УДК 519.21

И. В. Рудаков, А. В. Шляева

МОДЕЛИРОВАНИЕ ВХОДНЫХ ПОТОКОВ ДАННЫХ ДЛЯ СТОХАСТИЧЕСКИХ МОДЕЛЕЙ ДИСКРЕТНЫХ СИСТЕМ

Рассмотрена проблема моделирования входных данных, являющихся случайными величинами, для стохастических моделей дискретных систем. В качестве модели стационарного случайного процесса выбраны ARTA (AutoRegressive-To-Anything) процессы с произвольным безусловным распределением вероятности и структурой автокорреляции. Представлены алгоритмы оценки параметров ARTA-процесса по имеющейся реализации случайного процесса и генерации ARTA-процесса.

Одним из важнейших этапов создания стохастической имитационной модели сложной системы является моделирование потоков входных данных, зависящих от случайных факторов, причем неверная их формализация приводит к получению недостоверных результатов моделирования и принятию неверных решений [1].

Для моделирования входных данных на этапе построения концептуальной модели системы осуществляется сбор информации по моделируемым случайным воздействиям, в частности экспериментальных данных (реализаций рассматриваемых случайных величин), при этом данные выборки могут быть независимыми или зависимыми.

Существующие в настоящее время программные комплексы для автоматизации процесса моделирования входных данных (ExpertFit, Stat::Fit, Arena Input Analyzer) работают с независимыми выборками данных.

Часто при моделировании входные данные являются линейно-зависимыми, или автокоррелированными (в частности, автокоррелированными являются потоки заявок при моделировании компьютерных сетей), а автокорреляция во входном потоке может существенно повлиять на результаты моделирования. В таких случаях последовательность входных случайных величин моделируют как последовательность, полученную из одного и того же одномерного безусловного распределения, но с некоторой автокорреляцией между ее величинами. Таким образом, моделируемые величины одинаково распределенные, но не независимые. Большинство существующих методов формализации автокоррелированных данных основано на предположении о том, что известно безусловное распределение вероятностей исследуемых величин, причем до последнего времени не было автоматизированных процедур определения безусловного распределения по данным наблюдений.

Таким образом, актуальной становится задача создания программного комплекса для автоматизации моделирования входного потока данных по автокоррелированным данным наблюдений.

Рассмотрим один из методов описания автокоррелированных входных данных при известном безусловном распределении и структуре автокорреляции – ARTA (AutoRegressive-To-Anything) процессы [1, 2].

В работе [3] предложено использовать для построения модели автокоррелированной случайной величины ARTA-процесс — стационарный случайный процесс с произвольным маргинальным распределением вероятности F_X и структурой автокорреляции с конечным шагом p. Для моделирования ARTA-процесса сначала конструируется базовый процесс, т.е. стандартный авторегрессионный гауссов процесс $\{Z_t : t = 1, 2, \ldots, n\}$, из которого преобразованием $U_t = \Phi\left(Z_t\right)$ (где Φ – функция нормального распределения) получают последовательность $\{U_t : t = 1, 2, \ldots, n\}$ автокоррелированных случайных величин, в основе которых лежит равномерное распределение. К полученной последовательности применяют обратную функцию безусловного распределения $X_t = F_X^{-1}\left(U_t\right) = F_X^{-1}\left(\Phi\left[Z_t\right]\right)$ и получают требуемый случайный процесс $\{X_t : t = 1, 2, \ldots, n\}$. Данный подход используется для всех безусловных распределений, несмотря на то что для некоторых распределений обратную функцию распределения F_X^{-1} нельзя получить аналитически, а нужно рассчитывать приближенным численным методом.

Авторегрессионный процесс Z_t определяется как

$$Z_t = \sum_{h=1}^{p} \alpha_h Z_{t-h} + Y_t, \quad t = 1, 2, \ldots, n, \tag{1}$$

где $\alpha_h, h = 1, 2, \ldots, p$ — коэффициенты авторегрессии; Y_t — белый шум (случайная часть Z_t, причем дисперсия $D\left[Y_t\right] = \sigma_Y^2$).

