Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

ОТЧЕТ

к лабораторной работе №1

По курсу: «Моделирование»

Тема: ***«Генераторы случайных чисел»***

Студентка ИУ7-75Б
Оберган Т.М

Преподаватель
Рудаков И.В.

Москва, 2020 г.

Оглавление

Введение

Случайные числа — искусственно полученная последовательность реализаций случайной величины с заданным законом распределения.

Случайное число всегда рассматривается в контексте какой-то последовательности, которая характеризуется тем, что каждое число последовательности не зависит от всех остальных чисел последовательности. Случайные числа должны подчиняться равномерному распределению, то есть вероятность появления каждого числа равновероятна.

Случайные числа имеют применение в физике, например в исследованиях электронного шума, в инженерном деле и исследовании операций. Многие методы статистического анализа, методы Монте-Карло в физике и информатике требуют случайных чисел.

Аналитическая часть

В данном разделе будут рассмотрены основные типы генераторов случайных чисел.

Генераторы случайных чисел по способу получения чисел делятся на:

- аппаратные;
- табличные;
- алгоритмические.

Аппаратные генераторы

Аппаратные генераторы случайных чисел – это устройства, использующие для создания случайных чисел замеры параметров некоторых физических процессов. Как правило, аппаратный генератор случайных чисел состоит из источника энтропии и устройства, преобразующего значения, полученные с источника энтропии, в нужный формат. Разработка генераторов, использующих источники энтропии, генерирующих не коррелированные и статистически независимые числа – достаточно сложная задача.

Источниками энтропии могут быть:

- подбрасывание монеты;
- временные задержки между моментами излучения частиц в процессе радиоактивного распада;
- тепловые шумы при работе полупроводникового диода или резистора;
- частотные отклонения свободно работающего генератора частот;
- фотоэффект — испускание электронов веществом под действием света;
- звук от микрофона или видео с подключенной камеры;
- состояние некоторых блоков памяти компьютера.

Алгоритмические генераторы

Алгоритмический генератор является комбинацией физического генератора и детерминированного алгоритма. Такой генератор использует ограниченный набор данных, полученный с выхода физического генератора для создания длинной последовательности чисел преобразованиями исходных чисел.

Из-за дороговизны аппаратных генераторов случайных чисел в большинстве случаев, в качестве источника энтропии используются ресурсы вычислительной машины, на которой выполняется программа генерации ПСЧ. При отсутствии аппаратного генератора случайных чисел в качестве источника энтропии могут использоваться:

- состояние системных часов;
- время задержек между нажатиями клавиш клавиатуры или движениями мышки;
- содержимое буферов ввода/вывода;
- значения, получаемые при работе системы (время загрузки системы, сетевая активность и т. п.).

Сгенерированные числа с «привязкой» к подобным ИЭ будут менее случайны.

Рандомизация перемешиванием.

Допустим, имеются две последовательности ПСЧ, сгенерированные двумя разными методами. Тогда можно, например, использовать одну последовательность для изменения порядка другой.

Табличные генераторы

Табличные генераторы в качестве источника случайных чисел используют заранее подготовленные таблицы, содержащие проверенные некоррелированные числа и не являются генераторами в строгом понимании этого понятия. Недостатки такого способа очевидны: использование внешнего ресурса для хранения чисел, ограниченность последовательности, предопределенность значений.

Анализ выбранных методов

В данном разделе будут проанализированы методы, используемые в лабораторной.

Линейный конгруэнтный метод

Для осуществления генерации чисел данным методом, необходимо задать 4 числа:

$m > 0$, модуль

$0 \leq a \leq m$, множитель

$0 \leq c \leq m$, приращение

$0 \leq X_0 \leq m$, начальное число

Последовательность случайных чисел генерируется при помощи формулы:

$$X_{n+1} = (aX_n + c)\ mod\ m$$

Собственный табличный метод

Табличные ГСЧ в качестве источника случайных чисел используют специальным образом составленные таблицы, содержащие проверенные некоррелированные, то есть никак не зависящие друг от друга, цифры.

В качестве таблицы было предварительно сгенерировано и сохранено 7919 цифр. От сгенерированной последовательности зависит качество результата. В идеале в таблице должно быть одинаковое количество цифр каждого значения.

В придуманном алгоритме хранится указатель на текущую позицию в таблице $0 \leq pos < len(table)$. Для получения случайного числа происходит набор $n + dn$ случайных цифр (seq). Где n – максимальная длина требуемого случайного числа, dn – целое неотрицательное произвольное число. После завершения набора находится seq mod $(stop - start)$, где stop и start – границы диапазона генерации.

При получении очередной цифры указатель смещается (по модулю длины таблицы) на случайное число, полученное из источника энтропии (в данной ЛР – времени).

Критерий равномерности:

Пусть дана числовая последовательность X_n длины n. При оценке равномерности берется некоторое число d и для каждого r (где $0 \leq r < d$) подсчитывается количество случаев, когда элемент последовательности $X_i = r$ (где $0 \leq i < n$). После этого применяется критерий χ^2, в котором вычисляется статистика, имеющая следующий вид:

$$\chi^2 = \sum_{j=0}^{k-1} \frac{(n_j - E_j)^2}{E_j} \sim \chi^2_{k-1}$$

где $k = d$;

$p = \frac{1}{d}$ (для каждой категории 1-, 2-х и 3-х разрядных чисел);

$E_j = p * n$.

