Algorithms of Information Security: Key generation algorithms. Tutorial.

Olha Jurečková, Martin Jureček
{jurecolh,jurecmar}@fit.cvut.cz

Faculty of Information Technology
Czech Technical University in Prague

September 29, 2022

Pseudorandom bit generator (PRBG)

Definition

A *pseudorandom bit generator (PRBG)* is a deterministic algorithm which, given a truly random sequence of length k, outputs a binary sequence of length $l \geq k$ which "appears" to be random. The input to the PRBG is called the *seed*, while the output of the PRBG is called a *pseudorandom bit sequence*.

Hypothesis testing

A *statistical hypothesis,* denoted H_0, is an assertion about the distribution of one or more random variables.

A *test* of a statistical hypothesis is a procedure based upon observed values of the random variables that leads to the acceptance or rejection of the hypothesis H_0.

When we test hypotheses, we always compare two hypotheses. One of them we're testing, we called the *null hypothesis* H_0, and in contrast to it, we build the so-called *alternative hypothesis* H_A.

The test only provides a measure of the strength of the evidence provided by the data against the hypothesis; hence, the conclusion of the test is not defined, but rather probabilistic.

Hypothesis testing

Let $X = (X_1, \ldots, X_n)$ be a random sequence from some distribution $R(\theta)$, where θ is a parameter that can be multidimensional. Let $h(\theta)$ be a parametric function and k is a real constant. Then the null hypothesis be in the following form:

$$H_0 : h(\theta) = k.$$

The alternative hypothesis can be defined in the following three ways:

- Right-tailed alternative hypothesis: $H_A : h(\theta) > k$
- Left-tailed alternative hypothesis: $H_A : h(\theta) < k$
- Two-tailed alternative hypothesis: $H_A : h(\theta) \neq k$.

Hypothesis testing

If the result of the test corresponds with reality, then a correct decision has been made. However, if the result of the test does not correspond with reality, then an error has occurred.

There are two situations in which the decision is wrong. The null hypothesis may be true, whereas we reject H_0. On the other hand, the alternative hypothesis H_A may be true, whereas we do not reject H_0.

Two types of error are distinguished: type I error and type II error.

Table of error types

Reality	**Decision about null hypothesis (H_0)**	
	Do not reject H_0	Reject H_0
H_0 is true	true positive	type I error
H_0 is not true	type II error	true negative

Chi-Square Goodness-of-Fit Test

The χ^2 (chi-square) goodness of fit test is used to compare the observed distribution to an expected distribution in a situation where we have two or more categories in discrete data. In other words, it compares multiple observed proportions to expected probabilities. The chi-square test is defined for the hypothesis:

- H_0 : There is no significant difference between the observed and the expected value.
- H_A : There is a significant difference between the observed and the expected value.

Chi-Square Goodness-of-Fit Test

For the chi-square goodness-of-fit computation, the value of the test-statistic is

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = N \sum_{i=1}^{n} \frac{(\frac{O_i}{N} - p_i)^2}{p_i}$$

where

- O_i the number of observations of type i,
- N is the total number of observations,
- $E_i = N \cdot p_i$ the expected (theoretical) count of type i, asserted by the null hypothesis that the fraction of type i in the population is p_i and n is the number of cells in the table.

Chi-Square Goodness-of-Fit Test

The test can be used provided that all values of $N \cdot p_i$ are at least 5.

The random variable X^2 (see the previous slide) has an approximate χ^2 distribution of $n - 1$ degrees of freedom.

We reject the null hypothesis at the level of significance α, if $X^2 > {\chi^2}_{1-\alpha;n-1}$, where the value of ${\chi^2}_{1-\alpha;n-1}$ is a quantile χ^2 distribution of $n - 1$ degrees of freedom.

The chi-squared statistic can then be used to calculate a p-value by comparing the value of the statistic to a chi-squared distribution. The number of degrees of freedom is the number of values in the final calculation of a statistic that are free to vary.

Chi-Square Goodness-of-Fit Test

Example. Suppose a gambler rolls the dice 300 times with the following observed counts:

Value	1	2	3	4	5	6
Frequency	40	55	54	49	46	59

Is the die fair?

