УДК 519.245

ALIAS-МЕТОД ДЛЯ МОДЕЛИРОВАНИЯ ТАБЛИЧНО ЗАДАННЫХ РАСПРЕДЕЛЕНИЙ СЛУЧАЙНЫХ ВЕЛИЧИН

А. Н. Залялов
(ФГУП "РФЯЦ-ВНИИЭФ", г. Саров Нижегородской области)

Рассмотрен Alias-метод моделирования таблично заданных распределений случайных величин. Приведен оригинальный алгоритм предварительной обработки таблиц. Представлены формулы для моделирования многомерных таблично заданных распределений: дискретного, кусочно-линейного и кусочно-постоянного.

Ключевые слова: случайная величина, Alias-метод, таблично заданные распределения, метод Монте-Карло.

Введение

Для моделирования энергетического, углового и пространственного распределений источника частиц или его энергии в методиках переноса частиц обычно используются три типа табличного задания этих распределений: дискретное, кусочно-линейное и кусочно-постоянное, заданное либо гистограммой, либо в виде доли частиц в каждом интервале. В качестве основного шага моделирования таких распределений традиционно используется стандартный метод моделирования случайных величин [1], который заключается в следующем.

Пусть дискретное распределение случайной величины ξ задано таблицей, содержащей ее значения $x_1, \ldots, x_n$ и соответствующие вероятности $p_1, \ldots, p_n$; n — длина таблицы. Пусть η — случайная величина, равномерно распределенная на отрезке $[0, 1]$. Тогда если $\sum\limits_{k=1}^{m} p_k \le \eta < \sum\limits_{k=1}^{m+1} p_k$, $m = \overline{1, n-1}$, $\sum\limits_{k=1}^{n} p_k = 1$, то выбирается значение случайной величины $\xi = x_m$.

Отсюда следует, что при больших значениях n на выборку значения случайной величины ξ может тратиться значительное время, растущее пропорционально n. При этом основное время тратится на вычисление суммы $\sum\limits_{k=1}^{m} p_k$.

В монографии [1] рассматриваются примеры распределений, в которых для сокращения времени вычислений изменяется порядок вычисления сумм. В методике С-007 [2] и в книге [3] для этой цели используется метод деления отрезка пополам. В этом методе с ростом n время выборки растет как $\log_2 n$.

Если бы все p_k были равны $1/n$, то выборку случайного значения ξ можно было бы определить по формуле $\xi = x_{[n\eta]+1}$. Очевидно, что время выборки в этом случае не зависит от n.

В данной работе рассматривается Alias-метод, который по сравнению со стандартным методом обладает существенным преимуществом при моделировании распределений, заданных таблицами большой длины.

Alias-метод

В работах 70-х годов прошлого века [4, 5] Уолкер предложил оригинальный метод для выборки значения случайной величины из дискретного распределения с разными вероятностями, по скорости не уступающий методу выборки из дискретного распределения с равными вероятностями. Позже этот метод был независимо заново открыт Брауном, Мартином и Калаханом, назван Alias-методом и применен для выборки значений случайных величин из дискретных распределений в программе моделирования методом Монте-Карло переноса частиц [6]. После этого, хотя Alias-метод широко не использовался, он был распространен на непрерывные распределения [7—9].

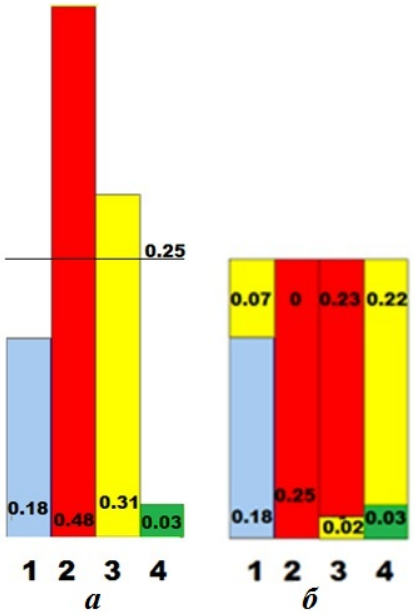

Рис. 1. Пример представлений распределения случайной величины: *а* — в исходном формате; *б* — в Alias-формате

