ПРОГРАММНАЯ ПОДДЕРЖКА МОДЕЛИРОВАНИЯ И ИССЛЕДОВАНИЯ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ ФУНКЦИЙ ОТ СЛУЧАЙНЫХ ВЕЛИЧИН

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

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.

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.

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 $Y = \varphi(\overline{X})$.

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.

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

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

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.

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. $Y = X_1 / X_2$, где $X_1, X_2 \in N(0,1)$ и независимы. Теоретическим законом распределения Y является стандартное распределение Коши с плотностью $f(y) = \dfrac{1}{\pi\left(1 + y^2\right)}$, $y \in (-\infty, +\infty)$. Гипотеза о согласии смоделированной выборки величины Y (10000 значений) с распределением Коши не отвергается по критериям согласия (χ^2 Пирсона при $k = 15$, Колмогорова, ω^2 Мизеса, Ω^2 Андерсона-Дарлинга).

Пример 2. 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 $D[Y] \approx \sum_{i=1}^{k} D_i \left(\prod_{\substack{j=1 \\ j \neq i}}^{k} M_j \right)^2$.

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]:

$$\alpha \frac{\theta_3}{2\theta_2 \Gamma(1/\theta_3)} \exp\left\{ -\left(\frac{|y - \theta_1|}{\theta_2} \right)^{\theta_3} \right\} + (1-\alpha) \frac{1}{\theta_1} \exp\left\{ \frac{y - \theta_4}{\theta_5} - \exp\left(\frac{y - \theta_4}{\theta_5} \right) \right\}.$$

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.

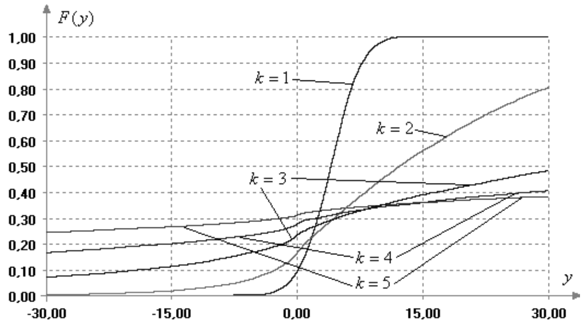

Рис.1. Эмпирические распределения произведений $k=\overline{1,5}$ нормальных величин с параметрами сдвига равным 4 и масштаба равным 3

Таким образом, 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.

Литература

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.