УДК 519.218.5

МОДЕЛИРОВАНИЕ СЛУЧАЙНЫХ ПОТОКОВ СОБЫТИЙ НА ОСНОВЕ ЭМПИРИЧЕСКИХ ДАННЫХ

Смирнов С. В. (Самара)

Концепция потока событий весьма плодотворна в исследованиях поведения сложных систем. Часто в актуальных задачах потоки событий, переводящие систему из одного состояния в другое, могут быть с достаточным основанием отнесены к простым рекуррентным потокам [1, 2]. Такие потоки полностью определяются законом распределения величины промежутков времени между событиями в потоке $\tau \in (0, \infty)$.

Построение подобной модели производится на основе статистических данных, которые представляют собой эмпирически зарегистрированную последовательность промежутков времени между событиями в реальном потоке:

- во-первых, проверяются гипотезы о стационарности этой последовательности и возможности расценивать члены этой последовательности как независимые реализации некоторой положительно определенной случайной величины. Проверка производится либо путем анализа физических условий возникновения потока (важно установить неизменность этих условий и выявить степень влияния факта наступления некоторого события на условия появления последующих событий потока), либо формальными методами (стационарность можно оценивать с помощью непараметрических статистических критериев типа серий, тренда и т.п.; необходимым условием независимости является отсутствие корреляции ряда наблюдений [3]);
- во-вторых, выбирается вид аппроксимирующей («теоретической») функции распределения для τ, который оказывает определяющее влияние на удобство, а иногда и саму возможность конструктивного использования модели потока в актуальном исследовании. Наиболее предпочтительными видами аппроксимирующих распределений для τ (и в аналитической работе, и в имитационных экспериментах) являются экспоненциальное и образуемые на его основе гипо- и гиперэкспоненциальные распределения [4-8], т.е. структурно простые члены семейства гиперэрланговских распределений, обладающего хорошими аппроксимирующими свойствами [9];
- наконец, находятся параметры аппроксимирующего закона распределения τ. Исходной информацией для этого служит эмпирический закон распределения (например, в форме гистограммы) и/или оценки его числовых характеристик. Вопрос о полноценности получаемой аппроксимации распределения для τ связан с теорией вероятностных метрик [3, 9]. На практике широко используется то обстоятельство, что распределение ограниченной случайной величины полностью определяется ее моментами. Поэтому в методе моментов [4, 9] параметры аппроксимирующего закона распределения для τ определяются из условия равенства важнейших моментов у теоретического и эмпирического распределений. При этом стараются ограничиться статистически надежно определяемыми моментами – математическим ожиданием и дисперсией: $E(\tau)$, $D(\tau)$.

В предлагаемой работе показывается совместное решение задач выбора и параметризации аппроксимирующего закона распределения для τ на основе эмпирических оценок среднего и дисперсии этой величины. Приводятся примеры соответствующих имитаторов рекуррентных потоков на языке GPSS.

Рассматриваемая комплексная задача сводится к определению положительно определенной случайной величины X, для математического ожидания $E(X)$ и дисперсии $D(X)$ которой выполняется система условий:

$$\left\{\begin{array}{l} E(X) = E(\tau), \\ D(X) = D(\tau). \end{array}\right\} \quad (1)$$

Вид условий (1) ограничивает выбор распределений для X двухпараметрическими (т.е. зависящими от двух параметров – q_1 и q_2) законами $F_X(x) = F_X(x, q_1, q_2)$ с моментами $E(X) = f_1(q_1, q_2)$, $D(X) = f_2(q_1, q_2)$. Поэтому система условий (1) является системой 2-х уравнений относительно 2-х неизвестных q_1 и q_2:

$$\left\{\begin{array}{l} f_1(q_1, q_2) = E(\tau), \\ f_2(q_1, q_2) = D(\tau). \end{array}\right\} \quad (2)$$

Для ряда известных двухпараметрических распределений положительно определенных случайных величин (Вейбулла, гамма-распределения, усеченного нормального, логарифмически нормального и др.) решение системы (2) существует и может быть получено для любых $0 < E(\tau)$, $D(\tau) < \infty$. Однако с точки зрения удобства использования соответствующих моделей в аналитических и имитационных исследованиях указанные законы распределения часто малоприемлемы как из-за некоторых принципиальных свойств этих моделей, так и в следствие громоздкости вычислений, сопровождающих использование таких распределений [10].

В [5] обоснован метод замены произвольного рекуррентного потока обобщенным потоком Эрланга k-го порядка, у которого случайный промежуток времени X между событиями является суммой экспоненциальных случайных величин X_i, $i = 1, 2, \ldots, k$, из которых $k - 1$ подчиняется экспоненциальному закону с параметром λ_0 и одна – экспоненциальному закону с параметром λ_1. Параметры λ_0 и λ_1 определяются из системы (2), которая для рассматриваемого распределения может быть приведена к виду

$$\left\{\begin{array}{l} \lambda_1 = \lambda_0/(1 - k + \lambda_0 E(\tau)), \\ (E(\tau)^2 - D(\tau))\lambda_0{}^2 + 2(1 - k)E(\tau)\lambda_0 + k(k-1) = 0. \end{array}\right\} \quad (3)$$

Структурный параметр k гипоэкспоненциального распределения, описывающего промежутки времени между событиями в обобщенном потоке Эрланга, определяется на основе эмпирических данных по формуле [5]:

$$k = \left\{\begin{array}{l} E(\tau)^2/D(\tau), \\ \mathbf{entier}(E(\tau)^2/D(\tau)) + 1, \text{если первое выражение дает дробное значение.} \end{array}\right\} \quad (4)$$

Тогда решением системы (3) являются две различные пары значений параметров λ_0 и λ_1 одинаково пригодные для моделирования.

