МАТЕМАТИЧЕСКОЕ И АЛГОРИТМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ МОДЕЛИРОВАНИЯ И ИССЛЕДОВАНИЯ МНОГОМЕРНЫХ ФУНКЦИЙ ОТ СИСТЕМ СЛУЧАЙНЫХ ВЕЛИЧИН

Огурцов Д.В., Лемешко Б.Ю.
НГТУ, Новосибирск
E-mail: dogurtsov@gmail.com

В практике статистического анализа возникает существенно больше постановок задач, чем предлагается решений в классической математической статистике. Разнообразие законов распределения и различная сложность функций от случайных величин и систем случайных величин делают применение классического аппарата для определения закона распределения функции очень трудоемкой задачей, а зачастую, задачей, не имеющей аналитического решения.

Вследствие этого предлагаются различные аппроксимации, которые позволяют достаточно просто найти числовые характеристики интересующего нас закона распределения функции. К сожалению, такие подходы накладывают слишком жесткие ограничения на решаемую задачу, которые в реальных условиях не выполняются.

В большинстве случаев отсутствие необходимых теоретических результатов объясняется сложностью и трудоемкостью получения решений аналитическими методами. Можно констатировать, что количество и уровень сложности задач, выдвигаемых практикой, возрастают настолько быстро, что ресурсы человеческого интеллекта, его производительность просто не в состоянии обеспечить решение такого множества задач без создания и использования соответствующих вычислительных технологий.

Вышесказанное подчеркивает необходимость развития компьютерных методов исследования статистических закономерностей, компьютерных методов исследования свойств оценок и статистик различных критериев проверки статистических гипотез, построения вероятностных моделей для исследуемых закономерностей [1]. Это позволяет с меньшими интеллектуальными затратами получать фундаментальные знания в области математической статистики и, следовательно, осуществлять корректные статистические выводы при анализе данных в различных прикладных областях.

Поэтому цель работы состояла в создании программного обеспечения, позволяющего моделировать различные функции от случайных величин и от систем независимых случайных величин, и в последующем построении для смоделированных закономерностей приближенных математических моделей законов.

Разработанное программное обеспечение позволяет:

– моделировать различные многомерные законы распределения с заданными параметрами;

– получать выборки функций случайной величины или систем независимых случайных величин. Вид функции от независимых случайных величин может быть "произвольным" и задается пользователем в режиме диалога.

Тестирование разработанной системы проводилось на функциях случайных величин, для которых известны результирующие законы распределения. В этом случае осуществлялась проверка простой гипотезы о согласии эмпирического распределения функции, полученного в результате моделирования, с известным теоретическим законом [2].

Пример 1. $y = x[1][1] + x[1][2]$, т.е. суммируются первая и вторая компонента вектора x, принадлежащего нормальному закону с вектором математических ожиданий $\Theta_o = [2, 1]^T$ и ковариационной матрицей $\Theta_1 = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}$. В данном случае y принадлежит нормальному закону: $y \in N(3, \sqrt{13})$. Как видно из таблицы 1 наблюдается хорошее согласие моделируемой функции с известным результатом.

Таблица 1. Результаты проверки согласия с распределением $N(3, \sqrt{13})$ выборки значений функции $y = x[1][1] + x[1][2]$

Критерий	Значение статистики	Достигнутый уровень значимости
χ^2 Пирсона при $k = 15$	8.9504	0.8342
Колмогорова	0.8664	0.4406
ω^2 Мизеса	0.1467	0.3997
Ω^2 Андерсона-Дарлинга	0.7589	0.5116

Пример 2. Моделирование вектора в соответствии с заданным нормальным законом: $y = x$, где вектор x принадлежит нормальному закону с вектором математических ожиданий $\Theta_o = [-1, 0.5, 1]^T$ и ковариационной матрицей $\Theta_1 = \begin{bmatrix} 2 & 0 & 0.1 \\ 0 & 1 & 0 \\ 0.1 & 0 & 4 \end{bmatrix}$. Результаты проверки гипотез о векторе математических ожиданий и ковариационной матрице с использованием классических критериев [3] приведены в таблице 2.

Таблица 2. Результаты проверки гипотез о векторе математических ожиданий и о ковариационной матрице в примере 2.

Критерий	Значение статистики	Достигнутый уровень значимости
со статистикой X^2	3.1511	0.2069
со статистикой T^2	1.0606	0.3463
со статистикой L_1	3.6665	0.2998
со статистикой L_2	6.8176	0.2346

Пример 3. $y = x[1][3] + x[2] * x[3]$, где $x[1][3]$ – соответствующая компонента вектора, принадлежащего нормальному закону с математическим ожиданием $\Theta_o = [-1, 0.5, 1]^T$ и ковариационной матрицей $\Theta_1 = \begin{bmatrix} 2 & 0 & 0.1 \\ 0 & 1 & 0 \\ 0.1 & 0 & 4 \end{bmatrix}$, а $x[2] * x[3]$ – скалярное произведение случайных векторов, подчиняющихся нормальному закону с математическим ожиданием $\Theta_o = [1, 0, 0, -1]^T$ и ковариационной матрицей $\Theta_1 = \begin{bmatrix} 4 & 0.1 & 0.3 & 0.9 \\ 0.1 & 1 & 0 & 0.2 \\ 0.3 & 0 & 3 & 0 \\ 0.9 & 0.2 & 0 & 5.5 \end{bmatrix}$. В данном случае имеем неизвестное распределение одномерной случайной величины y. Наиболее хорошо случайная величина y описывается L-распределением [4] с параметрами сдвига $\theta_0 = 1.4673$, масштаба $\theta_1 = 2.2775$ и параметрами формы $\theta_2 = 0.4440$, $\theta_3 = 0.3570$.

Таким образом, разработанное программное обеспечение позволяет моделировать многомерные случайные величины и функции от многомерных случайных величин. Использование статистического моделирования и специализированного программного обеспечения, примером которого является развиваемая система “Интервальная статистика” ISW [4], позволяет строить хорошие приближенные математические модели законов распределения функций от случайных величин, когда этот закон не удается найти аналитически.

Настоящие исследования выполнены при частичной поддержке РФФИ (проект № 09-01-00056-а), Федерального агентства по образованию в рамках АВЦП “Развитие научного потенциала высшей школы” и ФЦП “Научные и научно-педагогические кадры инновационной России”.

Литература

1. Ермаков С.М., Михайлов Г.А. Курс статистического моделирования. – М.: Наука, 1976. – 320 с.
2. Лемешко Б.Ю., Постовалов С.Н. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим: Методические рекомендации. Часть II. Непараметрические критерии. – Новосибирск: Изд-во НГТУ, 1999. – 85 с.
3. Лемешко Б.Ю., Помадин С.С. Корреляционный анализ наблюдений многомерных случайных величин при нарушении предположений о нормальности // Сибирский журнал индустриальной математики. 2002. - Т.5. - № 3. - С.115-130.
4. Лемешко Б.Ю., Постовалов С.Н. Компьютерные технологии анализа данных и исследования статистических закономерностей. – Новосибирск: Изд-во НГТУ, 2004. – 119 с.