ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«СИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ТЕЛЕКОММУНИКАЦИЙ И ИНФОРМАТИКИ»

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА
к нулевой лабораторной работе
«Оценка качества генератора псевдослучайных чисел»
по дисциплине «Моделирование»

Выполнил студент Федосеев Вячеслав Эдуардович
Ф.И.О.

Группы ИВ – 621

Работу принял ассистент кафедры ВС
Я. В. Петухова
подпись

Новосибирск – 2020

СОДЕРЖАНИЕ

ПОСТАНОВКА ЗАДАЧИ

В рамках лабораторной работы необходимо выбрать три генератора псевдослучайных чисел, оценить равномерность последовательностей псевдослучайных чисел на заданном отрезке с помощью критерий согласия Пирсона, оценить статистическую независимость генерируемых последовательностей с помощью автокорреляционной функции.

ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

Проверка гипотез о законе распределения

Во многих случаях закон распределении изучаемой случайной величины неизвестен, но есть основания предположить, что он имеет определенный вид: нормальный, показательный или равномерный.

Пусть необходимо проверить гипотезу H_0 о том, что с. в. X подчиняется определенному закону распределения, заданному функцией распределения $F_0(x)$, т. е. H_0: $F_X(x) = F_0(x)$. Альтернативная гипотеза H_1: $F_X(x) \neq F_0(x)$.

Для проверки гипотезы о распределении с. в. X нужно провести выборку, которую удобно оформить в виде статистического ряда:

Таблица 1 – статистический ряд

x_1	x_2	x_3	…	x_m
n_1	n_2	n_3	…	n_m

где $\sum_{i=1}^{n} n_i = n$ – объем выборки.

Требуется сделать заключение: согласуются ли результаты наблюдений с гипотезой. Для этого используется специально подобранная величина критерий согласия.

Критерий согласия – статистический критерий проверки гипотезы о предполагаемом законе распределения с эмпирическими данными.

Критерий согласия Пирсона или χ^2

Для проверки гипотезы H_0 поступают следующим образом. Область значений с. в. X разбивается на k интервалов δ_i и подсчитывают вероятности p_i $(i = 1, 2, \dots, k)$ попадания в δ_i интервал, используя формулу $P\{\alpha \le X \le \beta\} = F_0(\beta) - F_0(\alpha)$. Тогда число значений с. в. X, попавших δ_i интервал, можно рассчитать по формуле $n \cdot p_i$.

Таблица 2 – теоретический интервальный ряд

δ_1	δ_2	δ_3	…	δ_k
$n_1^{'} = n \cdot p_1$	$n_2^{'} = n \cdot p_2$	$n_3^{'} = n \cdot p_3$	…	$n_k^{'} = n \cdot p_k$

Таким образом, имеем статистический ряд распределения с. в. X и теоретический ряд распределения. Чтобы оценить степень расхождения эмпирических частот n_i и теоретических частот $n_i^{'}$ К. Пирсон предложил такую величину:

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i} = \sum_{i=1}^{k} \frac{{n_i}^2}{n \cdot p_i} - n$$

По теореме Пирсона, данная величина при $n \to \infty$ имеет χ^2-распределение с $m = k - r - 1$ с степенями свободы, где $k -$ число интервалов, $m -$ число параметров распределения. В частности для равномерного распределения $m = k - 3$.

Применение критерия Пирсона сводится к следующему:

1. Вычисляется величина $\chi^2_{набл}$;
2. Выбирается уровень значимости α критерия, по таблице χ^2-распределения находят квантиль $\chi^2_{\alpha,m}$;
3. Если $\chi^2_{набл} \le \chi^2_{\alpha,m}$, то гипотеза H_0 не противоречит опытным данным, иначе H_0 отвергается;

Необходимым условием применения критерия Пирсона является наличие в каждом интервале не менее 5 наблюдений.

Автокорреляционная функция

Автокорреляционная функция предназначена для оценка корреляции между сдвинутыми копиями последовательностей и отдельных подпоследовательностей.

Рассчитывается автокорреляция по следующей формуле:

$$\hat{\alpha}(\tau) = \sum_{i=1}^{n-\tau} \frac{(X_i - \overline{X}) \cdot (X_{i+\tau} - \overline{X})}{S_n^2 \cdot (n - \tau)}$$

где $\overline{X}$ – выборочное среднее, S_n^2 – выборочная дисперсия, n – кол-во чисел, τ – сдвиг последовательности.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Для экспериментов возьмем три генератора из математической библиотеки NumPy:

