УДК 519.6

А. В. Гусаров
кандидат технических наук, доцент
ФГБОУ ВО Рыбинский государственный авиационный технический университет имени П. А. Соловьева, г. Рыбинск Ярославской обл., Россия

К ВОПРОСУ ПРИМЕНЕНИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ ДЛЯ ИССЛЕДОВАНИЯ ГЕНЕРАТОРОВ ПСЕВДОСЛУЧАЙНЫХ ДВОИЧНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Аннотация. Одной из проблем, возникающих при исследовании генераторов псевдослучайных последовательностей, является определение показателей качества генерируемых последовательностей. Цель работы – показать, как можно определять показатели качества программных генераторов псевдослучайных последовательностей. Результаты работы проходят апробацию в процессе обучения и применяются в научных исследованиях.

Ключевые слова: псевдослучайные двоичные последовательности; математическая статистика; генератор ANSI X9.17.

DOI: 10.25206/2307-5430-2020-8-111-115

Псевдослучайные двоичные последовательности применяются в различных областях науки и техники. Для их генерации используются различные программные генераторы, одним из которых является генератор *ANSII X*9.17 [1, 2]. Однако возникает вопрос: насколько «хорош» этот генератор, т. е. каков процент качественных последовательностей из всех тех, которые он генерирует?

На первый взгляд ответить на этот вопрос просто, т. к. существуют многочисленные тесты [3 – 5], позволяющие определить, являются ли оцениваемые последовательности случайными. Однако как оценить, каков процент качественных последовательностей, формируемых генератором, в этих и других многочисленных источниках, включая классические работы по математической статистике, не указывается. Поэтому предлагается способ оценки генератора двоичных последовательностей, основанный на выполнении «классических» расчетов, использующих методы математической статистики.

Для наглядности будем генерировать каждый раз по 100 двоичных последовательностей в соответствии с американским стандартом *ANSII X*9.17. Для того, чтобы обеспечить статистическую правдоподобность, будем каждый раз обрабатывать по 30 групп последовательностей, каждая из которых включает в

себя 100 сгенерированных последовательностей.

Для этого сгенерируем 30 подгрупп SG_j последовательностей по 100 последовательностей в группе. Длина последовательности – 64 бита.

Полученные 30 подгрупп последовательностей по 100 последовательностей в группе назовем макроподгруппой. Всего будем формировать 50 макрогрупп.

Далее при помощи различных статистических тестов [6] определим количество неотбракованных двоичных последовательностей в каждой подгруппе. Индекс i соответствует номеру подгруппы ($i = 1 \div 30$), индекс j соответствует номеру макрогруппы ($j = 1 \div 50$). Двоичная последовательность будет считаться неотбракованной, если после прохождения статистических тестов она признается случайной.

Для макрогруппы рассчитаем среднее выборочное $\overline{X_j}$ по формуле

$$\overline{X_j} = \frac{\sum_{i=1}^{n} x_{ij}}{n}, \tag{1}$$

где n – объем макрогруппы (т. е. количество подгрупп из 100 последовательностей), $n = 30$; x_{ij} – количество неотбракованных последовательностей в одной выборке из $N = 100$ значений.

Процесс повторим 50 раз – для получения нужного количества точек на графике, хотя достаточно было и 30 точек. В результате получим ряд значений $\overline{X_{ij}}$ из 50 элементов (рис. 1).

Кроме того, эти значения используем для вычисления среднего значения SR_{ANSI} средних выборочных $\overline{X_j}$ и среднего квадратичного отклонения SKV_{ANSI} средних выборочных $\overline{X_j}$ макрогрупп последовательностей, полученных ранее:

$$SR_{ANSI} = \frac{\sum_{j=1}^{m} \overline{X_j}}{m}, \tag{2}$$

$$SKV_{ANSI} = \frac{\sum_{j=1}^{m} \left(\overline{X_j} - SR_{ANSI}\right)^2}{m-1}, \tag{3}$$

где SR_{ANSI} – среднее значение для генератора *ANSII X*9.17, полученное из значений средних выборочных $\overline{X_j}$ для каждой макрогруппы; SKV_{ANSI} – среднее квадратичное отклонение для генератора *ANSII X*9.17, полученное из значений средних

выборочных $\overline{X_j}$ для каждой макрогруппы; m – количество макрогрупп, m = 50; $\overline{X_j}$ – среднее выборочное для макрогруппы, расчет по формуле (1).

Рис. 1. График изменения среднего выборочного для неотбракованных последовательностей длиной 64 бита для генератора *ANSI X*9.17

Среднее значение SR_{ANSI} средних выборочных $\overline{X_j}$ (в процентах) составляет 88,42±0,46.

Зададимся доверительной вероятностью γ = 0,95, тогда в соответствии с таблицей значений функции Лапласа коэффициент доверия t_γ = 1,96. Точность оценки δ определим по известной формуле

$$\delta = \frac{t_\gamma \cdot \sigma}{\sqrt{m}}, \qquad (4)$$

где t_γ – коэффициент доверия, t_γ = 1,96; σ – среднее квадратичное отклонение, σ = SKV_{ANSI} = 0,46; m – объем выборки (количество макрогрупп), m = 50;

Подставив в (4) значения, получим, что δ = 0,13. Границы доверительного интервала будут следующие:

- нижняя: 88,42 – 0,13 = 88,29;
- верхняя: 88,42 + 0,13 = 88,55.

Таким образом, с вероятностью γ = 0,95 истинное значение SR_{ANSI} средних выборочных будет находиться в интервале от 88,29 до 88,55. Поэтому можно предположить, что не менее 88 % сгенерированных последовательностей будут признаны качественными, остальные – будут отбракованы.

