Testing the NIST Statistical Test Suite on artificial pseudorandom sequences

A. M. Zubkov, A. A. Serov

Steklov Mathematical Institute of Russian Academy of Sciences

Suzdal 2018

Generators of random and pseudo-random sequences are used in many fields of science and technology, including the cryptography. The most strict conditions on the quality of generated sequences are used in cryptography.

Shannon theory of secret communications: the ideal ciphers should generate ciphertexts which are indistinguishable from the equiprobable Bernoulli sequence.

Formally random sequence $\gamma = \gamma(\omega)$ is the function $\gamma\colon \Omega \to \{0,1\}^{\mathbb{Z}}$ defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

Random binary sequence $\gamma = \{\gamma_t\}_{t\in\mathbb{Z}}$ is equiprobable Bernoulli iff for all integers $n \geqslant 1$, $t_1 < \ldots < t_n$, $a_1, \ldots, a_n \in \{0, 1\}$

$$\mathbf{P}\{\gamma_{t_1} = a_1, \ldots, \gamma_{t_n} = a_n\} = \frac{1}{2^n} \qquad (*)$$

1. What means «indistinguishability»?

2. Is it possible for long concrete binary sequence to be «indistinguishable» from equiprobable Bernoulli sequence?

3. Is it possible to measure the level of «distinguishability»?

4. How robust is the property of «indistinguishability»?

etc.

Any concrete binary sequence $(a_1, \ldots, a_T)$ may be considered as a segment of realization of random sequence.
For each of 2^T binary sequences of length T the probability of its appearance in equiprobable Bernoulli sequence equals $\frac{1}{2^n}$.
So, under hypothesis of independence and equiprobability there are no reasons to consider some finite binary sequences as being «not so random» as other sequences.

A correct probabilistic interpretation of system of conditions (*) may be as follows: if there are a sequence of independent sequences $\gamma^{(n)} = \{\gamma_t^{(n)}\}_{t\in\mathbb{Z}}$, then the fraction of sequences $\gamma^{(1)}, \ldots, \gamma^{(T)}$ satisfying condition of the type

$$\mathbf{P}\{\gamma_{t_1} = a_1, \ldots, \gamma_{t_n} = a_n\} = \frac{1}{2^n}$$

tends to $\frac{1}{2^n}$ as $T \to \infty$.

When applying statistical tests to a concrete binary sequence usually this sequence is splitted into nonoverlapping segments.

These segments are supposed to be independent realizations of random sequence. (It is additional implicit hypothesis on the analysed sequence!)

The fractions of segments satisfying condition(s) of type (*) are computed and values of some statistics (functions of these fractions) are compared with «critical levels» obtained by means of limit theorems of probability theory.

If values of statistics contradict «critical levels», then the hypothesis used to compute these levels is rejected, in the opposite case the hypothesis is accepted.
More correct formulation of the result in the last case would be «not rejected». Our experiments support the validity of this note.

Statistical test packages

There are several popular packages of statistical tests which are distributed with open source codes (e. g. TESTU01, DIEHARD, NIST, SPRNG), or with closed source codes (e. g. Crypt-X).

From the statistical test packages listed above, the NIST statistical tests package was selected as one of the most popular, fully documented and actively used for generator certifications.

List of NIST Statistical Tests

Number	Test Name
1	Frequency
2	Block Frequency
3	Runs
4	Longest Run
5	Binary Matrix Rank
6	Discrete Fourier Transform
7	Non-overlapping Template Matching
8	Overlapping Template Matching
9	Universal Maurer's Test
10	Linear Complexity
11	Serial
12	Approximate Entropy
13	Cumulative Sums
14	Random Excursions
15	Random Excursions Variant

Frequency & Block Frequency tests

n — length of the tested bit string $\{\varepsilon_i\}_{i=1}^n$, $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{z} e^{-\frac{u^2}{2}} \mathrm{d}u$.

Frequency (Monobit) Test

$$s_{obs} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (2\varepsilon_i - 1), \quad \lim_{n\to\infty} \mathbf{P}\{s_{obs} \leqslant x\} = \Phi(x).$$

Frequency Test within a Block

M — length of each block, $N = \lfloor \frac{n}{M} \rfloor$ — number of non-overlapping blocks.

$\pi_i = \frac{\sum_{j=1}^{M} \varepsilon_{(i-1)M+j}}{M}$, $\quad \chi^2 = 4M \sum_{i=1}^{N} \left(\pi_i - \frac{1}{2}\right)^2$ weakly converges to chi-square distribution with N degrees of freedom as $M \to \infty$

Runs test

n — length of the tested bit string $\{\varepsilon_i\}_{i=1}^{n}$.

