ТЕСТИРОВАНИЕ СТАТИСТИЧЕСКИХ СВОЙСТВ ГЕНЕРАТОРОВ ПСЕВДОСЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Ташев Комил Ахматович
Ташкентский университет информационных технологий,
Проректор по научной работе и инновациям
E-mail: k.tashev@tuit.uz
Даминов Акмальбек Абдурасулович
Ташкентский университет информационных технологий, Магистр
E-mail: akmalbekdaminov1998@gmail.com

Аннотатция: *Генераторы псевдослучайных последовательностей могут использоваться для выполнения различных функций. Помимо прочего, они могут использоваться для решения важных задач, в том числе – при защите данных. Поэтому необходима надежная проверка свойств генераторов, в этой статье мы рассмотрели близость выходной последовательности к истинно случайной с точки зрения ее статистических свойств и непредсказуемость выходных значений.*
Ключевые слова: *ГПСП, ПСП, ИСП, гипотеза, Статистическое, NIST.*

TESTING THE STATISTICAL PROPERTIES OF PSEUDORANDOM SEQUENCE GENERATORS

Abstract: *Pseudorandom sequence generators can be used to perform various functions. Among other things, they can be used to solve important tasks, including data protection. Therefore, a reliable verification of the properties of generators is necessary, in this article we considered the proximity of the output sequence to a truly random one in terms of its statistical properties and the unpredictability of the output values.*
Key words: *PRSG, PRS, ISS, hypothesis, statistical, NIST.*

Существует два основных направления анализа псевдослучайных последовательностей:

– Криптографическое. Цель этого направления – поиск таких закономерностей исследуемой последовательности, чтобы по ее части можно было восстановить всю последовательность целиком.

– Статистическое. Это направление ориентировано на поиск отклонений статистических свойств псевдослучайной последовательности, на основе которых можно предсказывать последующие и предыдущие значения членов последовательности с вероятностью, большей 0,5.[1]

Таксономия тестов случайных и псевдослучайных чисел очень обширна. Из самых обобщенных соображений можно выделить два класса критериев: эмпирические критерии, при использовании которых вычисляются некоторые

статистики от групп чисел ПСП; и теоретические критерии, для которых анализ последовательности чисел производится теоретико-числовыми методами, над рекуррентными правилами, образующими ПСЧ.

Критерии, как правило, применяются к последовательности независимых друг от друга действительных равномерно распределенных чисел из интервала [0,1]: $X^n = (x_1, x_2, \dots, x_n)$. Некоторые из критериев предназначены для целочисленных последовательностей. В этом случае будем использовать вспомогательную последовательность $Y^n = (d\lfloor x_1 \rfloor, d\lfloor x_2 \rfloor, \dots, d\lfloor n \rfloor)$.[2]

Статистические тесты – это эмпирические тесты для оценки качества ГПСП и выявления их «слабых мест» путем расчета статистических характеристик ПСП и сравнения их с аналогичными характеристиками ИСП. Статистические тесты соединяют в себе вычислительные процедуры для нахождения статистики исследуемой последовательности и решающее правило проверки, с помощью которого по значениям статистики определяют, принять или отвергнуть нулевую гипотезу:

– если выборочное значение статистики принадлежит критической области, то нулевая гипотеза H_0 отвергается, так как при однократном испытании произошло событие, вероятность которого мала и равна α;

– если выборочное значение статистики попадает в допустимую область, то делается вывод, что данные испытания не противоречат выдвинутой

нулевой гипотезе H_0 и она принимается. Таким образом, алгоритм проверки статистических гипотез состоит изследующих шагов (рис.1):