Имеем распределение χ^2 с $k - 1$ степенями свободы.

После сравнения полученного значения χ^2 с теоретическим χ^2 можно сделать вывод о пригодности генератора для использования. При этом возможно три случая:

1. полученный χ^2 много больше любого теоретического χ^2 – гипотеза о случайности равномерного генератора не выполняется (разброс чисел слишком велик, чтобы быть случайным);
2. полученный χ^2 много меньше любого теоретического χ^2 – гипотеза не выполняется (разброс чисел слишком мал, чтобы быть случайным);
3. полученный χ^2 лежит между теоретическими значениями двух рядом стоящих столбцов – гипотеза о выполняется с вероятностью p (то есть в p случаях из 100).

Критерий промежутков между днями рождений

Предположим, что $(Y_1, Y_2, \dots, Y_m)$ – это дни рождения, где $0 \leq Y_k < m$. Расположим их в порядке неубывания $Y_1 \leq \cdots \leq Y_n$, определим n "промежутков"

$$S_1 = Y_2 - Y_1, \dots, S_{n-1} = Y_n - Y_{n-1}, S_n = Y_1 + m + Y_n$$

и, наконец, расположим промежутки в таком порядке: $S_1 \leq \cdots \leq S_n$. Пусть R – число равных промежутков, а именно – число индексов j, таких, что $1 < j \leq n$ и $S_j = S_{j-1}$.

В среднем, число одинаковых промежутков для выбранных m и n должно быть равным приблизительно 1. Далее, мы можем применить этот критерий много раз и воспользоваться χ^2-критерием с тремя степенями свободы, чтобы сравнить эмпирические значения R_j с правильным распределением. Так можно узнать, будет ли генератор вырабатывать приемлемые случайные промежутки между днями рождений.

Данный критерий примечателен тем, что на нём споткнулся генератор Фибоначчи с запаздыванием и его производные.

Результаты

Стандартная библиотека

N = 100; Диапазон: [0; 100]

p-value: 0.0452871

Критерий промежутков между днями рождения: R = [39, 37, 16, 8]

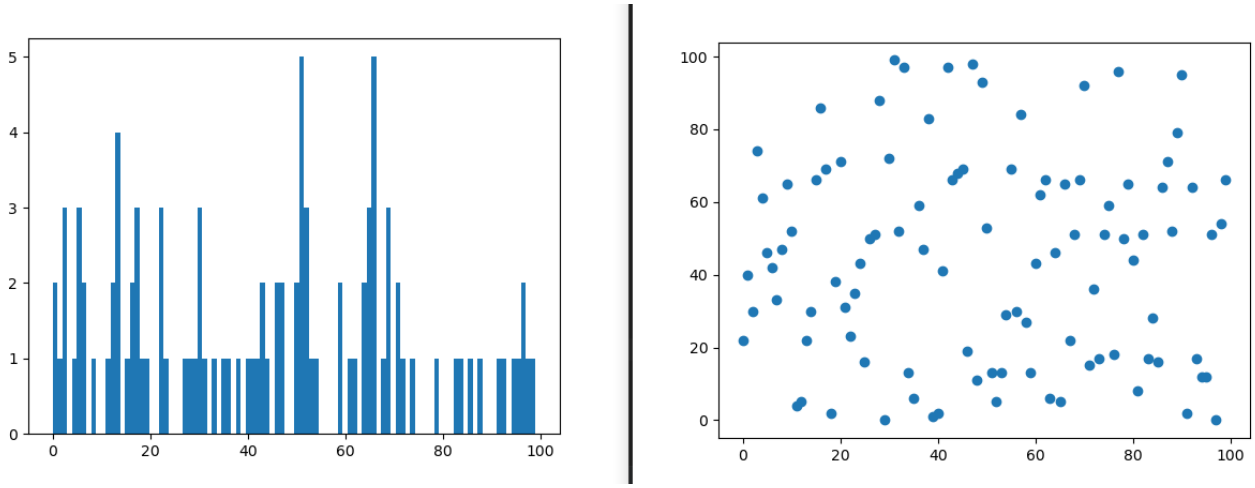

Рис. 1 – стандартная библиотека, N = 100, диапазон - [0; 100]

N = 1000; Диапазон: [0; 100]

p-value: 0.5723586

Критерий промежутков между днями рождения: R = [888, 100, 0, 0]

