ОЦЕНКА ДОСТОВЕРНОСТИ НЕЙРОСЕТЕВОЙ АВТОМАТИЗИРОВАННОЙ ЭКСПЕРТИЗЫ АВТОРСТВА РУКОПИСНОГО ПОЧЕРКА

***Качайкин Евгений Иванович**, г. Москва*
***Иванов Александр Иванович**, доктор технических наук, доцент, г. Пенза*
***Безяев Александр Викторович**, кандидат технических наук, г. Пенза*
***Перфилов Константин Александрович**, г. Пенза*

Рассматривается вопрос оценки достоверности решений, принимаемых приложением автоматизированной нейросетевой экспертизы рукописного почерка. В случае привлечения эксперта высокой квалификации для анализа авторства нет возможности оценить достоверность принимаемых экспертом решений. Иная ситуация возникает при использовании экспертизы с применением больших искусственных нейронных сетей. В этом случае дополнительные примеры рукописного слова, воспроизведенные автором и другими людьми, могут быть использованы для оценки вероятности ошибок первого и второго рода. Новые возможности возникают из-за того, что искусственная нейронная сеть Пирсона-Хэмминга способна обрабатывать большие объемы входных данных, преобразуя их в длинный код идентификации авторства.

***Ключевые слова**: почерковедение, достоверность, биометрические параметры, криминалистические приложения, нейронная сеть, экпертиза.*

RELIABILITY ESTIMATION OF THE AUTOMATED NEURAL NETWORK EXPERTISE OF AUTHORSHIP HAND-WRITTEN HANDWRITING

***Evgeny Kachaykin**, Moscow*
***Alexander Ivanov**, Doctor of Science (Tech), Associate Professor, Penza*
***Alexander Bezyaev**, Ph.D., Penza*
***Konstantin Perfilov**, Penza*

The question of an estimation of reliability of the decisions accepted by the appendix of automated neural network examination of hand-written handwriting is considered. In case of attraction of the expert of high qualification for the analysis of authorship there is no possibility to estimate reliability of decisions accepted by the expert. Other situation arises, at use of examination with application of the big artificial neural networks. In this case, the additional examples of a hand-written word reproduced by the author and other people can be used for an estimation of probability of errors of the first and second sort. New possibilities arise that the artificial neural network of Pirson-Hemming is capable to process great volumes of the entrance data, converting them in a long code of identification of authorship.

***Keywords**: graphology, reliability, biometrics, forensic applications, neural network, expertise.*

Введение

Почерковедческое исследование документов является одной из наиболее популярных экспертиз в гражданских и арбитражных судебных спорах. На данный момент экспертизу осуществляет человек, имеющий значительный опыт работы и анализирующий порядка 16 измеряемых биометрических параметров [1, 2]. Обращение к услугам человека-эксперта наряду с множеством положительных моментов обладает рядом недостатков. Во-первых, высококвалифицированный эксперт достаточно сильно загружен (для проведения экспертизы необходимы достаточно большие затраты времени), во-вторых, эксперт не дает оценок достоверности результатов осуществленной им экспертизы.

Следует подчеркнуть, что наряду с криминали-

стическими приложениями проверки авторства рукописного текста в 21 веке активно развиваются средства нейросетевой биометрической аутентификации по динамике рукописного слова-пароля [3, 4]. Задачи, решаемые при криминалистической экспертизе авторства рукописного слова (например, автографа или резолюции под документом) и биометрической аутентификации по рукописному слову-паролю похожи. Отличие состоит только в получении биометрических данных. При криминалистической почерковедческой экспертизе используются статические рукописные образы, оставленные на бумажном документе, а при биометрической аутентификации используются динамические данные о скоростях и ускорениях движения пера при воспроизведении человеком его автографа или рукописного слова-пароля.

Если анализируемый статический рукописный образ на документе отсканировать, то становится возможным выделить псевдодинамику его воспроизведения путем обхода траектории его воспроизведения [5] с постоянной скоростью. При этом мы получаем сотни низкоуровневых биометрических параметров, например, в виде коэффициентов двухмерного преобразования Фурье для пары функций $Y(t), X(t)$.