1. PCG64 – это 128-битная реализация конгруэнтного генератора перестановок O'Нила. Разработан в 2014 году, с помощью функции перестановок улучшает статистические свойства линейного конгруэнтного генератора по модулю 2. Период: 2^{128};
2. SFC64 – это 256-битная реализация малого быстрого хаотического PRNG Криса Доти-Хамфри. Разработан в 2007 году. Период: 2^{255};
3. Philox – это 64-битный генератор, который основан на счетчиках и более слабых криптографических функциях. Разработан в 2011 году. Упрощение и модификация блочного шифра Threefish с добавлением S-блока. Период: 2^{256} - 1;

Оценка равномерности последовательностей псевдослучайных чисел на заданном отрезке с помощью критерий согласия Пирсона

Кол-во чисел: N = 100000.
Кол-во интервалов: k = 50.
Кол-во свободных степеней: $m = k - 3 = 47$.
Уровень значимости $\alpha = 0.8$.
Квантиль $\chi^2_{кр} = \chi^2_{\alpha=0.8,\ m=47} = 54.9056$.

Рисунок 1 Гистограмма частот для PCG64, $\chi^2_{набл} = 53.5 \le \chi^2_{кр}$

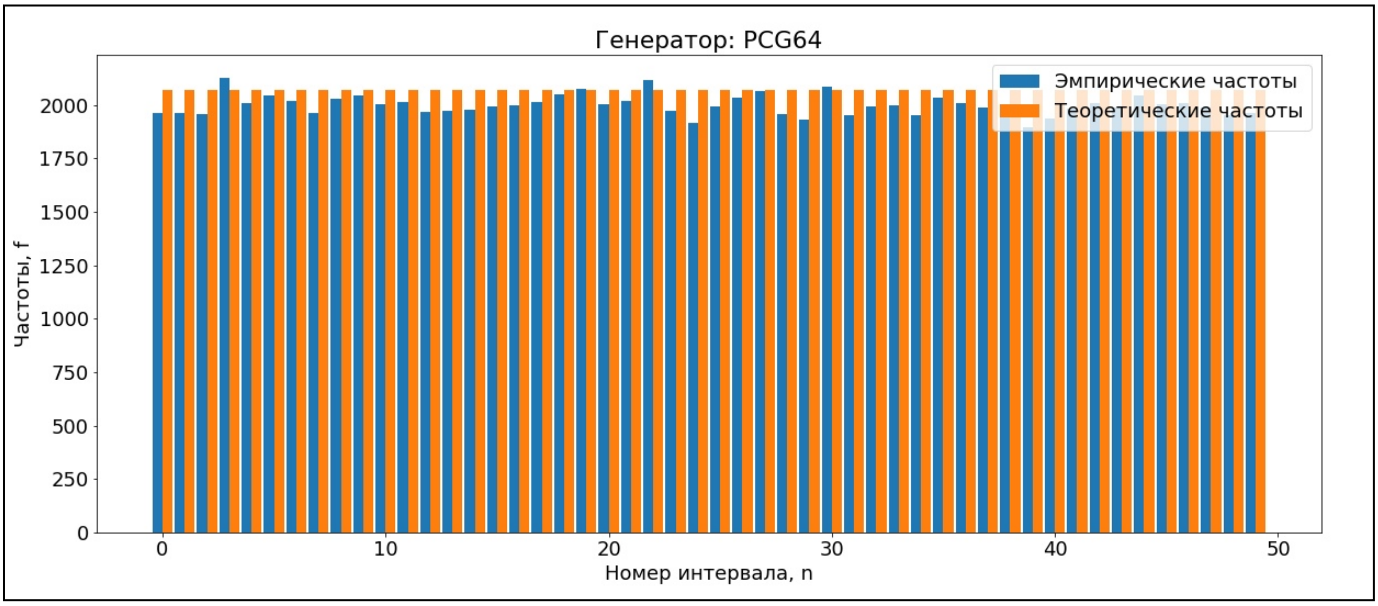


Рисунок 2. Гистограмма частот для SFC64, $\chi^2_{набл} = 41.8 \leq \chi^2_{кр}$.

