Master's Thesis

Study on a further improvement of Maurer's universal statistical test

Guidance

Professor Ken UMENO
Assistant Professor Atsushi IWASAKI

Yasunari HIKIMA

Department of Applied Mathematics and Physics

Graduate School of Informatics

Kyoto University

February 2020

Abstract

Maurer's universal statistical test is a hypothesis test for evaluating the randomness of a binary sequence and it is included in NIST SP 800-22 which is one of the most famous test suites. The test statistic of Maurer's test relates to the entropy of a tested sequence and hence the test can detect various defects of the sequence about randomness. It has been reported that flipping a part of bits in a sequence makes Maurer's test more sensitive. The test with flipping is called highly sensitive universal statistical test. To perform the highly sensitive test, the variance for the reference distribution is necessary, however, the theoretical value has not been derived. In this thesis, we theoretically derive the variance for the reference distribution of the highly sensitive test and investigate the validity for testing randomness.

Contents

1 Introduction

1.1 Random sequence

Throughout this thesis, we only consider a binary sequence. A random sequence is intuitively considered as a number sequence without any recognizable patterns or regularities, and it is not easy to define such properties mathematically. The reader may consider a sequence to be random if its each bit is independent and symmetrically distributed, but this description is inconsistent with an intuitive definition. When we intuitively say "random sequence", this term is used to represent a specific sequence, not to represent a sequence of random variables.

Several approaches to define a random sequence have been proposed. In Algorithmic information theory, several definitions have been proposed based on Kolmogorov complexity [1, 2, 3, 4]. Kolmogorov complexity of a finite sequence is defined as the minimal length of a program which generates the sequence with a given universal machine. Let s be a finite sequence and u be a universal machine. Then, Kolmogorov complexity of s for u, $K_u(s)$ is written as

$$K_u(s) := \min_{p:u(p)=s} r(p), \tag{1}$$

where p is a program that generates s by u and $r(p)$ is the length of the program p. Then, a finite binary sequence is regarded as random if its Kolmogorov complexity is almost equal to its length. In other words, a finite binary sequence which cannot be compressed is random. We can define randomness for given finite sequences by its notion, however, it is shown to be impossible to compute Kolmogorov complexity. Furthermore, this complexity depends on a choice of universal machine. Other definitions for randomness which is associated with the definition by Kolmogorov complexity have been proposed by Demuth [5], Martin-Löf [6, 7] and Schnorr [8, 9, 10].

In cryptography, an approach based on a notion of indistinguishability has been proposed. Let $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ be sequences of random variables. We say $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ are computationally indistinguishable if for any probabilistic polynomial time algorithm $\mathcal{A}$ and positive polynomial p, there exists k_0 such that

$$\left|\Pr[\mathcal{A}(X_k) = 1] - \Pr[\mathcal{A}(Y_k) = 1]\right| < \frac{1}{p(k)}, \tag{2}$$

for any $k > k_0$. With this computational indistinguishability, we can define a cryptographically secure pseudo random number generator (CSPRNG). Let g_n be a map from $\{0,1\}^n$ to $\{0,1\}^{h(n)}$, where $h(n)$ is a polynomial of n. We say that a sequence of maps $g = \{g_n\}_{n\geq 1}$ is CSPRNG if this sequence satisfies the following three properties:

- The relation $n < h(n)$ holds for any $n \geq 1$.
- For any $n \geq 1$ and input $x \in \{0,1\}^n$, there exists a deterministic algorithm for computing $g_n(x)$ in polynomial time depending on n.
- $\{g(U_n)\}_{n\geq 1}$ and $\{U_{h(n)}\}_{n\geq 1}$ are computationally indistinguishable, where U_n is a random variable uniformly distributed on $\{0,1\}^n$.

Pseudo random numbers in cryptography are referred to as sequences generated by a CSPRNG. Under some assumptions, several algorithms for CSPRNG such as Blum–Blum–Shub algorithm [11] and Blum–Micali algorithm [12] have been proposed. However, the definition based on computationally indistinguishability also has obstacles such as:

- For an arbitrary finite sequence, there exist CSPRNG and $x \in \{0, 1\}^*$ such that the CSPRNG outputs the sequence when x is input. Thus, it is meaningless to define "randomness" for a specific sequence by the above definition.
- It is impossible to verify whether a sequence satisfies Eq. (2) for any probabilistic polynomial time algorithm.
- There exists no sequence satisfying the definition if P = NP.