Кроме коэффициентов псевдодинамики могут быть выделены графемы (элементы рукописных знаков) [6], сочетания которых, в свою очередь, могут быть преобразованы в сотни статических биометрических параметров рукописных образов. Независимо от того, как получены сотни биометрических параметров статического автографа, эти данные могут быть использованы для обучения большой искусственной нейронной сети [7], которая позднее может быть использована при нейросетевой почерковедческой экспертизе.

Использование радиальных сетей Пирсона-Хэмминга при экспертизе рукописных автографов

Независимо от того, каким образом получены биометрические данные рукописных образов при экспертизе авторства, необходимо использовать какую-то нейронную сеть. В частности, может быть использована нейронная сеть, состоящая из 256 персептронов [8] и обученная по ГОСТ Р 52633.5 [9]. Еще одним вариантом может быть использование радиальной нейронной сети, работающей с 416 биометрическими параметрами, полученными из среды моделирования «БиоНейроАвтограф» [8]. В частности, может быть использована радиальная сеть Пирсона-Хэмминга, имеющая 256 радиальных нейронов. При этом каждый нейрон целесообразно строить случайным выбором половины входных биометрических данных (по 208 биометрических параметров). Обучим радиальную сеть по правилу Пирсона:

$$y_k(\overline{\nu}) = \frac{1}{208}\sum_{i=1}^{208}\frac{(E(\nu_i)-\nu_i)^2}{(\sigma(\nu_i))^2} \qquad (1),$$

где y_k – отклик, на выходе сумматора k-го радиального нейрона, $(E(\nu_i)$ – математическое ожидание i-го биометрического параметра, $\sigma(\nu_i)$ – стандартное отклонение -го биометрического параметра.

Для данных среды моделирования «БиоНейроАвтограф» [8] отклики на рукописный образ автографа «Свой» (рисунок 1), воспроизведенного рукой одного человека, изменяются в интервале от 0 до 6.08. То есть, автограф «Свой» следует признавать подлинным, если отклики всех 256 сумматоров радиальных нейронов или большей их части будут менее порога в 6.08.

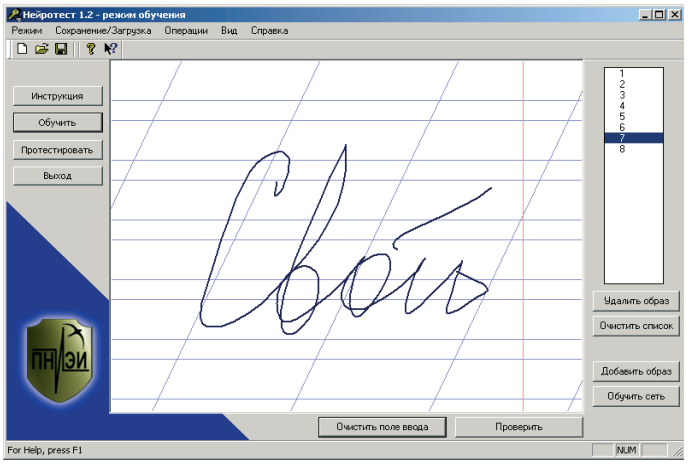

***Рис. 1** Экранная форма 7-го примера рукописного автографа «Свой» при обучении нейросети в среде моделирования «БиоНейроАвтограф» [8]*

Оценка вероятности ошибок первого и второго рода для одного нейрона

Основным преимуществом сетей Пирсона-Хэмминга является то, что распределение данных на выходах каждого радиального нейрона описывается зависимым хи-квадрат распределением [10] с дробным (фрактальным) числом степеней свободы. Число степеней свободы для подлинников автографа «Свой» и попыток подделки чужим оказываются разными и могут быть оценены как соответствующие математические ожидания биометрических данных:

$$\begin{cases} m_{k,\text{свой}} = E(y_{k,\text{свой}}), \\ m_{k,\text{чужой}} = E(y_{k,\text{чужой}}). \end{cases} \quad (2)$$

Свойство (2) выполняется для всех хи-квадрат распределений как с целыми, так и с дробными показателями числа степеней свободы-m. Численный эксперимент, проведенный в среде моделирования «БиоНейроАвтограф» [8] для данного рукописного образа «Свой» дал m=2.49, а воспроизведение другим человеком дает значение m=9.08. Соответствующие кривые хи-квадрат распределений для независимых данных приведены на рисунке 2.