Runs Test

$$V_n(obs) = \sum_{k=1}^{n-1} r(k) + 1, \quad \text{where } r(k) = \begin{cases} 0 & \text{if } \varepsilon_k = \varepsilon_{k+1}, \\ 1 & \text{otherwise,} \end{cases}$$

$$\text{the distribution of } \quad \frac{V_n(obs) - \frac{n}{2}}{\sqrt{n/2}}$$

converges to the standard normal distribution as $n \to \infty$

Test parameters

- The critical values of statistics in the NIST Statistical Test Suite were computed by means of limit theorems, and it was recommended that analysed sequences should have sufficiently large lengths. All segments of sequences that we have tested were of $33,554,431 = 2^{25} - 1$ bit length.

- The significance level $\alpha = 0.01$ determining the rule of acceptance/rejection of the hypothesis was selected by default.

Testing pseudorandom sequences generated by linear shift registers of the maximal period

Pseudorandom sequences of the maximal period were generated by the linear shift registers with feedbacks given by the following primitive polynomials of degrees 25 and 27 over $\mathrm{GF}(2)$:

$$
\begin{aligned}
f(x) &= x^{25} + x^3 + 1, \\
g(x) &= x^{27} + x^5 + x^2 + x + 1, \\
h(x) &= x^{27} + x^{19} + x^{18} + x^{17} + x^{11} + x^6 + 1, \\
m(x) &= x^{27} + x^{26} + x^{25} + x^{24} + x^{23} + x^{22} + x^{21} + x^{20} + x^{19} + x^{17} \\
&\quad + x^{15} + x^{13} + x^{11} + x^9 + x^7 + x^5 + x^3 + x + 1.
\end{aligned}
$$

Initial states of all linear shift registers were chosen to have only one nonzero bit, namely the most significant one.

Results of testing pseudorandom sequences generated by linear shift registers of the maximal period

The segments of the pseudorandom sequences obtained by the linear shift registers with the $g(x)$, $h(x)$ and $m(x)$ polynomials have successfully passed all the tests from the NIST Test Suite except for The Binary Matrix Rank Test (ranks of binary matrices of the size 32×32), The Discrete Fourier Transform (Spectral) Test and The Linear Complexity Test, where the P-values were less than 10^{-6}.

The full-period sequence corresponding to the polynomial $f(x)$, in addition to the listed tests, did not pass the Tests for the Longest Run-of-Ones in a Block (P-value $6 \cdot 10^{-6}$) and Maurer's "Universal Statistical"Test (P-value $1.91 \cdot 10^{-4}$).

Testing pseudorandom linear shift register sequences with additive noise

The additive binary noise applied to the output sequences of linear shift registers over $\mathrm{GF}(2)$ with the primitive polynomials $f(x)$, $g(x)$, $h(x)$ and $m(x)$ was produced by means of pseudorandom generator of the key stream cipher **A8**.

In the case of Bernoulli noise sequence with parameter $\frac{1}{4}$ the only test from the NIST Test Suite that detects nonrandomness in the disjoint $2^{25}-1$ bit segments of output sequences of linear shift registers with polynomials $f(x)$, $g(x)$, $h(x)$, $m(x)$ was The Discrete Fourier Transform (Spectral) Test; the corresponding P-values were smaller 10^{-6}.

In the case of Bernoulli noise sequence with parameter $\frac{3}{8}$ almost all sequences had passed **all tests** from the NIST Test Suite (with the exception of some sequences corresponding to the polynomial $f(x)$).

Testing the filtered output sequences of linear shift registers of the maximal period

To filter output sequences of linear shift registers given by primitive polynomials $f(x)$, $g(x)$, $h(x)$ and $m(x)$ over $\mathrm{GF}(2)$ we use the balanced Boolean function corresponding to the most significant bit of the nonlinear substitution SubBytes $S = (s_1, s_2, \ldots, s_8)\colon \mathrm{GF}(2)^8 \to \mathrm{GF}(2)^8$ of the AES symmetric block cipher algorithm.

For several variants of choosing arguments of the filter function filtered sequences failed to pass a number of tests of NIST Test Suite, and corresponding P-values in many cases were smaller than 10^{-6}.