Суть Alias-метода состоит в замене традиционной схемы выборки номера m на выборку этого номера из специально построенного распределения. Построение такого распределения называется приведением распределения к Alias-формату. В Alias-формате все n вероятностей распределения одинаковы и равны $p_m + \tilde{p}_{m^*} = 1/n$, где каждый элемент распределения с номером m состоит из двух частей: остатка исходной части p_m и донорского дополнения $\tilde{p}_{m^*}$. Примеры исходного и Alias-формата дискретного распределения вероятностей из четырех элементов приведены на рис. 1 (см. также цветную вкладку). Если в ходе приведения к Alias-формату вероятность донорского элемента станет меньше $1/n$, то для него будет определен свой донор (см. элемент 3 на рис. 1, *б*). Донор допускается только один. Донорские дополнения на рис. 1, *б* размещены в верхних частях столбцов гистограммы и имеют цвета донорских элементов (см. рис. 1, *а*). Внизу показаны остатки исходных частей элементов. Числами указаны соответствующие значения вероятностей. Для выборки значения случайной величины в Alias-методе нужно хранить для каждого элемента распределения значение $p_m n$ и номер m^* донорского элемента.

Алгоритм приведения распределения к Alias-формату

Рассмотрим теперь алгоритм приведения распределения к Alias-формату. Данный алгоритм был предложен и реализован автором в 2000 году. В 2011 году вышла книга С. М. Ермакова [3] с описанием метода Уолкера [9], однако алгоритмов в ней не содержится.

Рассмотрим распределение вероятностей $\sum\limits_{m=1}^{n} p_m = 1$ и упорядочим его по возрастанию значений p_m. Выравнивать вероятности будем по значению $c = 1/n$. Определяем разность $c - p_{m_1}$, где m_1 — номер первого элемента с минимальной вероятностью. Если эта разность положительна (она может быть равной нулю!), то, начиная со следующего элемента, ищем донора, обеспечивающего донорское дополнение до значения c. При этом вероятность донора уменьшается на $c - p_{m_1}$. Запомнив номер донора и исключив элемент с номером m_1 из рассмотрения, повторяем данный алгоритм для оставшихся $n - 1$ элементов.

Проиллюстрируем данный алгоритм на примере.

Пусть есть дискретное распределение вероятности из четырех элементов ($n = 4$), заданное в виде гистограммы (см. рис. 1, *а* и табл. 1). После упорядочения данных таблицы по возрастанию вероятности случайной величины получим табл. 2. В результате последовательного поиска донорского дополнения для каждого элемента, имеющего $p_m < 0{,}25$, получим табл. 3. На последнем шаге вернем таблицу к исходной последовательности элементов и рассчитаем для каждого элемента значение $p_m n$ и номер донора m^* (табл. 4).

Таблица 1

Заданное распределение случайной величины

Исходный номер элемента m	Вероятность p_m
1	0,18
2	0,48
3	0,31
4	0,03

Таблица 2

Упорядоченная по возрастанию вероятности таблица распределения случайной величины

Исходный номер элемента m	Вероятность p_m
4	0,03
1	0,18
3	0,31
2	0,48

Таблица 3

Упорядоченная по возрастанию вероятности таблица с донорскими дополнениями

Исходный номер элемента m	Номер донорского элемента m^*	Вероятность донорского дополнения	Вероятность остатка исходной части
4	3	0,22	0,03
1	3	0,07	0,18
3	2	0,23	0,02
2	0	0	0,25

Таблица 4

Данные в Alias-формате

Номер элемента m	Номер донора m^*	$p_m n$
1	3	0,72
2	0	1
3	2	0,08
4	3	0,12

Таким образом, данные приведены к Alias-формату и рассчитаны массивы значений $p_m n$ и m^*.

Алгоритмы моделирования случайной величины в Alias-формате

Дискретное распределение. Рассмотрим сначала базовый алгоритм выборки номера m для дискретного распределения, уже приведенного к Alias-формату. Он состоит в следующем:

1. Разыгрываем две равномерно распределенные на отрезке [0,1] случайные величины ζ и η.
2. Полагаем $m = [n\zeta] + 1$.
3. Если $\eta \le p_m n$, то выбираем номер m, иначе берем номер донора $m = m^*$.

