УДК 519.25;378.1

ИДЕНТИФИКАЦИЯ ПАРАМЕТРОВ СТАТИСТИЧЕСКОГО РАСПРЕДЕЛЕНИЯ РЕЗУЛЬТАТОВ ОБРАБОТКИ ПЕДАГОГИЧЕСКОГО ЭКСПЕРИМЕНТА

Брежнева О.А., Еремеев В.С., Кузьминов В.В.
Мелитопольский государственный педагогический университет им. Богдана Хмельницкого, г. Мелитополь
e-mail: eremeev@mdpu.org.ua.

Постановка проблемы. 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 педагогических экспериментов, говорят о востребованности этих методов на всех стадиях организации и проведения учебного процесса [3], [4], [5]. 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 изменчивости величины и ее корреляцию с другими величинами, в связи с чем идентификация параметров нормальных распределений, определяющих эмпирическую функцию распределения, имеет практический интерес.

Анализ последних достижений. Теория педагогического эксперимента изложена во многих трудах. Классические работы Ю. К. Бабанского [6], С. И. Архангельского [7] и других учёных [3], [4] развиты в более поздних исследованиях. Статистические исследования в педагогике базируются на предположении о выполнении известного закона распределения, что во многих случаях не отвечает действительности [5]. Поэтому создание методики изучения статистических данных в случае неизвестного закона является актуальным. Обработка экспериментальных данных во многих случаях свидетельствует об одномодальности распределения, хотя фактически оно является суперпозицией нескольких случайных величин. В литературе отсутствует информация по представлению эмпирической функции в виде суммы нескольких нормальных законов с различными математическими ожиданиями и дисперсиями. **Цель нашей работы** 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 величина x распределена нормально [2] 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 М 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 [8]. 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]:

$$\begin{cases}\dfrac{\partial K}{\partial M}=0\\[2mm] \dfrac{\partial K}{\partial \sigma}=0\end{cases} \qquad (3)$$

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

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

где

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

Для решения системы уравнений (4) можно использовать метод покоординатного спуска [10] и дихотомию [9]. В этом случае для вычисления интегралов следует воспользоваться формулами Ньютона-Котеса, в частности, формулой Симпсона [9], которую можно переписать 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 нескольких 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:

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

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

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

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

$$\begin{cases} \sum_{m=1}^{R} \sum_{i=1}^{N} (\frac{1}{\sigma_m \sqrt{2\pi}} \int_{-\infty}^{x_i} \alpha_m dt - F_e(x_i)) = 0 \\ \sum_{m=1}^{R} \sum_{i=1}^{N} [(\frac{1}{\sigma_m \sqrt{2\pi}} \int_{-\infty}^{x_i} \alpha_m dt - F_e(x_i)_i)(\int_{-\infty}^{x_i} (t - M_m)^2 \alpha_m dt - \sigma_m{}^2 \int_{-\infty}^{x_i} \alpha_m dt)] = 0 \end{cases} \qquad (8)$$

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 [10].

Выводы. Результаты педагогического эксперимента 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 [10].

Литература

1. Жалдак М.І. Теорія ймовірностей і математична статистика: Підручник для студентів фізико-матем. спеціальностей педагог. ун-тів. Вид.2, перероб 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. Первые шаги начинающего педагога. Сайт кафедры информатики Харьковского Национального Автодорожного Университета. [Электронный ресурс]. – Режим доступа к ресурсу: http://www.pervyi-shag.narod.ru.
4. Лузан П.Г. Основи науково педагогічних досліджень./ П.Г. Лузан, І.В. Сопівник, С.В. Виговська. Національний університет біоресурсів і природокористування України. – Київ, 2010. – 219 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». 2013. Вип 13. С. 44–51.
6. Бабанский Ю.К. Проблемы повышения эффективности педагогических исследований: (дидактический аспект)/ Ю.К. Бабанский. – М. : Педагогика. 1982. – 182 с.
7. Архангельский С. И. Лекции по научной ориентации учебного процесса в высшей школе / С. И. Архангельский.- М.: Высшая школа. 1976.-200 С.
8. Калиткин Н.Н. Численные методы / Н.Н. Калиткин – Москва: «Наука». 1978 – 512 с.
9. Ильин В. А. Основы математического анализа (в двух частях) / В.А.Ильин, Э. Г.Позняк — Москва: Физматлит, 2005. – 648 с.
10. Березин И.С. Методы вычислений. т. II / И.С. Березин, Н.П. Жидков - Москва: Государственное изд-во физико-математической литературы. 1962. – 620 с.

Анотація. Сформульована проблема ідентифікації параметрів нормальних складових стохастичною величини при обробці педагогічного експерименту для випадку одномодальних або многомодальним розподілу. Розглянуто можливість подання емпіричної функції у вигляді декількох нормальних законів. Як критерій точності перетворення емпіричної функції використана сума квадратів відхилень емпіричних даних від розрахованих на основі нормального розподілу. Запропоновано чисельний метод знаходження математичних очікувань і дисперсій.

Ключові слова: математичне очікування; дисперсія; метод найменших квадратів; нормальний розподіл; педагогічний експеримент; випадкова величина; емпірична функція розподілу.

Аннотация. Сформулирована проблема идентификации параметров нормальных составляющих стохастической величины при обработке педагогического 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 на основе нормального распределения. Предложен численный метод нахождения математических ожиданий и дисперсий.

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

Annotation. We formulate the problem of identifying the parameters of the normal components of the stochastic value of pedagogical experiment with the processing for the case of unimodal or multimodal distribution. The possibility of presenting the empirical function as several normal laws. As a criterion for the accuracy of the empirical conversion function used the sum of squared deviations from the empirical data calculated based on a normal distribution. A numerical method for finding the mathematical expectations and variances.

Keywords: expectation; dispersion; least square method; normal distribution; pedagogical experiment; random value; empirical distribution function.