ВЕСТНИК
ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
МОРСКОГО И РЕЧНОГО ФЛОТА ИМЕНИ АДМИРАЛА С. О. МАКАРОВА

АВТОМАТИЗАЦИЯ И УПРАВЛЕНИЕ ТЕХНОЛОГИЧЕСКИМИ ПРОЦЕССАМИ И ПРОИЗВОДСТВАМИ

DOI: 10.21821/2309-5180-2023-15-3-513-523

ALGORITHM FOR AUTOMATION OF THE METHOD OF TESTING STATISTICAL HYPOTHESES ACCORDING TO THE PEARSON CRITERION BY THE MEANS OF MATLAB

A. A. Chertkov

Admiral Makarov State University of Maritime and Inland Shipping,
St. Petersburg, Russian Federation

The purpose of the work is to develop and improve the methods of computer modeling and digitalization of procedures for testing statistical hypotheses according to consent criteria for use in high-tech software and measuring complexes of automated control systems. In most cases that are important for practice, the analysis of the statistical properties of distributions requires multiple checks performed using numerical methods, which require the tools of powerful software and computing environments. Examples of such cases are the procedures for finding information subsets of features in regression and statistical analysis, as well as identifying parameter deviations in automated control systems. It is noted that the tasks of processing experimental data do not always fit into the framework of the theory of normal processes and have exclusively such hypotheses. To test such hypotheses, modern mathematical statistics has developed criteria based on specially developed probabilistic rules that distinguish between zero and alternative hypotheses. Moreover, if for the testable (zero) hypothesis the distribution of statistics is known exactly or approximately, then for the alternative it has a greater degree of uncertainty and it (the alternative) itself is not a complete negation of the null hypothesis. As the zero and alternative hypotheses move away from each other, the statistics of the consent criterion begins to take a larger absolute value than with the null hypothesis, which allows us to build a critical area on the basis of estimating the set of this statistics values. However, the use of analytical methods for statistical analysis to determine the discrepancy between empirical and hypothetical statistics in practice is associated with the use, as a rule, of complex computational calculation algorithms and the implementation of high requirements for the assessment accuracy, which they cannot provide. To solve this problem, it is proposed to use numerical methods of computer modeling and integration of probability density of distribution according to the Simpson algorithm using the tools of built-in functions of the Statistics Toolbox of the MATLAB environment. The proposed method for automation of statistical analysis of zero and alternative hypotheses is implemented on the example of applying the Pearson criterion to verify compliance with the normal distribution of two random data samples, one of which meets the normal distribution law. The obtained graphic and numerical results of computer modeling confirm the correspondence of one of the testable distributions to a given normal distribution according to the null hypothesis.

Keywords: statistical hypothesis, consent criterion, distribution law, critical area, significance level, sample value, probability density.

For citation:
Chertkov, Alexandr A. "Algorithm for automation of the method of testing statistical hypotheses according to the Pearson criterion by the means of MATLAB." *Vestnik Gosudarstvennogo universiteta morskogo i rechnogo flota imeni admirala S. O. Makarova* 15.3 (2023): 513–523. DOI: 10.21821/2309-5180-2023-15-3-513-523.

УДК 681.5

АЛГОРИТМ АВТОМАТИЗАЦИИ МЕТОДА ПРОВЕРКИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ ПО КРИТЕРИЮ ПИРСОНА СРЕДСТВАМИ MATLAB

А. А. Чертков

ФГБОУ ВО «ГУМРФ имени адмирала С. О. Макарова»,
Санкт-Петербург, Российская Федерация

Целью исследования является совершенствование методов компьютерного моделирования и цифровизации процедур проверки статистических гипотез по критериям согласия для последующего использования в высокотехнологичных программно-измерительных комплексах автоматизированных систем управления. При этом в большинстве важных для практики случаев анализ статистических свойств распределений требует проведения множественных проверок, выполняемых с применением численных методов, для которых необходим инструментарий мощных программно-вычислительных сред. Примерами таких случаев являются процедуры нахождения информативных подмножеств признаков в регрессионном и статистическом анализе, а также выявление отклонений параметров в автоматизированных системах управления. Отмечается, что задачи обработки экспериментальных данных не всегда вписываются в рамки теории нормальных процессов, обладая признаками исключительно такого рода гипотез, для проверки которых в современной математической статистике разработаны критерии, основанные на специально разработанных вероятностных правилах, различающих проверяемую (нулевую) и альтернативную гипотезы. Причем, если для проверяемой гипотезы распределение статистики известно точно или приближенно, то для альтернативной оно обладает большей степенью неопределенности и сама она не является полным отрицанием нулевой гипотезы. По мере удаления нулевой и альтернативной гипотез друг от друга статистика критерия согласия начинает принимать большие по абсолютной величине значения, чем при нулевой гипотезе, что позволяет на основе оценки множества значений этой статистики строить критическую область. Однако использование аналитических методов статистического анализа для определения расхождения эмпирической и гипотетической статистик на практике сопряжено с применением, как правило, сложных вычислительных алгоритмов расчета и реализацией высоких требований к точности оценки, которые они обеспечить не могут. Для решения этой проблемы в работе предлагается использование численных методов компьютерного моделирования и интегрирования плотности вероятностей распределения по алгоритму Симпсона с применением инструментария встроенных функций Statistics Toolbox среды MATLAB. Предложенный метод автоматизации статистического анализа нулевой и альтернативной гипотез реализован на примере применения критерия Пирсона для проверки соответствия нормальному распределению двух случайных выборок данных, одна из которых отвечает нормальному закону распределения. Полученные графические и численные результаты компьютерного моделирования подтверждают соответствие одного из проверяемых распределений заданному нормальному распределению согласно нулевой гипотезе.

