Лекция 2 Моделирование случайных чисел.

Случайные числа и принцип их моделирования

При исследовании сложных систем с помощью компьютерного моделирования широко используются случайные события, случайные величины с заданными законами распределения вероятностей и различные случайные процессы. Основной способ компьютерной имитации случайных закономерностей любой природы сводится к моделированию последовательности случайных чисел с равномерным законом распределения на отрезке [0, 1] и к ее последующему функциональному преобразованию.

Выбор в качестве исходной или базовой последовательности случайных чисел, являющихся реализациями z_i случайной величины ξ равномерно распределенной на отрезке [0, 1], вызван следующими двумя факторами:

– проблемой моделирования случайных чисел с равномерным распределением ученые заинтересовались буквально с первых дней создания компьютеров и разработали достаточно эффективные алгоритмы их имитации;

– равномерное распределение является самым простым из случайных закономерностей и легко поддается математическим преобразованиям.

Случайная величина ξ подчинена равномерному закону распределения на отрезке [a, b], если ее функция плотности распределения представляет собой непрерывную функцию, постоянную на отрезке [a, b] и равную нулю вне этого отрезка

$$f(z)=\begin{cases}\dfrac{1}{b-a} & \text{при} \quad z\in[a,b],\\ 0 & \text{при} \quad z\notin[a,b].\end{cases}$$

Функция распределения случайной величины ξ имеет вид

$$F(z)=\begin{cases}0 & \text{при } z<a,\\ \dfrac{z-a}{b-a} & \text{при} \quad a\le z\le b,\\ 1 & \text{при} \quad z\ge b.\end{cases}$$

Математическое ожидание, дисперсия и среднеквадратическое отклонение соответственно, равны:

$$m_z=\int_a^b z\cdot f(z)dz=\frac{a+b}{2},$$

$$\sigma_z^2=\int_a^b (z-m)^2\cdot f(z)dz=\frac{(b-a)^2}{12},$$

$$\sigma_z=\frac{b-a}{2\sqrt{3}}.$$

В частном случае, когда величина ξ равномерно распределена на отрезке $[0,1]$, имеем:

$$f(z)=\begin{cases}1 & \text{при} \quad z\in[0,1],\\ 0 & \text{при} \quad z\notin[0,1].\end{cases}$$

$$F(z)=\begin{cases}0 & \text{при } z<0,\\ z & \text{при } 0\le z\le 1,\\ 1 & \text{при } z\ge 1.\end{cases}$$

$$m_z=\frac{1}{2},\ \sigma_z^2=\frac{1}{12},\ \sigma_z=\frac{1}{2\sqrt{3}}.$$

Учитывая важную роль для моделирования различных случайных закономерностей величины с равномерным распределением на отрезке [0,1], рассмотрим несколько методов ее компьютерной имитации. Все эти методы основаны на рекуррентных соотношениях и вырабатывают псевдослучайные числа.

Определение. Псевдослучайными числами называют реализации z случайной величины ξ, полученные с помощью математических выражений, в частности, рекуррентных соотношений.

Рассмотрим несколько наиболее известных методов моделирования псевдослучайных чисел. При этом в дальнейшем вместо термина "псевдослучайные числа" будем использовать термин "случайные числа", так как предполагаем, что большинство предлагаемых ниже прикладных алгоритмов моделирует последовательности с достаточно хорошими статистическими свойствами.

Метод усечения

Этот метод основывается на том, что разряды очередного случайного числа получаются путем отбрасывания или "усечения" части разрядов у результата нелинейного преобразования над одним или несколькими предыдущими числами.

Первый алгоритм для моделирования равномерно распределенных случайных чисел, использующий идею усечения, был предложен фон Нейманом и Метрополисом в 1946 году. Этот алгоритм получил название "срединных квадратов" и оперирует с $2k$ – значными числами.

Вычислительная процедура алгоритма состоит из следующих шагов:

Шаг 1. Положить $z_i=0,a_1,a_2,...,a_{2k}$.

Шаг 2. Возвести z_i в квадрат $z_i^2=0,b_1,b_2,...,b_k,b_{k+1},...,b_{3k},b_{3k+1},...,b_{4k}$.

Шаг 3. Отобрать средние $2k$ цифр полученного квадрата и считать их разрядами следующей реализации последовательности $z_{i+1}=0,b_{k+1},b_{k+2},...,b_{3k}$.

К сожалению, алгоритм срединных квадратов во многих случаях не дает статистически удовлетворительных результатов. Вырабатываемая последовательность имеет больше чем нужно чисел с малыми значениями, т.е. $m_z < 0{,}5$; часто наблюдается вырождение последовательности к нулю. Так, в серии экспериментов, проведенных американцем Дж. Форсайтом в начале пятидесятых годов, были получены следующие результаты. Из 16 начальных значений 12 привели к последовательностям, оканчивающимся циклом: 0,6100; 0,2100; 0,4100; 0,8100; 0,6100, а две – к вырождению последовательности.

Иногда в выработанной алгоритмом срединных квадратов последовательности вообще отсутствует случайность.

В настоящее время из-за этих недостатков алгоритм срединных квадратов не получил большого распространения и он для нас представляет лишь исторический интерес. Его прежняя популярность объяснялась простотой и оригинальностью.

Последователями Джон фон Неймана были предложены различные модификации этого алгоритма.

Но, тем не менее, в настоящее время почти все стандартные библиотечные программы моделирования последовательностей равномерно распределенных случайных чисел реализуют методы вычетов и суммирования.

