

**В. П. Кулагин**, д-р техн. наук, проф., руководитель научно-исследовательской лаборатории космических исследований в области технологий, систем и процессов, e-mail: vkulagin@hse.ru, Московский институт электроники и математики НИУ ВШЭ,

**А. И. Иванов**, д-р техн. наук, доц., нач. лаборатории биометрических и нейросетевых технологий АО "Пензенский научно-исследовательский электротехнический институт", г. Пенза, e-mail: ivan@pniei.penza.ru.,

**Ю. И. Серикова**, студент, e-mail: julia\_ska@yandex.ru  
ФБГОУ ВПО "Пензенский государственный университет"

## Корректировка методических и случайных составляющих погрешностей вычисления коэффициентов корреляции, возникающих на малых выборках биометрических данных

*Показано, что вычисление математических ожиданий, стандартных отклонений и коэффициентов корреляции дает значительные ошибки при использовании малых выборок. Ошибка вычисления коэффициентов корреляции существенно больше, чем ошибки вычисления математических ожиданий и стандартных отклонений. Причиной ошибок является квантование континуумов исходных данных через их представление небольшой выборкой. Приводятся графики распределения вероятностей появления ошибок квантования и их следствия — ошибок, возникающих при вычислении коэффициентов корреляции на малых выборках. Даны таблицы значений параметров для двух величин коэффициентов корреляции при разных размерах выборок биометрических данных. Доказано наличие методических ошибок измерительно-вычислительных процедур определения коэффициентов корреляции. Даны таблицы описания значений методических ошибок, обусловленных малой тестовой выборкой, и корректирующих методическую ошибку гипербола.*

**Ключевые слова:** методическая ошибка, коэффициент корреляции, малая выборка, обработка биометрических данных

### Постановка задачи

В настоящее время активно идут процессы информатизации современного общества. Появилась необходимость в создании программируемых автоматов, способных безошибочно узнавать своего хозяина и автоматически обучаться на малом числе примеров биометрического образа человека-хозяина. За рубежом для этой цели используют так называемые "нечеткие экстракторы" [1–4]. В России эти же задачи решаются с использованием искусственных нейронных сетей [5, 6]. Нейросетевые преобразователи биометрии в код доступа почти во всех отношениях оказываются лучше "нечетких экстракторов". Единственный показатель, по которому "нечеткие экстракторы" превосходят нейронные сети, — это объем обучающей выборки. Настраивать "нечеткие экстракторы" удается даже на одном примере распознаваемого биометрического образа, тогда как при обучении нейросетевого преобразователя алгоритмом ГОСТ Р 52633.5—2011 [6] требуется порядка 20 примеров биометрического образа "Свой". При этом стандартный алгоритм обучения [6] устойчив и имеет линейную вычислительную сложность. Низкая вычислительная сложность стандартного алгоритма обучения и его высокая устойчивость обусловлены тем, что он не является итерационным. Алгоритм построен на вычислении весовых коэффициентов нейронной сети путем преобразо-

вания математических ожиданий биометрических параметров и их стандартных отклонений:

$$\begin{cases} E(v) = \frac{1}{n} \sum_{i=1}^n v_i; \\ \sigma(v) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (E(v) - v_i)^2}, \end{cases} \quad (1)$$

где  $n$  — число примеров в обучающей выборке;  $i$  — номер примера в обучающей выборке;  $E(v)$  — математическое ожидание одного из контролируемых биометрических параметров ( $v$ );  $\sigma(v)$  — стандартное отклонение контролируемого биометрического параметра.

