МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

МОСКОВСКИЙ ЭНЕРГЕТИЧЕСКИЙ ИНСТИТУТ
(ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

В.Н. Балашов, А.Г. Гольцов

МОДЕЛИРОВАНИЕ ГЕНЕРАТОРОВ СЛУЧАЙНЫХ ЧИСЕЛ С РАВНОМЕРНЫМ ЗАКОНОМ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТИ

Рекомендации к лабораторным занятиям
по курсу "Моделирование

для студентов, обучающихся по направлению
"Информатика и вычислительная техника"

Москва МЭИ 2019

УДК
621.398
П692
УДК:681.3

Утверждено учебным управлением МЭИ
Подготовлено на кафедре вычислительных машин, систем и сетей.

Балашов В.Н., Гольцов А.Г.
Моделирование генераторов случайных чисел
Рекомендации к лабораторным занятиям. Методическое пособие по курсу "Моделирование" / – М.: Изд-во МЭИ, 2019, 40с.
Представлены описание лабораторной работы, выполняемой на ПЭВМ типа IBM PC , по моделированию генераторов случайных чисел. Лабораторная работа включает в себя теоретический материал, рекомендации по разработке и отладке программ генераторов равномерно распределенных случайных чисел, а также методы и рекомендации по статистической оценке распределений полученных последовательностей случайных чисел
Предназначено студентам, обучающимся по направлению подготовки/специальности: 09.03.01 Информатика и вычислительная техника, по образовательной программе: «Вычислительные машины, комплексы, системы и сети», изучающим курс "Моделирование".
Уровень образования: бакалавриат.
Работа выполняется по индивидуальным заданиям.

Введение

В процессе имитационного моделирования вычислительных систем и сетей применяются генераторы случайных чисел с заданным законом распределения. В настоящее время применяются в основном программные генераторы случайных чисел, входящие в форме стандартной функции в трансляторы с языков высокого уровня или в программы моделирования, например, MATLAB. В лабораторной работе исследуются два метода получения равномерно распределенных случайных чисел. Строго говоря, получаемые программным путем последовательности чисел являются псевдослучайными, так как при сохранении в алгоритме начальных значений (параметров и "затравок") генерируется та же последовательность чисел.

МЕТОДЫ ПОЛУЧЕНИЯ РАВНОМЕРНО РАСПРЕДЕЛЕННЫХ СЛУЧАЙНЫХ ЧИСЕЛ

Генераторы равномерно распределенных случайных чисел широко применяются в имитационном моделировании и являются основой алгоритмов получения случайных чисел с заданным законом распределения. Рассмотрим два таких метода.

1. Метод середин квадратов

В этом методе строится последовательность целых псевдослучайных чисел, которым ставится в соответствие числа на интервале [0, 1). Генератор работает на 4k разрядной сетке (k – целое число), например, на 28-ми разрядной сетке из двоичных чисел.

На первом шаге алгоритма задается произвольное число "затравка" x_0 разрядности 2k. Это число можно задать с клавиатуры или взять случайное число, полученное при помощи встроенной в транслятор языка программирования функции (для языка Паскаль Random (I: word): word); На каждом шаге алгоритма целое число x_i возводится в квадрат (получается число разрядности 4k), из него выбираются средние 2k разрядов и принимаются за следующее псевдослучайное число x_{i+1}.

Эта последовательность чисел преобразуется в дробную часть нормированной десятичной дроби, что превращает ее в последовательность псевдослучайных чисел с равномерным законом распределения на интервале [0,1).

Рассмотрим алгоритм метода середин квадратов на примере для калькулятора.

Зададимся 8-разрядной десятичной сеткой, пусть начальное число x_0 = 4824. Будем генерировать случайные числа в формате 0.xxxx, то есть с четырьмя десятичными знаками после запятой.

Получим следующую последовательность чисел

x_0=4824; x_0^2=23270976; u_0=0.4824;
x_1=2709; x_1^2=07338681; u_1=0.2709;
x_2=3386; x_2^2=11464996; u_2=0.3386;
x_3=4649; x_2^2 = 21613201; u_3=0.4649; ...

Нельзя априори утверждать, что полученная последовательность чисел u_i обладает необходимыми свойствами. Необходимо проверить статистическую гипотезу о том, что последовательность u_0 является последовательностью случайных чисел с равномерным распределением.

