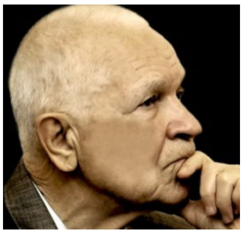
УДК 519.25;378.1

ИДЕНТИФИКАЦИЯ ПАРАМЕТРОВ СТАТИСТИЧЕСКОГО РАСПРЕДЕЛЕНИЯ РЕЗУЛЬТАТОВ ОБРАБОТКИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ



Еремеев Владимир Сергеевич,
профессор, доктор технических наук,

Кузьминов Виталий Викторович,
инженер,

Брежнева Ольга Александровна,
студентка 4-го курса,

Мелитопольский государственный педагогический университет имени Богдана Хмельницкого, г.Мелитополь.

Донева Ольга Викторовна
кандидат педагогических наук, доцент

Северо-Кавказский филиал федерального государственного бюджетного образовательного учреждения высшего образования «Белгородский государственный технологический университет им. В.Г. Шухова» г. Минеральные Воды

IDENTIFICATION OF PARAMETERS OF STATISTICAL DISTRIBUTION OF EXPERIMENTAL DATA

Eremeev Vladimir Sergeevich,
Professor, doctor of technical Sciences,

Kuzminov Vitaly Viktorovich,
engineer,

Brezhneva Olga Aleksandrovna
student of the 4th course

Melitopol state pedagogical University named after Bohdan Khmelnytsky, Melitopol.

Doneva Olga Victorovna
candidate of pedagogical Sciences, associate Professor

North Caucasian branch of Federal state budgetary educational institution of higher education "Belgorod state technological University named after. V. G. Shukhov " Mineralnye Vody

АННОТАЦИЯ

Сформулирована проблема идентификации параметров нормальных составляющих стохастической величины при статистической обработке эксперимента для случая одномодального или многомодального распределения. Рассмотрена возможность представления эмпирической функции в виде нескольких нормальных законов. В качестве критерия точности преобразования эмпирической функции использована сумма квадратов отклонений эмпирических данных от рассчитанных значений. Предложен численный метод нахождения математических ожиданий и дисперсий.

Ключевые слова: математическое ожидание; дисперсия; метод наименьших квадратов; нормальное распределение; случайная величина; эмпирическая функция распределения.

ABSTRACT

Formulated the problem of parameters identification of stochastic components normal value for the statistical processing of the experiment for different cases of distribution. Considered the possibility of presenting empirical functions in the form of several normal distributions. As a criterion of empirical accuracy of the transformation function used the sum of squares of deviations of empirical data from the calculated values. The numerical method of finding the mathematical expectations and dispersions.

Keywords: expectation; variance; least squares method; normal distribution; random variable; empirical distribution function.

Актуальность. Теория вероятности [1] широко применяется в процессе подготовки специалистов различного профиля, что предопределило использование статистических методов исследования во всех теоретических и прикладных науках [2]. Многочисленные публикации [3],[4], относящиеся к обработке экспериментов, говорят о востребованности этих методов на всех стадиях организации и проведения исследований [4], [5], [6]. Из теории вероятностей известно, что статистическая выборка может быть представлена в виде эмпирической функции распределения. Согласно предельной теореме теории вероятностей случайная величина во многих случаях с большой точностью описывается одним или несколькими нормальными законами. Анализ эмпирических распределений позволяет выявить характер изменения величины и ее корреляцию с другими величинами, в связи с чем разработка методов идентификации параметров нормальных распределений, определяющих эмпирическую функцию распределения, имеет практический интерес.

Постановка задач статьи. Теория статистической обработки экспериментальных данных изложена во многих трудах [3], [4], [5]. Во многих случаях предполагается выполнение известного закона распределения, которое в действительности является суперпозицией нескольких законов. [5]. Поэтому создание методики изучения статистических данных в случае неизвестного закона является актуальным. В литературе отсутствует информация по представлению эмпирической функции в виде суммы нескольких нормальных законов с различными математическими ожиданиями и дисперсиями. Цель статьи состоит в создании методики моделирования эмпирической функции в виде суммы нескольких нормально распределённых случайных величин.