Предположительно, стандартный алгоритм обучения искусственных нейронных сетей [6] может быть улучшен, если он будет усовершенствован путем использования большего числа статистических параметров. В частности, ожидается, что дополнительно при обучении нейронных сетей можно будет использовать коэффициенты парной корреляции биометрических данных [7], например, будет использоваться коэффициент корреляции между первым и вторым биометрическими параметрами:

$$r(v_1, v_2) = \frac{1}{n} \sum_{i=1}^n \frac{(E(v_1) - v_{1,i})(E(v_2) - v_{2,i})}{\sigma(v_1)\sigma(v_2)}. \quad (2)$$

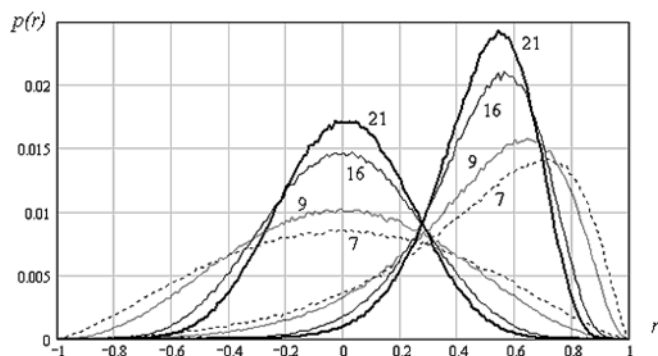


Рис. 1. Распределения значений коэффициентов корреляции для выборок из 7, 9, 16, 21 примеров при двух заданных значениях коэффициентов корреляции  $r = 0$  и  $r = 0,5$

Дополнительный учет коэффициентов корреляции [7], как дополнительных контролируемых биометрических параметров, позволяет увеличить число входов нейронной сети с величины  $N$  до величины  $(N + N^2/2)$ . Так, при числе контролируемых биометрических параметров  $N = 400$  появляется теоретическая возможность перейти от нейронных сетей с 400 входами к нейронным сетям с 80 000 входов.

К сожалению, на малых тестовых выборках, состоящих из  $n \approx 21$  примеров, ошибки вычисления коэффициентов корреляции (2) оказываются велики. Причиной больших ошибок является то, что при вычислении коэффициентов парной корреляции используются четыре статистических момента  $E(v_1)$ ,  $E(v_2)$ ,  $\sigma(v_1)$ ,  $\sigma(v_2)$ . Каждый из этих моментов будет иметь собственную погрешность вычисления  $\Delta E(v_1)$ ,  $\Delta E(v_2)$ ,  $\Delta\sigma(v_1)$ ,  $\Delta\sigma(v_2)$ . Любые вычисления склонны накапливать ошибки входящих в них величин. На рис. 1 приведены распределения значений коэффициентов корреляции, полученные при разных размерах тестовых выборок.

Из рис. 1 видно, что методические ошибки, обусловленные малым размером тестовой выборки, могут быть весьма значительны. Так, при 21 примере в тестовой выборке и отсутствии корреляции между параметрами вычисления по формуле (2) вместо ожидаемого значения  $r = 0$  можно получить значения  $r = \pm 0,65$ . Чем больше число примеров в выборке, тем точнее результат расчетов, однако увеличивать размер выборки не всегда возможно. Как следствие, актуальной становится задача повышения устойчивости вычислений по формуле (2) либо поиск других более устойчивых формул для оценки парной корреляции для малых выборок биометрических данных.

### Шумы квантования статистических функций биометрических данных, возникающие из-за малого числа примеров в тестовой выборке

Причиной ошибок вычислений является то, что биометрические данные по своей природе контину-

альны (непрерывны). Многие из них хорошо описываются нормальным законом распределения значений. Когда мы представляем континуум возможных значений всего 21 примером, возникает ошибка квантования статистических функций. В верхней части рис. 2 приведен график ступенчатой эмпирической функции вероятности биометрического параметра  $P(v)$ . Ее график монотонно растет с амплитудой каждого скачка —  $\Delta P(\Delta v) = \frac{1}{21}$ .

В нижней части рис. 2 приведена ошибка квантования или шум квантования непрерывных данных. Очевидно, что влияние шумов квантования может быть снижено путем применения специальных алгоритмов уменьшения (подавления) шумов.

Следует подчеркнуть, что хорошо изученные статистические критерии имеют простую интерпретацию, если их рассматривать как некоторую процедуру обработки шумов квантования. В частности, статистический критерий Колмогорова—Смирнова [8]

$$\sup_{-\infty < v < +\infty} |P(v) - \tilde{P}(v)| = \max\{\max|\Delta P(v)|\} \quad (3)$$

следует рассматривать как оценку сверху амплитуды шума квантования непрерывной функции вероятности  $P(v)$  или как глобальный максимум значений модуля ошибок квантования.