Недостатком метода середин квадратов является возможность получить на очередном шаге алгоритма нулевого значения u_i. Тогда и все последующие числа в последовательности также будут нулями. (Например:

5600203 4 → 0020 → 0000040 0 → 0004 → 0000001 6 → 0000

Поэтому при программной реализации метода середин квадратов необходимо проводить "проверку на ноль". Если в последовательности $\{u_i\}$ появляется нулевой элемент, то алгоритм необходимо выполнять сначала, задав новую затравку.

Приведем пример фрагмента программы, реализующей метод середин квадратов. Каждый вызов функции rrsch дает новое псевдослучайное число:

```
var xi: longint; {глобальная переменная; сюда изначально
                  должна быть помещена затравка}
function rrsch:real;
begin
  xi:=((xi*xi) shr 7) and $3FFF;      {28 бит - квадрат, 14 бит - число}
  if xi=0 then begin
    Random( …);          {заносит в RandSeed очередное псевдослучайное
число - затравку}
    xi:=RandSeed and $3FFF;
  end;
  rrsch:=xi/$4000;                    {двоичное 01000000 00000000}
end;
```

При работе на 28-ми разрядной сетке для двоичных чисел период последовательности псевдослучайных чисел обычно не превышает 200 – 500, что недостаточно для проверки закона распределения псевдослучайных чисел (необходимо получить последовательность из 15 – 20 тыс. чисел). Поэтому в программу необходимо включить проверку повторяемости элементов последовательности чисел и при появлении ранее встречающегося числа (например, x_2 – следующего за очередной затравкой), необходимо прервать программу и продолжить ее с новой затравкой.

Метод середин квадратов лучше работает на 64-ти разрядной сетке, при этом период последовательности обычно превышает 20 тыс. чисел.

2. Метод мультипликативного датчика

В этом методе очередное случайное число x_{i+1} получается как остаток от деления двух больших целых чисел. Это случайное число определяется следующей формулой

$$x_{i+1} = (ax_i+b) \bmod M;$$

где a, b, M - заданные целые числа,. x_i – предыдущее целое случайное число.

Для получения "хорошего" со статистической точки зрения распределения рекомендуется выбирать параметры M, a и b исходя из следующих соображений:

M — большое целое число, желательно простое;

a — целое число порядка корня квадратного из M, желательно простое, M не должно делиться на a;

a и M выбираются такими, чтобы их произведение не выходило за границы разрядной сетки для целых переменных;

b — целое число того же порядка, что и a.

В ходе вычислительных экспериментов замечено, что хорошие в статистическом смысле реализации мультипликативного датчика получаются, если целое число a выбрано так, что бы при делении на 8 в остатке оставалось бы 3 или 5.

Слагаемое b используется, чтобы устранить потенциальную опасность получения непрерывной последовательности нулей, начиная с определенного шага. Если M и a — простые числа, то такой опасности не существует.

Случайное число, принадлежащее интервалу [0,1), вычисляется по формуле:

$$u_i = x_i/M;$$

Понятно, что для начала работы генератора случайных чисел необходимо задать затравку x_0.

В качестве примера приведем фрагмент программы, реализующей мультипликативный датчик.

```
var xi: longint; {глобальная переменная; сюда изначально
                  должна быть помещена затравка}
function rrsch: real;
  {M=1000; a=37; b=1;}
begin
  xi := (37*xi+b) mod 1000; {изменение глобальной переменной}
  rrsch:=xi/1000;
end;
```

В этой программе параметры M. a, b выбраны исходя из возможностей ручного счета на калькуляторе. Для хорошо работающего генератора псевдослучайных чисел эти параметры должны быть очень большими.

Генератор создает следующую последовательность псевдослучайных чисел (N = 25 первых чисел)

x_i = {38, 407, 60, 221, 178, 587, 720, 641, 718, 567, 980, 261, 658, 347, 840, 810, 998, 927, 300, 101, 738, 307, 360, 321, …}.

u_i = {0.038, 0.407, 0.060, 0.221, 0.178, 0.587, 0.720, 0.641, 0.718, 0.567, 0.980, 0.261, 0.658, 0.347, 0.840, 0.081, 0.998, 0.927, 0.300, 0.101, 0.738, 0.307, 0.360, 0.321, 0.878, …}.

При выполнении работы необходимо получить 15 – 20 тыс. чисел.