Из алгоритма Alias-метода видно, что по скорости выполнения он сравним с алгоритмом выборки из дискретного распределения с равными вероятностями. Его эффективность по сравнению с традиционными методами повышается с увеличением размера выборки n, поскольку для Alias-метода время выборки не зависит от n, а, как отмечалось выше, в наиболее быстрых традиционных методах оно растет как $\log_2 n$.

Необходимо отметить перспективы использования Alias-метода на векторных ЭВМ, так как он теоретически позволяет всем процессам одновременно выполнять один и тот же код.

Кусочно-линейное распределение. По мнению американских авторов [8, 9], распространение Alias-метода на кусочно-линейные распределения сдерживалось до тех пор, пока не был обнаружен оригинальный метод моделирования случайных величин, имеющих кусочно-линейную плотность распределения. Этот метод был, по-видимому, впервые предложен в 1981 году Андросенко и Поповым [10]. Американским авторам он стал известен в 1989 году [8, 9].

Рассмотрим кратко этот метод.

Ненормированную кусочно-линейную плотность распределения $f(x)$ случайной величины x можно задать графически (рис. 2) или таблицей, содержащей значения x_0, x_1, x_1, x_2, $x_2, \ldots, x_{n-1}$, x_{n-1}, x_n и соответствующие значения f_0, f_{1l}, f_{1r}, f_{2l}, $f_{2r}, \ldots, f_{(n-1)l}$, $f_{(n-1)r}$, f_n, где f_{ml}, f_{mr} ($m = \overline{1, n-1}$) — соответственно левое и правое значения кусочно-линейной плотности распределения в точке разрыва x_m.

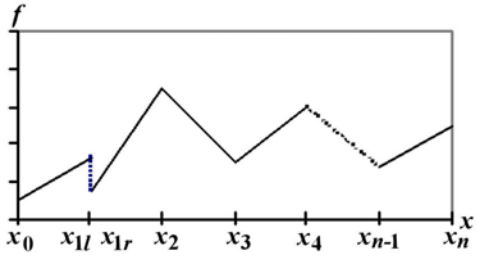

Рис. 2. Случайная величина, заданная ненормированной кусочно-линейной плотностью распределения

После вычисления и нормировки вероятностей $p_m = \frac{1}{2}\left(f_{(m-1),r} + f_{ml}\right)(x_m - x_{m-1})$ получим таблицу, аналогичную табл. 1, и, применив базовый Alias-алгоритм выборки, найдем номер m и соответствующий ему отрезок $[x_{m-1}, x_m]$, который содержит смоделированное значение ξ случайной величины.

Для вычисления случайной величины ξ Андросенко и Попов предложили следующую формулу:

$$\xi = \begin{cases} \zeta, & \eta \le f(\zeta); \\ x_{m-1} + x_m - \zeta, & \eta > f(\zeta). \end{cases} \tag{1}$$

Здесь ζ и η — независимые случайные величины, причем случайная величина ζ равномерно распределена на отрезке $\left[x_{m-1}, \dfrac{x_{m-1}+x_m}{2}\right]$, а случайная величина η равномерно распределена на отрезке $[0, f(x_{m-1}) + f(x_m)]$; $f(\zeta) = f(x_{m-1}) + \dfrac{f(x_m) - f(x_{m-1})}{x_m - x_{m-1}}(\zeta - x_{m-1})$.

В работах [8, 9] приведена другая формула для вычисления случайной величины ξ:

$$\xi = \begin{cases} (1-\zeta)\,x_{m-1} + \zeta x_m, & \text{если} \quad (f(x_{m-1}) + f(x_m))\,\eta \le (1-\zeta)\,f(x_{m-1}) + \zeta f(x_m); \\ \zeta x_{m-1} + (1-\zeta)\,x_m, & \text{если} \quad (f(x_{m-1}) + f(x_m))\,\eta > (1-\zeta)\,f(x_{m-1}) + \zeta f(x_m), \end{cases} \tag{2}$$

где независимые случайные величины ζ и η равномерно распределены на отрезке $[0, 1]$. В данной записи случайной величины ξ присутствуют элементы стохастической интерполяции аргумента $(1-\zeta)\,x_{m-1} + \zeta x_m$ и функции $(1-\zeta)\,f(x_{m-1}) + \zeta f(x_m)$, что удобно для обобщения на многомерный случай.