Из (4) и очевидного требования $k \geq 2$ ясно, что описанный способ моделирования может быть применен лишь при соблюдении условия $D(\tau) < E(\tau)^2$, т.е. тогда, когда выборочный коэффициент вариации величины промежутка времени между событиями в рекуррентном потоке не превышает единицы:

$$v(\tau) = \frac{\sqrt{D(\tau)}}{E(\tau)} < 1.$$

При $v(\tau) = 1$ наиболее приемлемой аппроксимацией реального потока следует признать простейший поток, в котором промежутки времени между соседними событиями распределены по экспоненциальному закону с параметром

$$\lambda_0 = \frac{1}{E(\tau)} = \frac{1}{\sqrt{D(\tau)}}.$$

Для аппроксимационного моделирования потоков с промежутками времени между событиями, характеризуемыми большей относительно величины математического ожидания дисперсией (т.е. когда $1 < \nu(\tau) < \infty$), в [11] предложено использовать специальный вид гиперэкспоненциального распределения, параметризация которого возможна для всего диапазона возможных значений $E(\tau)$ и $D(\tau)$ при $D(\tau) > E(\tau)^2$.

В общем случае схему образования гиперэкспоненциальной случайной величины X можно представить формулой

$$X = \begin{Bmatrix} X_0 \text{ с вероятностью } p_0, \\ X_1 \text{ с вероятностью } p_1, \\ \cdots \\ X_n \text{ с вероятностью } p_n, \end{Bmatrix}, \quad (5)$$

где $n > 0$, $\sum_{i=0}^{n} p_i = 1$, X_i – случайная величина, распределенная по экспоненциальному закону с параметром λ_i, $i = 0, 2,\ldots, n$.

Специальный вид гиперэкспоненциального распределения допускает параметризацию на основе эмпирических данных в смысле (2) и определяется следующей конкретизацией общей схемы (5): $n = 1$, $p_0 = 1 - p$, $p_1 = p$, $\lambda_1 = p\lambda_0$.

Другими словами, в качестве аппроксимирующего распределения для τ предлагается выбирать двухфазное гиперэкспоненциальное распределение с двумя независимыми параметрами $p \in (0, 1)$ и $\lambda_0 > 0$. В этом случае система уравнений (2) существует и имеет вид

$$\begin{Bmatrix} \lambda_0 = (2 - p)/E(\tau), \\ (\nu(\tau)^2 + 1)p^3 - (4\nu(\tau)^2 + 2)p^2 + (4\nu(\tau)^2 + 2) - 2 = 0. \end{Bmatrix} \quad (6)$$

Для отыскания решения (6) необходимо вначале найти корень кубического уравнения системы. В [11] доказано, что существует, по меньшей мере, один положительный корень этого уравнения, лежащий в интервале (0, 1). Кроме того, доказано, что функция $p(\nu(\tau))$ непрерывна и строго убывает на множестве $\nu(\tau) \in [1, \infty)$. Это позволяет определять подходящий корень p кубического уравнения в (6) по выборочному $\nu(\tau)$ путем интерполяции табличных значений функции $p(\nu(\tau))$ – см. таблицу 1 [11]. В силу такой однозначности аппроксимация распределения случайной величины τ при $\nu(\tau) > 1$ рассматриваемым специальным распределением будет предпочтительнее, чем использование двухфазного гиперэкспоненциального распределения согласно общей схемы (5), т.к. идентификация его трех параметров по эмпирическим оценкам моментов $E(\tau)$ и $D(\tau)$ предполагает наличие некоторого произвола [6, 12].

Таким образом, заключаем, что рациональный выбор аппроксимирующего распределения интервала времени между событиями в их рекуррентном потоке, основанный на эмпирических оценках среднего и дисперсии величины указанного интервала в моделируемом реальном потоке событий, включает следующие шаги:

- вычисление эмпирического коэффициента вариации $\nu(\tau)$;
- выбор в зависимости от величины $\nu(\tau)$ типа аппроксимирующего распределения: при $\nu(\tau) \in (0, 1)$ – гипоэкспоненциальное, при $\nu(\tau) \approx 1$ – экспоненциальное, при $\nu(\tau) > 1$ – гиперэкспоненциальное специального вида;
- выполнение параметризации аппроксимирующего распределения одним из рассмотренных способов в соответствии с типом этого распределения.