3. Период последовательности псевдослучайных чисел

Программные генераторы случайных чисел реализуются на компьютерах, обладающих ограниченной разрядной сеткой. Поэтому любая последовательность чисел, программно реализуемая компьютером, будет повторяться, начиная с некоторого числа.

Периодом генератора случайных чисел называют количество неповторяющихся чисел в последовательности $\{x_i\}$ при заданных параметрах и "затравке". Для генераторов случайных чисел важно, что бы эта последовательность была бы достаточно длинной.

Период генератора случайных чисел определяется экспериментально в ходе вычислительного эксперимента. Следует учитывать, что последовательность случайных чисел имеет апериодическую (начальную) и периодическую части. Длина апериодической части заранее не известна. Поэтому при определении периода генератора случайных чисел необходимо провести вычислительный эксперимент. Для определения периода необходимо последовательно выбрать ряд целых случайных чисел, полученных в результате работы процедуры генератора, а затем для каждого из них подсчитать количество вызовов процедуры генератора до того момента, когда в результате получится то же целое число.

Если выбранные числа относятся к апериодической части псевдослучайной последовательности, то полученные периоды будут разные. По мере увеличения номера выбранного числа величина периода перестанет изменяться, что означает, что получен действительный период псевдослучайной последовательности.

Если период псевдослучайной последовательности чисел оказывается коротким (меньше 10 тыс.), необходимо прервать выполнение программы и продолжить ее работу с новой затравкой.

Заметим, что определять период генератора случайных чисел по нормированной последовательности $\{u_i\}$ не следует, так как возникают ошибки, связанные с округлением.

4. **Оценка качества последовательности случайных чисел**

Оценка качества последовательности случайных чисел проводится методами математической статистики. Это стандартная задача проверки статистической гипотезы о типе распределения вероятности.

Допустим, что генератор создает последовательность случайных чисел с равномерным на интервале [0, 1) законом распределения вероятности (статистическая гипотеза). Определим вероятность этого предположения.

Для проверки гипотезы необходимо выполнить следующие действия:

1. разбить интервал [0, 1) на m частей (m = 20 – 50);
2. определить число v_j случайных чисел u_i , попадающих в каждый интервал;
3. определить теоретическое число w_j равномерно распределенных случайных чисел, которые должны попасть в эти интервалы;
4. вычислить значение критерия Пирсона χ^2;
5. по таблице распределения χ^2 найти вероятность правильности статистической гипотезы.

Проведем проверку статистической гипотезы о равномерности распределения последовательности случайных чисел.

Пример

Разобьем интервал $0 \le x < 1$ на m = 5 частей (в последовательности всего N = 25 чисел).

Построим гистограмму попаданий случайных чисел экспериментальной и теоретической последовательности в эти интервалы.

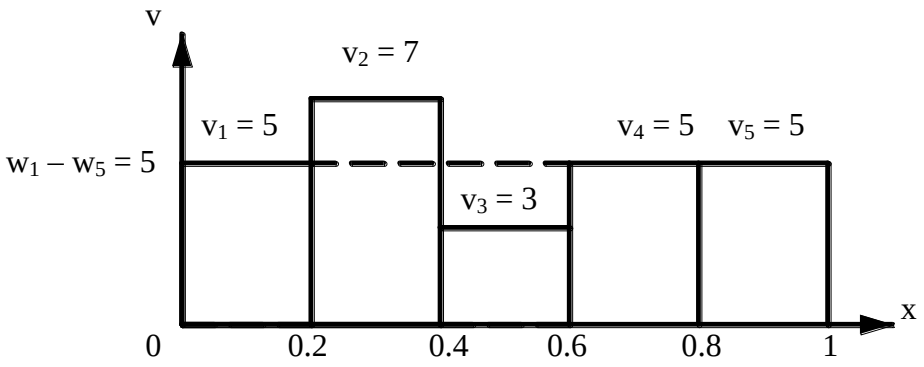


Рис. 1. Пример гистограммы попаданий случайных чисел в интервалы.
$v_1 – v_5$ – экспериментальная,
$w_1 – w_5$ – теоретическая.