С этих позиций статистический критерий Крамера фон Мизеса [8] является оценкой стандартного отклонения шума квантования непрерывной функции вероятности:

$$\begin{aligned} & \int_{-\infty}^{\infty} \{P(v) - \tilde{P}(v)\}^2 dv = \\ & = \int_{-\infty}^{\infty} \{E(\Delta P) - \Delta P(v)\}^2 dv = \sigma^2(\Delta P). \end{aligned} \quad (4)$$

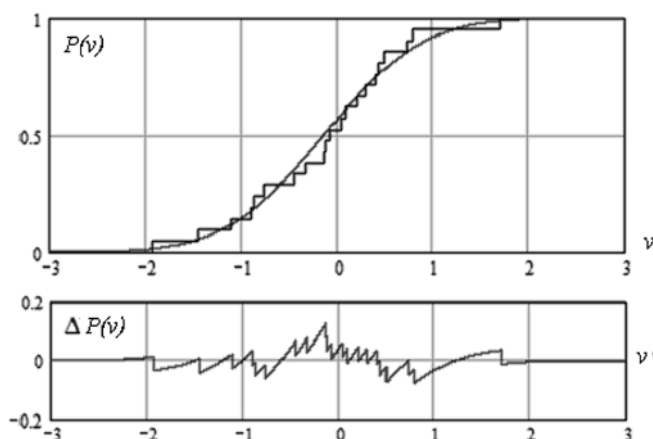


Рис. 2. Эффекты квантования непрерывной функции вероятности при представлении континуума данных 21 примером

По сути дела, каждый из известных критериев проверки статистических гипотез [8] можно рассматривать как некоторый способ подавления шумов квантования. В этом смысле можно сравнивать между собой различные статистические критерии по их способности подавлять шумы квантования. В частности, критерий Крамера фон Мизеса подавляет влияние шумов квантования много эффективнее, чем критерий хи-квадрат Пирсона.

Могут использоваться и другие процедуры снижения влияния шумов квантования, возникающих из-за малого объема тестовой выборки. За рубежом при обработке биометрических данных часто используют так называемый "бут-страп" метод [9], суть которого состоит в искусственном дополнении недостающих данных, например, их размещением в пустых интервалах эмпирической гистограммы. Хороший результат дает также сглаживание эмпирических гистограмм одномерными цифровыми фильтрами [10] и переход к многомерной обработке данных [11].

Можно предположить, что одномерные алгоритмы снижения влияния шумов квантования [9—11] могут быть обобщены до двухмерных алгоритмов более устойчивого вычисления коэффициентов парной корреляции. Таких алгоритмов вычисления коэффициентов корреляции может быть создано достаточно много. Так как каждый такой алгоритм вычисления коэффициентов корреляции будет иметь собственную погрешность, появляется возможность направленного поиска наиболее устойчивых алгоритмов вычислений, способных эффективно противодействовать шумам квантования.

### Моделирование зависимых равнокоррелированных данных

Известно, что моделировать случайные многомерные процессы крайне сложно [12]. Технически вполне возможно вычислить симметричную матрицу размерностью  $256 \times 256$ , описывающую корреляционные связи между выходами нейросетевого

преобразователя (между 256 входами нейронной сети, наблюдающей 256 биометрических параметров). Однако построить генератор случайных чисел, точно воспроизводящий корреляционные связи столь высокой размерности, технически невозможно.

Формально можно использовать 256 генераторов независимых случайных данных  $\xi$ , умножив их на некоторую связывающую матрицу  $A$ . Однако найти подобную матрицу, которая даст нужные корреляционные связи  $r(y_k, y_j)$ , достаточно трудно. Это обратная задача, которая относится к плохо обусловленным.