To summarize, it is not easy to define randomness to a specific finite sequence.

In spite of the difficulty of defining randomness, "random sequences" are widely used in many fields such as numerical simulations (e.g., Monte Carlo method), randomized algorithm (e.g., Simulated Annealing), and cryptography (e.g., key generation). In engineering applications, a sequence generated by a hardware (or physical) random number generator (HRNG) or a pseudo random number generator (PRNG) are extensively used. Note that PRNG does not mean CSPRNG. An HRNG is a device for generating numbers from a physical process such as thermal noise in a transistor, whereas a PRNG is a deterministic algorithm for generating numbers. In general, it is not easy to predict the output generated by an HRNG. On the other hand, a PRNG is important in practice because it generates a sequence much faster than an HRNG and a sequence can be reproductive by using the same initial value called seed. These advantages are useful in many situations such as numerical experiments. However, both generators are nothing but generate a number sequence regarded as a random sequence approximately which is required for applications. Hence, such sequences are necessary to be examined whether it satisfy the properties as "random sequences" or not.

1.2 Tests for randomness

As seen in the previous subsection, it is a hard task to evaluate whether a finite binary sequence satisfies "randomness" or not. In practice, a statistical hypothesis test has been extensively considered to evaluate the randomness of a binary sequence. A number of statistical tests has been proposed. In most of the cases, the randomness is evaluated by multiple statistical tests since one statistical test is designed to detect the specific defect of a binary sequence and cannot detect other types of defects.

There are some test suites such as TestU01 test suite [13], the BSI (Bundesamt fr Sicherheit in der Informationstechnik) test suite [14, 15], Marsaglia's DIEHARD test suite, Crypt-X statistical test suite [16] and FIPS 140-2 test suite [17].

NIST Special Publication 800-22 (NIST SP 800-22) [18, 19] proposed by National Institute of Standards and Technology (NIST) is one of the standard statistical test suites that was originally used for selecting Advanced Encryption Standard (AES) [20]. NIST SP 800-22 consists of fifteen kinds of statistical tests, and the null hypothesis $\mathcal{H}_0$ is that a given binary sequence is truly random.

We can regard a n-bit given binary sequence as a sample from a uniform distribution on $\{0,1\}^n$. Associated with the null hypothesis, the alternative hypothesis $\mathcal{H}_1$ is that the sequence is not random. NIST SP 800-22 specifies the evaluating process as follows. For a tested binary sequence of length n, a p-value is computed. If p-value is equal or larger than α, the null hypothesis $\mathcal{H}_0$ is accepted, where α is the significance level of the test. Repeat the same procedure for m sample sequences and obtain m p-values. NIST SP 800-22 recommends to perform additional statistical tests for the m p-values. The following two tests are specified under the null hypothesis that m p-values independently follow a uniform distribution in $[0,1]$.

1. (Proportion test) Let m_p be the number of sequences whose p-value satisfies p-value $\geq \alpha$ for the given m sequences. The null hypothesis is rejected if m_p lies outside the significant interval $[m(1-\alpha) - \xi\sigma, m(1-\alpha) + \xi\sigma]$, where $m(1-\alpha)$ and $\sigma = \sqrt{m\alpha(1-\alpha)}$ is the expected value and standard deviation of m_p, respectively.

2. (Uniformity test) The distribution of the m p-values against the uniform distribution on $[0,1]$ is tested with a Chi-Square goodness of fit test in k bins. This is again a statistical test, which yields a level-two p-value p_T. Given a significance level α_T, the null hypothesis is rejected if $p_T \leq \alpha_T$.

NIST SP 800-22 recommends to choose parameters as $m = 1000$, $\alpha = 0.01$, $\xi = 3$, $k = 10$ and $\alpha_T = 0.0001$.

1.3 Outline

This thesis is organized as follows.