Testing the pseudorandom sequences obtained by merging of outputs of two linear shift registers of maximal periods

We have considered two types of pseudorandom sequences constructed by segments of two binary linear shift register sequences with feedbacks defined by two primitive polynomials

A) The output sequence of the first register $\{x_1, x_2, \ldots\}$ corresponding to the polynomial $f(x)$ was splitted into adjacent segments of $L_1 = 25$ bits, the output sequence of the second register $\{y_1, y_2, \ldots\}$ corresponding to the polynomial $g(x)$ was similarly splitted into segments of $L_2 = 27$ bits. Further, the tested sequence $\{z_1, z_2, \ldots\}$ of the first type was constructed by merging the obtained segments of two sequences:

$$\{z_k\}_{k=0}^{2^{L_1}-1+\left[\frac{2^{L_1}-1}{L_1}\right]L_2} = \{x_1, \ldots, x_{L_1}, y_1, \ldots, y_{L_2}, x_{L_1+1}, \ldots, x_{2L_1}, y_{L_2+1}, \ldots, y_{2L_2}, \ldots\}.$$

Testing the pseudorandom sequences obtained by merging of outputs of two linear shift registers of maximal periods

B) The output register sequences were splitted into adjacent segments of a variable lengths according to the following rule:

- the first segment of the first register output sequence consists of $L_1 = L_1^* = 25$ sequential output bits,
- the first segment of the second register output sequence consists of

$$L_2^* = 16 + 2^3 x_{L_1^*-3} + 2^2 x_{L_1^*-2} + 2 x_{L_1^*-1} + x_{L_1^*}$$

  bits (a fixed value 16 was added with the integer formed by the last 4 bits of the already constructed sequence),
- the second segment of the first register consists of

$$L_1^* = 16 + 2^3 y_{L_2^*-3} + 2^2 y_{L_2^*-2} + 2 y_{L_2^*-1} + y_{L_2^*}$$

  bits (a fixed value 16 was added with the integer formed by the last 4 bits of the already constructed sequence),
- and so on.

The tested sequence $\{w_1, w_2, \ldots\}$ of the second type had the form

$$\{x_1, \ldots, x_{L_1}, y_1, \ldots, y_{L_2^*}, x_{L_1+1}, \ldots, x_{L_1+L_1^*}, y_{L_2^*+1}, \ldots\}.$$

Results of testing pseudorandom sequences obtained by merging of outputs of two linear shift registers of maximal periods

The first type sequences $\{z_1, z_2, \ldots\}$ had passed all tests with the exception of Discrete Fourier Transform (Spectral) Test: for this test P-values were smaller than 10^{-6}, while almost all second type sequence had passed all the tests with P-values being as a rule essentially larger than $\alpha = 0.01$.

Increasing the mean values of L_1^* and L_2^* up to 128 didn't change the result significantly (with some exceptions).

Testing the pseudorandom sequence generated by AES

The tested pseudorandom sequence was obtained by iterative application of the AES block cipher algorithm to the zero plain text with a 128-bit key in the cipher block chaining mode with initialization 128-bit vector all bits of which are nonzero except for the 7 lower bits; the key bits were fixed by zero and did not change during the iterative calculations. Each byte of the encrypted sequence was replaced by the corresponding bit depending on the byte value.

Four non-overlapping segments of the length $2^{25} - 1$ bits of the initial sequence of the length $2^{27} - 4$ bits passed all the tests from the NIST Test Suite in the aggregate, except for The Serial Test for the second segment only, where the P-value of one of two statistics turned out to be slightly less than the significance level $\alpha = 0.01$, namely 0.008415.

Testing the output sequence of shrinking generators composed of two linear shift registers

Two tested sequences were obtained by extracting from the output sequence of the first linear shift register (with feedback polynomial $g(x)$) all bits corresponding to the nonzero bits in the output sequence of the second linear shift register (with feedback polynomial $f(x)$ for the first type test sequence and the polynomial $h(x)$ for the second).

The first type sequence passed all the tests from the NIST Test Suite with the significance level $\alpha = 0.01$. The second type sequence passed all the tests except for the Serial Test: for this test P-values turned out to be smaller than 10^{-6}. Maybe this is the consequence of coincidence of orders of the source and control sequences.