Для решения поставленной задачи воспользуемся симметричной связывающей матрицей, которая имеет единичную диагональ и одинаковые элементы вне диагонали:

$$\begin{bmatrix} 1 & a & \dots & a \\ a & 1 & \dots & a \\ \dots & \dots & \dots & \dots \\ a & a & \dots & 1 \end{bmatrix} \times \begin{bmatrix} \xi_{1,i} \\ \xi_{2,i} \\ \dots \\ \xi_{m,i} \end{bmatrix} = \begin{bmatrix} y_{1,i} \\ y_{2,i} \\ \dots \\ y_{m,i} \end{bmatrix} \Rightarrow$$

$$\Rightarrow R_m = \begin{bmatrix} 1 & r & \dots & r \\ r & 1 & \dots & r \\ \dots & \dots & \dots & \dots \\ r & r & \dots & 1 \end{bmatrix}. \quad (5)$$

В этом случае данные оказываются равнокоррелированными. Если плавно изменять параметр  $a$ , связывающей матрицы  $A$  от 0 до 1, то коэффициенты равной коррелированности  $r$  также будут плавно меняться в пределах от 0 до 1.

### Оценка методической ошибки

Используя формулу (2), проведем вычисления, меняя значения коэффициента корреляции  $r$  при разных значениях размера выборки  $n$ . Усредненные результаты одного миллиона вычислений приведены в табл. 1.

Таблица 1

Значения математических ожиданий  $E(r)$  при разных размерах тестовой выборки  $n$

| $n$                                     | Число степеней свободы (размер тестовой выборки) |       |       |       |       |       |       |       |       |       |       |       |
|---|--|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|   | 7  | 8     | 9     | 10    | 11    | 12    | 14    | 16    | 21    | 32    | 64    |       |
| Равная коррелированность данных ( $r$ ) | 0,01   | 0,01  | 0,01  | 0,01  | 0,01  | 0,01  | 0,01  | 0,01  | 0,01  | 0,01  | 0,01  | 0,01  |
|   | 0,1  | 0,092 | 0,092 | 0,093 | 0,094 | 0,095 | 0,095 | 0,096 | 0,096 | 0,097 | 0,098 | 0,099 |
|   | 0,2  | 0,185 | 0,186 | 0,189 | 0,189 | 0,189 | 0,191 | 0,193 | 0,194 | 0,195 | 0,197 | 0,198 |
|   | 0,3  | 0,278 | 0,281 | 0,284 | 0,286 | 0,287 | 0,288 | 0,290 | 0,292 | 0,294 | 0,296 | 0,298 |
|   | 0,4  | 0,373 | 0,377 | 0,379 | 0,382 | 0,384 | 0,386 | 0,388 | 0,390 | 0,393 | 0,396 | 0,398 |
|   | 0,5  | 0,467 | 0,473 | 0,476 | 0,479 | 0,481 | 0,483 | 0,485 | 0,487 | 0,491 | 0,494 | 0,497 |
|   | 0,6  | 0,567 | 0,572 | 0,575 | 0,578 | 0,581 | 0,582 | 0,585 | 0,588 | 0,591 | 0,594 | 0,598 |
|   | 0,7  | 0,667 | 0,672 | 0,676 | 0,679 | 0,681 | 0,683 | 0,686 | 0,688 | 0,691 | 0,695 | 0,698 |
|   | 0,8  | 0,771 | 0,775 | 0,778 | 0,781 | 0,783 | 0,785 | 0,787 | 0,789 | 0,792 | 0,795 | 0,797 |
|   | 0,9  | 0,883 | 0,886 | 0,888 | 0,890 | 0,891 | 0,892 | 0,894 | 0,895 | 0,897 | 0,898 | 0,899 |
| 0,99                                    | 0,988  | 0,988 | 0,989 | 0,989 | 0,989 | 0,989 | 0,989 | 0,99  | 0,99  | 0,99  | 0,99  |       |

Из табл. 1 видно, что заданные при моделировании значения коэффициента корреляции (левый столбец таблицы) существенно отличаются от математических ожиданий результатов моделирования  $E(r)$ . Данная методическая ошибка  $\Delta r$  уменьшается с увеличением размера тестовой выборки (разность соответствующих значений левого и правого столбцов табл. 1 стремится к нулю).