In Section 2, we introduce the statistical tests proposed by Maurer in 1992 [21] and by Coron in 1999 [22]. These statistical tests are referred to as "Maurer's universal test" and "Coron's universal test", respectively. Coron's universal test is an improvement of Maurer's universal test. We also introduce the statistical test proposed by Yamamoto and Liu in 2016 [23]. The test is referred to as "highly sensitive universal statistical test" and the test is constructed on the basis of Maurer's universal test and Coron's universal test. In the highly sensitive test, a tested sequence is converted to another binary sequence as each bit of a tested sequence is stochastically flipped under a certain distribution. It is suggested in [23] that the converting make the test more sensitive.

In Section 3, we derive one and two dimensional distributions. We need these distribution to derive the variance for the reference distribution of the highly sensitive test. In existing literature, the theoretical results for a truly random sequence without flipping have been obtained. However, we cannot apply the results directly to the highly sensitive test since a tested sequence is biased.

In Section 4, we derive the theoretical variance for the reference distribution of the highly sensitive test. We also show some results of experiments in this section. Firstly, we show the variance can be computed accurately by the derived equation. Secondly, we show the fitted curve for computing the variance for effectively. Thirdly, we show the difference between the highly sensitive test with proposed parameter and the test with the existing value.

In Section 5, we conclude this thesis.

2 Universal statistical test

In this section, we introduce "Maurer's universal statistical test" [21], "Coron's universal statistical test' [22] and "highly sensitive universal statistical test" [23].

2.1 Maurer's universal statistical test

Maurer's universal test is one of the tests included in NIST SP 800-22 [18, 19]. Maurer's universal statistical test aims at detecting non-randomness based on the test statistic value which is relating to the source's entropy, and the non-randomness is evaluated whether the computed entropy attains the maximum or not. Unlike other types of statistical tests which is designed to detect specific defects, Maurer's universal test is able to detect a wide range of statistical defects.

Let us consider an information source S which generates a sequence $U_1, U_2, \ldots$. For each i, we regard U_i as a sample from a random variable. A source S is called a finite memory source if there exists a positive integer M such that the conditional probability of U_n, given $U_1, \ldots, U_{n-1}$, depends only on the most previous M bits, i.e.,

$$P_{U_n|U_{n-1}\ldots U_1}(u_n \mid u_{n-1} \ldots u_1) = P_{U_n|U_{n-1}\ldots U_{n-M}}(u_n \mid u_{n-1} \ldots u_{n-M}), \tag{3}$$

for $n > M$ and for every binary sequence $u_1, \ldots, u_n \in \{0,1\}^n$. The smallest M satisfying Eq. (3) is called the memory of the source. The probability distribution of U_n is thus determined by the source's state $\Lambda_n = U_{n-M}, \ldots, U_{n-1}$ at time n. Let $\Lambda_1 = U_0 \ldots U_{-M+1}$ be the initial state where $U_{-M+1} \ldots U_0$ are dummy random variables. Then, the source is called *stationary*, if the information source S satisfies the following relation in addition to Eq. (3),

$$P_{U_n|\Lambda_n}(u \mid \lambda) = P_{U_1|\Lambda_1}(u \mid \lambda), \tag{4}$$

for all $n > M$, $u \in \{0,1\}$ and $\lambda \in \{0,1\}^M$. Therefore, it depends only on the most previous M-bit sequence when a sequence is generated by a stationary source. On the other hand, each bit is independent of the previous sequence, and the probability of taking "1" or "0" is exactly the same as $\frac{1}{2}$ if we say a sequence is truly random in intuitive sense.

The formulation of Maurer's universal test is motivated by the universal source coding algorithms that has been proposed in [24, 25] and the procedure is described as follows. In the following, let B be the set $\{0,1\}$, and $x^n = x_1, x_2, \ldots, x_n \in B^n$ be a binary sequence of length n, where $x_i \in B$. The test takes as input three positive integers L, Q and K, and a binary sequence $s^n \in B^n$ generated by a tested source. The sequence is divided into adjacent non-overlapping blocks of length L. Then, the first Q blocks ($L \times Q$-bits) are used for initialization, and the remaining K blocks ($L \times K$-bits) are used for the test. Without loss of generality, we can assume that $n = L \times (Q + K)$ holds[1]. Let $b_k(x^n)$ be the k-th block of x^n, i.e., $b_k(x^n) = x_{L(k-1)+1}, x_{L(k-1)+2}, \ldots, x_{Lk}$. Considering the situation that the $(n-m)$-th block takes the same value with the n-th block and the blocks from $(n-m+1)$-th to $(n-1)$-th blocks do not take the value, we define an integer-valued variable $A_n(x^n)$ as m. Figure 1 illustrates an example for

[1] If the relation $n = L \times (Q + K)$ does not hold, then let K be $\lfloor \frac{n}{L} \rfloor - Q$.