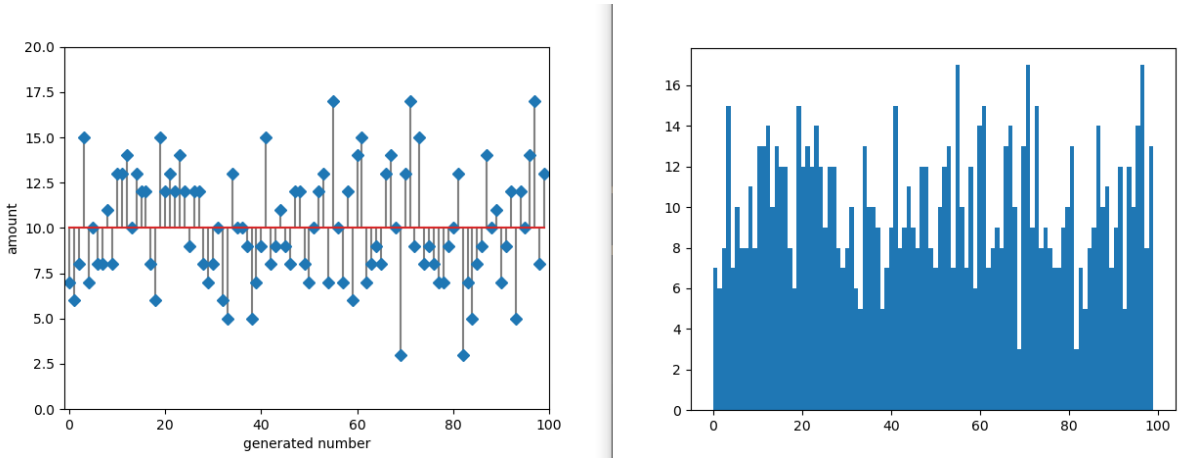

Рис. 2 – стандартная библиотека, N = 1000, диапазон - [0; 100]

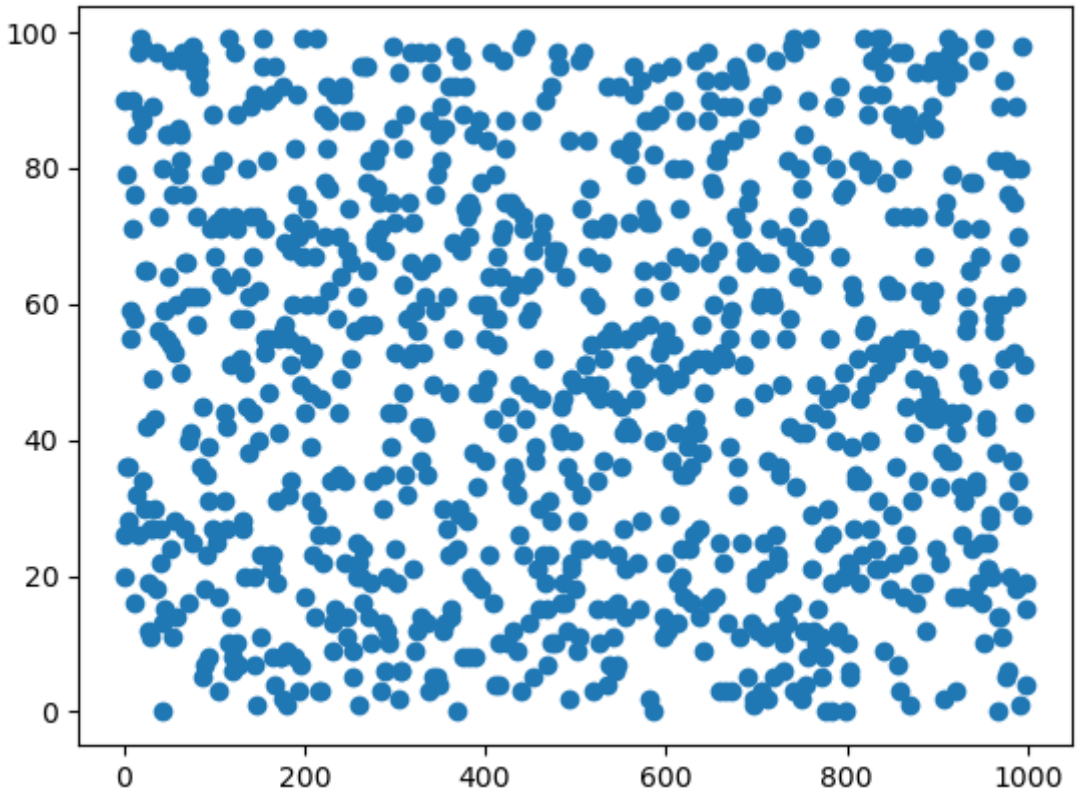

Рис. 3 – стандартная библиотека, N = 1000, диапазон - [0; 100]

N = 10000; Диапазон: [0; 100]

p-value: 0.3731133

Критерий промежутков между днями рождения: R = [9824, 100, 0, 0]