Таблица 1 – Зависимость вероятности p выбора фазы в схеме специального двухпараметрического гиперэкспоненциального распределения от величины необходимого коэффициента вариации $v(\tau)$

$v(\tau)$	p	$v(\tau)$	p	$v(\tau)$	p	$v(\tau)$	p
1.0000	1.000	1.8419	0.150	3.5377	0.040	15.811	0.002
1.0009	0.900	2.0527	0.120	3.7814	0.035	22.361	0.001
1.0069	0.800	2.2447	0.100	4.0839	0.030	31.623	0.0005
1.0226	0.700	2.3644	0.090	4.4732	0.025	70.711	0.0001
1.0530	0.600	2.5062	0.080	5.0008	0.020	100.00	0.00005
1.1055	0.500	2.6776	0.070	5.7740	0.015	223.61	0.00001
1.1924	0.400	2.8907	0.060	7.0713	0.010	707.11	0.000001
1.3384	0.300	3.0186	0.055	8.4517	0.007		
1.4498	0.250	3.1653	0.050	10.000	0.005		
1.6063	0.200	3.3359	0.045	12.910	0.003		

Предложенный подход к моделированию рекуррентных потоков событий на основе эмпирических данных позволяет легко формировать случайные потоки событий в имитационных моделях.

Например, разработчиками популярной среды имитационного моделирования GPSS предлагается табулированная функция XPDIS, определяющая датчик реализаций экспоненциально распределенной случайной величины с параметром $\lambda_0 = 1$. С помощью него можно легко моделировать простейшие потоки событий – рекуррентные потоки с экспоненциально распределенным интервалом времени между событиями, – характеризуемые произвольными параметрами интенсивности λ_0.

Входной поток транзактов в некоторой GPSS-модели, описанный как обобщенный поток Эрланга (т.е. рекуррентный поток с гипоэкспоненциально распределенным интервалом времени между событиями) с параметрами $k = 3$, $\lambda_0 = 0.005$, $\lambda_1 = 0.01$, может быть определен следующим образом:

```
      GENERATE   ,,,1
      ADVANCE    100,FN$XPDIS
RPT   ADVANCE    200,FN$XPDIS
      ADVANCE    200,FN$XPDIS
      SPLIT      1,RPT
      ...
```

Аналогично рекуррентный поток со специального вида гиперэкспоненциальным распределением интервала времени между событиями, конкретизированным параметрами $p = 0.005$, $\lambda_0 = 0.01$, можно задать следующими GPSS-блоками:

```
      GENERATE   ,,,1
NEXT  TRANSFER   .005,,PHS2
      ADVANCE    100,FN$XPDIS
      TRANSFER   OUTC
PHS2  ADVANCE    2000,FN$XPDIS
OUTC  SPLIT      1,NEXT
      ...
```

Литература

1. **Кокс Д., Смит В.** Теория восстановления. – М.: Сов. радио, 1967. – 298 с.
2. Словарь по кибернетике / Под ред. В.С. Михалевича. 2-е изд. – Киев: Главная редакция Украинской Советской Энциклопедии имени М. П. Бажана, 1989. – 751 с.

3. **Кобзарь А.И.** Прикладная математическая статистика. Для инженеров и научных работников. – М.: ФИЗМАТЛИТ, 2006. – 816 с.
4. **Венцель Е.С., Овчаров Л.А.** Теория случайных процессов и ее инженерные приложения: учебное пособие. 5-е изд., стер. — М.: ЮСТИЦИЯ, 2018. – 480 с.
5. **Тараканов К.В., Овчаров Л.А., Тырышкин А.Н.** Аналитические методы исследования систем. – М.: Сов. радио, 1974. – 240 с.
6. **Тарасов В.Н.** Исследование систем массового обслуживания с гиперэкспоненциальными входными распределениями // Проблемы передачи информации. – 2016. – Т. 52, выпуск 1. – С. 16–26.
7. **Рыжиков Ю.И., Уланов А.В.** Применение гиперэкспоненциальной аппроксимации в задачах расчета немарковских систем массового обслуживания // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2016. – №3. – С. 60-65.
8. **Томашевский В.Н., Жданова Е.Г.** Имитационное моделирование в среде GPSS. – М.: Бестселлер, 2003. – 416 с.
9. **Калашников В.В., Рачев С.Т.** Математические методы построения стохастических моделей обслуживания. – М.: Наука, 1988. – 312 с.
10. **Ivleva A., Smirnov S.** Comparison of Models of Positively Defined Random Variables // Proc. XXI Int. Conf. “Complex Systems: Control and Modeling Problems” CSCMP 2019. Eds.: S.A. Nikitov, D.E. Bykov, S.Yu. Borovik, Yu.E. Pleshivtseva. – IEEE Xplore. – P. 449-454.
11. **Смирнов С.В.** Моделирование «сверхнерегулярных» случайных величин по экспериментальным данным // Автоматизация научных исследований: Межвуз. сб. научн. трудов. – Куйбышев: КуАИ, 1988. – С. 52-57.
12. **Вишневский В.М.** Теоретические основы проектирования компьютерных сетей. – М.: Техносфера, 2003. – 512 с.