Solution. This problem can be set up as a goodness-of-fit problem. Let's first define the null and alternative hypotheses:

H_0: the die is fair

H_A: the die is not fair

If the cube is regular, then the probability of each value is $\frac{1}{6}$. That is, the expected frequencies of individual values are the same and equal to 50. The measured frequencies are $(X_1, X_2, X_3, X_4, X_5, X_6) = (40, 55, 51, 49, 46, 59)$. Next, we apply the goodness-of-fit test and calculate the value:

$$X^2 = \sum_{i=1}^{n} \frac{(X_i - mp_i)^2}{mp_i} = \sum_{i=1}^{6} \frac{(X_i - 50)^2}{50} =$$

$$= \frac{(40-50)^2}{50} + \frac{(55-50)^2}{50} + \frac{(51-50)^2}{50} + \frac{(49-50)^2}{50} +$$

$$+ \frac{(46-50)^2}{50} + \frac{(59-50)^2}{50} = 4.48$$

We choose the level of significance $\alpha = 0.05$, and we consider $n - 1 = 5$ degrees of freedom.
The value of the quantile $\chi^2_{1-\alpha;\, n-1}$ is $\chi^2_{0.95;\, 5} = 11.071$.
We have $X^2 < 11.071$. Therefore, we can not reject the null hypothesis H_0 at the given significance level.

Golomb's randomness postulates

Golomb's randomness postulates were one of the first attempts to establish some necessary conditions for a periodic pseudorandom sequence to look random.

Definition

Let $s = s_0, s_1, s_2, \ldots$ be an infinite sequence. The subsequence consisting of the first n terms of s is denoted by $s^n = s_0, s_1, s_2, \ldots, s_{n-1}$.

Definition

The sequence $s = s_0, s_1, s_2, \ldots$ is said to be *N-periodic* if $s_i = s_{i+N}$ for all $i \geq 0$. The sequence s is *periodic* if it is N-periodic for same positive integer N. The *period* of a periodic sequence s is the smallest positive integer N for which s is N-periodic. If s is a periodic sequence of period N, then the *cycle* of s is the subsequence s^N.

Definition

Let s be a sequence. A *run* of s is a subsequence of s consisting of consecutive 0's or consecutive 1's, which is neither preceded nor succeeded by the same symbol. A run of 0's is called a *gap*, while a run of 1's is called a *block*.

Definition

Let $s = s_0, s_1, s_2, \ldots$ be a periodic sequence of period N. The *autocorrelation function* of s is the integer-valued function $C(t)$ defined as

$$C(t) = \frac{1}{N} \sum_{i=0}^{N-1} (2s_i - 1)(2s_{i+t} - 1), \text{for } 0 \leq t \leq N - 1.$$

The autocorrelation function $C(t)$ measure the amount of similarity between the sequence s and a shift of s by t positions.

Definition

Let s be a periodic sequence of period N. *Golomb's randomness postulates* are the following.

- R1: In the cycle s^N of s, the number of 1's differs from the number of 0's by at most 1.
- R2: In the cycle, s^N, at least $\frac{1}{2}$ the runs have length 1, at least $\frac{1}{4}$ have length 2, at least $\frac{1}{8}$ have length 3, etc., as long as the number of runs so indicated exceeds 1. Moreover, for each of these lengths, there are (almost) equally many gaps and blocks.
- R3: The autocorrelation function $C(t)$ is two-valued. That is for some integer K,

$$N \cdot C(t) = \sum_{i=0}^{N-1} (2s_i - 1)(2s_{i+t} - 1) = \begin{cases} N, & \text{if } t = 0, \\ K, & \text{if } 1 \leq t \leq N-1 \end{cases}$$

Example 1. Let $s = 0, 1, 1, 0, 1, 1, 0, 1, 1, \ldots$ be a sequence. Find:

- period and a cycle of this sequence
- run and gap
- $C(0), C(1), C(2)$ a $C(3)$.

Solution.

1. $s = 0, 1, 1, 0, 1, 1, 0, 1, 1, \ldots$ is the sequence, then $s_i = s_{i+3} = 0, 1, 1$. Therefore, s is the periodic sequence of period 3, and cycle of the sequence s is the subsequence $s^3 = 0, 1, 1$.