Следует отметить, что приведенные кривые соответствуют неверной гипотезе об их независимости. В рамках гипотезы независимости плотности распределения хи-квадрат аналитически описываются через гамма - функции:

$$p(x^2) = \frac{1}{2^{\frac{m}{2}} \cdot \Gamma\left\{\frac{m}{2}\right\}} \cdot h^{\left\{\frac{m}{2}-1\right\}} \cdot \exp\left\{\frac{-x^2}{2}\right\} \quad (3)$$

Если бы гипотеза независимости была бы верной, то стандартное отклонение данных и число степеней свободы были бы связаны очень простой зависимостью:

$$\begin{cases} \sigma(y_{k,\text{свой}}) = 2m_{k,\text{свой}} = 2E(y_{k,\text{свой}}), \\ \sigma(y_{k,\text{чужой}}) = 2m_{k,\text{чужой}} = 2E(y_{k,\text{чужой}}). \end{cases} \quad (4)$$

Для реальных данных соотношение (4) не выполняется из-за того, что они оказываются сильно коррелированными.

Переход к хи-квадрат распределениям зависимых данных с целыми показателями чисел степеней свободы

Из теории известно [11], что моделировать случайные многомерные процессы крайне сложно. Технически вполне возможно вычислить симметричную матрицу корреляционных связей 208x208, описывающую корреляционные связи между биометрическими входными данными нейросетевого преобразователя. Однако построить генератор столь высокой размерности технически невозможно.

Формально можно использовать 208 генераторов независимых случайных данных, умножив их некоторую связывающую данные матрицу A. Однако найти нужную связывающую матрицу A, которая даст нужные корреляционные связи $r(\text{“}y_k\text{“},\text{“}y_j\text{“})$ трудно. Эта обратная задача относится к плохо обусловленным.

Так как задача не решается, ее нужно симметризовать. Для этой цели необходимо использовать симметричную связывающую матрицу, которая имеет единичную диагональ и одинаковые элементы вне диагонали:

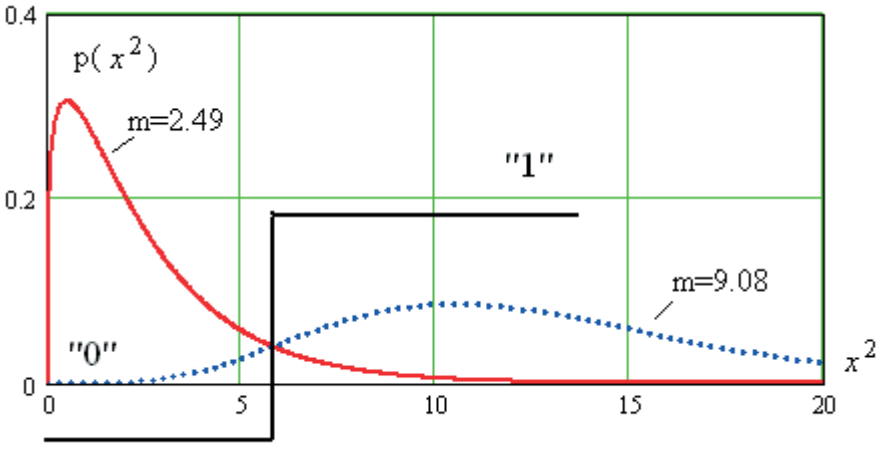

Рис. 2. *Кривые хи-квадрат распределений на выходе сумматора одного из 256 радиально базисных нейронов*

