DOI: 10.37791/2687-0649-2025-20-4-69-81

Статистический g-тест для генераторов случайных чисел

С.Т. Афанасьев[1]***, И.В. Нечта***[1*]
[1]*Сибирский государственный университет телекоммуникаций и информатики, Новосибирск, Россия*
[*]*ivannechta@gmail.com*

Аннотация. Генераторы случайных чисел порождают равномерное распределение чисел, причем между этими числами нет какой-либо зависимости, позволяющей предсказывать порождаемое число лучше, чем случайное угадывание. Такие генераторы применяются в различных задачах, например моделирование или обеспечение информационной безопасности. В качестве генераторов используются аппаратные устройства, преобразующие случайный физический процесс в поток данных. Другим типом генераторов являются программные, которые используют формулу или алгоритм для получения нового числа. Считается, что программные хоть и быстрее работают, но не обладают доказанным свойством случайности. Известен ряд случаев, когда в таких генераторах обнаруживались закономерности, что приводило к отказу от их использования. Для проверки генератора существуют наборы тестов. При успешном прохождении всех тестов генератор рекомендуют к использованию. Целью данной статьи является разработка алгоритма проверки битовых последовательностей на случайность. В настоящей работе предложен новый тест для генераторов случайных чисел, который может быть включен в общий набор статистических тестов, необходимых для проверки. В отличие от известных ранее тестов, базирующихся на энтропийном подходе, новый выявляет автокорреляцию сгенерированных данных. Показано, что предложенный алгоритм позволяет обнаружить отклонения от случайности у генераторов, которые ранее проходили известные статистические тесты. В ходе экспериментов тестировались выходные последовательности шифра, хэш-функции и некоторых генераторов псевдослучайных чисел. Как показал анализ результатов тестирования, некоторые генераторы имеют автокорреляцию в порождаемых данных. В данной работе не делается заявлений относительно уязвимостей шифров, а только ранжируются генераторы относительно друг друга. Идея теста заключается в том, что у любого генератора существует период, когда в последовательности порождены все числа из генерируемого множества (алфавита). В идеальном случае период будет приближаться к размеру алфавита или будет незначительно больше. Однако если некоторые символы повторяются чаще, то период станет значимо возрастать. Другими словами, удлинение периода может свидетельствовать либо об отклонении от равномерного распределения, либо о наличии автокорреляции в порождаемых данных. Последнее является областью интереса авторов статьи. В ходе исследования использовались элементы теории вероятностей и математической статистики. Эксперименты проведены на большом объеме сгенерированной последовательности чисел.

Ключевые слова: генераторы случайных чисел, статистические тесты на случайность, криптография, тесты NIST, равномерное распределение

Для цитирования: *Афанасьев С.Т., Нечта И.В.* Статистический g-тест для генераторов случайных чисел // Прикладная информатика. 2025. Т. 20. № 4. С. 69–81. DOI: 10.37791/2687-0649-2025-20-4-69-81

Statistical g-test for random number generators

S. Afanasyev[1], ***I. Nechta***[1*]
[1]*Siberian State University of Telecommunications and Information Science, Novosibirsk, Russia*
[*]*ivannechta@gmail.com*

Abstract. Random number generators produce numbers with the uniform distribution, and there is no correlation between these numbers that would allow to predict the generated number better than random guessing. Such generators are used in various tasks, such as modeling or information security. Hardware devices that transform a random physical process into a data stream are used as generators. Another type of generators are software generators that use a formula or algorithm to obtain a new number. It is believed that software generators, although they work faster, but do not have a proven property of randomness. There are a number of known cases when patterns were found in such generators, which led to the refusal to use them. There are test sets for checking the generator exists. If all tests are successfully passed, the generator is recommended for use. The purpose of this article is to develop an algorithm for testing bit sequences for randomness. In this paper, a new algorithm for testing random number generators is proposed, that can be included in the general set of statistical tests required for verification. Unlike previously known tests based on the entropy approach, the new test uses autocorrelation of the generated data. It is shown that the proposed test allows one to identify deviations from randomness in generators that have previously passed known statistical tests. During the experiments, the output sequences of the cipher, hash function and some pseudo-random number generators were tested. As the analysis of the test results showed, some generators have autocorrelation in the generated data. The idea of the test is that any generator has a period when all numbers from the generated set (alphabet) are appeared in the sequence. In the ideal case, the period will be close to the size of the alphabet or will be slightly larger. However, if some symbols are repeated more often, the period will increase significantly. In other words, an increase in the period may indicate either a deviation from a uniform distribution or the presence of autocorrelation in the generated data. The latter phenomenon is the area of interest this article. In this research we use elements of probability theory and mathematical statistics. The experiments were conducted on a large volume of the generated sequence of numbers.

Keywords: random number generators, statistical randomness tests, cryptography, NIST tests, uniform distribution

For citation: Afanasyev S., Nechta I. Statistical g-test for random number generators. *Prikladnaya informatika*=Journal of Applied Informatics, 2025, vol.20, no.4, pp.69-81 (in Russian). DOI: 10.37791/2687-0649-2025-20-4-69-81

Введение

Генераторы случайных чисел находят свое применение в различных областях науки и промышленности. Так, например, для моделирования требуются большие объемы данных, обладающих определенными статистическими свойствами – заданным распределением вероятности символов. Частным случаем является метод моделирования Монте-Карло. Существует множество инструментальных средств, решающих поставленную задачу, но часто для их запуска требуется подать на вход некоторый источник слу-