2. Let $s = 0, 1, 1, 0, 1, 1, 0, 1, 1, \ldots$ be a sequence. Then "0" is the run of the length 1 (i.e., gap), and "1,1" is the run of the length 2 (i.e., block).

3. Then period N is 3, i.e., $N = 3$.

$$C(t) = \frac{1}{N}\sum_{i=0}^{N-1}(2s_i - 1)(2s_{i+t} - 1), \text{for } 0 \leq t \leq N-1.$$

Then

$$C(0) = \frac{1}{3}\sum_{i=0}^{2}(2s_i - 1)(2s_{i+0} - 1) =$$

$$= \frac{1}{3}((2s_0 - 1)(2s_0 - 1) + (2s_1 - 1)(2s_1 - 1) + (2s_2 - 1)(2s_2 - 1)).$$

$$C(0) = \frac{1}{3}((2\cdot 0-1)(2\cdot 0-1)+(2\cdot 1-1)(2\cdot 1-1)+(2\cdot 1-1)(2\cdot 1-1)) =$$

$$= \frac{1}{3}(1+1+1) = 1.$$

$$C(1) = \frac{1}{3}((2\cdot 0-1)(2\cdot 1-1)+(2\cdot 1-1)(2\cdot 1-1)+(2\cdot 1-1)(2\cdot 0-1)) =$$

$$= \frac{1}{3}(-1+1-1) = -\frac{1}{3}.$$

$$C(2) = \frac{1}{3}((2\cdot 0-1)(2\cdot 1-1)+(2\cdot 1-1)(2\cdot 0-1)+(2\cdot 1-1)(2\cdot 1-1)) =$$

$$= \frac{1}{3}(-1-1+1) = -\frac{1}{3}.$$

$$C(3) = \frac{1}{3}((2\cdot 0-1)(2\cdot 0-1)+(2\cdot 1-1)(2\cdot 1-1)+(2\cdot 1-1)(2\cdot 1-1)) =$$

$$= \frac{1}{3}(1+1+1) = 1.$$

Example 2. Let $s = 1, 1, 0, 0, 1, 1, 0, 0, \ldots$ be a sequence. Find:

- period and a cycle of this sequence
- $C(0), C(1)$.

Solution.

1. $s = 1, 1, 0, 0, 1, 1, 0, 0, \ldots$ is the sequence, then $s_i = s_{i+4} = 1, 1, 0, 0$. Therefore, s is the periodic sequence of period 4, and cycle of the sequence s is the subsequence $s^4 = 1, 1, 0, 0$.

2. Then period N is 4, i.e., $N = 4$.

$$C(t) = \frac{1}{N}\sum_{i=0}^{N-1}(2s_i - 1)(2s_{i+t} - 1), \text{pro } 0 \leq t \leq N-1.$$

Then

$$C(0) = \frac{1}{4}\sum_{i=0}^{3}(2s_i - 1)(2s_{i+0} - 1) =$$

$$= \frac{1}{4}((2s_0-1)(2s_0-1)+(2s_1-1)(2s_1-1)+(2s_2-1)(2s_2-1))+(2s_3-1)(2s_3-1)).$$

$$C(0) = \frac{1}{4}((2\cdot1-1)(2\cdot1-1)+(2\cdot1-1)(2\cdot1-1))+(2\cdot0-1)(2\cdot0-1)+(2\cdot0-1)(2\cdot0-1) =$$

$$= \frac{1}{4}(1+1+1+1) = 1.$$

$$C(1) = \frac{1}{4}((2\cdot1-1)(2\cdot1-1)+(2\cdot1-1)(2\cdot0-1)+(2\cdot0-1)(2\cdot0-1)+(2\cdot0-1)(2\cdot1-1)) =$$

$$= \frac{1}{4}(1-1+1-1) = 0.$$

Example 3. Let s be a sequence with period $N = 15$ and cycle

$$s^{15} = 0, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1.$$

Verify Golomb's postulates of randomness.

Solution.

- R1: The number of 0's is 7, and the number of 1's is 8. The difference between the number of 0's and 1's is 8-7=1. Therefore, R1 is satisfied.

Solution.