Функция равномерного распределения плотности вероятности на интервале [0, 1) имеет следующий вид

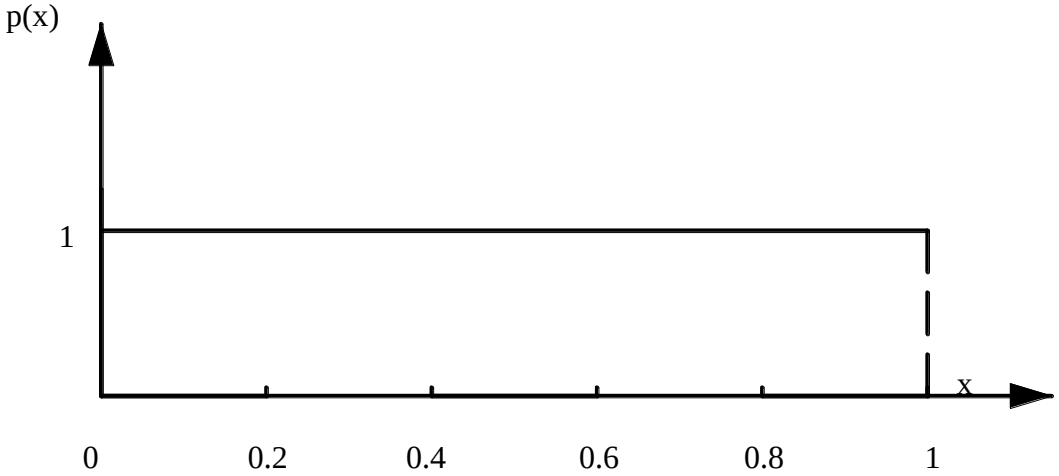


Рис. 2. Функция плотности вероятности для последовательности равномерно распределенных случайных чисел.

Функция распределения вероятностей связана с функцией плотности распределения вероятностей следующим интегралом

$$F(x) = \int_0^x p(t)dt = \int_0^x 1 \cdot dt = x; \qquad (1)$$

График функции равномерного распределения вероятности представлен на следующем рисунке.

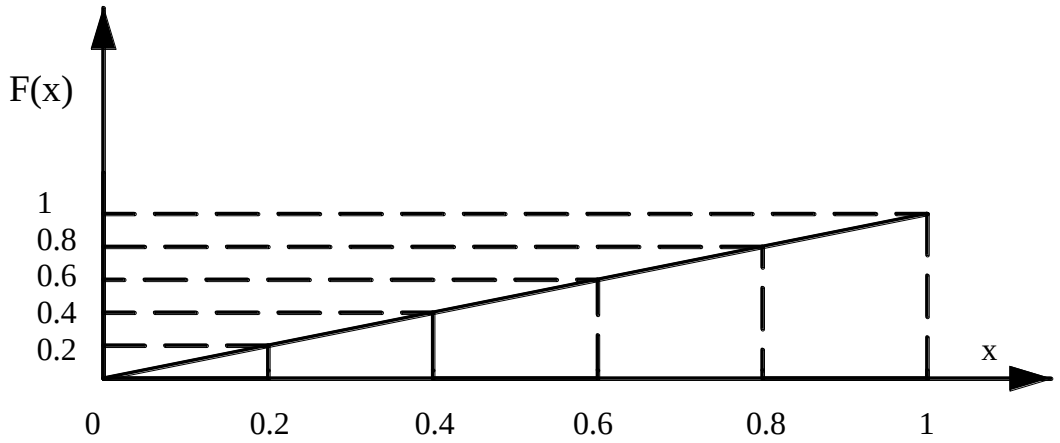


Рис. 3. Функция распределения вероятности для последовательности равномерно распределенных случайных чисел.

Вероятность попадания случайной величины в каждый интервал на оси x равна

$$\Delta P_i = F_{i+1} - F_i = 0.2; \qquad (2)$$

Тогда число случайных чисел последовательности длиной N, которые теоретически должны попасть в каждый интервал, равно

$$w_i = N \cdot \Delta P_i = 25 \cdot 0.2 = 5; \qquad (3)$$

Гистограмма теоретического распределения числа попаданий равномерно распределенных случайных чисел в интервалы показана на рисунке 1 пунктиром.

Понятно, что чем ближе теоретическая и экспериментальная гистограммы, тем ближе распределение вероятности случайных чисел экспериментальной и теоретической последовательности. Это простое соображение положено в основу критерия согласия Пирсона (критерия χ^2).