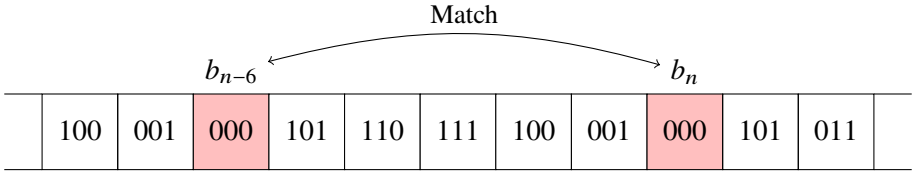

Figure 1: An example of the situation of $A_n = 6$ for $L = 3$.

$L = 3$. The test function f_M, which maps a binary sequence to a real number, is defined by

$$f_M(x^n) = \frac{1}{K} \sum_{n=Q+1}^{Q+K} \log_2 A_n(x^n), \tag{5}$$

where $A_n(x^n)$ is defined by

$$A_n(x^n) = \begin{cases} n, & \text{if } b_{n-m}(x^n) \neq b_n(x^n) \text{ for } 1 \leq m \leq n-1, \\ \min\{m \in \mathbb{N} \mid m \geq 1,\ b_{n-m}(x^n) = b_n(x^n)\}, & \text{otherwise,} \end{cases} \tag{6}$$

for $n = Q+1, Q+2, \ldots, Q+K$.

In order to evaluate the non-randomness of a given binary sequence, it is necessary to derive the expected value and the variance of the reference distribution for a truly random sequence. Note that a truly random sequence is a binary sequence generated by a uniform distribution on $\{0,1\}^n$. Under the assumption $Q \to \infty$, the expectation and the variance are given by

$$\mathbb{E}[f_M(R^n)] = 2^{-L} \sum_{i=1}^{\infty} (1 - 2^{-L})^{i-1} \log_2 i, \tag{7}$$

$$\sigma_M^2 = c_M(L,K)^2 \times \frac{\mathrm{Var}[\log_2 A_n(R^n)]}{K}, \tag{8}$$

where R^n denotes a truly random sequence of length n. In [21], the following approximation is empirically proposed:

$$c_M(L,K) \simeq 0.7 - \frac{0.8}{L} + \left(1.6 + \frac{12.8}{L}\right) K^{-4/L}. \tag{9}$$

The approximation above has been obtained by numerical simulations. In [26], the accurate expression of $c_M(L,K)$ has been obtained theoretically. However, it requires much cost to compute the $c_M(L,K)$ for given L and K, and hence an approximation of the theoretical form is given as follows:

$$c_M(L,K)^2 \simeq d_M(L) + \frac{e_M(L) \times 2^L}{K}, \tag{10}$$

where $d_M(L)$ and $e_M(L)$ are listed in [26] for $L = 3, 4, \ldots, 16$. The approximation in Eq. (10) is accurate for $K \geq 33 \times 2^L$ in practice. Note that $\mathrm{Var}[\log_2 A_n]$ in Eq. (8) can be computed by the

definition of variance by

$$\mathrm{Var}[\log_2 A_n(R^n)] = 2^{-L}\sum_{i=1}^{\infty}(1-2^{-L})^{i-1}(\log_2 i)^2 - (\mathbb{E}[f_M(R^n)])^2. \tag{11}$$