Гистограмма, отражающая распределение количества сгенерированных качественных последовательностей, приведена на рис. 2. Границы интервалов приведены в табл. 1.

Таблица 1

Границы интервалов

№ интервала	1	2	3	4	5	6
Диапазон	[78;	[79,6;	[81,2;	[82,8;	[84,4: 86)	[86; 87,6)

	79,6)	81,2)	82,8)	84,4)		
№ интервала	7	8	9	10	11	12
Диапазон	[87,6; 89,2)	[89,2; 90,8)	[90,8; 92,4)	[92,4; 94)	[94; 95,6)	[95,6; 97,2]

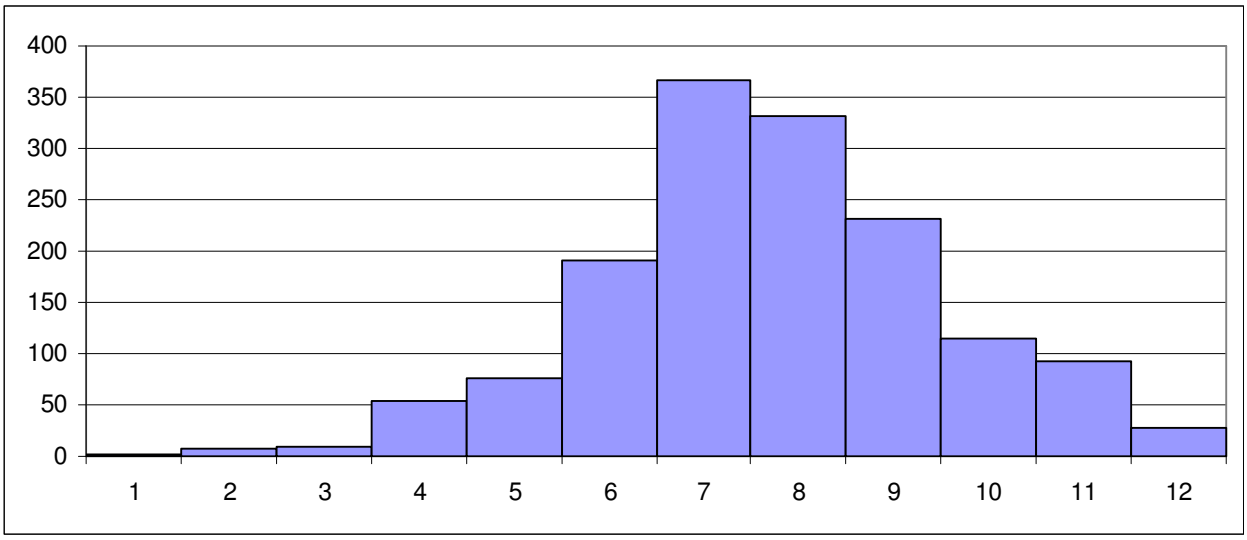

Рис. 2. Гистограмма, отражающая распределение количества сгенерированных качественных последовательностей длиной 64 бита для генератора *ANSI X*9.17

На основании вида гистограммы можно сделать вывод о нормальном законе распределения количества качественных последовательностей, сгенерированных генератором в соответствии с американским стандартом *ANSII X*9.17.

Одна из проблем, возникших в процессе работы, заключалась в том, что вычисления проводились по результатам «косвенных измерений», т. е. сгенерированная псевдослучайная двоичная последовательность сначала тестировалась, а результаты тестирования далее были обработаны статистическими методами.

В процессе получения результатов были сгенерированы 150000 псевдослучайных двоичных последовательностей, что позволило получить статистически значимый результат.

Данный способ оценки генератора псевдослучайной двоичной последовательности использовался для оценки качества генераторов, исследуемых магистрантами РГАТУ имени П. А. Соловьева во время работы в СКБ кафедры вычислительных систем.

Библиографический список

1. Генератор псевдослучайных чисел *ANSI X*9.17. [Электронный ресурс]: Электрон. текстовые, граф. дан. – *URL*:

https://studopedia.su/14_16630_generator-psevdosluchaynih-chisel-ANSI-X.html (дата обращения 02.09.2020).

2. *Triple DES* [Электронный ресурс]: Электрон. текстовые, граф. дан. – *URL*: https://ru.bmstu.wiki/Triple_DES (дата обращения: 02.09.2020).

3. Будько М. Б., Будько М. Ю., Гирик А. В., Грозов В. А. Методы генерации и тестирования случайных последовательностей – СПб: Университет ИТМО, 2019. – 70 с.

4. Тестирование случайных последовательностей [Электронный ресурс]: Электрон. текстовые, граф. дан. – *URL*: *https://ru.wikipedia.org/wiki/ Тестирование_случайных_последовательностей* (дата обращения 30.11.2019).

5. Статистические тесты *NIST* [Электронный ресурс]: Электрон. текстовые, граф. дан. – *URL*: *https://wiki2.org/ru/Cтатистические_тесты_NIST* (дата обращения: 20.12.2019).

6. Гусаров А. В., Хачикова Н. А. Анализ программных генераторов псевдослучайных последовательностей. Актуальные проблемы преподавания математических и естественно-научных дисциплин в образовательных организациях высшего образования. Материалы Всероссийской научно-методической конференции. – Кострома: Военная академия радиационной, химической и биологической защиты имени Маршала Советского Союза С. К. Тимошенко, 2020. – с. 140 – 148.

Сведения об авторах:

Александр Вячеславович Гусаров

E-mail alvgus@mail.ru; spin-code: 6260-2229.

Научные интересы: автоматизация технологических процессов, прикладная математика, информационная безопасность.