Одна из проблем данного подхода заключается в том, что структура автокорреляции входного процесса X_t не совпадает со структурой автокорреляции базового процесса Z_t. Естественно, возникает задача подбора структуры автокорреляции базового процесса, причем так, чтобы структура автокорреляции входного процесса совпала с необходимой.

В работе [3] показано, что если структура корреляции входного процесса известна, то задача распадается на p независимых задач подбора структуры автокорреляции — определения значений

$\rho_Z(h), h = 1,2,\ldots,p$ при известных значениях автокорреляций $\rho_X(h), h = 1,2,\ldots,p$.

Каждая задача подбора структуры автокорреляции заключается в решении уравнения

$$\rho_X(h) = \frac{\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} F_X^{-1}[\Phi(z_t)]\, F_X^{-1}[\Phi(z_{t-h})]\, \vartheta_{\rho_{Z(h)}}(z_t, z_{t-h})\, dz_t dz_{t-h} - \mu^2}{\sigma^2} \tag{2}$$

для всех $\rho_Z(h), h = 1,2,\ldots,p$, где $\vartheta_{\rho_{Z(h)}}$ — плотность вероятности стандартного двумерного нормального распределения с корреляцией $\rho_Z(h)$.

Данные уравнения не имеют аналитического решения в общем виде, однако существуют эффективные численные методы их решения [4–6]. Решение каждого такого уравнения требует численного интегрирования на каждом шаге численного поиска решения уравнения (2).

Таким образом, для описания входных данных с помощью ARTA-процессов необходимо знать структуру автокорреляции входного процесса и маргинальное распределение, лежащее в основе данных. Однако на практике задача заключается в описании некоторого случайного процесса с помощью ARTA-процесса только по известному ряду измерений. В работах [3, 7] предложено решение проблемы подбора параметров ARTA-процессов, заключающееся в подборе распределения из семейства распределений Джонсона. Предложенный метод работает для всех распределений.

В основе семейства распределений Джонсона лежит преобразование $f(x)$ исходной случайной величины X (заданной в некотором интервале), что позволяет рассматривать результат преобразования как стандартизованную случайную величину, распределенную по нормальному закону. Преобразование Джонсона в общем случае имеет следующий вид:

$$F(x) = \Phi\left[\gamma + \delta f\left(\frac{x-\xi}{\lambda}\right)\right], \tag{3}$$

где f — одна из функций: $f(x) = \ln x$ — логнормальное семейство распределений S_L при $\lambda = 1$, $\xi < X_{(1)}$; $f(x) = \sinh^{-1}(x)$ — неограниченное семейство распределений S_U при $\lambda > 0$; $f(x) = \ln\left[\dfrac{x}{1-x}\right]$ — ограниченное семейство распределений S_B при $\lambda > X_{(n)} - \xi$, $\xi < X_{(1)}$; $f(x) = x$ — нормальное семейство распределений S_N при $\lambda = 1, \xi = 0$.

Функцию f выбирают, исходя из оценки эксцесса и асимметрии.

Для распределений Джонсона найдены оценки параметров методами наименьших квадратов и моментов.

Оценка параметров ARTA-процесса сводится к подбору функции маргинального распределения, шага корреляции p и коэффициентов автокорреляции $\alpha_1, \alpha_2, \ldots, \alpha_p$. Рассмотрим оценку параметров ARTA-процесса при выборе маргинального распределения из семейства распределений Джонсона. Тогда ARTA-процесс можно представить как

$$X_t = F_X^{-1}\left[\Phi\left(Z_t\right)\right] = \xi + \lambda f^{-1}\left[\frac{Z_t - \gamma}{\delta}\right], \tag{4}$$

где $\xi, \lambda, \gamma, \delta$ — параметры распределения Джонсона.

Задача сводится к оценке следующих параметров: $\xi, \lambda, \gamma, \delta, f, p$ и $\alpha_1, \alpha_2, \ldots, \alpha_p$.