- R2: s^{15} has 8 runs. We have 4 runs of length 1 (2 gaps and 2 blocks): s_0, s_5, s_{13}, s_{14}. The number of gaps is 2 (s_0, s_{13}) and the number of blocks is 2 (s_5, s_{14}).
  The number of the runs of the length 2 is 2 $(s_1 s_2, s_3 s_4)$. The number of the gaps is 1 $(s_3 s_4)$, and the number of the blocks is 1 $(s_1 s_2)$.
  The number of the runs of the length 3 is 1 $(s_6 s_7 s_8)$. The number of the gaps is 1 $(s_6 s_7 s_8)$, and the number of the blocks is 0.
  The number of the runs of the length 4 is 1 $(s_9 s_{10} s_{11} s_{12})$. The number of gaps is 0, and the number of the blocks is $(s_9 s_{10} s_{11} s_{12})$.
  Total number of the runs is 4+2+1+1=8. Half the runs have the length 1, one-fourth runs have the length 2, and one-eighth runs have the length 3. Then R2 is satisfied.

- R3: Calculate $C(0)$.

$$C(0) = \frac{1}{15}\sum_{i=0}^{14}(2s_i - 1)(2s_{i+0} - 1) =$$

$$= \frac{1}{15}((2s_0 - 1)(2s_0 - 1) + (2s_1 - 1)(2s_1 - 1) + (2s_2 - 1)(2s_2 - 1) + \ldots +$$
$$+(2s_{13} - 1)(2s_{13} - 1) + (2s_{14} - 1)(2s_{14} - 1)) =$$
$$= \frac{1}{15}((2 \cdot 0 - 1)(2 \cdot 0 - 1) + (2 \cdot 1 - 1)(2 \cdot 1 - 1) +$$
$$+(2 \cdot 1 - 1)(2 \cdot 1 - 1) + (2 \cdot 0 - 1)(2 \cdot 0 - 1) +$$
$$+(2 \cdot 0 - 1)(2 \cdot 0 - 1) + (2 \cdot 1 - 1)(2 \cdot 1 - 1) +$$
$$+(2 \cdot 0 - 1)(2 \cdot 0 - 1) + (2 \cdot 0 - 1)(2 \cdot 0 - 1) +$$
$$+(2 \cdot 0 - 1)(2 \cdot 0 - 1) + (2 \cdot 1 - 1)(2 \cdot 1 - 1) +$$
$$+(2 \cdot 1 - 1)(2 \cdot 1 - 1) + (2 \cdot 1 - 1)(2 \cdot 1 - 1) +$$
$$+(2 \cdot 1 - 1)(2 \cdot 1 - 1) + (2 \cdot 0 - 1)(2 \cdot 0 - 1) +$$
$$+(2 \cdot 1 - 1)(2 \cdot 1 - 1)) =$$

$$= \frac{1}{15}(1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1) = 1.$$

$$C(0) = \frac{1}{15}(1+1+1+1+1+1+1+1+1+1+1+1+1+1+1) = 1.$$

$$C(1) = \frac{1}{15}(-1+1-1+1-1-1+1+1-1+1+1+1-1-1-1) = -\frac{1}{15}.$$

$$C(1) = C(2) = \ldots = C(14) = -\frac{1}{15}$$

$$15 \cdot C(t) = \begin{cases} 15, & t = 0 \text{ or } t = 15 \\ -1, & \text{for } 1 \leq t \leq 14 \end{cases}$$

Then R3 is satisfied.

Frequency test(monobit test)

Let $s = s_0, s_1, s_2, \ldots, s_{n-1}$ be a binary sequence of length n. The purpose of this test is to determine whether the number of 0's and 1's in s are approximately the same, as would be expected for a random sequence.

First, we count the number of 1's in the sequence s and their number denote as $n_1 = \sum_{i=0}^{n-1} s_i$. The number of 0's will be $n_0 = n - n_1$, where n is the length of the sequence s.

In a random sequence of bits of length n, the number of 1's is expected to be the same as the number of 0's and will be equal to $n_0 = n_1 = \frac{n}{2}$.

In other words, at zero hypothesis, we assume that the number of units of the sequence s is described by a random quantity with binomial distribution with parameters n and $p = \frac{1}{2}$. Next, we apply the Chi-square goodness of fit test.