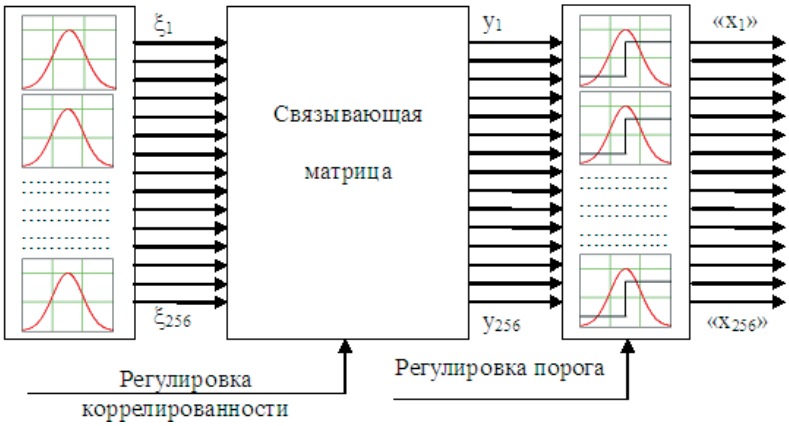

Рис. 3. *Блок-схема моделирования симметричных равно коррелированных биометрических данных и равно коррелированных кодов*

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

$$\Rightarrow \quad R_n = \begin{bmatrix} 1 & \tilde{r} & \cdots\cdots & \tilde{r} \\ \tilde{r} & 1 & \cdots\cdots & \tilde{r} \\ \cdots & \cdots & \cdots\cdots & \cdots \\ \tilde{r} & \tilde{r} & \cdots\cdots & 1 \end{bmatrix} \qquad (5)$$

В этом случае данные оказываются равно коррелированными. Если плавно изменять регулируемый параметр связывающей матрицы от 0 до 1, равная коррелированность также меняется от 0 до 1. Умножение непрерывных данных (континуумов) на связывающую матрицу порождает вектор непрерывных откликов – $\overline{y}$. Для того, что бы непрерывные данные преобразовать в дискретные данные необходимо использовать 208 компараторов. Блок-схема моделирования симметричных равно коррелированных биометрических данных и соответствующих им равно коррелированных кодов приведена на рисунке 3.

Следует отметить, что применение схемы моделирования рисунка (3) приводит к тому, что число степеней свободы хи-квадрат распределений зависимых данных остается целым m = 1, 2, 3, 4, ….. Примеры подобных распределений 3 и 4 степеней свободы даны на рисунке 4.

Пользуясь тем, что в блок-схеме моделирования (3) используется только одна регулировка (меняются одинаковые параметры связывающей матрицы), мы можем плавно менять кривые рас-

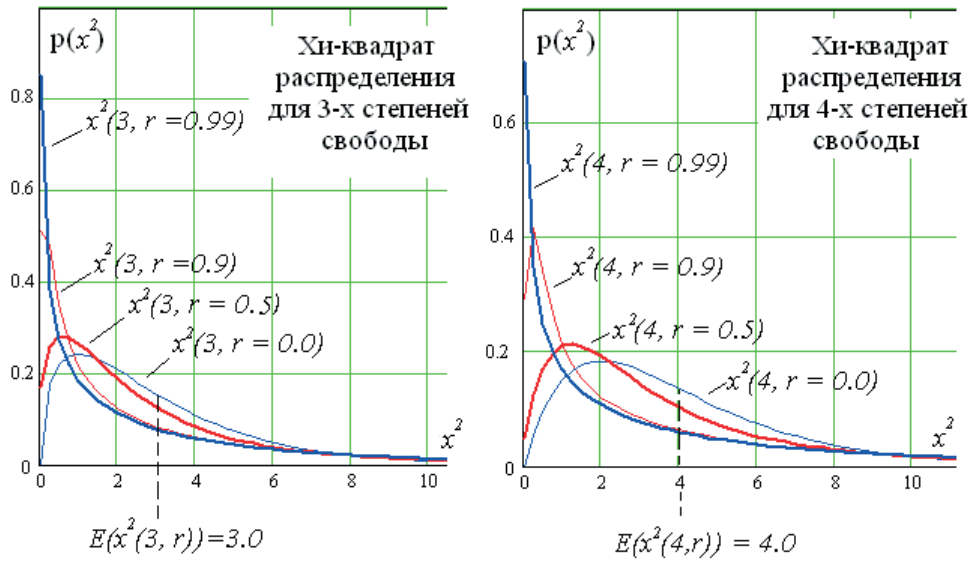

Рис. 4. *Кривые плотности χ^2 распределения для трех и четырех степеней свободы, полученные для разных значений коррелированности данных.*

пределений зависимых данных хи-квадрат и соответствующие им стандартные отклонения. Регулировка ведется до того момента, пока не появится полученное экспериментально значение $\sigma(y_k)$.

