ЛАБОРАТОРНАЯ РАБОТА 1
ПОСТРОЕНИЕ И ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК ДАТЧИКОВ БАЗОВЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Базовой случайной величиной (БСВ) в статистическом моделировании называют непрерывную случайную величину z, равномерно распределенную на интервале (0,1). Ее плотность распределения вероятностей (**п.р.в.**) имеет вид:

$$f(t) = 1,\ 0 < t < 1,$$

Математическое ожидание (м.о.) и дисперсия БСВ составляют

$$M(z) = \frac{1}{2},$$

$$D(z) = \frac{1}{12},$$

соответственно.

БСВ моделируется на ЭВМ с помощью *датчиков* БСВ. Датчик БСВ – это устройство или программа, выдающая по запросу одно или несколько независимых значений z_1 , ..., z_n БСВ.

Датчики БСВ могут быть трех типов: табличные, физические и программные.

Программный датчик БСВ обычно вычисляет значения z_1, z_2,..., по какой-либо рекуррентной формуле типа

$$z_i = f\,(z_n),$$

при заданном стартовом значении z_0.

Заданное значение z_0 полностью определяет всю последовательность реализаций z_1, z_2,..., поэтому z часто называют *псевдослучайной* величиной. Но ее статистические свойства идентичны свойствам "чисто случайной" последовательности, что и обеспечивает успех статистического моделирования.

Имея датчик БСВ z, можно промоделировать любые случайные факторы: непрерывные или дискретные случайные величины (как простые, так и многомерные), случайные события, случайные процессы и поля и т.д. Для этого достаточно соответствующим образом преобразовать последовательность z_1, z_2, ... Поэтому БСВ z и называют базовой.

Теоретически в качестве базовой можно было бы взять почти любую случайную величину (**с.в**.). Использование **с.в**. z с распределением обусловлено технологическими соображениями: простотой и экономичностью датчика, простотой преобразования z в другие случайные факторы, относительной простотой тестирования датчика.

Метод середины квадрата

Метод середины квадрата предложен для получения псевдослучайных чисел Д. фон Нейманом в 1946 г. Вот один из вариантов этого метода.

1. Возьмем произвольное 4-значное число.

2. Возведем полученное число в квадрат и, если необходимо, добавим к результату слева нули до 8-значного числа.

3. Возьмем четыре цифры из середины 8-значного в качестве нового случайного 4-значного числа.

4. Если нужны еще случайные числа, то перейдем к 2.

Например, если взять в качестве начального числа 1994, то из него получается следующая последовательность псевдослучайных чисел: 9760 2576 6357 4114 9249 5440 5936 2360 5696 4444 7491 1150 3225 4006 0480 2304 3084 5110 1121 2566 ...
Сам по себе метод середины квадрата не получил широкого распространения, так как выдает "больше чем нужно малых значений". Но открытый в нем принцип используется во многих, если не во всех, более поздних датчиках БСВ. Этот принцип состоит в вырезании нескольких цифр из результата какой-либо операции над числами.

Мультипликативный конгруэнтный метод

Так называемый мультипликативный конгруэнтный датчик БСВ задается двумя параметрами: модулем m и множителем k. Обычно это достаточно большие целые числа.
При заданных m, k числа z_1, z_2, ..., вычисляются по рекуррентной формуле:

$$A_i = (kA_{i-1}) \bmod m,\ i = 1, 2, ...,$$

$$z_i = A_i / m,$$

где m – модуль, k – множитель, A_0 – начальное значение, mod – операция вычисления остатка от деления kA_{i-1} на m.

Таким образом, A_1 определяется как остаток от деления kA_0 на m; A_2 - как остаток от деления kA_1 на m и т.д. Поскольку все числа A_i – это остатки от деления на m, то $0 \le A_i < m$. Разделив последнее неравенство на m, видим, что $0 \le A_i / m < 1$, т. е. $0 \le z_i < 1$.

Из неравенства $0 \le A_i < m$ вытекает также, что датчик (2.5) дает периодическую последовательность A_i. Действительно, число всех возможных остатков от 0 до m - 1 равно m и, рано или поздно, на каком-то шаге i обязательно появится значение A_i, уже встречавшееся ранее. С этого момента последовательность A_i “зациклится".
Длина периода T будет не больше m - 1. Например, если встретится остаток $A_i = 0$, то далее, согласно (2.5), будет $A_{i+1} = 0$, $A_{i+2} = 0$, ... , т.е. длина периода T = 1. Ненулевых же остатков в интервале $0 \le A_i < m$ всего m - 1, и, если все они войдут в период, будет $T = m - 1$. Это имеет место, например, при m = 13, k = 7; в этом случае ряд A_i выглядит так:

$$\underbrace{1, 7, 10, 5, 9, 11, 12, 6, 3, 8, 4, 2,}_{T = m - 1 = 12} \quad 1, 7, ...$$

Поскольку в качестве случайной можно использовать лишь подпоследовательность A_i внутри одного периода, то параметры датчика выбирают так, чтобы длина периода T была максимальной. С учетом ограничения $T \leq m - 1$ модуль m берут максимально возможным. Чтобы упростить вычисление остатков по (2.5), для двоичных ЭВМ часто берут $m = 2^n$. Рекомендуется также брать достаточно большой множитель k, причем взаимно простой с m.