Example 4. Let's we have some pseudorandom bit generator that produced the following sequence:

$$1110100111101100.$$

Test this generator with the Frequency Monobit Test.

Solution.
The tested sequence has length $n = 16$ and contains 6 zeros and 10 ones. For a random sequence, the expected number of ones is equal to the expected number of zeros, which is equal to $\frac{n}{2} = 8$. Let's define the null hypothesis H_0 and the alternative hypothesis H_A:

- H_0 : the number of ones is equal $\frac{n}{2}$
- H_A : the number of ones is not equal $\frac{n}{2}$.

Let's calculate the value

$$X^2 = \frac{(6-8)^2}{8} + \frac{(10-8)^2}{8} = 1.$$

We work at the significance level $\alpha = 0.05$ and consider 1 degree of freedom, then $\chi^2_{0.95;\,1} = 3.842$. The tested generator passed the Frequency monobit test because $X^2 = 1 < 3.842$.

Serial test(two-bit test)

Let $s = s_0, s_1, s_2, \ldots, s_{n-1}$ be a binary sequence of length n. The purpose of this test is to determine whether the number of occurrences of 00, 01, 10, and 11 as subsequences of s are approximately the same as would be expected for a random sequence.

Let n_0, n_1 denote the number of 0's and 1's in s, respectively, and let $n_{00}, n_{01}, n_{10}, n_{11}$ denote the number of occurrences of 00, 01, 10, 11 in s, respectively. Note that $n_{00} + n_{01} + n_{10} + n_{11} = (n-1)$ since the subsequences are allowed to overlap.

The statistic used is

$$X^2 = \frac{4}{(n-1)}(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2) - \frac{2}{n}(n_0^2 + n_1^2) + 1$$

which approximately follows a χ^2 distribution.

Example 5. Let's s be a sequence of length $n = 160$, which contains the following subsequence repeated 4 times:

$$1110001100010001010011101111001001001001.$$

Apply Frequency (monobit) and Serial test.

Solution.

Frequency (monobit) test.

Let's define the null hypothesis H_0 and the alternative hypothesis H_A::

- H_0 : the number of ones is equal $\frac{n}{2}$
- H_A : the number of ones is not equal $\frac{n}{2}$.

We will use the following statistics:

$$X^2 = \frac{(n_0 - \frac{n}{2})^2}{\frac{n}{2}} + \frac{(n_1 - \frac{n}{2})^2}{\frac{n}{2}} = \frac{(n_0 - n_1)^2}{n}.$$

We need to count the number of zeros and ones in a given sequence. $n_0 = 84$ and $n_1 = 76$. Then

$$X^2 = \frac{(84 - 80)^2}{80} + \frac{(76 - 80)^2}{80} = 0.4.$$

We work at a significance level of $\alpha = 0.05$ and consider 1 degrees of freedom. Quantile $\chi^2_{0.95;\,1}$ can be found in the relevant statistical table and is equal to $\chi^2_{0.95;\,1} = 3.8415$. We have $X^2 < 3.8415$. Therefore the tested sequence passed the Frequency test.

Serial test.

Next, we will look at the Serial test. We use the following notation:

- n_0 is the number of 0's in the given sequence
- n_1 is the number of 1's in the given sequence
- n_{00} is the number of subsequence 00 in the given sequence
- n_{01} is the number of subsequence 01 in the given sequence
- n_{10} is the number of subsequence 10 in the given sequence
- n_{11} is the number of subsequence 11 in the given sequence

We will use the following formula for the Serial test

$$X^2 = \frac{4}{(n-1)}(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2) - \frac{2}{n}(n_0^2 + n_1^2) + 1.$$

So we have $n_0 = 84$ and $n_1 = 76$. Then
$n_{00} = 44, n_{01} = 40, n_{10} = 40$ and $n_{11} = 35$. Then we get
$X^2 = 0.6252$.
We work at a significance level of $\alpha = 0.05$ and consider 2 degrees of freedom. Quantile $\chi^2_{0.95;\ 2} = 5.9915$.
The given sequence also passed the Serial test.