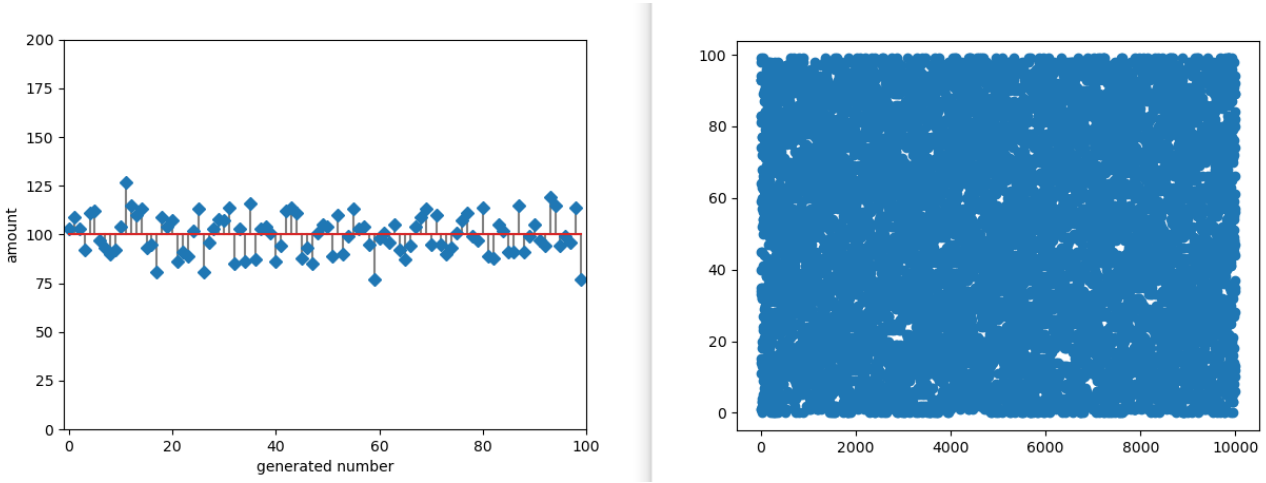

Рис. 4 – стандартная библиотека, N = 10000, диапазон - [0; 100]

Линейный конгруэнтный метод

В качестве параметров использованы: a = 106; c = 1283; X0 = 7

N = 100; Диапазон: [0; 99]

p-value: 1.0000000

Критерий промежутков между днями рождения: R = [9, 80, 10, 0]

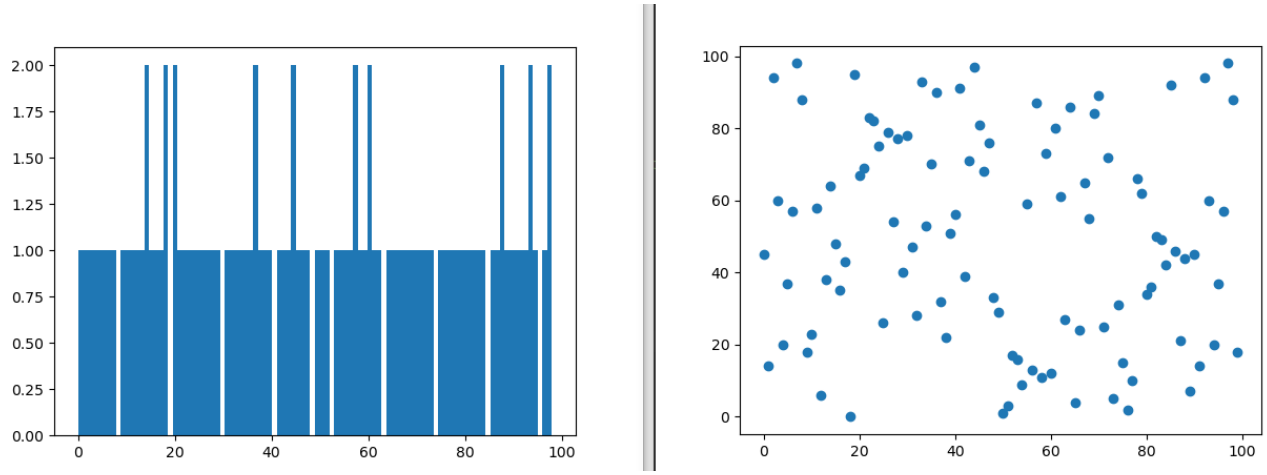

Рис. 5– визуальное отображение результата линейного конгруэнтного метода на N = 100, в диапазоне [0; 99]

N = 1000; Диапазон: [0; 99]

p-value: 0.1753649

Критерий промежутков между днями рождения: R = [899, 80, 10, 0]

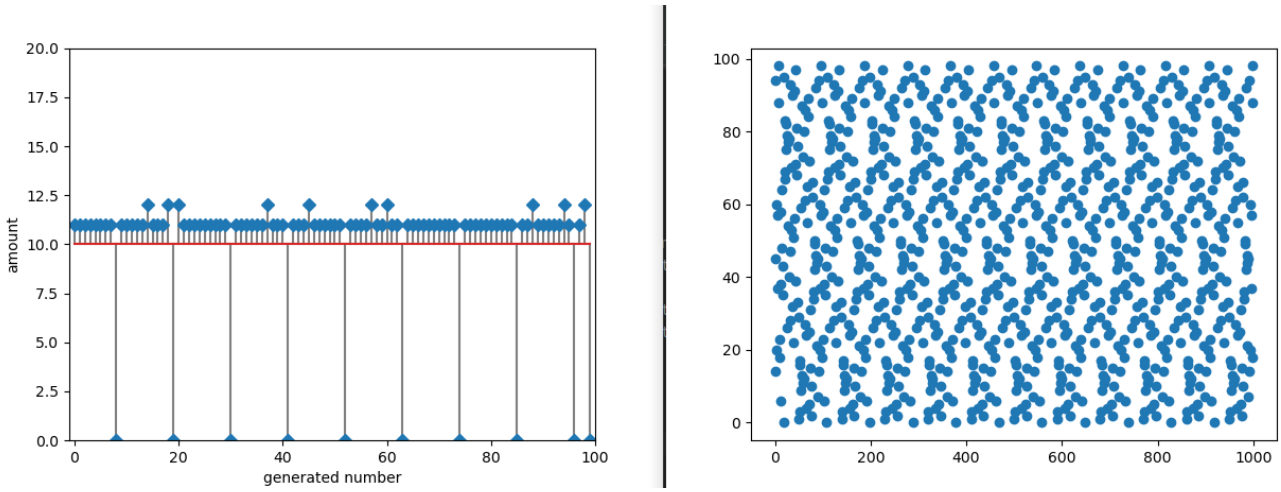

Рис. 6 – линейный конгруэнтный метод N = 1000, диапазон - [0; 99]

На рис. 6 заметен недостаток данного метода: некоторые числа так и не были ни разу сгенерированы, на правом графике явно виден паттерн генерации, последовательность не выглядит случайной.