При известных порядке p базового процесса и типе f распределения Джонсона для нахождения вектора параметров ARTA-процесса $\psi = (\lambda, \delta, \gamma, \xi, \alpha_1, \alpha_2, \ldots \alpha_p)$ необходимо минимизировать целевую функцию

$$\begin{aligned} S_D\left(\psi\right) = {} & \frac{1}{\left(n-p\right)^2}\times \\ & \times \sum_{t=p+1}^{n} \frac{\left(n-p+1\right)^2\left(n-p+2\right)}{\left(t-p\right)\left(n+1-t\right)}\left(\Phi\left[V_{(t)}\left(\psi\right)\right] - \frac{t-p}{n-p+1}\right)^2, \end{aligned} \tag{5}$$

где $V_t\left(\psi\right) = \dfrac{\gamma + \delta f\left[\dfrac{X_t - \xi}{\lambda}\right] - \sum\limits_{h=1}^{p}\alpha_h\left(\gamma + \delta f\left[\dfrac{X_{t-h} - \xi}{\lambda}\right]\right)}{g\left(p, \alpha\right)}$,

$g\left(p, \vec{\alpha}\right) \equiv \sigma_Y^2$, а $V_{(p+1)}\left(\psi\right) \leqslant V_{(p+2)}\left(\psi\right) \leqslant \ldots \leqslant V_{(n)}\left(\psi\right)$ — порядковые статистики случайных величин $V_t\left(\psi\right), t = p+1, p+2, \ldots, n$, в области

$$\Psi = \left\{ \left(\gamma, \delta, \lambda, \xi, \alpha_1, \alpha_2, \ldots, \alpha_p\right)' : \begin{array}{c} \delta\begin{cases} > 0, f \in \left\{S_U, S_B, S_L, S_N\right\}, \\ < \infty, f \in \left\{S_U, S_B\right\}, \end{cases} \\ \lambda\begin{cases} > 0, f = S_{U,} \\ > X_{(n)} - \xi, f = S_B, \\ = 1, f \in \left\{S_L, S_N\right\}, \end{cases} \\ \xi\begin{cases} < X_{(1)}, f \in \left\{S_L, S_B\right\}, \\ = 0, f = S_N, \end{cases} \\ \left| RootOf\left(1 - \sum\limits_{h=1}^{p}\alpha_h B^h = 0, B\right)\right| > 1 \end{array} \right\}, \tag{6}$$

где значения функции $RootOf$ — это все решения уравнения $1 - \sum\limits_{h=1}^{p}\alpha_h B^h = 0$ по переменной B.

Пусть $C : \Psi \to \Psi$ и $D : \Psi \to \Psi$ — отображения, задаваемые следующими отношениями:

$C\left(\psi\right) \equiv \arg\min_{\gamma,\delta,\lambda,\xi} S_D\left(\psi\left|x\right.\right)$по $\psi \in \Psi_C$;
$D\left(\psi\right) \equiv \arg\min_{\alpha_1,\alpha_2,\ldots,\alpha_p} S_D\left(\psi\left|x\right.\right)$по $\psi \in \Psi_D$,
где

$$\Psi_C = \left\{ (\gamma,\delta,\lambda,\xi)' : \begin{array}{c} \delta \left\{ \begin{array}{l} > 0, f \in \{S_U, S_B, S_L, S_N\}, \\ < \infty, f \in \{S_U, S_B\}, \end{array} \right. \\ \lambda \left\{ \begin{array}{l} > 0, f = S_U, \\ > X_{(n)} - \xi, f = S_B, \\ = 1, f \in \{S_L, S_N\}, \end{array} \right. \\ \xi \left\{ \begin{array}{l} < X_{(1)}, f \in \{S_L, S_B\}, \\ = 0, f = S_N \end{array} \right. \end{array} \right\};$$

$$\Psi_D = \left\{ (\alpha_1,\alpha_2,\ldots,\alpha_p)' : \left| RootOf\left(1 - \sum_{h=1}^{p} \alpha_h B^h = 0, B\right)\right| \geqslant 1 \right\}.$$

Для решения задачи минимизации используется следующий алгоритм [3].