Количественную оценку близости теоретического и экспериментального распределения вероятности случайных величин по критерию Пирсона находят следующим образом. Вычисляем значение χ^2 по следующей формуле

$$\chi^2 = \sum_{i=1}^{m} \frac{(v_i - w_i)^2}{w_i}; \qquad (4)$$

Для нашего примера значение χ^2 = 1.6; Дальше необходимо взять таблицу распределения χ^2 и по ней определить вероятность выдвинутой статистической гипотезы. В эту таблицу, кроме значения χ^2 входит второй параметр – число степеней свободы n. Число степеней свободы n на единицу меньше числа интервалов m на гистограмме.

$$n = m - 1; \qquad (5)$$

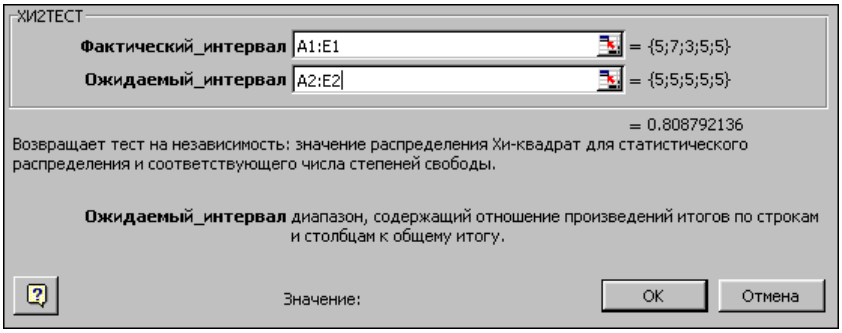
Для рассматриваемого примера число степеней свободы n = 4.

В результате по таблице находим, что вероятность выдвинутой статистической гипотезы равна 0.81 (81%).

Как относиться к этому результату?

Понятно, что если вероятность правильности статистической гипотезы мала (обычно меньше 70%), то результат нельзя считать хорошим. Аналогично обстоит дело и с очень высокой вероятностью правильности статистической гипотезы (обычно больше 95%). В этом случае, возможно, что последовательность чисел является не случайной, а является специально подобранным набором чисел.

Проверку статистической гипотезы по критерию χ^2 удобно проводить средствами программы Microsoft Excel. Для этого необходимо загрузить числа из экспериментальной и теоретической гистограмм в ячейки программы Excel и вызвать функцию ХИ2ТЕСТ. Для примера на рисунке 1 получим:



Если проверка статистической гипотезы проводится по таблице распределения χ^2, то необходимо выбрать число интервалов в гистограмме в соответствии с имеющейся таблицей. Целесообразно выбирать m из ряда чисел 21, 31, 41 и т.д., что позволяет получить удобные значения для числа степеней свободы n = 20, 30, 40 и т.д. Эти числа обычно указываются в таблице распределения χ^2

5. Практические советы по выполнению работы

Организация программы. При создании программ генераторов случайных чисел не следует записывать последовательности $\{x_i\}$ или $\{u_i\}$ в массив, сохраняя случайные числа для дальнейшей обработки. Во многих трансляторах языков программирования максимально допустимая длина массива недостаточно велика. Целесообразно организовать программу в виде цикла, в котором каждое случайное число заносится в гистограмму. Гистограмма экспериментального распределения случайных чисел является результатом работы генератора случайных чисел.

Операция взятия остатка от деления в С. В языке С существует бинарная операция взятия остатка от деления, аналогичная операции **mod** в языке Паскаль. Она обозначается знаком процента, например 25 % 10 (равно 5) или a % m (остаток от деления a на m).

Подсчет количества попаданий в интервалы гистограммы. Очевидно, что для интервала от *A* до *B*, разбитого на *L* равных отрезков, нумеруемых с нуля, номер отрезка *n*, которому принадлежит число *x* из интервала [*A*, *B*] определяется по формуле

$$n = \left[\frac{x - A}{B - A} \cdot L \right] \qquad 0 \le n \le L - 1,$$

где квадратные скобки означают округление до ближайшего меньшего целого. В связи с этим нет никакой необходимости применять для этой цели в программе цикл перебора границ всех *L* отрезков.

Вычисление коэффициента χ^2. При реализации программы на языке С следует помнить, что операция деления для целых чисел дает целый результат (в Паскале целочисленное деление обозначается оператором **div,** и поэтому такой проблемы не возникает). В связи с этим при подсчете значения критерия χ^2можно получить неверный (сильно заниженный) результат, если объявить все участвующие в расчете переменные целыми и проводить целочисленное деление. Целое значение χ^2 должно настораживать, так как обычно является следствием этой ошибки.

