JSIAM Letters Vol.2 (2010) pp.123–126 ©2010 Japan Society for Industrial and Applied Mathematics JSIAM Letters

On the pass rate of NIST statistical test suite for randomness

Akihiro Yamaguchi[1], Takaaki Seo[1] and Keisuke Yoshikawa[1]

[1] Department of Information and Systems Engineering, Fukuoka Institute of Technology, Wajiro 3-30-1, Higashi-ku, Fukuoka 811-0295, Japan

E-mail *aki@fit.ac.jp*

Received March 31, 2010, Accepted September 6, 2010

Abstract

In this paper, the pass rate of the NIST SP800-22 statistical test suite for the ideally true random sequences is analyzed by the simulation of statistical tests, and derived by the theoretical analysis under the assumption that there are no correlation among tests. As examples of chaos based system, Vector Stream Cipher (VSC128S) and the encryption system using Arnold's cat map are tested. The test results are compared with the theoretical one for the true random sequences and validity of presented analysis is discussed.

Keywords random number, chaos, randomness, statistical test, NIST SP800-22

Research Activity Group Applied Chaos

1. Introduction

As an application of chaos, random number generators and pseudo random number generators based on chaotic dynamics have been well investigated in various scientific and engineering fields including information security. For a correct and safety application, the randomness of generated sequences and its evaluation are very important. In order to evaluate the randomness, several statistical test suites have been proposed. NIST SP800-22 is one of the statistical test suites, and it was used for the evaluation of AES candidates [1, 2]. For chaos based random and pseudo random number generators, this test suite is also useful to evaluate the randomness of generated sequences.

NIST SP800-22 consists of 15 kinds of statistical tests and provides the criteria to determine whether given sequences are random or not for each statistical test. However, it was not mentioned in the criteria how many the ratio of passing all the 15 kinds of tests should be for the target generator to be regarded as the perfect random number generator. In this paper, the test suite NIST SP800-22 is focused on and its statistical properties for the idealized perfect random number generator are numerically analyzed. The results are compared with typical pseudo random number generators including chaos based system.

2. Statistical test of randomness

The random bit sequence should be independent and unpredictable. These are also characteristics of chaotic dynamics. Therefore, chaotic dynamics is one of the candidates for the basic mechanism to produce random bit sequences. Since randomness is a probabilistic property, it can be characterized and described in terms of probability.

There are various statistical tests that can be applied to a sequence to attempt to compare and evaluate the sequence to a true random sequence. The Special Publication (SP) 800-22 revision 1 proposed by National Institute of Standards and Technology (NIST) is a statistical test suite that consists of the 15 kinds of statistical tests of randomness [1]. The list of 15 tests and tests parameters are shown in Tables 1 and 2.

According to [1], the basic testing process common to each test is explained in the following. Targets of test are binary sequences of '0' and '1' with the length n. Here, the number of tested sequence is m and it is called sample size. At a test for one sequence, a statistics called P-value is calculated from the tested sequence. The P-value is the probability that a perfect random number generator would have produced a sequence less random than the tested sequence. It is determined whether the tested sequence is random or not random by the testing hypothesis. The null hypothesis H_0 under test is that the sequence being tested is random, and the alternative hypothesis H_1 is that the sequence being tested is not random. If P-value $\geq \alpha$, the null hypothesis is accepted (i.e., the target is random), and otherwise rejected (i.e., the target is not random), where $\alpha = 0.01$ is a significance level of the testing hypothesis.

For one test, m sequences are tested and m P-values are obtained. Thus, the m decisions of randomness are obtained for each test. These are only individual decisions for each sequence. As a holistic interpretation of these results, NIST adopted to include the following two conditions to determine the target sequences are holistically random or not random.

Condition 1 *Let ξ be a proportion of accepted sequences for the tests and it is called the pass rate of sequences. If ξ is in the acceptable interval*

$$\hat{p} - 3\sqrt{\frac{\hat{p}(1-\hat{p})}{m}} \leq \xi \leq \hat{p} + 3\sqrt{\frac{\hat{p}(1-\hat{p})}{m}}, \qquad (1)$$

Table 1. List of NIST SP800-22 statistical tests.

