department of Informatics, Information Technology and Information Security, Lipetsk State Pedagogical University named after P.P. Semenov-Tyan-Shan (Lipetsk, Russian Federation), yearn@bk.ru.

Andrew G. Bannykh, Postgraduate Student, Department of Information Security Systems Technology, Penza State University (Penza, Russian Federation), ibst@pgzgu.ru.

Received October 31, 2017

УДК 519.24, 53, 57.017 DOI: 10.14529/jcem170405

УСЛОВИЯ ПОЛУЧЕНИЯ ДИСКРЕТНОГО СПЕКТР ЭКСЦЕССА СТАТИСТИЧЕСКИХ РАСПРЕДЕЛЕНИЙ БИОМЕТРИЧЕСКИХ ДАННЫХ ДЛЯ МАЛЫХ ВЫБОРОК

В.И. Волчихин, А.И. Иванов, А.И. Газин, А.Г. Банных

Целью работы является усиление мощности статистических критериев на малых тестовых выборках. Предложено воспользоваться средствами имитационного моделирования и численно получать плотность распределения значений статистического эксцесс критерия для малых выборок. При синхронизации интервалов гистограммы с математическим ожиданием выборки спектр состояний эксцесс критерия становится дискретным. Ранее была построена хи-квадрат молекула Пирсона для малых выборок, созданная с использованием статистических моментов второго порядка. В данной статье показано, что подобные конструкции оказываются работоспособны и для статистических моментов четвертого порядка. Хи-квадрат математическая молекула Пирсона и математическая эксцесс молекула являются аналогами. Сделано предположение, что математических молекул, похожих по их свойствам на реальные физические молекулы, бесконечно много. Уравнения Шредингера не уникальны, для каждой математической молекулы может быть построен их аналог. Можно ожидать синтеза математических молекул с внутренними многомерными континуумами состояний кэлектроновь и их отображениями в виде выходных дискретных спектров состояний для 6, 8 и более высоких четных статистических моментов.

Ключевые слова: квантовая суперпозиция, хи-квадрат критерий Пирсона, дискретный спектр состояний, статистический анализ малых выборок.

Литература

1. Корн, Г. Справочник по математической статистике для научных работников и инженеров / Г. Корн, Т. Корн. – М.: Наука, 1974.

2. 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.

3. Akhmetov, B. The family of chi-square molecules pearson: software-continuum quantum accelerators of high-dimensional calculations / B. Akhmetov, A. Ivanov, A. Gilmutdinov, A. Bezyaev, Y. Funtikova // 15th International Conference on Control, Automation and Systems (ICCAS 2015). October 13-16, 2015. – Busan, Korea. – 2015. – P. 1337–1341.

4. 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. – С. 22–29.

5. Волчихин, В.И. Перспектива создания циклической континуально-квантовой хи-квадрат машины для проверки статистических гипотез на малых тестовых выборках биометрических данных и данных иной природы / В.И. Волчихин, А.И. Иванов, Д.В. Пащенко, Б.Б. Ахметов, С.Е. Вятчанин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2017. – №. 1(41). – С. 5–15.

6. 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. – С. 57–63.

7. Иванов, А.И. Многомерная нейросетевая обработка биометрических данных с программным воспроизведением эффектов квантовой суперпозиции / А.И. Иванов. – Пенза: Изд-во «ПНИЭИ», 2016.

8. Нильсон, М. Квантовые вычисления и квантовая информация / М. Нильсон, И. Чанг. – М.: Мир, 2006.

9. Волчихин, В.И. Быстрый алгоритм симметризации корреляционных связей биометрических данных высокой размерности / В.И. Волчихин, Б.Б. Ахметов, А.И. Иванов // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2016. – № 1(37). – С. 5–15.

Волчихин Владимир Иванович, доктор технических наук, профессор, президент, Пензенский государственный университет (г. Пенза, Российская Федерация), president@pnzgu.ru.

Иванов Александр Иванович, доктор технических наук, доцент, лаборатория биометрических и нейросетевых технологий, Пензенский научно-исследовательский электротехнический институт (г. Пенза, Российская Федерация), ivan@pniei.penza.ru.

Газин Алексей Иванович, кандидат технических наук, кафедра информатики, информационных технологий и защиты информации, Липецкий государственный педагогический университет имени П.П. Семенова-Тян-Шанского (г. Липецк, Российская Федерация), yearn@bk.ru.

Банных Андрей Григорьевич, аспирант, кафедра техники систем информационной безопасности, Пензенский государственный университет (г. Пенза, Российская Федерация), ibst@pgzgu.ru.

Поступила в редакцию 31 октября 2017 г.