Основная часть

1. Одномодальный случай. Пусть *F*e*(x)* – некоторое эмпирическое распределение [2] случайной величины *x*, заданное таблично, т.е. известен набор значений $\{(x_i, P_i), i = 1.,.., N\}$, где $P_i = F_e(x_i) = P(x < x_i)$.

Предположим, что значение *x* распределено нормально с плотностью

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-M)^2}{2\sigma^2}} \qquad (1)$$

Задача идентификации параметров распределения состоит в нахождении математического ожидания M и среднеквадратического отклонения σ. Для решения поставленной задачи необходимо выбрать критерий близости теоретической F(x) и экспериментальной F_e(x) функций распределения. На практике одним из наиболее часто используемых методов нахождения неизвестных параметров является метод наименьших квадратов [6]. В качестве критерия точности моделирования эмпирической функции нормальным законом рассмотрим сумму квадратов отклонений:

$$K = \sum_{i=1}^{N} (F(x_i) - F_e(x_i))^2 .$$

Задача моделирования сводится к поиску параметров стационарных точек и проверке их на экстремальность. Критерий близости с учётом (1) представим в виде:

$$K = \sum_{i=1}^{N} \left(\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_i} e^{-\frac{(t-M)^2}{2\sigma^2}} dt - F_e(x_i)\right)^2 \qquad (2)$$

Составим условие нахождения стационарных точек [9]:

$$\begin{cases} \dfrac{\partial K}{\partial M} = 0 \\ \dfrac{\partial K}{\partial \sigma} = 0 \end{cases} \qquad (3)$$

Подставив (2) в (3) и пренебрегая ненулевыми множителями в левых частях (3), после упрощений получим

$$\begin{cases} \sum_{i=1}^{N} \left(\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_i} \alpha dt - F_e(x_i)\right) = 0 \\ \sum_{i=1}^{N} \left[\left(\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_i} \alpha dt - F_e(x_i)\right)\left(\int_{-\infty}^{x_i} (t-M)^2 \alpha dt - \sigma^2 \int_{-\infty}^{x_i} \alpha dt\right)\right] = 0 \end{cases} \qquad (4)$$

где

$$\alpha = e^{-\frac{(t-M)^2}{2\sigma^2}}$$

Для решения системы уравнений (4) можно использовать метод покоординатного спуска [7] и дихотомию [8]. В этом случае для вычисления интегралов следует воспользоваться формулами Ньютона-Котеса, в частности, формулой Симпсона [8], которую перепишем в виде, удобном для реализации на ЭВМ:

$$\int_{a}^{b} f(x)dx = \frac{b-a}{12n}\left(f(a)+f(b)+4\sum_{k=0}^{2n-1} f\left(a+\frac{b-a}{4n}(2k+1)\right)+2\sum_{k=0}^{2n-1} f\left(a+\frac{b-a}{2n}k\right)\right)$$

где n – натуральное число.

2. **Многомодальный случай**. Предположение о возможности представления одномодальной эмпирической функции нормальным законом (1) не всегда правомочно, а при многомодальном распределении приводит к ошибочным результатам. В этом случае необходимо перейти к моделированию с использованием нескольких законов вида (1). Пусть функция распределения является суммой нормальных распределений с различными математическими ожиданиями M_m и дисперсиями σ^2_m (m=1,2...R):

$$f_m(x) = \frac{1}{\sigma_m\sqrt{2\pi}} e^{-\frac{(x-M_m)^2}{2\sigma_m{}^2}} \qquad (5)$$

В этом случае критерий точности моделирования эмпирической функции совокупностями нормальных законов (5) является двойная сумма:

$$K = \sum_{m=1}^{R}\sum_{i=1}^{N}(F(x_i)-F_e(x_i))^2 \cdot$$

Для рассматриваемой задачи идентификации критерий близости (2) преобразуется к виду:

$$K = \sum_{m=1}^{R}\sum_{i=1}^{N}\left(\frac{1}{\sigma_m\sqrt{2\pi}}\int_{-\infty}^{x_i} e^{-\frac{(t-M_m)^2}{2\sigma_m{}^2}}dt - F_e(x_i)\right)^2 \qquad (6)$$