### Построчная корректировка методических ошибок

Рассмотрим метод коррекции методической ошибки. Если рассматривать строки табл. 1, то можно заметить, что значения элементов строки удобно аппроксимировать гиперболой степени 1,5:

$$\Delta r(n) = \left( \alpha + \frac{\lambda}{n^{1,5}} \right), \quad (6)$$

где  $n$  — число опытов в знаменателе гиперболы;  $\alpha$  и  $\lambda$  — подбираемые при аппроксимации коэффициенты.

Мультипликативная корректировка методической ошибки может быть выполнена на основе выражения

$$\tilde{r}(n) = r(n) \left( 1 + \alpha + \frac{\lambda}{n^{1,5}} \right), \quad (7)$$

где  $r(n)$  — значение коэффициента корреляции, вычисленное по формуле (2),  $\tilde{r}(n)$  — значение коэффициента корреляции с учетом компенсации методической ошибки. Результаты аппроксимации данных гиперболой приведены в табл. 2.

Примеры реальных графиков методических ошибок (сплошные линии) и их приближений (пунктирные линии), полученные с использованием данных табл. 2, приведены на рис. 3.

В итоге можно отметить, что корректировка методической погрешности вычислений возможна как при использовании соответствующих корректирующих таблиц, так и при использовании аналитической аппроксимации в виде гипербол (6), (7).

### Корректировка методической ошибки при фиксированных объемах тестовой выборки

Рассмотрим другой способ корректировки методической ошибки, основанный на аппроксимации значений столбцов табл. 1. На рис. 4 приведены соответствующие графики.

Приближение данных численного моделирования (пунктирные линии рис. 4) осуществлено с использованием взвешенной функции бета-распределения значений [8]:

$$\Delta r = \text{sign}(r) \frac{\lambda(\alpha + \beta + 1)!}{\alpha! \beta!} |r|^\alpha (1 - |r|)^\beta, \quad (8)$$

где  $\lambda, \alpha, \beta$  — подбираемые при аппроксимации параметры,  $\text{sign}(r)$  — знак модуля учитываемого значения коэффициента парной корреляции.

Данные о параметрах аппроксимации методической ошибки вида (8) для разных размеров выборки  $n$  приведены в табл. 3.

Из рис. 4 видно, что корректировка с использованием аппроксимации (8) позволяет уменьшить

Таблица 2

Данные аппроксимации методической ошибки гиперболой при объемах выборки от 7 до 64 примеров

| $\tilde{r}(n)$ | 0,01 | 0,1    | 0,2    | 0,3    | 0,4   | 0,5   | 0,6   | 0,7   | 0,8   | 0,9    | 0,99  |
|----------------|------|--------|--------|--------|-------|-------|-------|-------|-------|--------|-------|
| $\alpha$       | 0,00 | 0,0008 | 0,0013 | 0,0008 | 0,001 | 0,002 | 0,002 | 0,002 | 0,002 | 0,0003 | 0,00  |
| $\lambda$      | 0,00 | 0,18   | 0,29   | 0,45   | 0,5   | 0,6   | 0,6   | 0,6   | 0,52  | 0,3    | 0,035 |

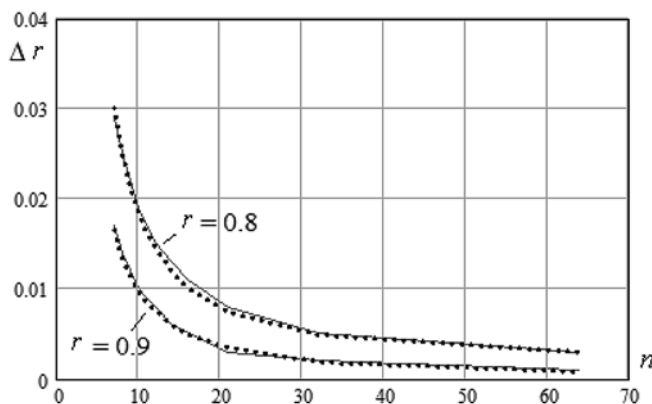


Рис. 3. Примеры графиков методических ошибок (аппроксимация дана пунктиром) как функции размеров тестовой выборки