Ключевые слова: статистическая гипотеза, критерий согласия, закон распределения, критическая область, уровень значимости, выборочное значение, плотность вероятностей.

Для цитирования:
Чертков А. А. Алгоритм автоматизации метода проверки статистических гипотез по критерию Пирсона средствами MATLAB / А. А. Чертков // Вестник Государственного университета морского и речного флота имени адмирала С. О. Макарова. — 2023. — Т. 15. — № 3. — С. 513–523. DOI: 10.21821/2309-5180-2023-15-3-513-523.

Введение (Introduction)

Анализ статистических свойств распределений требует проведения множественных проверок статистических гипотез, составляющих часть актуальной проблемы сопряжения (стыковки) статистических процедур [1]–[3]. Причем при обработке однородных групп данных наблюдений, не представляющих всей их совокупности, наличие взаимной корреляции среди них приводит к тому, что закон их внутригруппового распределения не будет строго нормальным (гауссовым). С увеличением объема выборки корреляция будет уменьшаться, но ненормальность сохранится. При этом групповое распределение будет стремиться не к нормальному, а к усеченному нормальному распределению. Использование в этом случае алгоритмов регрессионного анализа [4] будет некорректно. С точки зрения практики статистического анализа аналитические методы [5]–[6] вычисления вероятностей (плотности вероятностей) распределения с применением операций интегрирования (дифференцирования) не реализуются в явном виде в средствах цифровой обработки данных. Они могут быть реализованы в них только косвенно (в неявном виде) через операции суммирования / вычитания с аппроксимацией функций в интервалах приближения с применением, как правило, сложных вычислительных алгоритмов в пошаговом режиме.

В связи с этим большой интерес при анализе статистических свойств распределений вызывают *численные методы* [7]–[9], которые до появления мощных вычислительных средств находили в математической статистике ограниченное применение. Развитый инструментарий этих сред в большинстве важных для практики случаев позволяет применять весь каталог статистических функций Statistics Toolbox [10] и решать классические задачи статистического анализа [3] с максимальной эффективностью. С помощью численного моделирования можно получить наиболее надежный и относительно простой в реализации вычислительный аппарат для исследования законов распределений и их изменчивости в зависимости от различных факторов и оценок формирования результатов наблюдений. Как известно из [11], на основе компьютерного моделирования можно строить для конкретной ситуации модели распределений любой исследуемой статистики, выявляя закономерности их изменения с ростом объемов выборок [1] и изменением размерности данных.

Проверка статистических гипотез на основе параметрических или непараметрических критериев представляет одну из основных задач [4], [12] статистического анализа. Параметрические критерии применяются при полностью определенном гипотетическом распределении для проверки гипотез о равенстве средних значений двух выборок по критерию Стьюдента (t-критерию) или их дисперсий по критерию Фишера. Причем без предварительной проверки вида распределения применять параметрические критерии нецелесообразно из-за возможных ошибок при проверке основных гипотез, поскольку полностью определенное распределение на практике встречается довольно редко. В связи с этим перед применением параметрических критериев следует определиться с видом (функцией) исследуемого распределения, используя в данном случае непараметрические критерии.

