Система имитационного моделирования и исследование распределений функций от систем случайных величин

Борис Ю. Лемешко[1], Дмитрий В. Огурцов[1]
[1]*Новосибирский Государственный Технический Университет*

Аннотация **– На основании разработанного программного обеспечения методами статистического моделирования исследуются законы распределения вероятностей различных функций от случайных величин, подчиняющихся различным одномерным законам распределения. Показывается эффективность методики при исследовании**
Ключевые слова **– статистическое моделирование, функция случайных величин, закон распределения функции случайных величин.**

I. ВВЕДЕНИЕ

В ПРАКТИКЕ статистического анализа количество постановок задач оказывается существенно богаче, 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, зачастую 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.

В большинстве случаев отсутствие необходимых теоретических результатов объясняется сложностью и трудоемкостью получения решений аналитическими методами.

Развитие компьютерных методов исследования статистических закономерностей, свойств оценок и статистик различных критериев проверки статистических гипотез, построения вероятностных моделей для исследуемых закономерностей 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.

II. ПОСТАНОВКА ЗАДАЧИ

Цель настоящей работы заключалась в разработке программного обеспечения, позволяющего моделировать эмпирические законы распределения для произвольных функций от случайных величин, подчиняющихся различным законам распределений, с дальнейшим их исследованием.

Для решения поставленной задачи в работе использовался метод статистических испытаний, в основе которого лежит метод Монте-Карло.

III. ТЕОРИЯ

Требуется определить вероятностные характеристики величины Y, непосредственно недоступной для измерения, на основании доступных для многократных измерений величин X_1, X_2, …, X_k. Предполагается, что

$$Y = \varphi(X_1, X_2, ..., X_k),$$

где $\varphi(\cdot)$ – некоторая известная функция. Или в векторной форме $Y = \varphi(\overline{X})$. Предполагается, что закон распределения вектора $\overline{X}$ или, в случае независимости его компонент, законы распределения X_1, X_2, …, X_k (или законы распределения ошибок измерений) известны или могут быть найдены (построены) на основании результатов статистического анализа.

Классический подход определения закона распределения вероятностей функции от системы случайных величин предполагает знание совместной плотности распределения $f(x_1, x_2, ..., x_k)$ системы случайных величин X_1, X_2, …, X_k.

Пусть $X : \Omega \to R^n$ – случайная величина, и $g : R^n \to R^n$ – непрерывно дифференцируемая функция такая, что $J_g(x) \neq 0$, $\forall x \in R^n$, где $J_g(x)$ – якобиан функции g в точке x. Тогда случайная величина также абсолютно непрерывна, и её плотность имеет вид:

$$f_Y(y) = f_X(g^{-1}(y))\left|J_{g^{-1}}(y)\right|.$$

Однако аналитическое решение с помощью классического подхода удается найти только для некоторых частных случаев функций $Y = \varphi(\overline{X})$ и плотностей $f(x_1, x_2, ..., x_k)$.

Вследствие этого, например, при определении вероятностных характеристик результатов косвенных измерений, описываемых моделью $Y = \varphi(\overline{X})$, в случае некоррелированности компонент X_1, X_2, …, X_k вектора $\overline{X}$ измеряемых величин рекомендуется [1] линеаризация модели

$$Y \approx \varphi(\overline{M}) + (\overline{X} - \overline{M})^T \nabla\varphi(\overline{M}), \qquad (1)$$

где $\overline{M}$ – вектор математических ожиданий $\overline{X}$, $\nabla\varphi(\cdot)$ – градиент функции.

Такой подход позволяет достаточно просто определить характеристики случайной величины Y. К сожалению, данный подход оказывается эффективным также в относительно редких случаях при близости функции $\varphi(\overline{X})$ к линейной.

Покажем эффективность метода статистических испытаний для исследования вероятностных закономерностей.

IV. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

В [2] на примере функции $Y = X_1 / X_2$ демонстрируется различие в решениях, полученных при использовании классического подхода и в результате линеаризации, подчеркиваются недопустимо большие ошибки, к которым приводит применение метода линеаризации. В то же время в [3,4] демонстрируется эффективность и широкий круг задач, в который может применяться метод статистических испытаний.

Рассмотрим ряд примеров, показывающих точность моделирования.

Пример 1. $y = 2x_1 - 5x_2 + 3x_3$, где $x_1 \in N(3,7)$, $x_2 \in N(2,1)$, $x_3 \in N(3,5)$. Теоретическим распределением является нормальный закон $y \in N(20, \sqrt{446})$. Достигнутые уровни значимости по всем применяемым критериям свидетельствуют об очень хорошем согласии полученного в результате моделирования эмпирического распределения с теоретическим распределением (см. табл. 1).

Пример 2. $y = x_1 x_2$, где $x_1 \in N(0,1)$, $x_1 \in Exp(0,1)$. В данном случае теоретическое распределение функции неизвестно и проводилась идентификация закона распределения (проверялась сложная гипотеза). Наилучшей моделью для эмпирического распределения этого произведения оказался нормальный закон с параметрами сдвига θ_1=20.1906 и масштаба θ_2=9.2237. Результаты проверки согласия представлены в таблице 2.