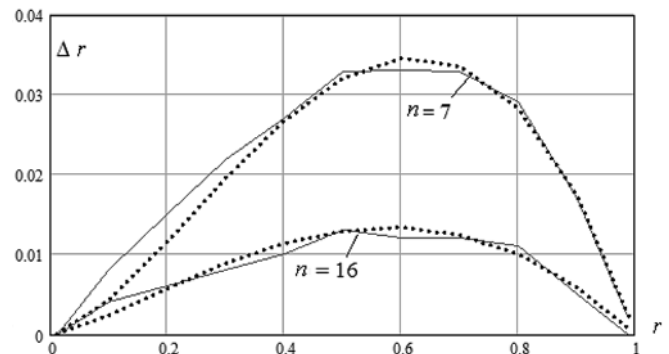


Рис. 4. Примеры графиков методических ошибок (аппроксимация дана пунктиром) как функции значений коэффициентов корреляции

Данные аппроксимации методической ошибки взвешенным бета-распределением для разных объемов выборки  $n$ 

| $n$       | 7     | 8     | 9      | 10     | 11    | 12     | 14    | 16     | 21     | 32     | 64     |
|-----------|-------|-------|--------|--------|-------|--------|-------|--------|--------|--------|--------|
| $\lambda$ | 0,021 | 0,018 | 0,0166 | 0,0147 | 0,013 | 0,0114 | 0,010 | 0,0084 | 0,0061 | 0,0043 | 0,0021 |
| $\alpha$  | 2,45  | 2,45  | 2,45   | 2,45   | 2,45  | 2,45   | 2,45  | 2,45   | 2,45   | 2,45   | 2,45   |
| $\beta$   | 1,86  | 1,88  | 1,90   | 1,93   | 1,96  | 1,98   | 1,99  | 2,11   | 2,04   | 2,06   | 2,08   |

методическую ошибку от 5 до 10 раз, если далее применять аддитивную коррекцию:

$$\tilde{r}(n) = r(n) + \text{sign}(r) \frac{\lambda(\alpha + \beta + 1)!}{\alpha! \beta!} |r(n)|^\alpha (1 - |r(n)|)^\beta. \quad (9)$$

Объединение корректора методической погрешности по строкам табл. 1 в виде соотношения (7) с корректором методической ошибки по столбцам табл. 1 в виде выражения (9) позволяет получить двухмерный цифровой корректор методических погрешностей определения коэффициентов корреляции, вычисляемых на малых тестовых выборках объемом от 7 до 64 примеров. На данный момент сравнение цифровых корректоров, полученных сплайн-аппроксимацией исходной табл. 1 с аналитическими приближениями (7) и (9), не проведено. Необходимы дополнительные исследования.

### Заключение

Чем меньше выборка примеров, представляющих биометрический образ, тем больше амплитуда шумов квантования (см. рис. 2). Кажется, что это тупик и обучать преобразователи биометрия—код на малых выборках невозможно. На самом деле это не так. Люди и животные способны обучаться на очень малом числе примеров нового образа. Это связано с тем, что их естественные нейронные сети обладают очень большой входной размерностью и способны учитывать до 10 000 параметров. Чем выше размерность решаемой задачи, тем эффективнее оказываются подавление случайных составляющих погрешностей наблюдаемых значений параметров нейросетевыми преобразователями биометрия—код [5–7]. В данной статье показано, что учет множества парных коэффициентов корреляции теоретически позволяет поднять размерность решаемой нейронной сетью задачи с 400 до 80 000 контролируемых параметров. Такой 20-кратный рост размерности решаемой задачи должен в конечном итоге обеспечить значительное повышение подавления случайных составляющих погрешностей исходных данных.