Известные из [2], [3] непараметрические критерии статистики, такие как критерий знаков, критерии Колмогорова, Смирнова, Крамера – Мизеса, Андерсона – Дарлинга и др., не зависят от вида наблюдаемого закона распределения («свободны» от вида распределения) выборок (в частности, от его параметров) и характеризуются несущественной зависимостью распределения статистики от объема выборки. В данном случае отклонение реального распределения статистики от предельного незначительно и практически не оказывает влияние на результаты статистического вывода. Выбор этих критериев не требует большого объема выборки и знания закона распределения, но связан с выполнением ряда условий [13], [14]. В связи с этим они не являются универсальными критериями. Наиболее известным универсальным критерием согласия является *критерий Пирсона*[1] χ^2 [15], [16], позволяющий получить оценку значимости различий между фактическим (выявленным) количеством исходов, попадающих в каждый интервал, и теоретическим количеством, которое можно ожидать в тех же интервалах при справедливости нулевой гипотезы. Иными словами, метод позволяет оценить статистическую значимость различий двух или нескольких относительных показателей (частот, долей).

Распределения статистик классических непараметрических критериев согласия известны и не зависят от вида наблюдаемого закона распределения (в частности, от его параметров), т. е. они являются «свободными от распределения». Это достоинство предопределило широкое использование их в различных приложениях.

Методы и материалы (Methods and Materials)

Оценка репрезентативности выборки с определенной вероятностью позволяет судить о границах генеральной выборки, сделать некоторые предположения об отдельных параметрах случайной величины, таких как математическое ожидание, дисперсия и др. Но при этом нельзя сделать вывод о характере распределения случайной величины. Полную характеристику случайной величины дает закон ее распределения. Однако об этом законе в генеральной совокупности можно получить информацию только из частичных выборок. Поэтому необходимо иметь критерии оценки соответствия статистических законов распределения, полученных из частичных выборок, теоретическим законам распределения, характеризующим генеральную совокупность. Такими критериями являются *критерии согласия*. Так как нормальное распределение встречается довольно часто, как правило, проверяют гипотезу о соответствии выборочного распределения нормальному. Из множества критериев согласия о распределениях наиболее известным является критерий χ^2 («хи-квадрат») Пирсона.

[1] Предложен и разработан в 1900 г. основателем математической статистики английским ученым К. Пирсоном для анализа таблиц сопряженности.

В ходе эксперимента нередко возникает необходимость проверки гипотезы о виде распределения исследуемой выборки. Пусть по выборке $x_1, \ldots, x_n$ из некоторой генеральной совокупности нужно проверить гипотезу о том, что генеральная совокупность имеет заданное распределение. Полагаем, что $f_\xi(x)$ — плотность вероятности генеральной совокупности, $f_0(x, \theta_1, \ldots, \theta_m)$ — гипотетическая плотность вероятности, известная с точностью до m параметров $\theta_1, \ldots, \theta_m$, где m может быть равно нулю. Необходимо проверить двухальтернативную непараметрическую сложную гипотезу:

$$\{H_0: f_\xi(x) = f_0(x, \theta_1, \ldots, \theta_m), H_1: f_\xi(x) \neq f_0(x, \theta_1, \ldots, \theta_m)\}. \quad (1)$$

Для этого критерием χ^2 множество возможных значений случайной величины ξ разбивается на r интервалов и подсчитывается количество выборочных значений W_i, попавших в каждый интервал (как при построении гистограммы). В качестве критерия согласия используется мера расхождения эмпирических частот относительно теоретических, представляющая собой величину (статистику):

$$\chi^2_{\text{набл}} = \sum_{i=1}^{r} \frac{(W_i - np_i)^2}{np_i}, \quad (2)$$

где p_i — гипотетическая вероятность попадания случайной величины ξ в i-й интервал.

Гипотетическая вероятность определяется по формуле, интегрирование в которой выполняется по i-му интервалу r:

$$p_i = \int_i f_0(x, \theta_1, \theta_2, \ldots, \theta_m)dx. \quad (3)$$

Здесь $f_0(x, \theta_1, \ldots, \theta_m)$ — гипотетическая плотность вероятности, в которую вместо неизвестных параметров подставлены их оценки $\theta_1, \ldots, \theta_m$.

Для определения числа интервалов существует несколько рекомендаций, которые сводятся, например, к использованию одной из двух зависимостей:
первой:

$$r = 1{,}87(n - 1)^{2/5} \quad (4)$$

или второй:

$$r = b \lg(n). \quad (5)$$

Здесь n — объем исследуемой выборки;

$b = 3 \ldots 5$.

В случае выполнения гипотезы H_0 статистика (4) имеет распределение, которое при $n \to \infty$ приближается к распределению хи-квадрат (χ^2) с $r - m - 1$ степенями свободы.

Критерий значимости для проверки этой гипотезы — правосторонний критерий вида

$$P(\chi^2_{\text{набл}} > \chi^2_{\text{крит}}) = \alpha, \quad (6)$$

где $\chi^2_{\text{набл}}$, $\chi^2_{\text{крит}}$ — соответственно наблюдаемое (эмпирическое) и критическое значение критерия хи-квадрат (χ^2).