No.	Test Name
1	The Frequency (Monobit) Test
2	Frequency Test within a Block
3	The Runs Test
4	Tests for the Longest-Run-of-Ones in a Block
5	The Binary Matrix Rank Test
6	The Discrete Fourier Transform (Spectral) Test
7	The Non-overlapping Template Matching Test
8	The Overlapping Template Matching Test
9	Maurer's "Universal Statistical" Test
10	The Linear Complexity Test
11	The Serial Test
12	The Approximate Entropy Test
13	The Cumulative Sums (Cusums) Test
14	The Random Excursions Test
15	The Random Excursions Variant Test

Table 2. Parameters used for NIST SP800-22 test suite.

Test Name	Block Length
Frequency Test within a Block	20,000
The Non-overlapping Template Matching Test	9
The Overlapping Template Matching Test	9
Maurer's "Universal Statistical" Test	7
The Linear Complexity Test	500
The Serial Test	10
The Approximate Entropy Test	10

Table 3. Count of passing each test and passing all of the 15 tests for 100 times iterations ($m = 1,000$).

Test No.	Number of sub-tests	Count of passing tests out of 100 repetitions			
		VSC128S	CatMap2D	AES	SHA-1
1	1	100	100	100	100
2	1	100	100	100	100
3	1	100	100	99	100
4	1	100	99	100	99
5	1	100	99	100	100
6	1	100	95	96	97
7	148	66	64	54	53
8	1	98	99	98	98
9	1	98	97	95	97
10	1	100	99	99	100
11	2	99	100	99	100
12	1	99	100	100	99
13	2	100	100	100	99
14	8	96	95	96	91
15	18	95	95	96	98
All(1-15)	188 (total)	56	55	43	41

the sequences are random, otherwise not random, where $\hat{p} = \alpha - 1$.

This interval corresponds to the three times of the standard deviation of ξ for the true random sequences that are produced by the perfect random number generator.

Condition 2 *The distribution of P-values is examined to ensure uniformity. If the obtained P-values are uniform, the sequences are random, otherwise not random.*

Uniformity is determined by the χ^2-test on the obtained P-values. The interval between 0 and 1 is divided into 10 sub-intervals, and the P-values that lie within each i-th sub-interval are counted as F_i. Then a P-value is calculated as

$$\text{P-value}_T = \text{igamc}\left(\frac{9}{2}, \frac{\chi^2}{2}\right), \tag{2}$$

where igamc is the incomplete gamma function and

$$\chi^2 = \sum_{i=1}^{10} \frac{\left(F_i - \frac{m}{10}\right)^2}{\frac{m}{10}}. \tag{3}$$

If the condition

$$\text{P-value}_T \geq \alpha_T = 0.0001 \tag{4}$$

is satisfied, then the P-values can be considered to be uniformly distributed.

Results of applying the NIST statistical test suite are shown in Table 3. Here, Vector Stream Cipher (VSC128S) [3] and the encryption system using two dimensional cat map (CatMap2D) [4] are tested as examples of chaos based generator. VSC128S is a stream cipher designed by ChaosWare, and it consists of the combination of 8 pseudo chaotic one dimensional maps with 32bits operation. CatMap2D is also a steam cipher, and it adopts combined 4 two dimensional Arnold's cat maps instead of 8 one dimensional maps in VSC128S. In comparison, SHA-1 and AES are also tested. The former is a hash function and hash results of plane texts are tested as a pseudo random sequence. The latter is a typical standard encryption system and the encrypted texts are tested.

For each system, NIST statistical test suite is applied 100 times for $m = 1,000$ binary sequences with length $n = 1,000,000$, and the number of passing each kind of test and the number of passing all of the 15 tests are counted. These tests are performed by the test suite program version 2.0b provided by NIST with the correction of the assessment condition and the discrete Fourier transform test. In the original program (version 2.0b), the former condition that corresponds to (1) was described only for the case $m = 100$. The latter test was corrected according to Kim et al. [5] since the variance of theoretical distribution was not yet corrected in NIST SP800-22 revision 1.

As a result, the rate of passing all 15 tests is around 55% for VSC128S and CatMap2D, and around 40% for SHA-1 and AES, respectively. It is to be noticed that there are obvious differences between their pass rates of all of the tests.