Шаг инициализации.

Пусть $k = 1$, ψ_0 — начальное приближение вектора параметров из области $cl\Psi$.

Основные шаги.

Шаг 1. Пусть $\psi_k \in C\left(\psi_{k-1}\right)$. Увеличить k на единицу, перейти к шагу 2.

Шаг 2. Пусть $\psi_k \in D\left(\psi_{k-1}\right)$. Если достигнуто условие останова — останов. Иначе увеличить k на единицу, перейти к шагу 1.

Необходимо определить условия останова для поиска минимума целевой функции. В работе [7] предложены два следующих условия останова:

$$\left|S_D\left(\psi_k\left|x\right.\right) - S_D\left(\psi_{k-1}\left|x\right.\right)\right| \leqslant AbsoluteErrorRequest \text{ и}$$
$$\left|S_D\left(\psi_k\left|x\right.\right) - S_D\left(\psi_{k-1}\left|x\right.\right)\right| \leqslant S_D\left(\psi_{k-1}\left|x\right.\right) \times RelativeErrorRequest.$$

Данный подход к оценке параметров ARTA-процессов применим не только для распределений Джонсона, но и для любого непрерывного F_X и вектора параметров ψ, определенного в некоторой области Ψ, обладающего определенными свойствами [7].

Приведенный алгоритм указывает на то, что известен тип распределения Джонсона f и порядок базового процесса автокорреляции p. Рассмотрим определение этих параметров.

Для достаточно больших размеров выборки методы подбора преобразования Джонсона для выборки независимых и одинаково распределенных данных правильно определяют преобразование Джонсона f и для автокоррелированной выборки в случае, когда автокорреляция достаточна слабая. Однако при увеличении автокорреляции

Рис. 1. Алгоритм оценки параметров ARTA-процесса

можно определить неправильное преобразование Джонсона. Одно из возможных решений заключается в том, чтобы подобрать параметры для всех семейств распределения и сравнить результаты проверок по критериям согласия [3].

После определения типа преобразования Джонсона с помощью алгоритма AS99 [8] получаем параметры распределения Джонсона, которые можно использовать в качестве начального приближения для алгоритма поиска минимума целевой функции.

На практике выбирают порядок автокорреляции не более 5. Например, в работе [9] утверждается, что обычно порядок зависимости меньше либо равен трем.

Алгоритм оценки параметров ARTA-процесса представлен на рис. 1, алгоритм моделирования — на рис. 2.

Для автоматизации моделирования потока входных данных по набору автокоррелированных данных был создан программный комплекс для оценки параметров и генерации значений ARTA-процессов. В программный комплекс включены функции моделирования входных данных по выборке независимых и одинаково распределенных данных [10].

Программный комплекс реализует следующие функции:

- определение независимости выборки входных данных [10];
- моделирование входных данных в случае выборки независимых и одинаково распределенных данных [10];
- оценку параметров ARTA-процесса по имеющейся выборке (для автокоррелированных данных);

● генерацию реализации ARTA-процесса нужного объема с заданными параметрами (для автокоррелированных данных).

На всех этапах работы предусмотрен вывод графической информации. Также существует функция проверки правильности гипотезы по критериям Колмогорова–Смирнова, Андерсона–Дарлинга и хи-квадрат (только для независимых и одинаково распределенных данных).

Разработанный программный комплекс позволяет моделировать входные данные, являющиеся случайными величинами, для стохастических имитационных моделей дискретных систем по имеющейся выборке — реализации искомой случайной величины. Характерными примерами таких входных случайных величин являются время обработки заявки и время между поступлениями заявок в системах массового обслуживания. Данный подход был успешно применен для моделирования измерения давления на поточной линии производства пластика, а также для моделирования продаж крупного предприятия транспортной промышленности [3].