ТАБЛИЦА 1
РЕЗУЛЬТАТЫ ПРОВЕРКИ СОГЛАСИЯ ВЫБОРКИ ВЕЛИЧИНЫ Y ИЗ ПРИМЕРА 1 С РАСПРЕДЕЛЕНИЕМ $N(20, \sqrt{446})$

Критерий	Значение статистики	Достигнутый уровень значимости
χ^2 Пирсона при $k=15$	12.226	0.5881
Колмогорова	0.5218	0.9482
ω^2 Мизеса	0.0381	0.9422
Ω^2 Андерсона-Дарлинга	0.2682	0.9599

ТАБЛИЦА 2
РЕЗУЛЬТАТЫ ПРОВЕРКИ СОГЛАСИЯ ВЫБОРКИ ВЕЛИЧИНЫ Y ИЗ ПРИМЕРА 2 С РАСПРЕДЕЛЕНИЕМ $N(20.1906, 9.2237)$

Критерий	Значение статистики	Достигнутый уровень значимости
χ^2 Пирсона при $k=15$	15.47	0.1619
Колмогорова	0.5751	0.4920
ω^2 Мизеса	0.0349	0.6222
Ω^2 Андерсона-Дарлинга	0.2637	0.5508

Пусть $Y = \prod_{i=1}^{k} X_i$, где X_i – взаимно некоррелированные случайные величины с математическим ожиданием M_i и дисперсией D_i. Согласно (1) $Y \approx \sum_{i=1}^{k} X_i \prod_{\substack{j=1 \\ j\neq i}}^{k} M_j - (k-1)\prod_{j=1}^{k} M_j$, математическое ожидание $E[Y] \approx \prod_{j=1}^{k} M_i$ и дисперсия $D[Y] \approx \sum_{i=1}^{k} D_i \left(\prod_{\substack{j=1 \\ j\neq i}}^{k} M_j \right)^2$.

Распределения произведений исследовались при различных законах распределения X_i.

В случае принадлежности X_i стандартным нормальным законам для $k=\overline{2,5}$ применение линеаризации невозможно по очевидной причине: дисперсия оказывается нулевой. Полученные в результате моделирования распределения Y в данном случае представляют собой асимметричные законы с нулевой медианой. Эти распределения не удается адекватно описать какой-то одной параметрической моделью закона, однако они достаточно хорошо аппроксимируются смесями вида [3,4]:

$$\alpha \frac{\theta_3}{2\theta_2 \Gamma(1/\theta_3)} \exp\left\{-\left(\frac{|y-\theta_1|}{\theta_2}\right)^{\theta_3}\right\} + (1-\alpha) \times$$

$$\frac{1}{\theta_1} \exp\left\{\frac{y-\theta_4}{\theta_5} - \exp\left(\frac{y-\theta_4}{\theta_5}\right)\right\}.$$

В случае нормальных законов с различными параметрами картина существенно меняется. На рис. 1 представлены эмпирические распределения такого рода произведений случайных величин, но принадлежащих нормальному закону с параметрами сдвига равным 4 и масштаба равным 3.

Распределения Y для этого случая при $k=\overline{2,4}$ неплохо описываются смесями двух, а при $k=5$ – трех параметрических моделей.

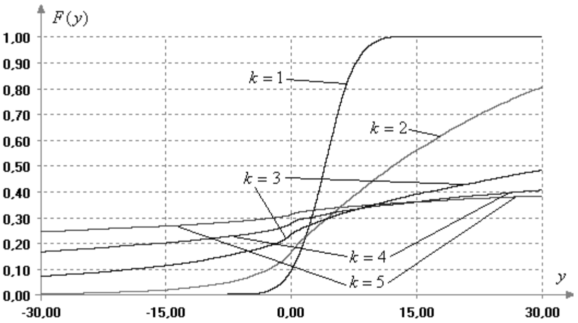

Рис.1. Эмпирические распределения произведений $k=\overline{1,5}$ нормальных величин с параметрами сдвига равным 4 и масштаба равным 3

V. ЗАКЛЮЧЕНИЕ

Таким образом, методы статистического моделирования в совокупности с программным обеспечением, позволяющим строить приближенные математические модели для полученных эмпирических распределений (в том числе в виде смесей различных параметрических законов), представляют собой эффективный инструмент для изучения законов распределения функций от случайных величин, для исследования вероятностных закономерностей, проявляющихся в различных приложениях технического характера.

Распределения функций от случайных величин X_i зависят не только от вида законов распределений X_i и могут меняться в широких пределах в зависимости от параметров этих законов [3, 4]. Используя методы статистического моделирования для исследования закона распределения Y, можно либо построить приближенную модель, аппроксимирующую этот закон в конкретном случае, либо выяснить условия, обеспечивающие обоснованность применения линеаризации.

Использование статистического моделирования и специализированного программного обеспечения, примером которого является развиваемая система "Интервальная статистика" ISW [5], позволяет строить хорошие приближенные математические модели законов распределения функций случайных величин (в том числе, в форме смесей параметрических моделей законов), когда этот закон не удается найти аналитически.

Работа выполнена при частичной поддержке РФФИ (проект № 06-01-0059а).

СПИСОК ЛИТЕРАТУРЫ

[1] МИ 2083-90. ГСИ. Измерения косвенные. Определение результатов измерений и оценивание их погрешностей.

[2] Левин С.Ф. Схема приведения в методе косвенного измерения // Измерительная техника, 2004. – № 3. – С.5-9.

[3] Лемешко Б.Ю., Огурцов Д.В. Статистическое моделирование как эффективный инструмент для исследования законов распределения функций случайных величин // Метрология. 2007. – № 5. – С. 3-13.

[4] Lemeshko B.Yu., Ogurtsov D.V. Statistical modeling as an effective instrument for investigating the distribution laws of functions of random quantities // Measurement Techniques, 2007. V.50, № 6. – P. 593-600

[5] Лемешко Б.Ю., Постовалов С.Н. Компьютерные технологии анализа данных и исследования статистических закономерностей. – Новосибирск: Изд-во НГТУ, 2004. – 119 с.

Лемешко Борис Юрьевич	Профессор кафедры прикладной математики Новосибирского государственного технического университета (НГТУ), д.т.н., профессор.
Огурцов Дмитрий Викторович	Магистр прикладной математики и информатики, в 2008 г. закончил НГТУ.