To implement the test, it is necessary to set the parameters. In [21], the study recommends to set parameters $L\,Q$ and K as $6 \le L, \le 16$, $Q \ge 10 \times 2^L$ and $K \ge 1000 \times 2^L$, respectively. The study also insists that rejection rate ρ should be chosen as $\rho \in [0.001, 0.01]$. Then, it is concluded that the null hypothesis of Maurer's test[2] is rejected if either $f_M(x^n) < t_1$ or $f_M(x^n) > t_2$ holds, where the thresholds t_1 and t_2 are written as

$$\begin{aligned} t_1 &= \mathbb{E}[f_M(R^n)] - y\sigma_M, \\ t_2 &= \mathbb{E}[f_M(R^n)] + y\sigma_M. \end{aligned} \tag{12}$$

Using the complementary error function erfc defined by

$$\mathrm{erfc}(z) = \frac{2}{\sqrt{\pi}}\int_z^{\infty} \mathrm{e}^{-u^2}\,\mathrm{d}u, \tag{13}$$

the value y in Eq. (12) is given as $\mathrm{erfc}(y) = \frac{\rho}{2}$. Notice that it is implicitly assumed that $f_M(R^n)$ follows a normal distribution.

In NIST SP 800-22, the non-randomness is evaluated by the p-value shown as

$$p\text{-value}_M = \mathrm{erfc}\left(\left|\frac{f_M(x^n) - \mathbb{E}[f_M(R^n)]}{\sqrt{2}\sigma_M}\right|\right), \tag{14}$$

where erfc is the complementary error function defined by Then, the null hypothesis of Maurer's test is rejected if $p\text{-value}_M < \alpha$, where α is a significance level.

There is much concern in the asymptotic relation between the Maurer's test statistic and the source's per-bit entropy. In [21], the expected value of the test statistic for a truly random sequence $\mathbb{E}[f_M(R^n)]$ is closely related to the entropy of blocks. It has been shown that the following relation holds:

$$\lim_{L\to\infty}[\mathbb{E}[f_M(R^n)] - L] = C, \tag{15}$$

where C is a constant whose value is equal to $-\frac{\ln 2}{\gamma} \simeq -0.8327$ and γ is Euler's constant [27]. We provide the proof of $C = -\frac{\ln 2}{\gamma}$ is given in Appendix A. Let us consider a binary sequence $U^n_{\mathrm{BMS}_p}$ generated by a binary memoryless source BMS_p. Note that a sequence $U^n_{\mathrm{BMS}_p}$ follows a distribution on $\{0,1\}^n$ taking each bit to be "1" with probability $p \in (0,1)$ independently. For a binary sequence $U^n_{\mathrm{BMS}_p} \in B^n$, the following relation

$$\lim_{L\to\infty}\left[\mathbb{E}[f_M(U^n_{\mathrm{BMS}_p})] - L \times H(p)\right] = C \tag{16}$$

[2]A null hypothesis of Maurer's test is that a binary sequence of length n follows a uniform distribution on $\{0,1\}^n$

holds for any $p \in (0, 1)$, where H is the binary entropy function corresponding to $\Pr[X = 1] = p$. Here, X is a random variable. In [21], a similar result has been studied to show for a binary sequence U_s^n generated by every ergodic stationary source S. The study in [26] develops the idea and proves that the following relation holds:

$$\lim_{L\to\infty} \left[\mathbb{E}[f_M(U_s^n)] - K_L\right] = C, \tag{17}$$

where K_L is equivalent to the entropy of L bit blocks defined by

$$K_L = -\sum_{b\in B^n} \Pr[b] \log_2 \Pr[b]. \tag{18}$$

Other asymptotic relations between Maurer's test statistic and a source's entropy can be found in [28, 29, 30, 31, 32].

2.2 Coron's universal statistical test

As seen in the previous subsection, Maurer's universal test can detect a wide range of statistic defects modeled by an ergodic statistic source with finite memory, however, the test only provides an asymptotic measure of the source's entropy. To address the problem, a modified version of Maurer's universal test called "Coron's universal test" has been proposed in [22]. In this test, the expectation of test statistic value is exactly equal to the source's entropy. The main procedure is the same as Maurer's test except for the test statistic. The test function f_C, which maps a binary sequence to a real number, is defined by

$$f_C(x^n) = \frac{1}{K} \sum_{n=Q+1}^{Q+K} g(A_n(x^n)), \tag{19}$$

where $A_n(x^n)$ is defined by Eq. (6) and $g : \mathbb{N} \to \mathbb{R}$ is given by

$$g(m) = (\log_2 \mathrm{e}) \sum_{k=1}^{m-1} \frac{1}{k}, \tag{20}$$

for $m \geq 2$. Note that we set $g(1) = 0$. For a binary sequence U_s^n generated by an ergodic stationary source S, the expected value is exactly equal to the entropy of L blocks of S, i.e.,

$$\mathbb{E}[f_C(U_s^n)] = K_L, \tag{21}$$

where K_L corresponds to the entropy of blocks given by Eq. (18).