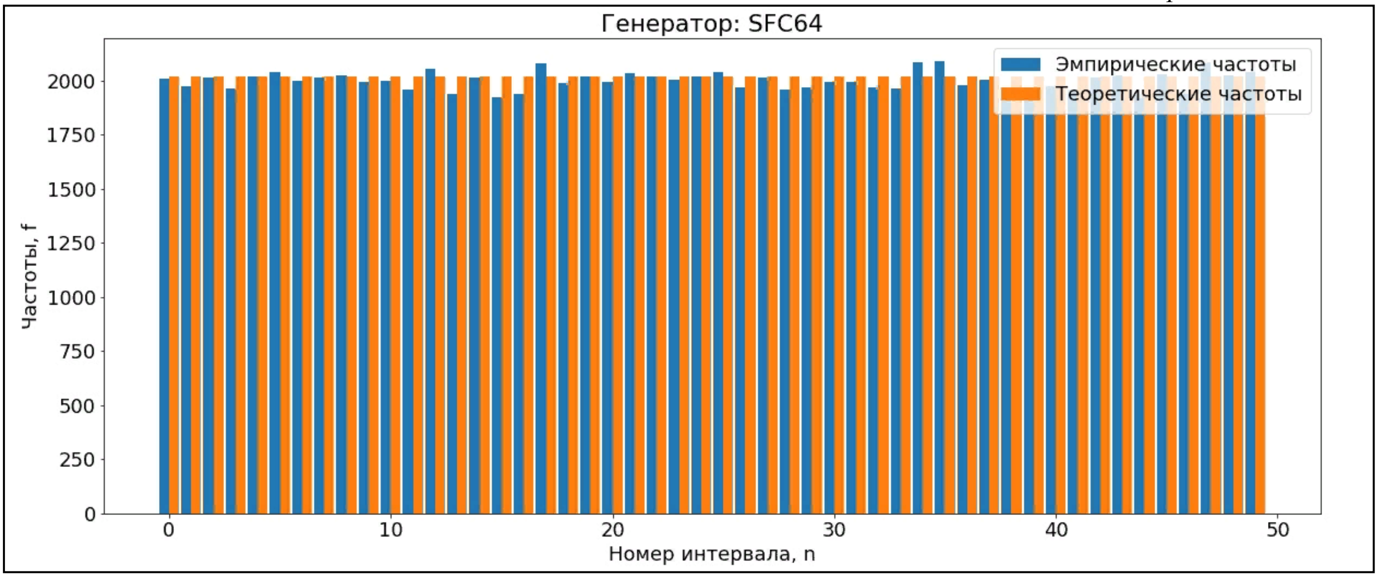


Рисунок 3. Гистограмма частот для Philox, $\chi^2_{набл} = 46.0 \leq \chi^2_{кр}$

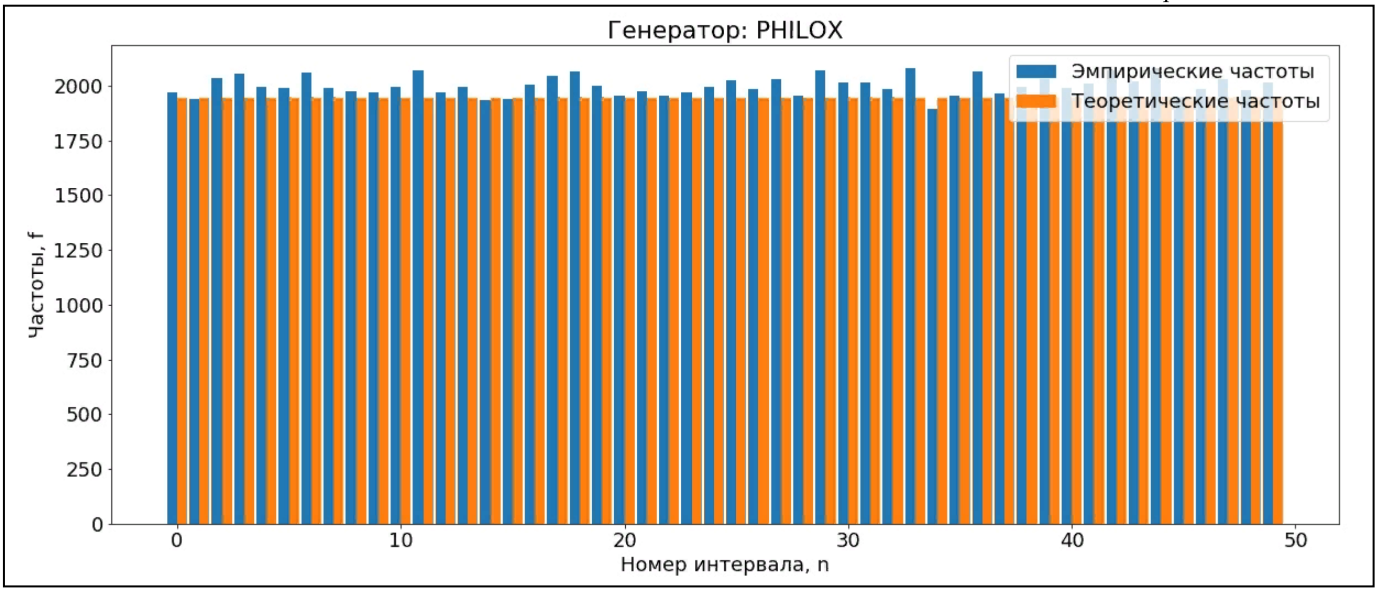
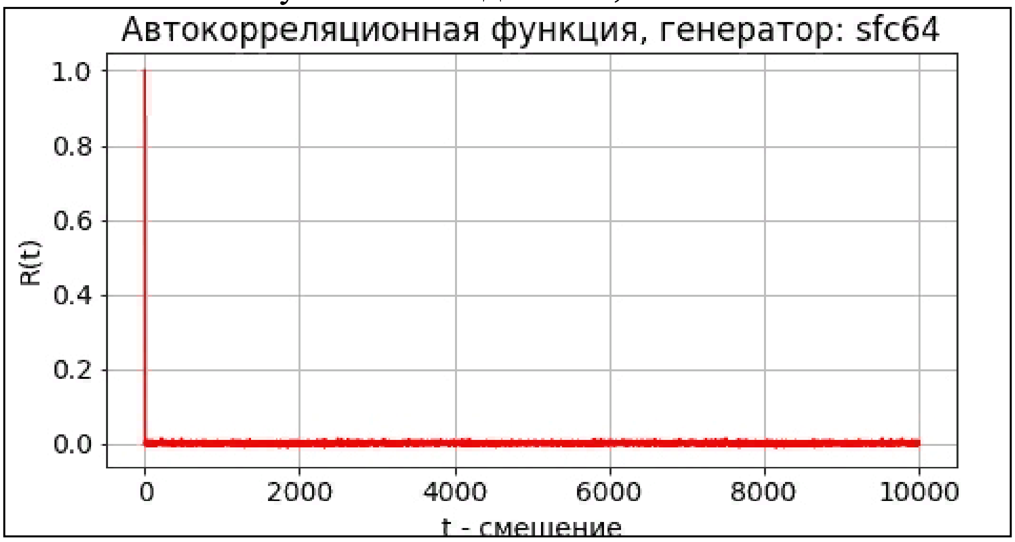