N = 10000; Диапазон: [0; 100]

p-value: 0.0000000

Критерий промежутков между днями рождения: R = [9576, 0, 0, 25]

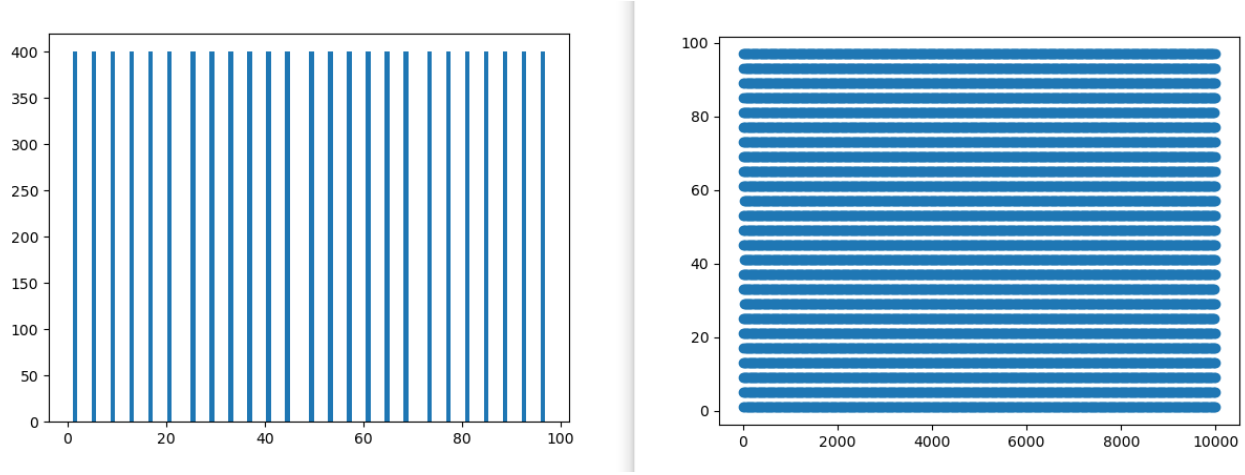

Рис. 7 – линейный конгруэнтный метод N = 10000, диапазон - [0; 100]

Рис. 7 демонстрирует, что на некоторых значениях параметров данный алгоритм показывает критически неудовлетворительный результат.

Собственная реализация табличного метода

Таблица 1 – подсчет количества определенной цифры в наборах данных

Цифра:	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*Всего*
Data1	788	770	818	791	824	813	781	796	747	791	7919
Data2	291	367	369	371	369	369	368	367	367	362	3600
Data3	0	0	0	0	0	1575	1610	16571	1587	1576	7919

*7919 – простое число

***Data1**; N = 100; Диапазон: [0; 100]*

p-value: 0.9584750

Критерий промежутков между днями рождения: R = [30, 51, 13, 6]

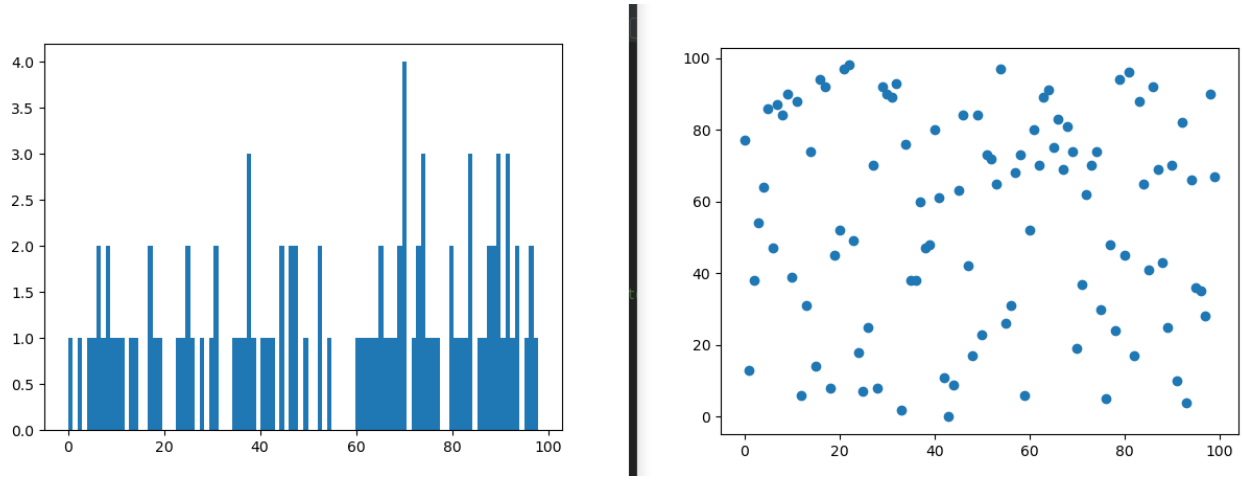

Рис. 8 – табличный метод N = 100, диапазон - [0; 100]

***Data1**; N = 1000; Диапазон: [0; 100]*

p-value: 0.1720090

Критерий промежутков между днями рождения: R = [889, 100, 0, 0]