Метод вычетов (конгруэнтный метод)

Этот метод был предложен Д.Лемером в 1948 году и в общем случае основывается на линейной формуле вида

$$z^*_{n+1} = az^*_n + c(\text{mod } m), \qquad (1)$$

где z^*_0, a, c и m – неотрицательные целые числа.

Запись (1) означает, что z^*_{n+1} равно остатку, полученному при делении выражения $az^*_n + c$ на m или, другими словами, z^*_{n+1} – это наименьший положительный вычет $az^*_n + c$ по модулю m. Формула (1) при любых значениях ее параметров может дать лишь конечное множество целых случайных чисел, после которого последовательность начинает повторяться. Это вытекает из очевидного ограничения $P < m$, где P – период последовательности.

Частным случаем выражения (1) является формула

$$z^*_{n+1} = az^*_n (\text{mod } m), \qquad (2)$$

полученная при $c = 0$. Эта формула моделирует случайные последовательности несколько быстрее, но с относительно меньшим периодом.

Параметры a, c, m и начальное значение z^*_0 надо выбирать так, чтобы обеспечить максимальный период последовательности, максимальную скорость ее генерирования и минимальную корреляцию между моделируемыми числами.

В настоящее время установлено, что для численной реализации метода вычетов удобной является формула (2), в которой $m = 2^b$, где b – число двоичных цифр в машинном слове. При этом максимальный период последовательности, равный $P = m/4$, можно получить, если выполняются следующие условия:

а) z^*_0 – любое целое положительное нечетное число;

б) $a = 8t \pm 3$, где t – любое положительное число.

При моделировании последовательности случайных чисел с помощью формулы (1) можно достичь максимальной длины периода, равной m. Длина периода последовательности, полученной по формуле (1), равна m тогда и только тогда, когда:

а) c и m – взаимно простые числа;

б) $a - 1$ кратно r для любого простого r, являющегося делителем m;

в) $a - 1$ кратно 4, если m кратно 4.

Так как период имеет длину m, каждое число от 0 до $m - 1$ будет встречаться в моделируемой последовательности ровно один раз. Поэтому в данном случае выбор z^*_0 не влияет на длину периода.

В заключении следует отметить, что по формулам (1) и (2) моделируются последовательности случайных чисел с равномерным распределением на отрезке $[0, m]$, а чтобы получить числа на отрезке $[0, 1]$ необходимо эти формулы дополнить выражением

$$z_{n+1} = \frac{z^*_{n+1}}{m}. \qquad (3)$$

Метод суммирования

Метод суммирования был предложен Ван Вейнгарденом и использует общую линейную формулу вида

$$z_{k+j} = \sum_{l=0}^{k-1} a_l z_{j+l} + c(\mathrm{mod}\, m),\ j = 1,2,\ldots, \qquad (4)$$

Этот метод позволяет получить более длительные периоды в последовательности равномерно распределенных случайных чисел по сравнению с методом вычетов, так как здесь совпадение отдельных членов последовательности не приводит к совпадению всех последующих ее элементов. Последнее возможно лишь в случае повторения всего множества участвующих в формуле (4) чисел. Кроме того, числа последовательности, полученные по методу суммирования, имеют малую корреляцию.

Наиболее простую зависимость (4) можно получить, приняв

$$a_0 = a_1 = 1,\ a_2 = a_3 = \ldots = a_{k-1} = c = 0 \ .$$

Тогда

$$z_{j+1} = z_j + z_{j-1} (\mathrm{mod}\ \ m). \qquad (5)$$

Это выражение получило название формулы Фибоначчи и имело достаточно широкое применение в начале пятидесятых годов. Однако числа, генерируемые формулой Фибоначчи, были недостаточно случайными. Исключение составил лишь алгоритм Дэвиса, которому удачным выбором первых двух начальных значений z_0 и z_1 удалось получить хорошую последовательность равномерно распределенных случайных чисел.

Алгоритм Дэвиса использует формулу

$$z_{j+1} = z_j + z_{j-1} (\mathrm{mod}\ 4), \qquad (6)$$

где $z_0 = \pi, z_1 = 5^{17} * 2^{-42} = 0{,}542101887$.

В настоящее время наибольшее распространение среди алгоритмов метода суммирования получили аддитивные алгоритмы, использующие формулу

$$z_{j+1} = z_j + z_{j-k} (\mathrm{mod}\ m),$$

где k – достаточно большое число (k >16).

При соответствующем выборе $z_0, z_1, z_2, z_3, \ldots, z_k$ эта формула является хорошим источником случайных чисел.

Экспериментальное определение периода и длины отрезка апериодичности случайной последовательности

Как следует из вышеизложенного, характерной чертой всех методов моделирования случайных чисел является то, что при реализации их на компьютере они порождают периодические последовательности. Действительно, в коде любого компьютера можно записать лишь конечное число всех чисел, заключенных на отрезке [0, 1], поэтому рано или

поздно какое-нибудь значение z_L совпадает с одним из предыдущих значений z_l. Тогда справедливо следующее равенство:

$$z_{L+i} = z_{l+i}, \qquad i = 1, 2, \ldots \tag{7}$$

Контрольные вопросы:

1. Какие числа называются случайными и почему их часто называют псевдослучайными?
2. Сформулируйте правила выбора параметров усеченного алгоритма метода вычетов, обеспечивающие максимальную длину последовательности.