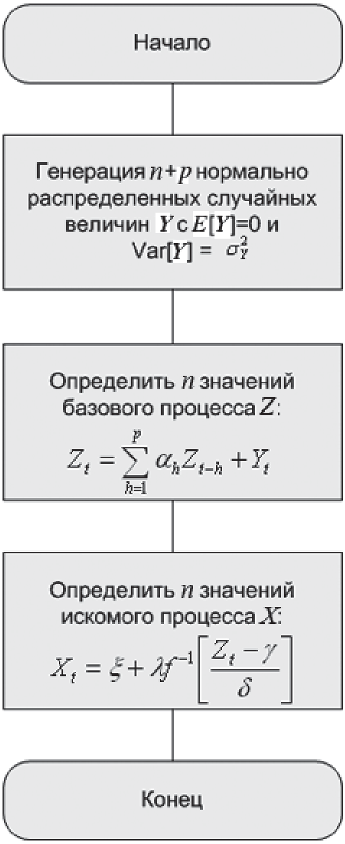

Рис. 2. Алгоритм генерации ARTA-процесса

СПИСОК ЛИТЕРАТУРЫ

1. К е л ь т о н В., Л о у А. Имитационное моделирование. – СПб.: Питер, 2004. – 847 с.
2. Ш л я е в а А. В. Использование ARTA- и TES-процессов для моделирования входных данных стохастических имитационных моделей // Сб. тр. научно-практич. семинара “Новые информационные технологии в автоматизированных системах-9”. – М.: МГИЭМ, 2006. – С. 178–180.
3. B i l l e r B., N e l s o n B. L. Parameter Estimation for ARTA Processes // Proceedings of the 2002 Winter Simulation Conference. – 2002. – P. 255–262.
4. C a r i o M. C., N e l s o n B. L. Numerical methods for fitting and simulating autoregressive-to-anything processes // INFORMS Journal on Computing. – 1998. – No. 10. – P. 72–81.
5. C h e n H. Initialization for NORTA: Generation of random vectors with specified marginals and correlations // INFORMS Journal on Computing. – 2001. – № 13. – P. 312–331.
6. S o n g W. T., H s i a o L. and C h e n Y. Generating pseudorandom time series with specified marginal distributions // European Journal of Operational Research. – 1996. – No. 93. – P. 1–22.

7. B i l l e r B., N e l s o n B. L. Fitting time series input processes for simulation // Operations Research. – 2005. – No. 53. – P. 549–559.
8. H i l l I. D., H i l l R., H o l d e r R. L. Fitting Johnson curves by moments // Applied Statistics. – 1976. – No. 25. – P. 180–189.
9. W e i W. W. S. Time Series Analysis: Univariate and Multivariate Methods. – New York: Addison Wesley, 1990.
10. Р у д а к о в И. В., Ш л я е в а А. В. Моделирование входных данных для стохастических имитационных моделей систем // Информационные технологии. – М.: Новые технологии, 2006. – 11 с.

Статья поступила в редакцию 24.04.2007

Игорь Владимирович Рудаков родился в 1958 г., окончил в 1981 году МВТУ им. Н.Э. Баумана. Канд. техн. наук, доцент кафедры “Программное обеспечение ЭВМ и информационные технологии” МГТУ им. Н.Э. Баумана. Автор более 50 научных работ в области вычислительной техники и программного обеспечения.

I.V. Rudakov (b. 1958) graduated from the Bauman Moscow Higher Technical School in 1981. Ph. D. (Eng.), assoc. professor of “Computer Software and Information Technologies” department of the Bauman Moscow State Technical University. Author of more than 50 publications in the field of computing technology and software.

Шляева Анна Викторовна родилась в 1984 г., окончила МГТУ им. Н.Э. Баумана в 2007 г. Автор трех научных работ в области численных методов и моделирования.

A.V. Shlyaeva (b. 1984) graduated from the Bauman Moscow State Technical University in 2007. Author of 3 publications in the field of numerical methods and simulation.