It is necessary to derive the expected value and the variance of the reference distribution just like Maurer's test. From Eq. (21), the expected value for R^n is obtained by

$$\mathbb{E}[f_C(R^n)] = L, \tag{22}$$

since $K_L = L$ when the source is the binary symmetric source. The variance is also obtained as

$$\sigma_C^2 = c_C(L, K)^2 \times \frac{\mathrm{Var}[g(A_n)]}{K}. \tag{23}$$

In [22], the following approximation is empirically proposed as

$$c_C(L, K) \simeq d_C(L) + \frac{e_C(L) \times 2^L}{K}, \tag{24}$$

where $d_C(L)$ and $e_C(L)$ are listed in [22] for $L = 3, 4, \ldots, 16$. Note that $\mathrm{Var}[g(A_n)]$ can be calculated by the definition of variance as

$$\begin{aligned} \mathrm{Var}[g(A_n)] &= \mathbb{E}[\{g(A_n)\}^2] - (\mathbb{E}[g(A_n)])^2 \\ &= 2^{-L} \sum_{i=2}^{\infty} (1 - 2^{-L})^{i-1} \left(\sum_{k=1}^{i-1} \frac{\log_2 \mathrm{e}}{k} \right)^2 - L^2. \end{aligned} \tag{25}$$

To obtain the second equality in the above equations, we use the relation $\mathbb{E}[g(A_n)] = \mathbb{E}[f(x^n)] = L$. When we consider a binary sequence $U^n_{\mathrm{BMS}_p}$ generated by a binary memoryless source BMS_p, we exactly have

$$\mathbb{E}[f_C(U^n_{\mathrm{BMS}_p})] = L \times H(p). \tag{26}$$

2.3 Highly sensitive universal statistical test

In the previous subsections, we have seen that Maurer's and Coron's universal statistical test can detect the non-randomness of a binary sequence. Both tests evaluate the non-randomness of a binary sequence whether the relation (15) or (22) holds for $p = 0.5$, where $q = \Pr[x_i = 1]$. It has been suggested in [23] that the deviation from $p = 0.5$ cannot be detected with high sensitivity since the derivative of the binary entropy function is equal to 0 at $p = 0.5$, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}p} H(p) \bigg|_{p=0.5} = \log_2 \frac{1-p}{p} \bigg|_{p=0.5} = 0. \tag{27}$$

In [23], the universal test called "highly sensitive universal statistical test" has been proposed. This test is constructed on the basis of Maurer's and Coron's universal statistical tests. In the highly sensitive test, a given binary sequence $x_1, x_2, \ldots$ with $q \simeq 0.5$ is converted into another binary sequence $\hat{x}_1, \hat{x}_2, \ldots$ with $\hat{q} \neq 0.5$ by

$$\begin{aligned} \Pr[\hat{x}_i = 0 \mid x_i = 0] &= 1, \\ \Pr[\hat{x}_i = 1 \mid x_i = 1] &= \beta, \end{aligned} \tag{28}$$

where $\hat{q} = \Pr[\hat{x}_i = 1]$ and $\beta \in (0, 1)$. Note that "0" in a sequence is not converted, and "1" in a sequence is flipped into "0" in probability with $1 - \beta$. By Eq. (28), a given sequence is converted into another binary sequence with $\hat{q} = 0.5\beta$. In [23], the non-randomness can be detected by applying Coron's universal test for a converted binary sequence. Numerical experiments show the effectiveness of the highly sensitive test and $\beta = 0.66$ maximizes the effectiveness.