Если все параметры гипотетической плотности вероятности известны, то необходимо считать $m = 0$, т. е. следует воспользоваться таблицами распределения χ^2 с $(r - 1)$ степенями свободы.

К. Пирсон показал, что в соответствии с методом наименьших квадратов мера расхождения между экспериментом и теорией обладает чрезвычайно важным свойством: χ^2-распределение выражается через наблюдаемые и ожидаемые частоты для всех r интервалов, при этом χ^2-распределение имеет $k = r - 1$ степень свободы. Число степеней свободы k зависит от того, использовались ли рассматриваемые экспериментальные данные для получения параметров распределения. Выявление параметров распределения понижает число степеней свободы на число выявленных параметров m.

Таким образом, рассматриваемый критерий — параметрический, он учитывает как значение, так и число параметров распределения. Если рассматривать нормальное распределение, то оно, как известно, описывается двумя параметрами: математическим ожиданием и дисперсией ($m = 2$), отсюда число степеней свободы $k = r - m - 1 = r - 3$.

Алгоритм автоматизации проверки с помощью средств MATLAB по критерию согласия Пирсона сводится к следующим этапам:

1-й этап — выдвигают гипотезу (теоретическое распределение) о виде экспериментального распределения и устанавливают уровень значимости α, по которому определяют критическую область.

2-й этап — задается массив случайных чисел, вид распределения которого проверяется исходя из следующих предпосылок:

– если требуется получить подтверждение о согласии с нулевой гипотезой, то в качестве проверяемой задается выборка с нормальным законом распределения вероятностей с помощью функции *x* = *normrnd*(*mu*, *sigma*, *n*, 1), где *mu*, *sigma* — исходные значения, соответственно, математического ожидания и среднеквадратического отклонения;

– если нужно получить обратное утверждение (согласие с альтернативной гипотезой), то в качестве проверяемой выборки генерируется массив случайных чисел по любому другому закону, кроме нормального, например, близкому к нему по форме закону Стьюдента, с помощью функции *x* = *trnd*(*V*, *n*, 1), где *V* = 1,6–2,6.

3-й этап — выборка объема *n* разбивается на *r* интервалов группирования, где *r* рассчитывается по формуле (4) или (5). При этом необходимо, чтобы $r > 7$, а каждый интервал содержал не менее пяти выборочных значений. Определяется размах выборки $R = \max(x) - \min(x)$ и длина одного интервала по формуле $L = R/r$.

4-й этап — выполняется подсчет числа W_i наблюдаемых значений исследуемой выборки (эмпирических частот), попавших в каждый интервал, который выполняется автоматически с помощью функции построения гистограммы [*W*, *t*] = *hist*(*x*, *r*).

5-й этап — с помощью оператора суммирования элементов массива *sum*(*x*) определяются статистические оценки выборочных данных для математического ожидания *Mx* = (*sum*(*x*)) / *n* и среднеквадратического отклонения *d* = *sqrt*(*sum*((*x* – *Mx*).^2)/(*n* – 1)), которые используются в качестве параметров распределения. При этом число степеней свободы устанавливается $k = r - m - 1$, где $m = 2$.

6-й этап — строится гистограмма статистического (для исследуемой выборки значений) и кривая теоретического (нормального) законов распределения, характеризующих генеральную совокупность данных; кривую нормального закона распределения, отвечающего гипотезе H_0, строим с помощью встроенной функции *f* = *normpdf*(*x*, *Mx*, *d*).

Приведем фрагмент программы, позволяющей построить гистограмму исследуемого и получить кривую теоретического распределений на одном графике:

```
f=normpdf(x, Mx, d);
w=W/n;
j = 1;
for i=1:2:2*r
   a(i)=(xmin-L)+L*j;
   a(i+1)=(xmin-L)+L*(j+1)-0.001;
   b(i)=w(j)/L;
   b(i+1)=w(j)/L;
   j=j+1;
end
plot(a, b,'b', x, f,'r');
ylabel('Плотность')
```

7-й этап. С целью определения наблюдаемой статистики $\chi^2_{\text{набл}}$ по формуле (2) выполняется расчет гипотетической вероятности p_i путем интегрирования, согласно (3), функции плотности вероятностей f_0 в каждом *i*-м интервале группирования ($i = 1, \ldots, r$). С применением вычислительных средств эта операция реализуется только численными методами, из которых наибольшую точность до 10^{-6} имеет метод Симпсона (с аппроксимацией подынтегральной функции в узловых точках отрезками парабол). В MATLAB операция интегрирования методом Симпсона реализуется с помощью встроенной функции

$$Q = quad(FUN, A, B),$$

где A, B — нижний и верхний пределы интегрирования; FUN — подынтегральная функция, которая описана в виде в файл-функции «*myfun.m*».