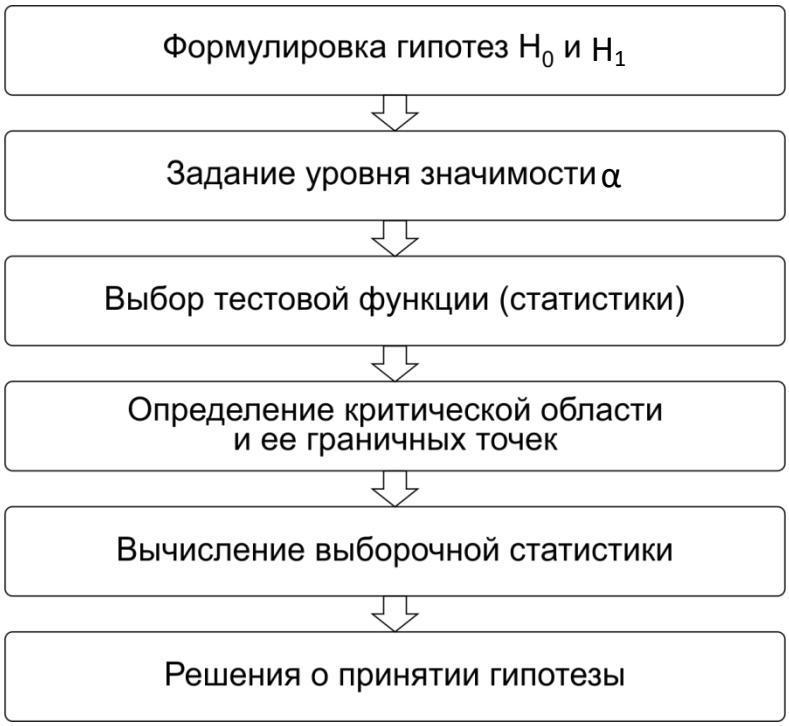

Рисунок 1 – Алгоритм проверки статистических гипотез.

Для того чтобы сделать вывод о прохождении теста, проверка этих гипотез

выполняется с помощью различных статистических критериев – правил, в соответствии с которыми принимается или отклоняется нулевая гипотеза. Ниже перечислены различные типы таких критериев в порядке нарастания их надежности:

– Пороговое значение. Величина вычисленной статистики сравнивается с некоторым пороговым значением, и, если статистика, например, превосходит его, тест считается пройденным.

– Доверительный интервал. Тест считается пройденным, если величина статистики теста попадает в определенный доверительный интервал, зависящий от принятого уровня значимости.

– Вероятностный подход. Набор значений статистики теста считается набором значений случайной величины с заданным законом распределения.

Последний вариант зарекомендовал себя наиболее эффективным и надежным. Именно он используется во многих пакетах статистических тестов.

При принятии решения о том, был ли пройден тест, возможны два типа ошибок. Возникновение ошибки первого рода означает, что тестируемая псевдослучайная последовательность на самом деле является случайной, но верная нулевая гипотеза H_0 отклоняется. Вероятность такой ошибки равна уровнюзначимости α, который задается до начала тестирования. Уровень значимости α – это вероятность того, что тестирование покажет неслучайность последовательности, тогда как фактически она является случайной. Соответственно, вероятность принятия правильного решения составляет (1–α). Обычно в практических задачах приемлемым считается значение уровня значимости 0,05. Однако для целей криптографии используют более строгие значения α (как правило, из интервала [0,001; 0,01]).[3]

Ошибка второго рода (β) означает принятие гипотезы о случайности рассматриваемой последовательности, когда последовательность в действительности неслучайна. С точки зрения криптографии такая ошибка более критична. Величина ошибки второго рода определяет мощность критерия – вероятность того, что нулевая гипотеза будет отклонена при верной альтернативной гипотезе. Между двумя этими видами ошибок существует взаимозависимость: чем меньше α, тем больше β, и наоборот.

Статистика теста построена так, что ее меньшие значения соответствуют дефектам псевдослучайной последовательности – отклонениям от истинной случайности.

Обычно для удобства восприятия результатов тестирования вычисленная с помощью эталонного распределения вероятностей тестовая статистика преобразуется в так называемое значение *p-value.* Это значение трактуется как вероятность того, при заданном уровне значимости идеальный генератор случайных последовательностей может произвести последовательность, менее случайную, чем исследуемая. Такое событие тем менее вероятно, чем меньше значение *p-value*. Выполнение условия *p-value*>=α означает успешное прохождение теста.

Перечислим несколько наиболее известных инструментов статистического тестирования псевдослучайны последовательностей [3]:

- подборка Кендалла и Бабингтон-Смита;
- тесты Д. Кнута;
- тесты DIEHARD (Дж. Марсалья);
- пакет тестов NIST (A. Rukhin и др.);
- пакет TestU01 (П. Л'Экуйе);
- Crypt-XS (Helen Gustafson);
- John Walker (Autodesk, Inc.), ENT;
- Dieharder (Robert G. Brown).

Следует понимать, что тестирование не может заменить криптоанализ. Тем не менее, оно является обязательным этапом анализа стойкости криптографического генератора. В условиях существования большого числа различных статистических тестов, как широко и давно распространенных, так и новых, важен обоснованный выбор, связанный со спецификой решаемых задач защиты информации.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ:

1. Слеповичев И.И. Генераторы псевдослучайных чисел. – Саратов: СГУ, 2017.– 100 с.
2. Смарт Н. Криптография. – М.: Техносфера, 2005. – 528 с.
3. М.Б. Будько. А.В. Гирик . В.А. Грозов. М.Ю. Будько. Методы генерации и тестированияслучайных последовательностей. – Санкт-Петербург, 2019. 45с