Существуют подробные рекомендации по выбору параметров m, k и начального значения A_0. Заметим, однако, что в настоящее время не известны правила, которые гарантировали бы высокое качество датчика без его специального статистического тестирования.

Датчик называют мультипликативно-конгруэнтным потому, что он использует две основные операции – умножение (англ. multiplication) и вычисление остатка (в теории чисел – получение конгруэнтного числа). Можно было бы поэтому перевести его название и как "множительно-остатковый датчик".

Обратим внимание также и на то, что операция вычисления остатка воплощает здесь упоминавшийся в п. 2.2 неймановский принцип вытаскивания цифр. Это становится очевидным, если записывать числа в системе счисления с основанием m. Тогда операция $X \bmod m$ означает выбор последней цифры из числа X. Для $m = 2^n$ операция $X \bmod m$.

Тестирование равномерности

Обозначим равномерное распределение вероятностей на интервале (0,1) через $R[0,1]$. Тогда утверждение, что БСВ z имеет распределение $R[0,1]$, можно кратко записать в виде $z \sim R[0,1]$.

С помощью статистических тестов проверяют два свойства датчика, делающих его точной моделью идеальной БСВ, – это *равномерность* распределения чисел z_i, выдаваемых датчиком на интервале (0,1), и их статистическая *независимость*. При этом числа z_i рассматривают как реализации некоторой **с.в.**, т.е. как статистическую выборку.

Достаточно простым методом проверки равномерности распределения является частотный тест. Он основан на законе больших чисел и выполняется по следующему алгоритму.

1. Разобьем интервал (0,1) на K равных отрезков (например, $K = 10$).
2. Сгенерируем n чисел $z_1,..., z_n$ с помощью тестируемого датчика БСВ (например, $n = 100$).
3. Подсчитаем, сколько чисел попало в каждый из k отрезков, т.е. найдем числа попаданий $n_1,...,n_k$.

4. Рассчитаем относительные частоты попаданий в отрезки: $\hat{p} = \frac{n_1}{n}, \dots, \hat{p} = \frac{n_k}{n}$

5. Построим гистограмму частот $\hat{p}_1, \dots, \hat{p}_k$ на K отрезках интервала (0,1).
6. Повторим действия (2) – (5) для большего значения n (например, для n =10 000).

7. Оценим по полученным гистограммам сходимость каждой частоты $\hat{p}_i$ к вероятности $p = 1/K$ того, что БСВ попадет в i-й отрезок. Согласно закону больших чисел должно быть

$$\hat{p}_i \xrightarrow{p} \frac{1}{K}, \qquad n \to \infty$$

Это значит, что высоты столбиков во второй гистограмме должны в целом быть ближе к уровню $1/K$, чем в первой.

Тестирование датчика на равномерность можно совместить с оцениванием **м.о**. и дисперсии **с.в.** Оценки $\hat{M}$ и $\hat{D}$ для **м.о.** и дисперсии рассчитываются соответственно по формулам:

$$\hat{M} = \frac{1}{n}\sum_{i=1}^{n} z_i \qquad \hat{D} = \frac{1}{n}\sum_{i=1}^{n} z_i^2 - (\hat{M})^2$$

С ростом n оценки $\hat{M}$ и $\hat{D}$ должны сходиться по вероятности к точным значениям $M(z) = 1/2$, $D(z) = 1/12 = 0.08333...$.

Тестирование независимости

Простейшую проверку статистической независимости реализаций $z_1, z_2, \dots$, можно осуществить, оценивая корреляцию между числами z_i и z_{i+s}, отстоящими друг от друга на шаг $s > 1$.

Для вывода формулы, по которой можно рассчитать коэффициент корреляции чисел z_i и z_{i+s}, рассмотрим две произвольные СВ x, y. Коэффициент корреляции определяется для них формулой:

$$R(x,y) = \frac{M(xy) - M(x)M(y)}{\sqrt{D(x)D(y)}}$$

С ростом n оценка R' должна приближаться к нулю, в противном случае датчик БСВ не отвечает требованию независимости.

Конечно, если R' сходится к нулю, то это еще не гарантирует наличие независимости, но все же один из тестов оказывается успешно выдержанным. При желании всегда можно продолжить испытания датчика другими методами.

ЗАДАНИЕ

Написать программы, реализующие рассмотренные методы построения датчиков случайных величин (разрядность чисел – не менее 8)

Выполнить статистическое исследование датчиков.

Сравнить результаты.