3. Simulation of NIST SP800-22

NIST SP800-22 is based on the testing hypothesis using P-values. Since the P-values uniformly distribute in the interval between 0 and 1 for the ideally true random sequences, their statistical tests can be simulated by the Monte Carlo method in which uniformly distributed artificial P-values are produced randomly. The procedure of the simulation of statistical test suite is follows.

Step 1 An artificial P-value that uniformly distributes in the interval between 0 and 1, is produced randomly. This P-value corresponds to the p-value calculated for the ideally true random sequence.

Step 2 The condition P-value $> \alpha = 0.01$ is examined for the produced P-value.

Step 3 Steps 1 and 2 are iterated m times and the pass rate ξ and the P-value $_T$ for one test are calculated. Conditions 1 and 2 are examined and pass or non-pass of one test is determined.

Step 4 Steps 1 to 3 are iterated K times, where K corresponds to the number of statistical tests (including sub-tests). Then, pass or non-pass of all of K tests is determined.

Some of the statistical tests constituted NIST SP800-22 have several sub-tests. Therefore, the actual number of tests is the total number of sub-tests. For an example, the non-overlapping template marching test consists of 148 sub-tests corresponding to different templates when the template length is 9. The number of sub-tests for each test is also shown in Table 3, and the total number K of tests is 188 for the parameters in Table 2.

As results of this simulation, pass rates $\hat{P}_{C1}$, $\hat{P}_{C2}$, $\hat{P}_{Pass1}$, and $\hat{P}_{PassK}$ could be obtained, where $\hat{P}_{C1}$ denotes the pass rate of Condition 1, $\hat{P}_{C2}$ denotes the pass rate of Condition 2, $\hat{P}_{Pass1}$ denotes the pass rate of one test, and $\hat{P}_{PassK}$ denotes the pass rate of all of K tests. Furthermore, the correlation coefficient that is denoted by $\hat{\rho}_{C1,C2}$ between Conditions 1 and 2 could also be estimated.

4. Probability of passing all tests

The probability of passing all tests was firstly analyzed by Okutomi et al. for Condition 1 [6]. Their analysis was, however, insufficient to obtain the probability of passing all tests, since the upper limit of the acceptable interval for Condition 1 (Eq. (1)) was ignored and Condition 2 was not concerned. For more precise analysis, this paper focuses on the probability of passing all tests for both of Conditions 1 and 2.

For one test, pass or non-pass is determined by Conditions 1 and 2, as previously mentioned in Section 2. Since the latter is also the testing hypothesis for obtained P-values, the pass rate P_{C2} of Condition 2 is determined by its significance level α_T such that $P_{C2} = 1 - \alpha_T = 0.9999$.

On the other hand, the pass rate P_{C1} of Condition 1 is determined by the distribution of ξ that is the proportion of accepted sequences. The probability that one true random sequence passes the test is $\hat{p} = 1 - \alpha$ by the definition of testing hypothesis. Since the probability that k out of m true random sequences pass the test, obeys the binomial distribution, it is obtained as

$$P(k;m) = {}_m\mathrm{C}_k \times \hat{p}^k \times (1-\hat{p})^{m-k}, \tag{5}$$

where k/m corresponds to ξ. The average of ξ is

$$\mu_\xi = \hat{p}, \tag{6}$$

and the standard deviation of ξ is

$$\sigma_\xi = \sqrt{\frac{\hat{p}(1-\hat{p})}{m}}. \tag{7}$$

Therefore, the pass rate of Condition 1 is exactly obtained as

$$P_{C1} = \sum_{k=k_0}^{k_1} P(k;m), \tag{8}$$

where

$$k_0 = \max(\lceil(\mu_\xi - 3\sigma_\xi) \times m\rceil, 0), \tag{9}$$

and

$$k_1 = \min(\lfloor(\mu_\xi + 3\sigma_\xi) \times m\rfloor, m). \tag{10}$$

Here, the range $k_0 \le k \le k_1$ corresponds to the range of acceptance in (1). Since this range corresponds to three times of the standard deviation σ_ξ, the Gaussian approximation of P_{C1} is also obtained by the error function as $\mathrm{erf}(3/\sqrt{2})$. In the analysis given by Okutomi et al. [6], the upper limit k_1 was fixed to m. Therefore, in the case of $m = 1,000$, their obtained probability $P_{C1} = 0.996712$ is slightly larger than the correct probability $P_{C1} = 0.99666846$ (Table 4-(a)).