Переход к распределениям с дробным показателем степеней свободы зависимых биометрических данных

Для перехода к дробным (фрактальным) показателям необходимо учитывать расстояния до ближайших целых чисел степеней свободы [10]:

$$p_{\chi^2}(m_k, r) = (m_k) - a_0) \cdot p_{\chi^2}(a_0, m_k, r) + \\ + (a_0 + 1 - m_k) \cdot p_{\chi^2}((a_0 + 1)\ m_k, r) \quad (6)$$

где параметр a_0 – это ближайшее целое число, меньшее или равное вычисленному значению - $E(y_k) = m_k$;

$$a_0 = \text{floor}\ (\ m_k\ )) \quad (7)$$

где операция floor (.) отбрасывает дробную часть числа - m_k.

Так для числа степеней свободы $m = 2.49 \approx 2.5$ нужно построить семейство кривых с разным уровнем корреляции данных для $m = 2$ и такое же семейство кривых для $m = 3$ (смотри левую часть рисунка 4). Для того, чтобы получить распределения хи-квадрат для числа степеней свободы 2.5 потребуется усреднить кривые, полученные для 2 и 3 степеней свободы. Усреднение работает, так как дробный показатель числа степеней свободы 2.5 оказывается одинаково удален от ближайших целых чисел показателя степени свободы.

Заключение

Таким образом, мы научились строить хи-квадрат распределения для зависимых биометрических данных, то есть мы можем решить задачу оценки вероятности ошибок первого и второго рода почерковедческой нейросетевой экспертизы для каждого из нейронов сети Пирсона-Хэмминга. Следует подчеркнуть, что такая уникальная возможность возникает только для радиально базисных нейронных сетей Пирсона-Хэмминга. Только для этого класса сетей точно известная функция распределения значений выходных данных (хи-квадрат-функция зависимых данных). Для любых других нейросетевых решений статистики выходных данных будут описываться своими (неизвестными заранее) законами распределения значений.

Еще одним важнейшим преимуществом сетей Пирсона-Хэмминга является то, что их результат легко интерпретируется. В идеале рукописный образ «Свой» должен давать кодовый отклик, состоящий из 256 нулей. Чем больше единиц будет в выходном коде, тем выше вероятность, что анализируемый образ является подделкой.

Литература:

1. Почерковедение и почерковедческая экспертиза / под ред. Серегина. В.В. — Волгоград: ВА МВД России, 2002. — ISBN 5-7899-0234-0
2. Ищенко Е.П., Топорков А.А.. Криминалистика. — М.: Контракт, 2006. — ISBN 5-900785-58-0
3. Ахметов Б.С., Иванов А.И., Фунтиков В.А., Безяев А.В., Малыгина Е.А. Технология использования больших нейронных сетей для преобразования нечетких биометрических данных в код ключа доступа. Монография, Казахстан, г. Алматы, ТОО «Издательство LEM», 2014 г. -144 с., находится в открытом доступе (http://portal.kaznu.kz/files/publicate/2014-06-27-11940.pdf)
4. Ахметов Б.С., Надеев Д.Н., Фунтиков В.А., Иванов А.И., Малыгин А.Ю. Оценка рисков высоконадежной биометрии. Монография. Алматы: Из-во КазНТУ им. К.И. Сатпаева, 2014 г.- 108 с.
5. Иванов А.И., Андреев Д.Ю., Воячек С.А., Елфимов А.В. Описание патента RU 2390843 «Способ распознавания знаков». МКИ: G06K 9/62. Заявка: 2008117180/09 от 29.04.2008. Опубликовано: 27.05.2010 Бюл. № 15.
6. Качайкин Е.И., Андреев Д.Ю. Алгоритм выделения графических примитивов на изображении рукописного текста. Пенза-2014, Том 9, с. 55-57, Трудов конференции «БИТ» (http://www.pniei.penza.ru/RV-conf/T9/c.55).
7. Елфимов А.В., Воячек С.А., Качайкин Е.И., Куликов С.В. Обучение нейросетевого идентификатора авторства рукописных текстов // Нейрокомпьютеры: разработка, применение. 2009. № 6. С. 17–21.
8. Среда моделирования «БиоНейроАвтограф» размещена на сайте ОАО «ПНИЭИ» http://пниэи.рф/activity/science/noc.htm. Продукт создан лабораторией биометрических и нейросетевых технологий ОАО «ПНИЭИ» в период 2009-2014 г.г. для свободного использования университетами России, Белоруссии, Казахстана.
9. ГОСТ Р 52633.5-2011 «Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа».