В работах [8, 9] авторами заявлено, что вычисление по формулам (1), (2) требует меньших затрат машинного времени, чем по существовавшему ранее методу, основанному на решении квадратного уравнения.

Кусочно-постоянное распределение. Рассмотрим теперь использование Alias-метода для кусочно-постоянного распределения, заданного либо гистограммой, либо в виде ненормированной доли частиц в каждом интервале (рис. 3).

Здесь $x_0, x_1, \ldots, x_n$ — значения аргумента; $h_1, h_2, \ldots, h_n$ — значения ненормированной кусочно-постоянной плотности распределения, представленной в виде гистограммы; $p_1, p_2, \ldots, p_n$ — значения ненормированной вероятности того, что фазовая координата частицы находится в определенном интервале.

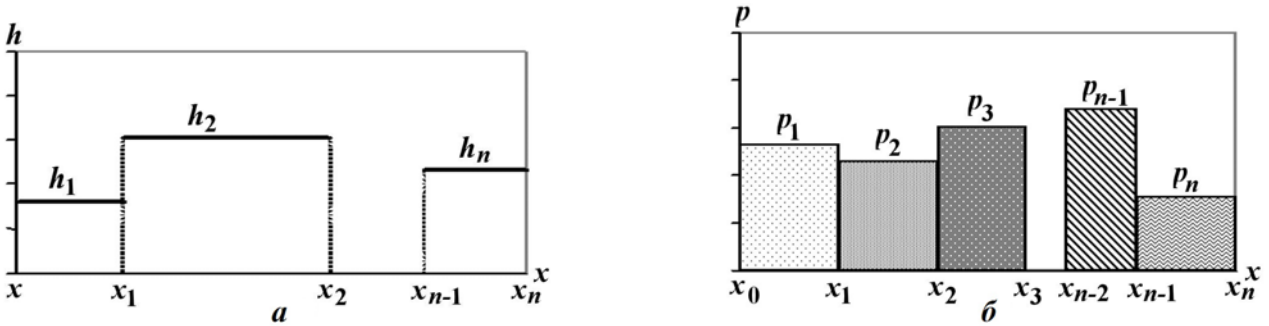

Рис. 3. Случайная величина, заданная ненормированной кусочно-постоянной плотностью распределения в виде гистограммы ($а$) или долей частиц в определенных интервалах ($б$)

После вычисления вероятностей $p_m = h_m (x_m - x_{m-1})$ они нормируются, чтобы была получена таблица, аналогичная табл. 1. Применив базовый Alias-алгоритм выборки, находим номер m и соответствующий ему отрезок $[x_{m-1}, x_m]$, которому будет принадлежать смоделированное значение ξ случайной величины, вычисляемое по формуле $\xi = x_{m-1} + \eta(x_m - x_{m-1})$, где η — равномерно распределенная на отрезке [0, 1] случайная величина.

Обобщение Alias-метода для многомерных распределений

Иногда при моделировании энергетического, углового и пространственного распределений источника частиц требуется задать совместное двумерное или трехмерное распределение из перечисленных выше распределений. Поэтому рассмотрим, как с помощью Alias-метода можно моделировать многомерное распределение $P(x_1, x_2, \ldots, x_n)$.

Сначала многомерное распределение переводится в одномерное путем следующего преобразования номеров элементов:

$$k = i_1 + (i_2 - 1) l_1 + (i_3 - 1) l_1 l_2 + \ldots + (i_n - 1) l_1 l_2 \ldots l_{n-1}.$$

Здесь $i_m = 1, \ldots, l_m, (m = \overline{1, n})$, l_m — длина таблицы распределения по направлению измерения m.