6. Задания к лабораторной работе

- Построить генератор случайных чисел с равномерным законом распределения на интервале [0,1) по методу середин квадратов или мультипликативного датчика в соответствии с заданием. Написать и отладить программу, реализующую генератор на языке Паскаль или Си. Получить выборку неповторяющихся псевдослучайных чисел объемом не меньше 20 тыс.

- Определить период генератора случайных чисел. Если он меньше 10000, то изменить исходные данные.

- Провести анализ качества последовательности случайных чисел по критерию Пирсона.

- Составить краткий отчет, включающий текст программы, скриншоты, теоретическую и экспериментальную гистограмму (скриншот, гистограмму, построенную в EXEL), вероятность принятия статистической гипотезы о равномерном распределении случайных чисел.

Отчет должен включать:

1. Фамилию, номер группы, номер варианта студента.
2. Формулировку задания.
3. Параметры исследуемой последовательности из задания.
4. Текст программы. Генератор должен быть реализован в виде функции языка программирования. В программе использовать только целочисленные операции для расчета вспомогательной последовательности целых чисел.
5. Результаты анализа генератора.
6. Вывод о достигнутой или не достигнутой цели работы.

Отчет направляется преподавателю для проверки. После проверки и исправления ошибок проводится защита лабораторной работы.

Таблица вариантов задания для группы А7 -

№	Метод для генератора случайных чисел
1	Середин квадратов
2	Середин квадратов
3	Середин квадратов
4	Середин квадратов
5	Середин квадратов
6	Середин квадратов
7	Мультипликативный датчик
8	Мультипликативный датчик
9	Мультипликативный датчик
10	Мультипликативный датчик
11	Мультипликативный датчик
12	Мультипликативный датчик
13	Середин квадратов
14	Середин квадратов
14	Середин квадратов
16	Середин квадратов
17	Середин квадратов
18	Середин квадратов
19	Мультипликативный датчик
20	Мультипликативный датчик
21	Мультипликативный датчик
22	Мультипликативный датчик
23	Мультипликативный датчик
24	Мультипликативный датчик
25	Середин квадратов

Таблица вариантов задания для группы А8 -

№	Метод для генератора случайных чисел
1	Середин квадратов
2	Середин квадратов
3	Середин квадратов
4	Середин квадратов
5	Середин квадратов
6	Середин квадратов
7	Мультипликативный датчик
8	Мультипликативный датчик
9	Мультипликативный датчик
10	Мультипликативный датчик
11	Мультипликативный датчик
12	Мультипликативный датчик
13	Середин квадратов
14	Середин квадратов
14	Середин квадратов
16	Середин квадратов
17	Середин квадратов
18	Середин квадратов
19	Мультипликативный датчик
20	Мультипликативный датчик
21	Мультипликативный датчик
22	Мультипликативный датчик
23	Мультипликативный датчик
24	Мультипликативный датчик
25	Середин квадратов

Таблица вариантов задания для группы А4 -

№	Метод для генератора случайных чисел
1	Середин квадратов
2	Середин квадратов
3	Середин квадратов
4	Середин квадратов
5	Середин квадратов
6	Середин квадратов
7	Мультипликативный датчик
8	Мультипликативный датчик
9	Мультипликативный датчик
10	Мультипликативный датчик
11	Мультипликативный датчик
12	Мультипликативный датчик
13	Середин квадратов
14	Середин квадратов
15	Середин квадратов
16	Середин квадратов
17	Середин квадратов
18	Середин квадратов
19	Мультипликативный датчик
20	Мультипликативный датчик
21	Мультипликативный датчик
22	Мультипликативный датчик
23	Мультипликативный датчик
24	Мультипликативный датчик
25	Середин квадратов

Таблица вариантов задания для группы А12 -

№	Метод для генератора случайных чисел
1	Середин квадратов
2	Середин квадратов
3	Мультипликативный датчик
4	Мультипликативный датчик
5	Середин квадратов
6	Середин квадратов
7	Мультипликативный датчик
8	Середин квадратов
9	Середин квадратов
10	Мультипликативный датчик
11	Мультипликативный датчик
12	Мультипликативный датчик
13	Середин квадратов
14	Середин квадратов
15	Середин квадратов
16	Середин квадратов
17	Середин квадратов
18	Середин квадратов
19	Мультипликативный датчик
20	Мультипликативный датчик
21	Середин квадратов
22	Середин квадратов
23	Мультипликативный датчик
24	Мультипликативный датчик
25	Середин квадратов