10. Безяев А.В., Иванов А.И., Фунтикова Ю.В. Оптимизация структуры самокорректирующегося био-кода, хранящего синдромы ошибок в виде фрагментов хеш-функций. «Вестник Уральского федерального округа. Безопасность в информационной сфере», 2014 г. № 3(13) с. 4-14.
11. Шалыгин А.С., Палагин Ю.И. Прикладные методы статистического моделирования. Л.: Машиностроение, 1986 г., 320 с.

References:

1. Pocherkovedenie i pocherkovedcheskaya ekspertiza / pod red. Seregina. V.V. — Volgograd: VA MVD Rossii, 2002. — ISBN 5-7899-0234-0
2. Ischenko E.P., Toporkov A.A. Kriminalistika. — M.: Kontrakt, 2006. — ISBN 5-900785-58-0
3. Ahmetov B.S., Ivanov A.I., Funtikov V.A., Bezyaev A.V., Malyigina E.A. Tehnologiya ispolzovaniya bolshih neyronnyih setey dlya preobrazovaniya nechetkih biometricheskih dannyih v kod klyucha dostupa. Monografiya, Kazahstan, g. Almatyi, TOO «Izdatelstvo LEM», 2014 g. -144 c., nahoditsya v otkryitom dostupe (http://portal.kazntu.kz/files/publicate/2014-06-27-11940.pdf)
4. Ahmetov B.S., Nadeev D.N., Funtikov V.A., Ivanov A.I., Malyigin A.Yu. Otsenka riskov vyisokonadezhnoy biometrii. Monografiya. Almatyi: Iz-vo KazNTU im. K.I. Satpaeva, 2014 g.- 108 s.
5. Ivanov A.I., Andreev D.Yu., Voyachek S.A., Elfimov A.V. Opisanie patenta RU 2390843 «Sposob raspoznavaniya znakov». MKI: G06K 9/62. Zayavka: 2008117180/09 ot 29.04.2008. Opublikovano: 27.05.2010 Byul. № 15.
6. Kachaykin E.I., Andreev D.Yu. Algoritm vyideleniya graficheskih primitivov na izobrazhenii rukopisnogo teksta. Penza-2014, Tom 9, s. 55-57, Trudov konferentsii «BIT» (http://www.pniei.penza.ru/RV-conf/T9/S55).
7. Elfimov A.V., Voyachek S.A., Kachaykin E.I., Kulikov S.V. Obuchenie neyrosetevogo identifikatora avtorstva rukopisnyih tekstov // Neyrokompyuteryi: razrabotka, primenenie. 2009. № 6. S. 17–21.
8. Sreda modelirovaniya «BioNeyroAvtograf» razmeschena na sayte OAO «PNIEI» http://пниэи.рф/activity/science/noc.htm. Produkt sozdan laboratoriey biometricheskih i neyrosetevyih tehnologiy OAO «PNIEI» v period 2009-2014 g.g. dlya svobodnogo ispolzovaniya universitetami Rossii, Belorussii, Kazahstana.
9. GOST R 52633.5-2011 «Zaschita informatsii. Tehnika zaschityi informatsii. Avtomaticheskoe obuchenie neyrosetevyih preobrazovateley biometriya-kod dostupa».
10. Bezyaev A.V., Ivanov A.I., Funtikova Yu.V. Optimizatsiya strukturyi samokorrektiruyuschegosya bio-koda, hranyaschego sindromyi oshibok v vide fragmentov hesh-funktsiy. «Vestnik Uralskogo federalnogo okruga. Bezopasnost v informatsionnoy sfere» 2014 g. № 3(13) s. 4-14.
11. Shalyigin A.S., Palagin Yu.I. Prikladnyie metodyi statisticheskogo modelirovaniya. L.: Mashinostroenie, 1986 g., 320 s.