Poker test

- In the Poker test, the tested sequence $x = (x_1, \ldots, x_m)$ is divided into the sub-sequences $(x_{ni+1}, \ldots, x_{ni+n}), i = 0, 1, \ldots, \frac{m}{n} - 1$, length n(assuming, that n divides m), which we then assign to k categories.
- In this test, we use the same categories that are used in the poker card game.

Example 6. Let's s be a sequence of lenght $n = 1000$. We divide the tested sequence of elements into subsequences of length $n = 4$ and assign each subsequence to one of the following $k = 5$ categories:

all different	$abcd$
pair	$aabc$
two pairs	$aabb$
three of a kind	$aaab$
four of a kind	$aaaa$

Assume that the elements of the tested sequence are from the set $\{0, 1, \ldots, 9\}$, i.e., they are digits.

Poker test

We also know that the following number of subsequences belong to each category.

all different	560
pair	394
two pairs	32
three of a kind	13
four of a kind	1

Apply the Poker test.

Poker test

Solution.

- For the above five categories, we calculate the probabilities that the random subsequence of digits of length 4 belongs to the given category:

$$
\begin{aligned}
P_1 &= \Pr(\text{all different}) = 1 \cdot \frac{9}{10} \cdot \frac{8}{10} \cdot \frac{7}{10} \cdot \binom{4}{0} = 0.504 \\
P_2 &= \Pr(\text{pair}) = 1 \cdot \frac{1}{10} \cdot \frac{9}{10} \cdot \frac{8}{10} \cdot \binom{4}{2} = 0.432 \\
P_3 &= \Pr(\text{two pairs}) = 1 \cdot \frac{1}{10} \cdot \frac{9}{10} \cdot \frac{1}{10} \cdot \binom{4}{2} \cdot \binom{4}{2} \cdot \frac{1}{2} = 0.027 \\
P_4 &= \Pr(\text{three of a kind}) = 1 \cdot \frac{1}{10} \cdot \frac{1}{10} \cdot \frac{9}{10} \cdot \binom{4}{3} = 0.036 \\
P_5 &= \Pr(\text{four of a kind}) = 1 \cdot \frac{1}{10} \cdot \frac{1}{10} \cdot \frac{1}{10} \cdot \binom{4}{4} = 0.001
\end{aligned}
$$

Poker test

- For $i = 1, \ldots, 5$ we calculate the expected numbers $E_i = \frac{m}{n} \cdot P_i$ of subsequences in individual categories. We will denote the measured numbers of the subsequences of the tested sequence of digits for the respective categories as $X_i, i = 1, \ldots, 5$.
- Then, we apply χ^2 goodness-of-fit test. Let us define the null hypothesis H_0 and the alternative hypothesis H_A :

  H_0: the measured values X_i match the expected values E_i

  H_A: the measured values X_i do not match the expected values E_i.
- Next, we calculate the value of the test statistics:

$$X^2 = \sum_{i=1}^{5} \frac{(X_i - E_i)^2}{E_i}.$$

Poker test

- $X^2 = \frac{(560-504)^2}{504} + \frac{(394-432)^2}{432} + \frac{(32-27)^2}{27} + \frac{(13-36)^2}{36} + \frac{(1-1)^2}{1}.$
-
$$X^2 = 25.1852.$$
- We decide to reject or not reject the null hypothesis H_0 based on the value of the quantile $\chi^2_{1-\alpha;\ f}$.
- When testing the hypothesis, we choose the significance level $\alpha = 0,05$ and consider $f = 4$ degrees of freedom.
- Quantile $\chi^2_{1-\alpha;\ f}$ is then equal to $\chi^2_{0.95;\ 6} = 9.49$.
- We reject the null hypothesis at the significance level of 0.05 since $X^2 > \chi^2_{0.95;\ 4}$.

Runs test

The Runs test aims to determine whether the number of runs (of either zeros or ones) of various lengths in the sequence s is as expected for a random sequence.

The expected number of gaps (or blocks) of length i in a random sequence of length n is $e_i = \frac{(n-i+3)}{2^{i+2}}$.

Let k be equal to the largest integer i for which $e_i \geq 5$. Let B_i, G_i be the number of blocks and gaps, respectively, of length i in s for each $i, 1 \leq i \leq k$. The statistic used is

$$X^2 = \sum_{i=1}^{k} \frac{(B_i - e_i)^2}{e_i} + \sum_{i=1}^{k} \frac{(G_i - e_i)^2}{e_i}$$

which approximately follows a χ^2 distribution with $2k - 2$ degrees of freedom.