Затем номер k разыгрывается с помощью базового Alias-метода и выполняется обратное преобразование номеров:

$$\begin{aligned}
&i_n = \left[\frac{k - 0{,}5}{l_1 l_2 \ldots l_{n-1}}\right] + 1;\\
&i_{n-1} = \left[\frac{k - 0{,}5}{l_1 l_2 \ldots l_{n-2}}\right] - l_{n-1}(i_n - 1) + 1;\\
&\ldots\ldots;\\
&i_2 = \left[\frac{k - 0{,}5}{l_1}\right] - l_2 l_3 \ldots l_{n-1}(i_n - 1) - l_2 l_3 \ldots l_{n-2}(i_{n-1} - 1) - \ldots - l_2 (i_3 - 1) + 1;\\
&i_1 = k - l_1 l_2 \ldots l_{n-1}(i_n - 1) - l_1 l_2 \ldots l_{n-2}(i_{n-1} - 1) - \ldots - l_1 l_2 (i_3 - 1) - l_1 (i_2 - 1).
\end{aligned}$$

После определения многомерного номера $(i_1, i_2, \ldots, i_n)$ моделирование случайного вектора $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)$, имеющего закон распределения $P(x_1, x_2, \ldots, x_n)$, осуществляется следующим образом:

– для дискретного распределения $\boldsymbol{\xi} = (x_{i_1}, x_{i_2}, \ldots, x_{i_n})$;

– для кусочно-линейного распределения

$$\xi_1 = \begin{cases} (1 - \zeta_1) x_{i_1 - 1} + \zeta_1 x_{i_1}; & \text{если} \quad (f(x_{i_1 - 1}) + f(x_{i_1}))\,\eta_1 \le (1 - \zeta_1) f(x_{i_1 - 1}) + \zeta_1 f(x_{i_1}), \\ \zeta_1 x_{i_1 - 1} + (1 - \zeta_1) x_{i_1}, & \text{если} \quad (f(x_{i_1 - 1}) + f(x_{i_1}))\,\eta_1 > (1 - \zeta_1) f(x_{i_1 - 1}) + \zeta_1 f(x_{i_1}); \end{cases}$$

$\ldots\ldots\ldots\ldots\ldots\ldots;$

$$\xi_n = \begin{cases} (1 - \zeta_n) x_{i_n - 1} + \zeta_n x_{i_n}; & \text{если} \quad (f(x_{i_n - 1}) + f(x_{i_n}))\,\eta_n \le (1 - \zeta_n) f(x_{i_n - 1}) + \zeta_n f(x_{i_n}), \\ \zeta_n x_{i_n - 1} + (1 - \zeta_n) x_{i_n}, & \text{если} \quad (f(x_{i_n - 1}) + f(x_{i_n}))\,\eta_n > (1 - \zeta_n) f(x_{i_n - 1}) + \zeta_n f(x_{i_n}), \end{cases}$$

где $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \ldots, \zeta_n)$, $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_n)$ — случайные векторы, равномерно распределенные на единичном n-мерном кубе. При этом предполагается, что многомерное кусочно-линейное распределение таково, что через любые точки $f(x_{i_1 - 1})$, $f(x_{i_2 - 1}), \ldots, f(x_{i_n - 1})$, $f(x_{i_1})$, $f(x_{i_2})$, $\ldots, f(x_{i_n})$ можно провести $(n-1)$-мерную плоскость;

– для кусочно-постоянного распределения

$$\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_n), \quad \xi_j = x_{i_j - 1} + \eta_j \left(x_{i_j} - x_{i_j - 1}\right), \quad j = 1, 2, \ldots, n,$$

где η_j — равномерно распределенная на отрезке $[0, 1]$ случайная величина.

Результаты тестовых расчетов выборки случайной величины

Пусть имеется плотность нормального распределения $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ с параметрами $x = [0, 10]$, $\mu = 3$, $\sigma = 2$, заданная в виде таблиц кусочно-линейной функции. Длина таблиц варьировалась числом интервалов по x: 100, 500, 1 000, 10 000, 100 000. Сравнивались классический метод перебора интервалов и Alias-метод. По каждому методу в цикле моделировалась выборка из 10 млн случайных величин. В результате (табл. 5) было установлено, что применение Alias-метода заведомо выгоднее по времени выполнения, чем традиционные методы, при длине таблицы распределения больше 1 000. Данный метод также успешно используется для моделирования экспериментальных источников, содержащих 1 млн частиц и более.

