

МОДЕЛИРОВАНИЕ "СВЕРХНЕРЕГУЛЯРНЫХ" СЛУЧАЙНЫХ ВЕЛИЧИН
ПО ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ

(г. Куйбышев)

Практически при экспериментальном исследовании случайных величин надежно могут быть оценены лишь некоторые их моменты; обычно ограничиваются выборочными оценками среднего и дисперсии [1, 2]. Поэтому построение моделей случайных величин по таким экспериментальным данным фактически предполагает такую аппроксимацию реальной случайной величины, которая обеспечивает совпадение среднего и дисперсии у модели и прототипа. Несомненный интерес подобная аппроксимация представляет при условии неотрицательности исследуемой случайной величины (примерами таких случайных величин являются различные интервалы времени: промежутки между событиями в потоках событий различной природы, наработка на отказ и т.п.) и требование простоты получения реализаций случайной величины по ее модели (последнее приобретает особую важность при разработке на основе экспериментальных данных имитационных моделей исследуемых систем).

Пусть в результате статистической обработки экспериментальных данных получены оценки математического ожидания и дисперсии некоторой положительно определенной случайной величины: соответственно m_X и D_X . Рассмотрим задачу замены этой реальной случайной величины гипотетической случайной величиной с теми же математическим ожиданием и дисперсией.

Эта задача сводится к определению случайной величины T , для математического ожидания $M\{T\}$ и дисперсии $D\{T\}$ которой выполняется система условий:

$$\begin{cases} M\{T\} = m_X, \\ D\{T\} = D_X. \end{cases} \quad (I)$$

Счевидно, вид условий (I) ограничивает выбор распределений двухпараметрическими (т.е. зависящими от двух параметров - P_1 и P_2) распределениями: $F_T(t) = F_T(t, P_1, P_2)$, с ограниченными $M\{T\} = \varphi(P_1, P_2)$ и $D\{T\} = \varphi_2(P_1, P_2)$. При таком выборе распределения T - система условий (I) является системой двух уравнений относительно двух неизвестных P_1 и P_2 :

$$\begin{cases} \mathcal{Y}_1(P_1, P_2) = m_x, \\ \mathcal{Y}_2(P_1, P_2) = D_x. \end{cases} \quad (2)$$

Можно показать, что для важнейших двухпараметрических распределений положительно определенных случайных величин (Вейбулла, гамма-распределения, усеченного нормального, логарифмически-нормального) решение системы (2) существует и может быть вычислено для любых величин оценок $0 < m_x, D_x < \infty$. Однако, с точки зрения удовлетворения дополнительному требованию простоты получения случайных чисел с подобными распределениями, указанные двухпараметрические законы оказываются малопримемыми для моделирования главным образом из-за громоздкости соответствующих вычислительных методов и существенных затрат машинного времени при их использовании [3].

Действительно, в простейшем случае для распределения Вейбулла в соответствии с методом обратных функций значения случайной величины вычисляются по формуле

$$t = \left(-\frac{1}{\alpha} \ln r \right)^{\frac{1}{\alpha}},$$

где $\alpha, \lambda > 0$ - параметры распределения Вейбулла, а r - реализация равномерно распределенной на интервале $[0, 1]$ случайной величины. Операция возведения в степень с вещественным в общем случае показателем является при реализации на ЭВМ наиболее медленной арифметической операцией, предполагающей обычно вызов библиотечной подпрограммы [4]. В некоторых популярных языках программирования и моделирования встроенная операция возведения в степень вообще отсутствует, как например в Модуле-2 [5] и *GPSS* [6].

В работе [7] предложен метод замены реального потока с ограниченным последствием обобщенным потоком Эрланга k -го порядка, у которого случайный промежуток времени T между событиями является суммой экспоненциальных случайных величин T_i , $i = \overline{1, k}$, $k-1$ из которых подчиняются экспоненциальному закону распределения с параметром λ_0 и одна - экспоненциальному закону с параметром λ_1 .

Параметры λ_0, λ_1 распределения величины T , т.е. параметры гипоекспоненциального распределения [8], определяются из системы уравнений (2). Причем решением системы (2), содержащей в данном случае квадратное уравнение, являются две различные пары значений параметров $(\lambda_0^{(1)}, \lambda_1^{(1)})$ и $(\lambda_0^{(2)}, \lambda_1^{(2)})$, одинаково пригодные для моделирования. Параметр k в системе (2) явно отсутствует, но задает структуру выражений \mathcal{Y}_1 и \mathcal{Y}_2 . Его определение по экспериментальным данным производится по формуле

$$k = \left\{ \begin{array}{l} \frac{m_x^2}{D_x}, \\ \left\lceil \frac{m_x^2}{D_x} \right\rceil + 1, \end{array} \right. \quad \text{если первое выражение дает} \quad (3)$$

дробное значение.