The probability of passing one test P_{Pass1} is the probability that both Conditions 1 and 2 are simultaneously satisfied. If Conditions 1 and 2 have no correlation, P_{Pass1} is obtained as the direct product

$$P_{Pass1} = P_{C1} \times P_{C2}. \tag{11}$$

Furthermore, if results for the K tests in the test suite, are not correlated to each other, the pass rate of all of the K tests, P_{PassK}, is also obtained as

$$P_{PassK} = \prod_{k=1}^{K} P_{Pass1} = (P_{Pass1})^K. \tag{12}$$

Table 4. Theoretical values of pass rates obtained by the binomial distribution and the Gaussian distribution ($K = 188$).

(a) Binomial distribution.

m	P_{C1}	P_{Pass1}	$P_{PassK} = (P_{Pass1})^K$
100	98.162596%	98.152780%	3.003929%
500	99.479195%	99.469247%	36.770592%
1000	99.666846%	99.656880%	52.404673%
10000	99.689933%	99.679964%	54.736967%

(b) Gaussian distribution.

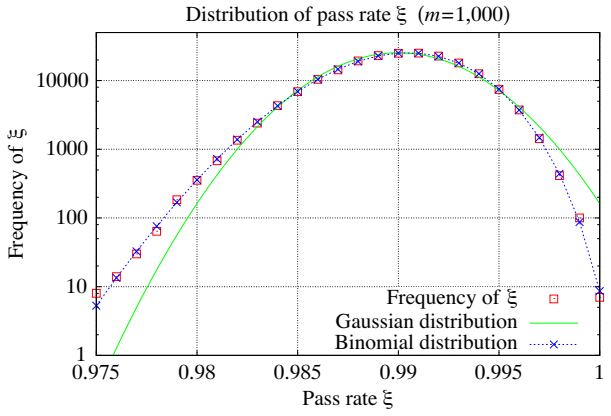
m	P_{C1}	P_{Pass1}	$P_{PassK} = (P_{Pass1})^K$
—	99.730020%	99.720047%	59.034450%

Table 5. Estimated pass rates and correlation coefficient between Conditions 1 and 2 ($K = 188$).

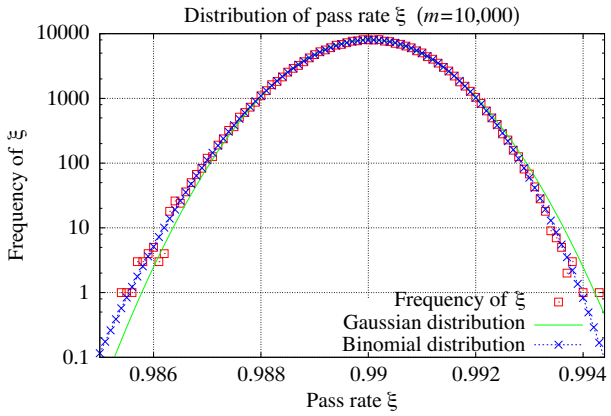
m	$\hat{P}_{C1}$	$\hat{P}_{C2}$	$\hat{P}_{Pass1}$	$\hat{\rho}_{C1,C2}$	$\hat{P}_{PassK}$
100	98.17%	99.992%	98.16%	−0.0012	3.20±0.18%
500	99.43%	99.990%	99.42%	−0.0008	35.81±0.48%
1000	99.66%	99.991%	99.65%	−0.0006	52.08±0.50%
10000	99.67%	99.993%	99.66%	−0.0005	54.78±0.50%

5. Numerical results and discussions

Results of numerical simulation are shown in Table 5, where $K = 188$, $\hat{P}_{C1}$, $\hat{P}_{C2}$, $\hat{P}_{Pass1}$ and $\hat{\rho}_{C1,C2}$ were estimated by the 100,000 times simulations of Steps 1 to 3, and $\hat{P}_{PassK}$ was estimated by the 10,000 times simulations of Steps 1 to 4. The standard error of $\hat{P}_{PassK}$



(a) $m = 1,000$.