Условие нахождения стационарных точек состоит из 2m уравнений:

$$\begin{cases} \dfrac{\partial K}{\partial M_m} = 0 \\ \dfrac{\partial K}{\partial \sigma_m} = 0 \end{cases} \qquad (7)$$

Из (6) в (7) после упрощений получим

$$\begin{cases} \sum\limits_{m=1}^{R}\sum\limits_{i=1}^{N}\left(\dfrac{1}{\sigma_m\sqrt{2\pi}}\int\limits_{-\infty}^{x_i}\alpha_m dt - F_e(x_i)\right)=0 \\ \sum\limits_{m=1}^{R}\sum\limits_{i=1}^{N}[(\dfrac{1}{\sigma_m\sqrt{2\pi}}\int\limits_{-\infty}^{x_i}\alpha_m dt - F_e(x_i)_i)(\int\limits_{-\infty}^{x_i}(t-M_m)^2\alpha_m dt - \sigma_m{}^2\int\limits_{-\infty}^{x_i}\alpha_m \end{cases} \qquad (8)$$

где

$$\alpha_m = e^{-\frac{(t-M_m)^2}{2\sigma_m{}^2}}$$

Система уравнений (8) также может быть решена с использованием численного метода покоординатного спуска [7].

Выводы и перспективы дальнейших исследований. Результаты экспериментальных данных во многих случаях представляются в виде эмпирической функции. Определение параметров этой функции является актуальной задачей. Практическая реализация идентификации параметров невозможна без использования численных методов. В настоящей работе предложен алгоритм статистического анализа, который позволяет моделировать эмпирическую функцию в виде нормального закона (1) или в виде суперпозиции нескольких нормальных законов с различными математическими ожиданиями и дисперсиями (формула (5)). Расчёт параметров для m нормальных распределений сводится к решению системы из 2m уравнений (8) с использованием численного метода покоординатного спуска [7]. Статистическое распределение экспериментальных данных не исчерпывается нормальным законом (1) или суперпозицией нормальных законов (5), поэтому рассмотрение других случаев распределения имеет теоретический и практический интерес.

СПИСОК ЛИТЕРАТУРЫ

1. Жалдак М. І. Теорія ймовірностей і математична статистика: Підручник для студентів фізико-матем. спеціальностей педагог. університетів. Вид.2, переробл і доп./ М. І. Жалдак, Н. М.Кузьмина, Г. О. Михалін. - Полтава: «Довкілля-К», 2010. - 500 С.
2. Єремєєв В. С. Теорія ймовірностей та математична статистика. Навчальний посібник / В. С. Єремєєв, Д. О. Сосновських, О. В. Тітова - Мелитополь: ТОВ «Видавничий будинок», 2009. – 187 с.
3. Шашков В.Б. Обработка экспериментальных данных и построение эмпирических формул. Курс лекций: Учебное пособие. - Оренбург: ГОУ ОГУ, 2005.-150 с.
4. Губин В.И. Статистические методы обработки экспериментальных данных / В.И. Губин, В.Н. Осташков. Учебное пособие. Тюмень. Изд. “ТюмГНГУ”». – 202 с.
5. Еремеев В.С. Статистическая обработка эксперимента в случае неизвестной функции распределения. /В.С. Еремеев, В.В. Кузьминов. Журнал Херсонского государственного педагогического университета «Інформаційні технології в освіті». 2013г.,вип 13, с.44-51.
6. Калиткин Н.Н. Численные методы / Н.Н. Калиткин – Москва: «Наука». 1978 – 512 с.
7. Березин И.С. Методы вычислений. т. II / И.С. Березин, Н.П. Жидков - Москва: Государственное изд-во физико-математической литературы. 1962.- 620 с.
8. Ильин В. А. Основы математического анализа (в двух частях) / В.А.Ильин, Э. Г.Позняк — Москва: Физматлит, 2005. — 648 с.
9.Eremeev V.S., Kurbatov V.L., Gulynina E.V. Determination of unknown parameters of mathematical model using the experimental design. Journal of Theoretical and Applied Information Technology 31st August 2015. Vol.78. No.3 © 2005 - 2015 JATIT & LLS. All rights reserved.(JTAIT(JATIT)-LLS - Journal of Theoretical and Applied Information Technology (ISSN19928645-Pakistan-Scopus), 00, 357480). pp.464-472.