Рисунок 4. ACF для PCG64, N=100000

Рисунок 5. ACF для Philox, N=100000

Рисунок 5. ACF для SFC, N=100000



ВЫВОД

В ходе работы были изучены методы тестирования качества работы генератора псевдослучайных чисел: критерий Пирсона и автокорреляция. С помощью этих методов были протестированы три генератора псевдослучайных чисел: PCG64, SFC64, Philox. Реализация данных алгоритмов поставляется математической библиотекой NumPy. Для наглядности результаты экспериментов были визуализированы в виде частотных гистограмм и графиков автокорреляционной функции. Генератор SFC64 имеет наименьшее $\chi^2_{\text{набл}}$ из всех имеющихся, это хорошо видно из гистограмм – среднее отклонение эмпирических частот от теоретических меньше, чем у остальных генераторов, следовательно, данный генератор в проведенном эксперименте выдал более равномерное распределение, чем другие генераторы. Значение $\chi^2_{\text{набл}}$, используемое в критерии Пирсона для всех генераторов при уровне значимости 0.8 не превышало критического значения $\chi^2_{\text{кр}}$, следовательно, гипотезу о равномерном распределении нельзя отвергнуть. Автокорреляционная функция во всех случаях колеблется около 0 с очень малой окрестностью, т. е. принимает как положительные значения, так и отрицательные близкие нулю значения, поэтому статистическая взаимосвязь между исходной и сдвинутой последовательностью пренебрежимо мала.

ПРИЛОЖЕНИЕ

Листинг

```
import numpy as np
import pandas as pd
import scipy.stats as st
import matplotlib.pyplot as plt
def segmentation(xi, X):
    k = len(xi) - 1

    frequencies = {i: 0 for i in range(k)}
    for i in np.searchsorted(xi[1:], X):
        frequencies[i] = frequencies[i] + 1

    df = pd.DataFrame({
        'a': xi[:-1],
        'b': xi[1:],
        'n': list(frequencies.values())
    })

    return df

def chi_squared_test(df):
    k = df.n.size
    N = np.sum(df.n)
    si = np.power(df.n, 2) * k
    return np.sum(si) / N – N

def autocorrelation(row, offset):
    mean, var, N = np.mean(row), np.var(row), len(row)

    r_numerator = np.sum(
        [(row[idx] - mean) * \
         (row[(idx + offset) % N] - mean) \
         for idx, xi in enumerate(row[:offset])]
    )

    r_denumerator = var * (N – offset)
    return r_numerator / r_denumerator

def expected_data(df):
    dx = (df.iloc[0].b - df.iloc[0].a) / 2
    z = [x + dx for x in df.a]

    mean = np.average(z, weights=df.n)
    var = np.average((z - mean) ** 2, weights=df.n)
    std = np.sqrt(var)

    N = np.sum(df.n)
    a_t = mean - np.sqrt(3) * std
    b_t = mean + np.sqrt(3) * std
    vp = [(v.b - v.a) / b_t - a_t  for i, v in df.iterrows()]
    n = [int(N * pi) for pi in vp]

    return pd.DataFrame({'a': df.a, 'b': df.b, 'n': n })
```