Table shows the results of almost all performed experiments

Testing	LSR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
LSR output	f	+	+	+	−	−	−	+	+	−	−	−	+	+	+	+
	g	+	+	+	+	−	−	+	+	+	−	±	+	+	+	+
	h	+	+	+	+	−	−	+	+	+	−	±	+	+	+	+
	m	+	+	+	+	−	−	+	+	+	−	+	+	+	+	+
LSR + additive noise with parameter $1/4$	f	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+
	g	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+
	h	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+
	m	+	+	+	−	+	−	+	+	+	+	+	+	+	+	+
with parameter $3/8$	f	+	+	+	+	+	±	+	+	+	+	+	+	+	−	−
	g	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	h	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	m	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
$1,3,9,13,17,21,23,25$	f	+	−	−	−	+	−	−	−	−	+	−	−	+	+	+
filtered LSR arguments of f_1 $1,3,9,13,17,21,25,27$	g	+	−	+	+	+	−	−	+	−	+	−	−	+	+	+
	h	+	−	+	+	+	+	−	+	−	+	−	−	+	+	+
	m	+	−	+	+	+	−	−	+	−	+	−	−	+	+	+
$1,3,9,14,17,21,22,24$	f	+	+	+	+	+	−	−	−	+	+	−	−	+	+	+
filtered LSR arguments of f_1 $1,3,9,14,17,21,23,26$	g	+	+	+	+	+	−	−	+	+	+	−	−	+	+	+
	h	+	−	+	+	+	±	−	−	+	+	−	−	+	+	+
	m	+	−	+	+	+	−	−	−	+	+	−	−	+	+	+
merging of segments with average length	$f,g,24$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	$f,g,64$	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+
	$f,g,96$	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+
	$F,g,64$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	$F,g,128$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	$h,g,128$	+	+	+	+	+	+	+	±	+	+	+	+	+	+	+
fixed length 25, 27	f, g	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+
PRS	AES	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PRS obtained by shrinking	g, f	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	g, h	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Some notes on the robustness of «indistinguishability» from the Bernoulli sequence

The majority of tests in the NIST suite is based on statistics which under the null hypothesis on the equiprobable Bernoulli sequence have asymptotically normal distributions. If test statistics is computed for sample of size N, then usually the fractions of different outcomes have asymptotically normal distributions with mean aN and variance bN for some constants a and b. Therefore the intervals of admissible values of concrete frequencies have the lengths of the order $\sqrt{N}$. So, if some concrete binary sequence $a = (a_1, \ldots, a_N)$ is not rejected by the test, then there exist concrete binary sequences $a' = (a'_1, \ldots, a'_N)$ rejected by the test which differs from a only in $O(\sqrt{N})$ bits.

Moreover, there exists tests (not included into standard packages) for which the set of inadmissible sequences of length $N \to \infty$ is significantly more dense.
An example of such test may be based on the maximal length of coinciding subsequences which for the sequence $\gamma = (\gamma_1, \ldots, \gamma_N)$ is computed by the formula

$$M(\gamma) = \max \left\{ m : \bigcap_{1 \leqslant i < j < N-m} \{(\gamma_{i+1}, \ldots, \gamma_{i+m}) = (\gamma_{j+1}, \ldots, \gamma_{j+m})\} \neq \varnothing \right\}.$$

Assertion. *If* $(\gamma_1, \ldots, \gamma_N)$ *is equiprobable Bernoulli sequence, then*

$$\mathbf{P}\{M(\gamma) > 2\log_2 N - u\} > 1 - \frac{1}{2^{u-2}}, \quad u > 2,$$

$$\mathbf{P}\{M(\gamma) > 2\log_2 N + v\} \leqslant \frac{1}{2^v}, \quad v > 0.$$

This statement shows that:

a) if $\gamma = (\gamma_1, \ldots, \gamma_N)$ is the equiprobable Bernoulli sequence, then the test based on the rule $M(\gamma) < 2\log_2 N + |\log_2 \varepsilon|$ will reject γ with probability not exceeding ε (it is estimate of the error probability of the test),

b) almost every realization of equiprobable Bernoulli sequence adopted by the test may be converted into a sequence rejected by the test by means of inverting only $O(|\log_2 \varepsilon|)$ elements.

So, the distance from almost each «indistinguishable» sequence to the set of «distinguishable» sequences is very small.

Clearly, two sequences which differ in small number of bits with unknown numbers should not have significantly different cryptographic quality.

Therefore, «indistinguishability» by tests of all types may be too strict criterion for cryptography.

Nevertheless, statistical tests applied to cryptographic sequences should not be bounded by only standard statistical criteria, but should account also for specific cryptographic demands.

Conclusions

The set of experiments with different non-random pseudorandom sequences showed that the NIST Test Suite (and very probably other packages) may detect some deviations of properties of analyzed sequences from that of equiprobable Bernoulli sequences, but may fail to detect non-randomness of deterministic sequences with not very complex artificial irregularities.

Thank you for attention!