Фрагмент программы, иллюстрирующей численное интегрирование с помощью функции *quad*:

```
% Определение теоретической вероятности по методу Симпсона:
a1=xmin; b1=a1+L;% верхний и нижний пределы интегрирования
for j=1: r
   s=0;% начальная сумма
   n1=10;% количество разбиений каждого участка
   L1=(b1-a1)/n1;% шаг разбиения участка на интервалы
   for i=a1:2*L1: b1-2*L1
s=quad(@(x)myfun(x, Mx, d), a1, b1);
   end
   p(j)=s;% теоретическая вероятность для j-го интервала
   a1=a1+L;
   b1=b1+L;
end
```

8-й этап — на основании (2) рассчитывается значение наблюдаемой статистики $\chi^2_{\text{набл}}$ расхождения эмпирических частот относительно теоретических. Затем рассчитывают критическое значение критерия хи-квадрат ($\chi^2_{\text{крит}}$) с применением обратной функции кумулятивного распределения хи-квадрата, имеющей синтаксис *chi2inv*(P, V), где $P = (1 - \alpha)$ — вероятность того, что статистика критерия попадает в область принятия нулевой гипотезы; V — число степеней свободы.

В качестве операторной функции статистика $\chi^2_{\text{крит}}$ записывается в следующем виде:

$$Chietabl = chi2inv(1 - alpha, k).$$

9-й этап — в завершение процедуры проверки применяют критерий согласия Пирсона (χ^2-критерий) согласно следующему правилу: если выполняется условие $\chi^2_{\text{набл}} \leq \chi^2_{\text{крит}}$, то эмпирическое (наблюдаемое) значение статистики не попадает в критическую область и, следовательно, гипотезу H_0 принимают, в обратном случае при $\chi^2_{\text{набл}} > \chi^2_{\text{крит}}$ значение наблюдаемой статистики попадает в критическую область и гипотезу H_0 отвергают.

В заключение следует отметить, что критерий хи-квадрат — это критерий, который асимптотически верен, т. е. выборочное распределение можно сделать сколь угодно близким к распределению хи-квадрат путем увеличения размера выборки.

Результаты (Results)

На основе предложенного алгоритма автоматизации на *первом этапе* эксперимента была выполнена проверка гипотезы о нормальном распределении генеральной совокупности данных, в качестве которой генерировался массив случайных чисел с нормальным законом распределения вероятностей с помощью функции $x = normrnd(mu, sigma, n, 1)$. Оценка статистической значимости различий распределений (проверяемого и нормального) выполнялась с помощью расчета относительных значений эмпирических и теоретических частот.

С целью оценки соответствия (близости) проверяемого распределения нормальному на рис. 1 приведены графики, демонстрирующие степень аппроксимации функции, описывающей плотность вероятностей нормального распределения гистограммой исследуемого распределения выборки.

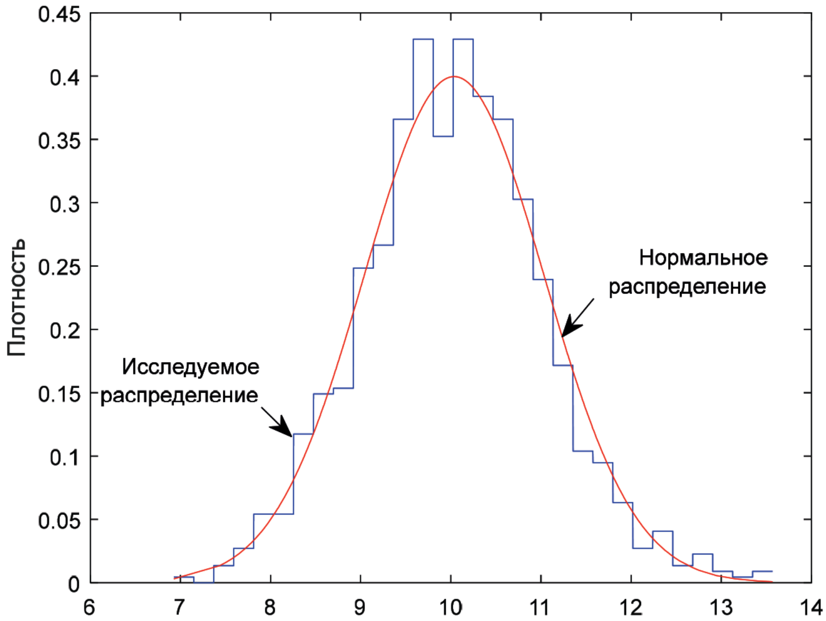

Рис. 1. Аппроксимация теоретической кривой нормального распределения гистограммой исследуемой выборки с нормальным распределением