Определение случайной величины T , распределенной по гипотезно-экспоненциальному закону, указывает весьма простой и достаточно эффективный способ моделирования ее реализаций: получение необходимого количества случайных чисел с экспоненциальным распределением и вычисление их суммы. Однако из формулы (3) и очевидного условия $k \geq 2$ ясно, что описанный способ моделирования величины T может быть применен только при соблюдении условия $D_x < m_x^2$, т.е. при условии, что выборочный коэффициент вариации реальной случайной величины не превышает единицы:

$$V_x = \frac{\sqrt{D_x}}{m_x} < 1.$$

При $V_x = 1$ наиболее приемлемой аппроксимацией распределения реальной случайной величины следует признать экспоненциальное распределение с параметром $\lambda_0 = \frac{1}{m_x} = \frac{1}{\sqrt{D_x}}$.

На практике нередки случаи наблюдения "сверхнерегулярных" случайных величин, оценка коэффициента вариации которых превышает единицу. Например, результаты измерений показывают, что коэффициенты вариации распределения времени занятия центрального процессора различными программами в мультипрограммных вычислительных системах в основном значительно больше единицы [8]. Такая же картина оказывается характерной для случайной величины промежутка времени между обращениями страничных программ к каждой из страниц их адресного пространства [9]. Поэтому актуальна задача указания простого и "быстрого" моделирования случайной величины с $V_x > 1$.

В качестве аппроксимирующего может быть предложено гиперэкспоненциальное распределение, которое всегда обеспечивает коэффициент вариации больший чем единица [8].

Механизм образования гиперэкспоненциально распределенной случайной величины T можно представить формулой

$$T = \left\{ \begin{array}{ll} T_0 & \text{с вероятностью } p_0 \\ T_1 & \text{с вероятностью } p_1 \\ \vdots & \vdots \\ T_n & \text{с вероятностью } p_n \end{array} \right\},$$

где $n > 0$, $\sum_{i=0}^n p_i = 1$, T_i - случайная величина, распределенная по экспоненциальному закону с параметром λ_i , $i = \overline{0, n}$.

Выражения для математического ожидания и дисперсии гиперэкспоненциальной случайной величины имеют вид

$$M\{T\} = \sum_{i=0}^n \frac{\rho_i}{\lambda_i}, \quad (4)$$

$$D\{T\} = 2 \sum_{i=0}^n \frac{\rho_i}{\lambda_i^2} - \left(\sum_{i=0}^n \frac{\rho_i}{\lambda_i} \right)^2, \quad (5)$$

Моделирование реализации случайной величины T , распределенной по гиперэкспоненциальному закону, очевидно, включает два этапа. На первом моделируется значение случайного номера N с дискретным распределением $P(N=i) = \rho_i$, $i = \overline{0, n}$. На втором этапе генерируется случайное число, экспоненциально распределенное с параметром λ_i , которое и фиксируется как реализация T . Таким образом, алгоритм получения случайных чисел с гиперэкспоненциальным распределением основан на использовании экспоненциально распределенных случайных чисел и не требует большого объема вычислений.

Трудность использования гиперэкспоненциального распределения для аппроксимации распределения случайной величины с точностью до совпадения математических ожиданий и дисперсии у модели и прототипа заключается в многопараметрическом характере гиперэкспоненциального закона распределения; количество параметров пропорционально величине структурного параметра n .

В простейшем случае $n = 1$ гиперэкспоненциальное распределение зависит от трех параметров: ρ , λ_0 , λ_1 (ρ определяет ρ_0 и ρ_1 ; например, $\rho_1 = \rho$, $\rho_0 = 1 - \rho$). Дальнейшее сокращение числа параметров до двух связано либо с фиксированием значения одного из них, либо требует введения однозначной зависимости одного из параметров от другого. Выбор любого из этих вариантов потенциально способен снизить мощность моделирования гиперэкспоненциального распределения, понимаемую как возможность аппроксимации распределения любой положительно определенной случайной величины с точностью до совпадения математического ожидания и дисперсии у модели и прототипа при коэффициенте вариации большем единицы.

Рассмотрим гиперэкспоненциальное распределение с $n = 1$, независимыми параметрами которого являются $\rho \in (0, 1)$ и $\lambda_0 > 0$, а зависимыми $\rho_0 = 1 - \rho$, $\rho_1 = \rho$, $\lambda_1 = \rho \lambda_0$. Исследуем возможность определения параметров ρ и λ_0 по экспериментальным данным и мощность предлагаемой модели в установленном смысле. Подставляя результат вычисления (4), (5) для введенного распределения в систему (2) и производя несложные преобразования, получим

$$\left\{ \begin{aligned} \lambda_0 &= \frac{2-\rho}{m_x}, \end{aligned} \right. \quad (5)$$

$$\left\{ \begin{aligned} (V_x^2 + 1)\rho^3 - (4V_x^2 + 2)\rho^2 + (4V_x^2 + 2)\rho - 2 = 0. \end{aligned} \right. \quad (7)$$

Поскольку предполагается, что $V_x = D_x/m_x^2 > 1$, уравнение (7) в соответствии с правилом знаков Декарта [11] имеет, по крайней мере, один положительный корень. С другой стороны, такой корень принадлежит интервалу $(0, 1)$, так как при $\rho = 0$ и $\rho = 1$ значения правой части уравнения (7) равны соответственно -2 и $V_x^2 - 1 > 0$.