The null hypothesis under the highly sensitive test $\mathcal{H}_0$ is that a binary sequence of length n being tested follows a uniform distribution on $\{0, 1\}^n$. As an implicit assumption, an additional hypothesis $\widetilde{\mathcal{H}}_0$ that a random number used for flipping is ideal needs to be considered. Hence,

the null hypothesis under the highly sensitive test $\overline{\mathcal{H}}_0 := \mathcal{H}_0 \wedge \widetilde{\mathcal{H}}_0$ is that a converted binary sequence is considered to be generated by a distribution on $\{0,1\}^n$ taking each bit to be "1" with probability $\hat{q}$ independently. If the null hypothesis $\overline{\mathcal{H}}_0$ is rejected, then the hypothesis $\mathcal{H}_0$ would be rejected. Otherwise, there is no evidence for rejecting the null hypothesis $\overline{\mathcal{H}}_0$. The algorithm of the highly sensitive test is described in Algorithm 1.

Algorithm 1 The procedure of highly sensitive universal statistical test

1: Set parameters α, L, Q, K and β.
2: Convert a given binary sequence s^n into $\hat{s}^n$ by Eq. (28).
3: Divide a converted binary sequence $\hat{s}^n$ into adjacent non-overlapping blocks of length L, and compute $A_n(\hat{s}^n)$ for $n = Q+1, Q+2, \ldots, Q+K$ by Eq.(6).
4: Compute a test statistic value $f_C(\hat{s}^n)$ by Eq. (19).
5: Compute a p -value by

$$p\text{-value} = \operatorname{erfc}\left(\left|\frac{f_C(\hat{s}^n) - L \times H(0.5\beta)}{\sqrt{2}\sigma_C(0.5\beta)}\right|\right). \tag{29}$$

6: Reject $\mathcal{H}_0$ if p-value $< \alpha$; else accept $\overline{H}_0$.

To implement the highly sensitive test, it is necessary to derive the expected value and the variance of reference distribution of a binary sequence. By Eq. (26), the expected value has been obtained, whereas the variance has not been analyzed theoretically. Then, a value obtained by simulation is used as the variance in [23]. The accurate variance should be derived to improve the reliability of the highly sensitive test.

3 Distribution

In this section, we consider the distribution of $A_n(\hat{x}^n)$, where $\hat{x}^n$ is an n-bit random variable. For each i, we have

$$\Pr[(\hat{x}^n)_i = 1] = \hat{q}, \tag{30}$$

where $(\hat{x}^n)_i$ is the i-th bit of $\hat{x}^n$. Each bit $(\hat{x}^n)_i$ is independent from other bits.

For simplicity, we write $A_n(\hat{x}^n)$ as A_n unless specified. In the following, we consider an assumption $Q \to \infty$, and due to the situation, the index of A_n should be replaced as illustrated in Figure 2. Then, a sequence of $\{A_k\}_{k=1}^{K}$ follows a stationary ergodic process, that is, the joint distribution of $\{A_k\}_{k=n}^{n+m}$ only depends on m. In this section, we derive a marginal distribution of A_n and a joint distribution of (A_n, A_{n+k}) necessary for calculating the variance of the reference distribution of the highly sensitive test.

3.1 Derivation of marginal distribution

We consider the event of $\langle A_n = i \rangle$ for $i \geq 1$. The event occurs when n-th block coincides $(n-i)$-th block and do not coincide other blocks between n-th and $(n-i)$-th blocks as illustrated in Figure 3. Let $\mathcal{M}$ be such an event. Then, $\mathcal{M}$ is written as

$$\mathcal{M} = \langle b_{n-i} = b_n, b_{n-i+1} \neq b_n, \ldots, b_{n-1} \neq b_n \rangle, \tag{31}$$

where b_k is the k-th block of a sequence $\hat{x}^n$. Then, we can derive the probability of occurring $\langle A_n = i \rangle$ under the assumption that the blocks are statistically independent and identically distributed as

$$\Pr[A_n = i] = \sum_{r=0}^{L} \Pr[\mathcal{M} \mid \ell(b_n) = r] \times \Pr[\ell(b_n) = r], \tag{32}$$

where $\ell(b)$ denotes the number of "1" included in the block $b \in \{0,1\}^L$. We also have