(b) $m = 10,000$.

Fig. 1. Distribution of pass rate ξ for 200,000 samples.

is also shown as an error range of estimation. In the case that $m = 1,000$, the estimated pass ratio $\hat{P}_{PassK}$ is around 52%, and this value is close to the pass rate for VSC128S and CatMap2D shown in Table 3. Since the correlation coefficient $\hat{\rho}_{C1,C2}$ is too small, the pass rate of Condition 1 and the pass rate of Condition 2 are almost independent of each other.

The distributions of ξ were also obtained in cases of $m = 1,000$ and $m = 10,000$ to examine its convergence to the Gaussian distribution. The results are shown in Fig. 1, where the red squares represent the distribution of ξ, the green solid lines represent the Gaussian distribution, and the blue crosses represent the binomial distribution. For the Gaussian distribution and the binomial distribution, their averages and variances are μ_ξ and σ_ξ^2, respectively. These results indicate that the distribution of ξ obeys the binomial distribution, and it approaches asymptotically to the Gaussian distribution when the sample size m is increased.

The pass rates P_{C1}, P_{Pass1} and P_{PassK} directly calculated by the (8), (11), (12) and Gaussian approximation, are also shown in Table 4. The pass rate P_{C2} is 0.9999 as previously mentioned. These theoretical values given by the binomial distribution, agree with the estimated values shown in Table 5, and they converge to the values given by the Gaussian distribution as increasing the sample size m.

As shown in Table 4-(a), the pass rate of all of the K tests, P_{PassK}, is almost 52% for the true random sequences in the case that $K = 188$ and the sample size $m = 1000$. Furthermore, its limit of m to infinity, is almost 59% (Table 4-(b)) that is given by the Gaussian approximation, since the binomial distribution converges to the Gaussian distribution for sufficiently large m. These results indicate that the sufficient number of repetition of the test suite is necessary to examine the proportion to pass all of the tests.

In this paper, two kinds of independencies are assumed for the analysis of pass rates. One is the independency between Conditions 1 and 2 at one test, and the other is the independency among the tests constituted the test suite. The former independency is supported by the results for the correlation coefficients shown in Table 5. Although plausibility of the latter independency could not be mentioned here, if there are some positive correlations among the actual tests, the pass rate of all of the tests P_{PassK} is expected to take larger value than (12). Even in such case, (12) and its values shown in Table 4 might be useful guideline for the decision of the perfect random number generator, since they give the lower bound of P_{PassK}.

6. Conclusions

The pass rate of the NIST SP800-22 statistical test suite for the ideally true random sequences was analyzed under the assumption that there are no correlations among tests. The obtained pass rate was close to the results for actual pseudo random number generators based on chaos. An analysis including the correlation among tests is one of the future works.

Acknowledgments

One of the authors (A. Y.) wishes to express his gratitude to Dr. K. Umeno, Mr. H. Terai and Dr. S. J. Kim for their interesting and fruitful discussions. Authors (T. S. and K. Y.) had joined this research when they were students of Fukuoka Institute of Technology. Their current affiliations are Miyazaki Jyoho Center (T. S.) and Techno Systems Co. (K. Y.).

References

[1] A. Rukhin et al., A statistical test suite for random and pseudorandom number generators for cryptographic applications, NIST Special Publication 800-22 Revision 1, 2008.

[2] J. Soto and L. Bassham, Randomness testing of the advanced encryption standard finalist candidates, NIST IR 6483, http://csrc.nist.gov/publications/nistir/ir6483.pdf, 2000.

[3] K. Umeno, Specification of VSC128S (in Japanese), http://www.chaosware.com/vsc128s.pdf, 2004.

[4] T. Araki and A. Yamaguchi, Statistical analysis of the VSC encryption system using two dimensional cat map (in Japanese), in: Proc. of 55th NCTAM, pp.185–186, 2006.

[5] S. J. Kim, K. Umeno and A. Hasegawa, On the NIST statistical test suite for randomness, IEICE Tech. Rep., **103**(499) (2003), 21–27.

[6] H. Okutomi and K. Nakamura, A study on rational judgement method of randomness property using NIST randomness test (NIST SP. 800-22) (in Japanese), IEICE Trans. A, **J93-A** (2010), 11–22.