Приведем численные результаты выборочных значений параметров проверяемого распределения, а также наблюдаемого и табличного значений критерия хи-квадрат (χ^2) на первом этапе эксперимента:

```
    Максимальные и минимальные значения в массиве:

xmax =
    13.5699

xmin =
    6.9278

r =
    30

Выборочное среднее:

Mx =
    10.0365

Выборочное среднеквадратичное отклонение:

d =
  0.9984

Наблюдаемое значение критерия хи-квадрат

chie =
     34.2031
```

```
Количество степеней свободы

k =
   27

Критическое значение критерия хи-квадрат:

chietabl =
       40.1133

Вывод. Распределение проверяемой выборки подчиняется нормальному закону
>>
```

На основе сравнения полученных на первом этапе эксперимента результатов численных оценок наблюдаемого $\chi^2_{набл} = 34.2031$ и табличного (критического) $\chi^2_{крит} = 40.1133$ значений критерия хи-квадрат (χ^2) следует, что условие для принятия гипотезы H_0: $\chi^2_{набл} < \chi^2_{крит}$, выполняется. Следовательно, гипотеза H_0 о нормальном распределении проверяемой выборки данных подтверждается.

На втором этапе эксперимента производилась проверка гипотезы H_0 о нормальном распределении выборки данных, в качестве которой генерировался массив случайных чисел с отличным от нормального законом распределения, но близким к нему по форме распределением Стьюдента, с помощью функции $x = trnd(2.5, n, 1)$. Соответствующие этому этапу графики исследуемого и нормального распределений, аналогичные рис. 1, приведены на рис. 2.

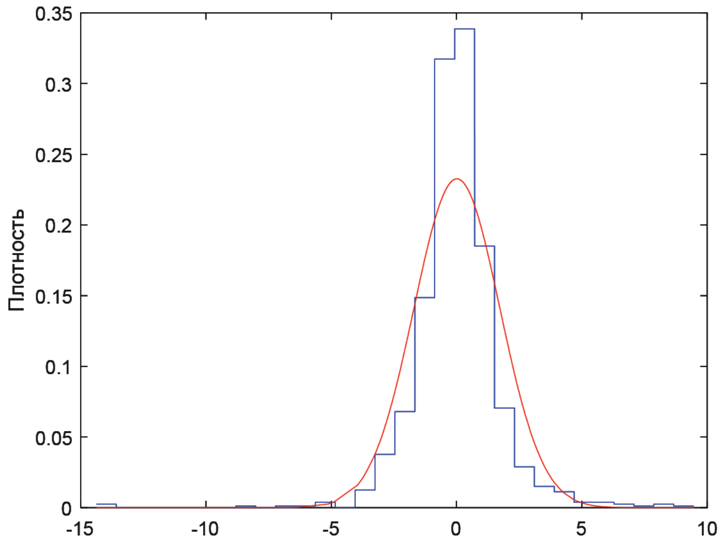

Рис. 2. Аппроксимация теоретической кривой нормального распределения гистограммой исследуемой выборки с распределением по закону Стьюдента

Приведем численные оценки тех же самых параметров, но для отличного от нормального проверяемого закона распределения, а также вывод, подтверждающий непринадлежность проверяемого распределения нормальному:

```
xmax =
        22.8397

xmin =
        -11.8021
```

```
r =
    30

Выборочное среднее:

Mx =
     −0.0138

Выборочное среднеквадратичное отклонение:

d =
    2.0300

Наблюдаемое значение критерия хи-квадрат

chie =
       3.5223e+12
     Количество степеней свободы

k =
    27

Критическое значение критерия хи-квадрат:

chietabl =
           40.1133
     Вывод. Распределение проверяемой выборки не подчиняется нормальному закону.
```

На основе сравнения полученных на втором этапе эксперимента результатов численных оценок наблюдаемого $\chi^2_{набл} = 3.5223 \cdot 10^{12}$ и табличного (критического) $\chi^2_{крит} = 40.1133$ значений критерия хи-квадрат (χ^2) следует, что условие для принятия гипотезы H_0: $\chi^2_{набл} < \chi^2_{крит}$, не выполняется. Отсюда следует, что гипотеза H_0 о нормальном распределении проверяемой выборки данных отклоняется и принимается альтернативная гипотеза.