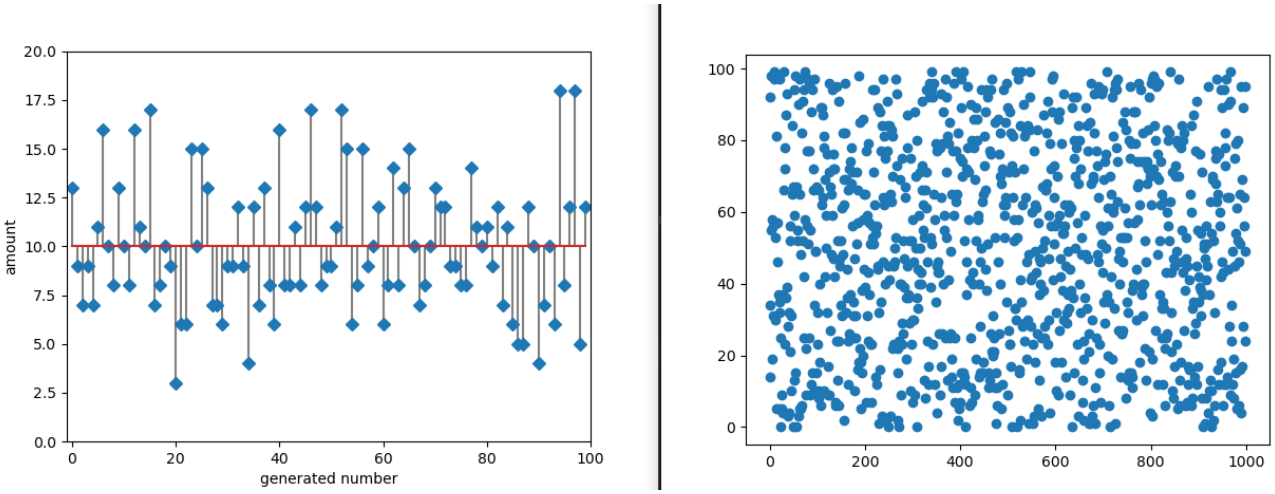

Рис. 9 – табличный метод N = 1000, диапазон - [0; 100]

На рис. 9 видно, что генерируемые числа вполне случайны.

При диапазоне, равному простому числу табличный метод показал результаты лучшие, чем в обратном случае (см рис. 10 и рис. 11).

***Data1**; N = 10000; Диапазон: [0; 101]*

p-value: 0.4874561

Критерий промежутков между днями рождения: R = [9803, 101, 0, 0]

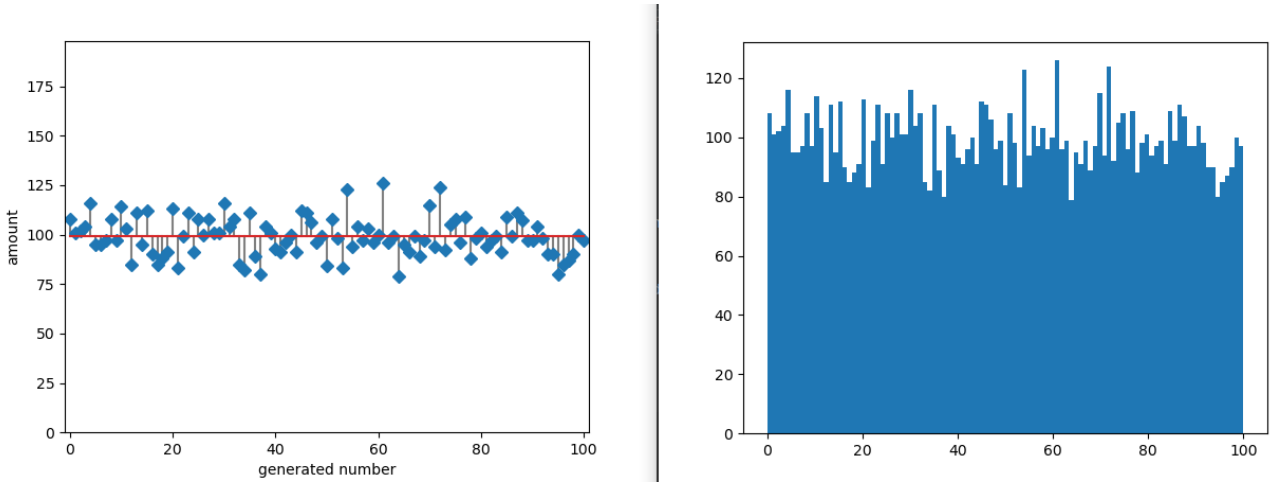

Рис. 10 – табличный метод N = 10000, диапазон - [0; 101]

На рисунках 11 и 12 можно заметить влияние неравномерного распределения цифр в Data2 и зависимость от длины таблицы.