Example 7. Let s be a sequence of the length $n = 160$:

$$0100001011000010011001010101001110011101 1$$

$$0010010101110100001011010111000001111000$$

$$1011100000111011100010010010010111011$$

$$0000011010011001100010100111101101001000 11.$$

Apply the Runs test.

Solution.

First, we find the largest integer k for which $e_i \geq 5$. In our case $e_4 \approx 2.5 < 5$. so we end up with e_3.

Next, we count B_i, G_i the number of blocks or gaps of length i in the sequence s for each $i, 1 \leq i \leq 3$.

$$0 \mid 1 \mid 0000 \mid 1 \mid 0 \mid 11 \mid 0000 \mid 1 \mid 00 \mid 11 \mid 00 \mid$$

$$\mid 1 \mid 0 \mid 1 \mid 0 \mid 1 \mid 0 \mid 1 \mid 00 \mid 111 \mid 00 \mid 111 \mid$$

$$\mid 0 \mid 11 \mid 00 \mid 1 \mid 00 \mid 1 \mid 0 \mid 1 \mid 0 \mid 111 \mid 0 \mid 1 \mid$$

$$\mid 0000 \mid 1 \mid 0 \mid 11 \mid 0 \mid 1 \mid 0 \mid 111 \mid 00000 \mid 1111 \mid 000 \mid$$

$$\mid 1 \mid 0 \mid 111 \mid 00000 \mid 111 \mid 0 \mid 111 \mid 000 \mid 1 \mid 00 \mid 1 \mid$$

$$\mid 00 \mid 1 \mid 00 \mid 1 \mid 0 \mid 111 \mid 0 \mid 11 \mid 00000 \mid 11 \mid 0 \mid$$

$$\mid 1 \mid 00 \mid 11 \mid 00 \mid 11 \mid 000 \mid 1 \mid 0 \mid 1 \mid 00 \mid 1111 \mid$$

$$\mid 0 \mid 11 \mid 0 \mid 1 \mid 00 \mid 1 \mid 000 \mid 11 \mid$$

i	B_i	G_i	e_i
1	23	20	20.25
2	10	13	10.0625
3	8	4	5

The test statistic is as follows:

$$X^2 = \sum_{i=1}^{3} \frac{(B_i - e_i)^2}{e_i} + \sum_{i=1}^{3} \frac{(G_i - e_i)^2}{e_i}$$

$$X^2 = \frac{(23 - 20,25)^2}{20,25} + \frac{(10 - 10,0625)^2}{10,0625} + \frac{(8-5)^2}{5} +$$
$$+ \frac{(20 - 20,25)^2}{20,25} + \frac{(13 - 10,0625)^2}{10,0625} + \frac{(4-5)^2}{5}.$$

So we have $X^2 = 3.23446$. We work at a significance level of $\alpha = 0.05$ and consider $2k - 2$ degrees of freedom, then $\chi^2_{0.95;\,4} = 9.48773$. The tested sequence passes the runs test since $X^2 = 3.23446 < 9.48773$.

Autocorrelation test

The purpose of this test is to check for correlations between the sequence s and (non-cyclic) shifted versions of it. Let d be a fixed integer, $1 \leq d \leq \lfloor \frac{n}{2} \rfloor$. The number of bits in s not equal to their $d-$shifts is $A(d) = \sum_{i=0}^{n-d-1} s_i \oplus s_{i+d}$, where $\oplus$ denotes the XOR operator. The statistic used is

$$X^2 = 2\frac{(A(d) - \frac{n-d}{2})}{\sqrt{n-d}}$$

which approximately follows a $N(0,1)$ distribution if $n - d \geq 10$. Since small values of $A(d)$ are as unexpected as large values of $A(d)$, a two-tailed test should be used.

Example 8. Let's have some pseudorandom bit generator that produced the following sequence of length 24:

$$110010010000111111011010.$$

Test this generator with the Autocorrelation test, using $d = 4$.