Обсуждение (Discussion)

Результаты численного моделирования оцениваемых по критерию Пирсона параметров, а также аппроксимации гипотетического нормального распределения исследуемыми распределениями выборок генерируемых массивов данных, представленные на рис. 1 и 2, получены в автоматическом режиме и полностью соответствуют ожидаемым результатам проверки гипотезы H_0: свидетельствуют о ее подтверждении в первом случае (при нормальном распределении выборки) и ее отклонении — во втором случае (при отличном от нормального распределении выборки). Это подтверждает работоспособность алгоритма автоматизации и его точность.

Заключение (Conclusion)

В работе приведены результаты исследований, связанные с построением и практической реализацией алгоритма автоматизации метода проверки статистических гипотез, позволяющего получить оцениваемые параметры распределений по критерию согласия Пирсона. Корректность и реализуемость предлагаемых технических решений подтверждена на примере эксперимента, состоящего из двух сценариев проверки основной гипотезы о нормальном распределении двух генерируемых массивов данных одного и того же объема, один из которых имеет нормальное, а другой — отличное от нормального закона распределение. Полученные оценки и решения по критерию Пирсона о принадлежности распределений проверяемых массивов данных нормальному закону распределения полностью соответствуют ожидаемым результатам.

Предложенный алгоритм автоматизации и численного моделирования процессов проверки статистических гипотез применительно к критерию согласия Пирсона (хи-квадрат) позволяет широко использовать инструментарий статистических функций MATLAB для реализации вероятностной и статистической моделей закона распределения и принимать обоснованные решения при обработке данных в системах автоматизированного управления и измерения на объектах водного транспорта. Кроме того, предложенные решения находятся в основе дальнейшего развития и совершенствования методов компьютерного моделирования и цифровизации процедур проверки статистических гипотез по критериям согласия для использования в высокотехнологичных программно-измерительных комплексах автоматизированных систем управления. Это связано с тем, что в большинстве важных для практики случаев анализ статистических свойств распределений требует проведения множественных проверок, выполняемых с применением численных методов, для которых требуется инструментарий мощных программно-вычислительных сред. Примерами таких случаев являются процедуры нахождения информативных подмножеств признаков в регрессионном и статистическом анализе, а также выявление отклонений параметров в автоматизированных системах управления.

СПИСОК ЛИТЕРАТУРЫ

1. *Antonov A. V.* Regarding the planning of testing scope for new equipment samples / A. V. Antonov, V. A. Chepurko, V. E. Chekhovich, V. F. Ukraintsev // Dependability. — 2016. — № 3. — Pp. 3–7. DOI: 10.21683/1729-2640-2016-16-3-3-7.

2. *Лемешко Б. Ю.* Сравнительный анализ мощности критериев согласия при близких конкурирующих гипотезах. I. Проверка простых гипотез / Б. Ю. Лемешко, С. Б. Лемешко, С. Н. Постовалов // Сибирский журнал индустриальной математики. — 2008. — Т. 11. — № 2 (34). — С. 96–111.

3. *Лемешко Б. Ю.* Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход / Б. Ю. Лемешко, С. Б. Лемешко, С. Н. Постовалов, Е. В. Чимитова. — Новосибирск: Изд-во НГТУ, 2011. — 888 с.

4. *Мещеряков В. В.* Задачи по статистике и регрессионному анализу с Matlab / В. В. Мещеряков. — М.: Диалог-МИФИ, 2009. — 448 с.

5. *Кулаичев А. П.* Методы и средства комплексного анализа данных / А. П. Кулаичев. — М.: ФОРУМ, ИНФРА-М, 2006. — 512 с.

6. *Льюис К. Д.* Методы прогнозирования статистических данных / К. Д. Льюис. — М.: Финансы и статистика, 1986. — 133 с.

7. *Мэтьюз Д. Г.* Численные методы. Использование MATLAB / Д. Г. Мэтьюз, К. Д. Финк. — М.: Вильямс, 2001. — 720 с

8. *Лапчик М. П.* Численные методы / М. П. Лапчик, М. И. Рагулина, Е. К. Хеннер. — М.: ИЦ «Академия», 2007. — 384 с.

9. *Кетков Ю. Л.* MATLAB 7: программирование, численные методы / Ю. Л. Кетков, А. Ю. Кетков, М. М. Шульц. — СПб.: БХВ-Петербург, 2005. — 752 с.

10. *Иглин С. П.* Математические расчеты на базе MATLAB / С. П. Иглин. — СПб.: БХВ-Петербург, 2005. — 629 с.

11. *Поршнев С. В.* Компьютерный анализ и интерпретация эмпирических зависимостей / С. В. Поршнев, Е. В. Овечкина, М. В. Мащенко, А. В. Каплан, В. Е. Каплан. — М.: ООО «Бином-Пресс», 2010. — 336 с.

12. *Гмурман В. Е.* Теория вероятностей и математическая статистика / В. Е. Гмурман. — 12-е изд., перераб. — М.: Высш. обр., 2006. — 478 с.