Приведенные в данной статье примеры показывают, что методическая составляющая погрешности вычисления коэффициентов корреляции может быть снижена от 5 до 10 раз. После корректировки методических составляющих погрешности доминирующей оказывается случайная погрешность,

которая подавляется сглаживанием данных малой выборки по каждому контролируемому параметру [10] или по всей совокупности учитываемых параметров за счет применения рекордно устойчивого алгоритма обучения искусственных нейронных сетей [6], обладающего к тому же линейной вычислительной сложностью. Применять иные алгоритмы обучения нельзя, так как они имеют более высокую вычислительную сложность и неустойчивы. На данный момент только алгоритм обучения ГОСТ Р 52633.5—2011 [6] способен обучать большие нейронные сети, имеющие от 400 до 80 000 входов на малых выборках (от 12 до 21 примера) за приемлемое время.

### Список литературы

1. **Dodis Y., Reyzin L., Smith A.** Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy // EUROCRYPT, April 13, 2004. P. 523–540.
2. **Monrose F., Reiter M., Li Q., Wetzel S.** Cryptographic key generation from voice // Proc. IEEE Symp. on Security and Privacy, pp. 202–213, 2001.
3. **Ramírez-Ruiz J., Pfeiffer C., Nolasco-Flores J.** Cryptographic Keys Generation Using FingerCodes // Advances in Artificial Intelligence — IBERAMIA-SBIA 2006 (LNCS 4140), p. 178–187, 2006.
4. **Hao F., Anderson R., Daugman J.** Crypto with Biometrics Effectively // IEEE TRANSACTIONS ON COMPUTERS. 2006. Vol. 55, N. 9. P. 1073–1074.
5. **Волчихин В. И., Иванов А. И., Фунтиков В. А.** Быстрые алгоритмы обучения нейросетевых механизмов биометрико-криптографической защиты информации. Пенза: Изд-во Пензенского государственного университета, 2005. 273 с.
6. **ГОСТ Р 52633.5—2011.** Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия — код доступа.
7. **Иванов А. И., Ложников П. С., Качайкин Е. И.** Идентификация подлинности рукописных автографов сетями Байеса—Хэмминга и сетями квадратичных форм // Вопросы защиты информации, 2015. № 2. С. 28–34.
8. **Кобзарь А. И.** Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2006. 816 с.
9. **Болл Руд, Коннел Джонатан Х., Панканти Шарат, Ратха Налини К., Сеньор Эндрю У.** Руководство по биометрии / Пер. с англ. М.: Техносфера, 2007. 368 с.
10. **Серикова Н. И., Иванов А. И., Качалин С. В.** Биометрическая статистика: сглаживание гистограмм, построенных на малой обучающей выборке // Вестник СибГАУ. 2014. № 3 (55). С. 146–150.
11. **Волчихин В. И., Иванов А. И., Серикова Н. И., Фунтикова Ю. В.** Эффект снижения размера тестовой выборки за счет перехода к многомерному статистическому анализу биометрических данных // Изв. высших учебных заведений. Поволжский регион. Технические науки. Пенза: ПГУ. 2015, № 2. С. 86–91.
12. **Шальгин А. С., Палагин Ю. И.** Прикладные методы статистического моделирования. Л.: Машиностроение, 1986. 320 с.

**V. P. Kulagin**, Professor, Head of Research Laboratory, e-mail vkulagi@hse.ru,  
Moscow Institute of Electronics and Mathematics, Higher School of Economics, Russia, Moscow  
**A. I. Ivanov**, Associate Professor, Head of Laboratory, JSC "Penza Research Electrotechnical Institute"  
Russia, Penza, (841-2) 59-33-10, e-mail ivan@pniei.penza.ru.  
**Yu. I. Serikova**, Undergraduate 1st year, FBGOU VPO "Penza State University", Russia,  
Penza, e-mail: gosh64@mail.ru

## Correction of Methodical and Casual Components of Errors of Calculation of the Coefficients of Correlation Arising on Small Selections of Biometric Data