Solution.
The tested sequence has length $n = 24$ and for $d = 4$ it holds that $1 \leq 4 \leq \lfloor \frac{24}{2} \rfloor$. So $n - d - 1 = 19$. We also know that the first 20 elements of the sequence are as follows:

$$11001001000011111101$$

and the 20 elements of the sequence starting with element 5 (i.e., shifted by 4 elements) are as follows:

$$10010000111111011010.$$

We perform the XOR operation and get the following sequence:

$$01011001111100100111.$$

The number of ones in the last sequence is 12, so the 12 members of the sequence are not equal to their shift by 4 elements, i.e., $A(4) = 12$.

Let's calculate the value of the test statistic using the following formula:

$$X^2 = 2\frac{(A(d) - \frac{n-d}{2})}{\sqrt{n-d}}$$

So we get

$$X^2 = 2\frac{(12 - \frac{24-4}{2})}{\sqrt{24-4}} = \frac{4}{\sqrt{20}} = 0.8944.$$

We know that $n - d = 24 - 4 = 20 \geq 10$ and therefore our test statistic has a normalized normal distribution $N(0, 1)$. We work at the significance level $\alpha = 0.05$ and find the critical value c for a given α. We use a two-tailed test, and therefore the generator fails the test if: $\mid X^2 \mid > c$.

From the tables, we find that our critical values for a two-tailed test and $\alpha = 0.05$ are as follows:

$$x_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96 = c$$

$$x_{\frac{\alpha}{2}} = z_{0.025} = -1.96 = -c$$

We compare the test statistic $X^2 = 0.8944$ with the critical values from the tables, and we see that $0.8944 < 1.96$ and $0.8944 > -1.96$. Thus, the tested generator passes the Autocorrelation test.

Example 9. LCG. Implement the LCG with the following parameters:

- $x_0 = 20170705$ is the seed
- $a = 742938285$
- $e = 31$
- $m = 2^e - 1$ modulus

Generate a sequence of length $N = 100000$ using the given LCG. Next, implement the Frequency (monobit) test and verify the generated sequence.

Linear congruential(LCG) PRBG

Algorithm 1 Algorithm LCG PRBG

Input: $N-$length generated sequence

Output: $x_1, x_2, \ldots, x_N \in \mathbb{Z}_2$ a pseudorandom bit sequence

1: select a random integer x_0 (the seed) and select parameters a, c, m, where a, c are in the interval $[1, m]$
2: **for** $i = 1$ **to** l **do**
3: $\quad x_i = ax_{i-1} + c \bmod m$
4: **end for**
5: **return** the output sequence $x_1, x_2, \ldots, x_l$

Example 10.

Let $s = 1, 0, 1, 1, 0, 1, 1, 0, 1, \ldots$ be a sequence. Find:

- period and a cycle of this sequence
- $C(0), C(1)$.

Solution.

1. $s = 1, 0, 1, 1, 0, 1, 1, 0, 1, \ldots$ is a sequence, then $s_i = s_{i+3} = 1, 0, 1$. Therefore, s is the periodic sequence of period 3, and cycle of the sequence s is the subsequences$s^3 = 1, 0, 1$.

2. Then the period N is 3, ie., $N = 3$.

$$C(t) = \frac{1}{N}\sum_{i=0}^{N-1}(2s_i - 1)(2s_{i+t} - 1), \text{for } 0 \leq t \leq N-1.$$

Then

$$C(0) = \frac{1}{3}\sum_{i=0}^{2}(2s_i - 1)(2s_{i+0} - 1) =$$

$$= \frac{1}{3}((2s_0 - 1)(2s_0 - 1) + (2s_1 - 1)(2s_1 - 1) + (2s_2 - 1)(2s_2 - 1)).$$

$$C(0) = \frac{1}{3}((2\cdot 1-1)(2\cdot 1-1)+(2\cdot 0-1)(2\cdot 0-1)+(2\cdot 1-1)(2\cdot 1-1)) =$$

$$= \frac{1}{3}(1+1+1) = 1.$$

$$C(1) = \frac{1}{3}((2\cdot 1-1)(2\cdot 0-1)+(2\cdot 0-1)(2\cdot 1-1)+(2\cdot 1-1)(2\cdot 1-1)) =$$

$$= \frac{1}{3}(-1-1+1) = -\frac{1}{3}.$$