13. *Corder G. W.* Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach / G. W. Corder, D. I. Foreman. — New York: Wiley, 2009. — 264 p.

14. *Tsybakov A. B.* Introduction to Nonparametric Estimation / A. B. Tsybakov. — Springer, 2008. — 224 p.

15. *Greenwood P. E.* A guide to chi-squared testing / P. E. Greenwood, M. S. Nikulin. — Wiley-Interscience, 1996. — 304 p.

16. *Bagdonavicius V. B*. Chi-square goodness-of-fit test for right censored data / V. B. Bagdonavicius, M. S. Nikulin // International Journal of Applied Mathematics and Statistics. — 2011. — Vol. 24. — Is. SI-11A. — Pp. 30–50.

REFERENCES

1. Antonov, Alexander V., Valery A. Chepurko, Vladimir E. Chekhovich, and Vladimir F. Ukraintsev. "Regarding the planning of testing scope for new equipment samples." *Dependability* 3 (2016): 3–7. DOI: 10.21683/1729-2640-2016-16-3-3-7.

2. Lemeshko, B. Yu., S. B. Lemeshko, and S. N. Postovalov. "Sravnitel'nyi analiz moshchnosti kriteriev soglasiya pri blizkikh konkuriruyushchikh gipotezakh. I. Proverka prostykh gipotez." *Journal of Applied and Industrial Mathematics* 11.2(34) (2008): 96–111.

3. Lemeshko, B. Yu., S. B. Lemeshko, S. N. Postovalov, and E. V. Chimitova. *Statistical data analysis, simulation and study of probability regularities. computer approach*. Novosibirsk: Izd-vo NGTU, 2011.

4. Meshcheryakov, V. V. *Zadachi po statistike i regressionnomu analizu s Matlab*. M.: Dialog-MIFI, 2009.

5. Kulaichev, A. P. *Metody i sredstva kompleksnogo analiza dannykh*. M.: FORUM, INFRA-M, 2006.

6. Lewis, C. D. *Industrial and Business Forecasting Methods*. Butterworth-Heinemann, 1982.

7. Mathews, John, and Kurtis Fink. *Numerical Methods Using Matlab*. — 4th Edition. — Pearson, 2003.

8. Lapchik, M. P., M. I. Ragulina, and E. K. Khenner. *Chislennye metody*. M.: Iz-datel'skii tsentr «Akademiya», 2007.

9. Ketkov, Yu. L., A. Yu. Ketkov, and M. M. Shul'ts. *MATLAB 7: programmirovanie, chislennye metody*. SPb.: BKhV-Peterburg, 2005.

10. Iglin, S. P. *Matematicheskie raschety na baze MATLAB*. SPb.: BKhV-Peterburg, 2005.

11. Porshnev, S. V., E. V. Ovechkina, M. V. Mashchenko, A. V. Kaplan, and V. E. Kaplan. *Komp'yuternyi analiz i interpretatsiya empiricheskikh zavisimostei*. M.: OOO «Binom-Press», 2010.

12. Gmurman, V. E. *Teoriya veroyatnostei i matematicheskaya statistika*. 12nd ed. M.: Vyssh. obr., 2006.

13. Corder, Gregory W., and Dale I. Foreman. *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach*. New York: Wiley, 2009.

14. Tsybakov, Alexandre B. *Introduction to Nonparametric Estimation*. Springer, 2008.

15. Greenwood, Priscilla E., and Michael S. Nikulin. *A guide to chi-squared testing*. Wiley-Interscience, 1996.

16. Bagdonavicius, V. B., and M. S. Nikulin. "Chi-square goodness-of-fit test for right censored data." *International Journal of Applied Mathematics and Statistics* 24.SI-11A (2011): 30–50.

ИНФОРМАЦИЯ ОБ АВТОРЕ

Чертков Александр Александрович —
доктор технических наук, доцент
ФГБОУ ВО «ГУМРФ имени адмирала
С. О. Макарова»
198035, Российская Федерация, Санкт-Петербург,
ул. Двинская, 5/7
e-mail: *chertkov51@mail.ru*,
kaf _electricautomatic@gumrf.ru

INFORMATION ABOUT THE AUTHOR

Chertkov, Alexandr A. —
Dr. of Technical Sciences, associate professor
Admiral Makarov State University of Maritime
and Inland Shipping
5/7 Dvinskaya Str., St. Petersburg, 198035,
Russian Federation
e-mail: *chertkov51@mail.ru*,
kaf _electricautomatic@gumrf.ru

Статья поступила в редакцию 27 февраля 2023 г.
Received: February 27, 2023.