***Data1**; N = 10000; Диапазон: [0; 100]*

p-value: 0.0017969

Критерий промежутков между днями рождения: R = [9797, 100, 0, 0]

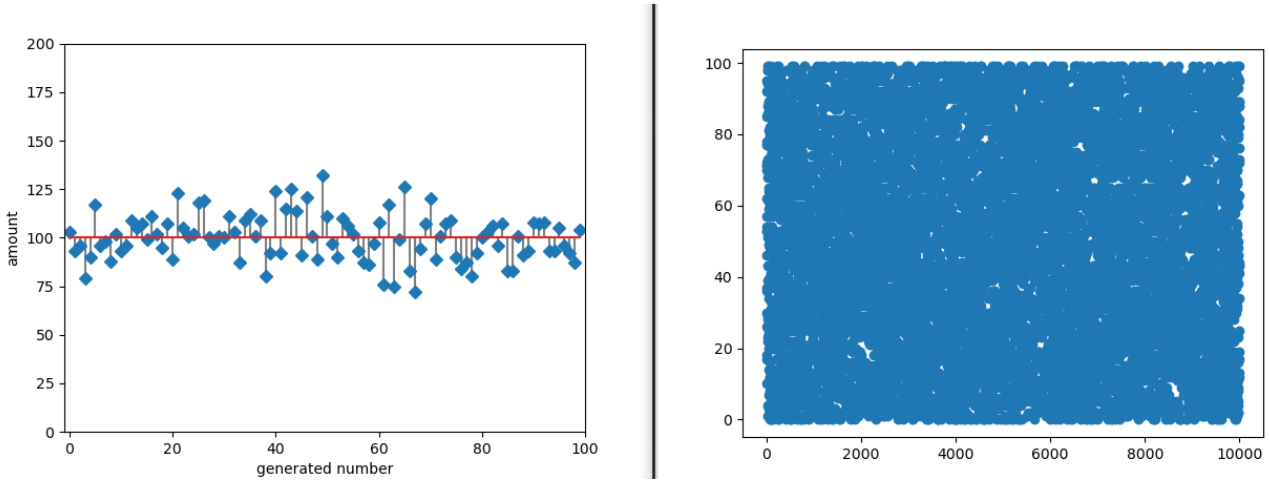

Рис. 11 – табличный метод N = 10000, диапазон - [0; 100]

***Data2**; N = 10000; Диапазон: [0; 100]*

p-value: 0.0000000

Критерий промежутков между днями рождения: R = [9796, 100, 0, 0]

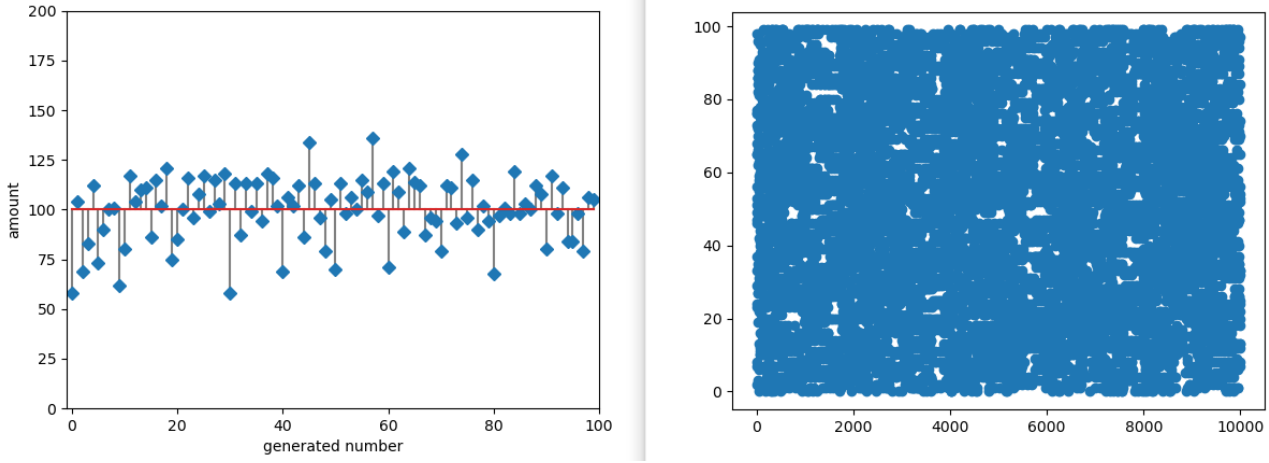

Рис. 12 – табличный метод N = 10000, диапазон - [0; 100]

На рисунках 13 и 14 показан результат работы табличного метода на таблице, содержащей только цифры 5-9.

***Data3**; N = 10000; Диапазон: [0; 101]*

p-value: 0.0000000

Критерий промежутков между днями рождения: R = [9620, 72, 8, 1]