*The paper makes it clear that calculating expectation values, standard deviations and correlation rates gives appreciable errors when using small samples. The correlation rate miscalculation far exceeds those of expectation values and standard deviations. Such errors happen all along of the given data continua quantization through their representation via small sample. We give the probability distribution plots of quantization errors and errors arising from calculating correlation rates in small samples. The allowance for association of correlation rate miscalculations with the test sample sizes makes it possible to apply for simulation modelling of several variables conditioned upon their equal mutual correlation. The table lists the expectation values for correlation rates derived from different sizes of test sample. These values indicate the presence of significant systematic errors arising while evaluating the correlation rates. Small samples show appreciable systematic error, yet decreasing rapidly with the increase in a test sample size. The paper stands for correcting systematic error in the additive and / or multiplicative form, as well as considers two approaches to correction data analytic description by adjusting the detected systematic error. The first approach to analytic description provides for approximating correction data rows by hyperbolic curves with dim fractional exponent. The second approach to analytic description provides for approximating correction data columns by applying to the analytical function describing beta distribution.*

**Keywords:** methodological error, coefficient of correlation, small sample, processing of biometric data

### References

1. **Dodis Y., Reyzin L., Smith A.** Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy, Data April 13, *EUROCRYPT*, 2004, pp. 523–540.
2. **Monrose F., Reiter M., Li Q., Wetzel S.** Cryptographic key generation from voice, *Proc. IEEE Symp. on Security and Privacy*, 2001.
3. **Ramirez-Ruiz J., Pfeiffer C., Nolzco-Flores J.** Cryptographic Keys Generation Using FingerCodes, *Advances in Artificial Intelligence — IBERAMIA-SBIA 2006* (LNCS 4140), 2006, pp. 178–187.
4. **Feng Hao, Ross Anderson, John Daugman.** Crypto with Biometrics Effectively, *IEEE TRANSACTIONS ON COMPUTERS*, 2006, vol. 55, no. 9.
5. **Volchihin V. I., Ivanov A. I., Funtikov V. A.** *Bystrye algoritmy obucheniia nei rosetevykh mehanizmov biometriko-kriptograficheskoi zashchity informacii.* Penza—2005 g. Izdatel'stvo Penzenskogo gosudarstvennogo universiteta, 2005, pp. 273.
6. **GOST R 52633.5—2011** *Zashchita informacii. Tekhnika zashchity informacii. Avtomaticheskoe obuchenie nei rosetevykh preobrazovatelei biometrii — kod dostupa.*
7. **Ivanov A. I., Lozhnikov P. S., Kachai'kin E. I.** Identifikatsiia podlinnosti rukopisnykh avtografov setiami Bai'esa-KHe'mminga i setiami kvadrachnykh form, *Voprosy zashchity informacii*, 2015, no. 2, pp. 28–34.
8. **Kobzar' A. I.** *Prikladnaia matematicheskaia statistika. Dlia inzhenerov i nauchnykh rabotneykov.* Moscow: FIZMATLIT, 2006, 816 p.
9. **Boll Rud, Konnel Jonathan Kh., Pankanti Sharat, Ratha Nalini K., Sen'or Andrew U.** *Rukovodstvo po biometrii.* Moscow: Tekhnosfera, 2007, pp. 368.
10. **Serikova N. I., Ivanov A. I., Kachalin S. V.** Biometricheskaia statistika: sglazhivanie gistogramm, postroennykh na maloi obuchaiushchei vyborke, *Vestnyk SibGAU*, 2014, no. 3 (55), pp. 146–150.
11. **Volchihin V. I., Ivanov A. I., Serikova N. I., Funtikova Iu. V.** E'fekt snizheniia razmera testovoi vyborki za schet perehoda k mnogomernomu statisticheskomu analizu biometricheskikh dannykh, *Izvestiia vysshikh uchebnykh zavedenii. Povolzhskii region. Tekhnicheskie nauki.* Penza: PGU, 2015, no. 2, pp. 86–91.
12. **Shalygin A. S., Palagin Iu. I.** *Prikladnye metody statisticheskogo modelirovaniia.* Leningrad: Mashinostroenie, 1986 g. 320 p.