Функция $\rho(V_x)$ непрерывна и строго убывает на интервале $[1, \infty)$. Этот факт нетрудно проверить, исследуя выражение (7), предварительно преобразовав его к виду

$$V_x^2 = \frac{-\rho^3 + 2\rho^2 - 2\rho + 2}{\rho^3 - 4\rho^2 + 4\rho}. \quad (8)$$

Анализ функций третьей степени [10] в числителе и знаменателе правой части выражения (8) позволяет установить, что с ростом ρ от 0 до 2/3 V_x^2 монотонно убывает, так как строго убывает числитель, а знаменатель строго растет. При изменении ρ от 2/3 до 1 знаменатель также начинает убывать, но, как показывают расчеты, медленнее, чем числитель. В конечном итоге получаем, что функция $V_x(\rho)$ непрерывна и строго убывает на полуинтервале $(0, 1)$. Следовательно, интересующая нас функция $\rho(V_x)$ непрерывна и строго убывает на множестве $[1, \infty) \ni V_x$.

Таким образом, для любых m_x и D_x при $D_x > m_x^2$, решая уравнение (7), устанавливаем значение параметра ρ , а затем из уравнения (5) - значение параметра λ_0 введенного двухпараметрического гиперэкспоненциального распределения: $\rho \in (0, 1)$, $\lambda_0 > 0$. Другими словами, предложенное распределение может быть использовано для моделирования положительно определенных случайных величин с заданными математическим ожиданием и дисперсией, когда коэффициент вариации превышает единицу.

Для отыскания корней кубического уравнения (7) можно воспользоваться либо известными алгоритмами, либо библиотечными программами ЭВМ. Кроме того, отмеченные свойства функции $\rho(V_x)$ позволяют рекомендовать определять ρ по выборочному V_x путем интерполяции заранее вычисленных табличных значений $\rho(V_x)$.*

* Автор благодарит и.а.Борисовскую за помощь в подготовке статьи.

Вероятность смещения ρ в двухпараметрическом гиперэкспоненциальном распределении

V_x	ρ	V_x	ρ	V_x	ρ	V_x	ρ
I,0000	I,000	I,84I9	0,150	3,5377	0,040	I5,8II	0,002
I,0009	0,900	2,0527	0,120	3,7814	0,035	22,36I	0,00I
I,0069	0,800	2,2447	0,100	4,0839	0,030	3I,623	0,0005
I,0226	0,700	2,3644	0,090	4,4732	0,025	70,7II	0,000I
I,0530	0,600	2,5052	0,080	5,0008	0,020	I00,00	0,00005
I,1055	0,500	2,6776	0,070	5,7740	0,015	223,6I	0,0000I
I,1924	0,400	2,8907	0,060	7,07I3	0,010	707,II	0,00000I
I,3364	0,300	3,0186	0,055	8,4517	0,007		
I,4498	0,250	3,1653	0,050	I0,000	0,005		
I,5053	0,200	3,3359	0,045	I2,910	0,003		

Библиографический список

1. Кюн Ю. Описательная и индуктивная статистика. М.: Финансы и статистика, 1981. - 126 с.
2. Шаракшанэ А.С., Железнов И.Г. Испытания сложных систем. М.: Высшая школа, 1974. - 184 с.
3. Иванова В.М. Случайные числа и их применение. М.: Финансы и статистика, 1984. - III с.
4. Ван Тассел Д. Стиль, разработка, эффективность, отладка и испытание программ. М.: Мир, 1985. - 332 с.
5. Вирт Н. Программирование на языке Модуля-2. М.: Мир, 1987. - 224 с.
6. Шрайбер Т.Дж. Моделирование на *GPS*. М.: Машиностроение, 1980. - 592 с.
7. Тараканов К.В., Овчаров Л.А., Тырышкин А.Н. Аналитические методы исследования систем. М.: Сов. радио, 1974. - 240 с.
8. Феррари Д. Сценка производительности вычислительных систем. М.: Мир, 1981. - 576 с.
9. Авен С.И., Коган Я.А. Управление вычислительным процессом в ЭВМ. Алгоритмы и модели. М.: Энергия, 1978. - 240 с.