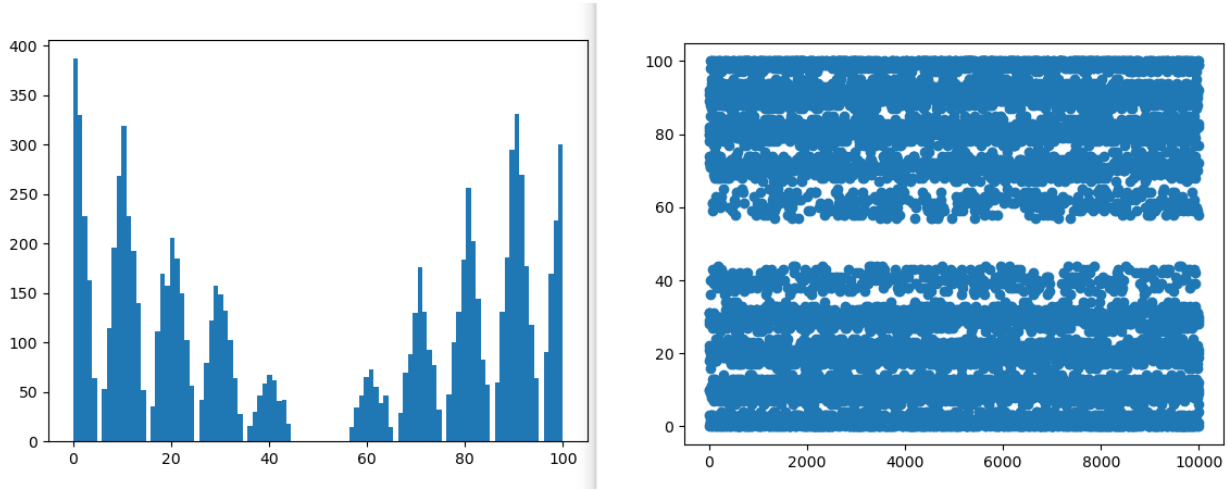

Рис. 13 – табличный метод N = 10000, диапазон - [0; 101]

***Data3**; N = 10000; Диапазон: [0; 100]*

p-value: 0.0000000

Критерий промежутков между днями рождения: R = [9587, 20, 0, 5]

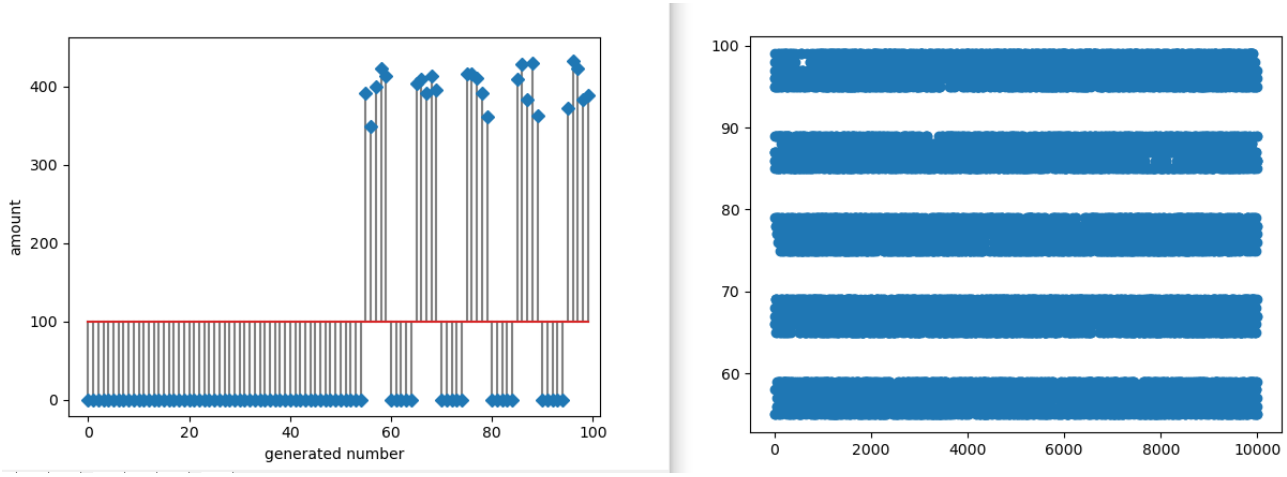

Рис. 14 – табличный метод N = 10000, диапазон - [0; 100]

Вывод

В ходе выполнения данной лабораторной работы были рассмотрены основные типы генераторов случайных чисел, выбраны и реализованы алгоритмические и табличные методы генерации случайных чисел, реализована проверка генерируемых последовательностей по критерию равномерности и критерию промежутков между днями рождения и визуальное отображение полученных результатов генерации.

В ходе анализа полученных результатов установилось, что линейный конгруэнтный метод не подходит для генерации случайных чисел ввиду повторяющихся подпоследовательностей, неудовлетворению критериям.

Реализованный табличный метод генерации проходит критерии, но уступает стандартному библиотечному методу.

Список литературы

[1] Владислав Ткачук. ГПСЧ: Линейный конгруэнтный метод. 31.07.2019. [Электронный ресурс]. – Режим доступа: http://blog.tkachuk.su/2019/07/31

[2] Дональд Кнут. Искусство программирования. Том 2. Получисленные алгоритмы — 3-е изд. — М.: «Вильямс», 2007. — с. 93-94

[3] Слеповичев И.И., «Генераторы псевдослучайных чисел», 2017

[4] Критерий согласия Пирсона χ2 (Хи-квадрат). [Электронный ресурс]. – Режим доступа: https://statanaliz.info/statistica/proverka-gipotez/kriterij-soglasiya-pirsona-khi-kvadrat/ . Проверено 27.10.2020.

[5] Wikipedia, Chi-square distribution.). [Электронный ресурс]. – Режим доступа: https://en.wikipedia.org/wiki/Chi-square_distribution. Проверено 27.10.2020.

[6] SciPy.org, chisquare function. [Электронный ресурс]. – Режим доступа: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chisquare.html. Проверено 29.10.2020.

[7] MathWorks, chi2cdf function. [Электронный ресурс]. – Режим доступа: https://www.mathworks.com/help/stats/chi2cdf.html#mw_75ebb453-75d8-4969-adf4-778bb0e09b68. Проверено 29.10.2020.