Таблица 5

Времена выборки (в мс) 10 млн случайных величин для классического и Alias-методов

Метод	Длина таблицы				
	100	500	1 000	10 000	100 000
Классический метод	77	148	244	1 954	19 058
Alias-метод	64	64	65	65	74
Ускорение	1,2	2,3	3,7	30	257

Заключение

В статье описано использование Alias-метода для моделирования таблично заданных случайных величин. Представлен оригинальный вариант модификации исходных данных и приведения их к Alias-формату. Рассмотрены алгоритмы для моделирования многомерных таблично заданных распределений: дискретного, кусочно-линейного и кусочно-постоянного.

Alias-метод применялся автором для моделирования таблиц спектрально-углового и пространственного распределений точек рождения тормозных квантов в модели учета тормозного излучения TTBIAS [11]. В методике С-007 [2] с использованием Alias-метода моделируется угол упругого рассеяния электронов и позитронов из дифференциальных сечений, заданных таблицами большой длины.

Список литературы

1. *Ермаков С. М., Михайлов Г. А.* Статистическое моделирование. М.: Наука, 1982.
2. *Житник А. К., Донской Е. Н., Огнев С. П., Горбунов А. В., Залялов А. Н., Иванов Н. В., Малькин А. Г., Рослов В. И., Семенова Т. В., Субботин А. Н.* Методика С-007 решения методом Монте-Карло связанных линейных уравнений переноса нейтронов, гамма-квантов, электронов и позитронов // Вопросы атомной науки и техники. Сер.: Математическое моделирование физических процессов. 2011. Вып. 1. С. 17—24.
3. *Ермаков С. М.* Метод Монте-Карло в вычислительной математике: Вводный курс. С.-Пб.: Невский диалект; М.: БИНОМ. Лаборатория знаний, 2011.
4. *Walker A. J.* New fast method for generating discrete random numbers with arbitrary frequency distributions // Electronic Letters. 1974. Vol. 10. P. 127—128.
5. *Walker A. J.* An efficient method for generating discrete random variables with general distributions // ACM Trans. Math. Software. 1977. Vol. 3, No 3. P. 253—256.
6. *Brown F. B., Martin W. R., Calaham D. A.* A discrete sampling method for vectorized Monte Carlo calculations // Trans. Am. Nucl. Soc. 1981. Vol. 38. P. 354—355.

7. *Wilderman S. J.* Vectorized algorithms for Monte Carlo simulation of kilovolt electron and proton transport: Ph. D. thesis. The University of Michigan, 1990.
8. *Rathkopf J. F., Smidt R. K., Edwards A. L.* The Alias Method: A Fast, Efficient Monte Carlo Sampling Technique. Report UCRL-JC-105535. Lawrence Livermore National Laboratory, 1990.
9. *Edwards A. L., Rathkopf J. F., Smidt R. K.* Extending the Alias Monte Carlo Sampling Method to General Distributions. Report UCRL-JC-104791. Lawrence Livermore National Laboratory, 1991.
10. *Андросенко П. А., Попов Г. В.* Эффективный метод моделирования распределения Клейна—Нишины—Тамма // Журнал вычисл. мат. и мат. физ. 1981. Т. 21, № 4. С. 1056.
11. *Donskoy E. N., Zalyalov A. N.* Bremsstrahlung account in photon transport // Parallel Computational Fluid Dynamics. Advanced Numerical Methods, Software and Applications / Ed. by B. Chetverushkin, A. Ecer, J. Periaux, N. Satofuka, P. Fox. Elsevier Science, 2004. P. 349—355.

Статья поступила в редакцию 13.11.17.

THE ALIAS METHOD FOR SIMULATING DISTRIBUTION TABLES OF RANDOM VARIABLES / A. N. Zalyalov (FSUE "RFNC-VNIIEF", Sarov, N. Novgorod region).

The paper considers the Alias method of simulating distribution tables of random variables. The original algorithm of preprocessing tables is described. Formulas are given for the simulation of multidimensional tabular distributions: discrete, piecewise linear, and piecewise constant (specified either by a histogram, or as a fraction of particles within each range of values) distributions.

Keywords: random variable, Alias method, distribution tables, the Monte Carlo method.