$$\Pr[\mathcal{M} \mid \ell(b_n) = r] = w_r \times (1 - w_r)^{i-1}, \tag{33}$$

$$\Pr[\ell(b_n) = r] = \binom{L}{r} w_r, \tag{34}$$

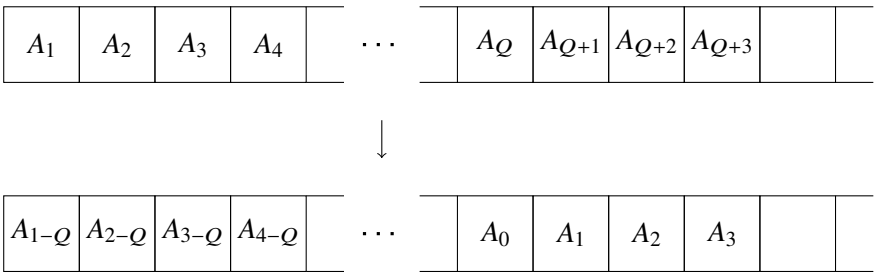

Figure 2: Replacement of the index

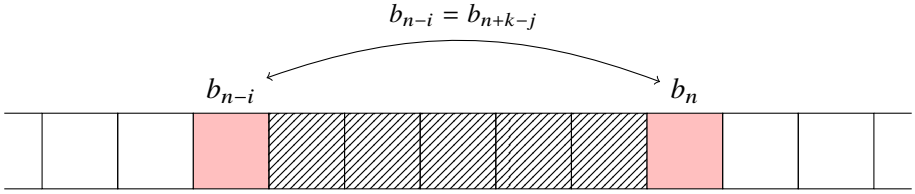

Figure 3: The arrangement of blocks in the case of $A_n = i$.

where $w_r = \hat{q}^r(1-\hat{q})^{L-r}$ and $\binom{L}{r} = \frac{L!}{r!(L-r)!}$ is a binomial coefficient. By combining Eqs. (32)–(34), the following relation is obtained as

$$\Pr[A_n = i] = \sum_{r=0}^{L} \binom{L}{r} w_r^2 (1-w_r)^{i-1}, \tag{35}$$

for $i \geq 1$.

3.2 Derivation of joint distribution

We consider the event $\langle A_n = i,\ A_{n+k} = j\rangle$ for $i \geq 1$ and $j \geq 1$. It has been shown in [26] that the following relation holds for a truly random sequence R^N

$$\begin{aligned}
&\Pr[A_n(R^N) = i,\ A_{n+k}(R^N) = j] \\
&= \begin{cases}
2^{-2L}(1-2^{-L})^{i+j-2} & (1 \leq j \leq k-1) \\
2^{-2L}(1-2^{-L})^{i+k-2} & (j = k) \\
2^{-2L}(1-2^{-L})^{i-j+2k-1}\left(1-2^{-L+1}\right)^{j-k-1} & (k+1 \leq j \leq k+i-1) \\
0 & (j = k+i) \\
2^{-2L}(1-2^{-L})^{-i+j-1}\left(1-2^{-L+1}\right)^{i-1} & (j \geq k+i+1)
\end{cases}.
\end{aligned} \tag{36}$$

In this subsection, we derive the joint distribution of $(A_n(\hat{x}^n), A_{n+k}(\hat{x}^n))$ holding for any $\hat{q} \in (0,1)$.

3.2.1 Case of $1 \leq j \leq k-1$

When $1 \leq j \leq k-1$, the events $\langle A_n = i\rangle$ and $\langle A_{n+k} = j\rangle$ are independent each other as illustrated in Figure 4, since there are no overlapping between blocks from b_{n-i} to b_n and blocks from b_{n+k-j} to b_{n+k}. Thus, we obtain the joint distribution as

$$\begin{aligned}
\Pr[A_n = i, A_{n+k} = j] =& \Pr[A_n = i] \times \Pr[A_{n+k} = j] \\
=& \left(\sum_{r=0}^{L} \binom{L}{r} w_r^2 (1-w_r)^{i-1}\right) \times \left(\sum_{r=0}^{L} \binom{L}{r} w_r^2 (1-w_r)^{j-1}\right).
\end{aligned} \tag{37}$$

In the above relations, the second equality is obtained from Eq. (35). Recall that $w_r = \hat{q}^r(1-\hat{q})^{L-r}$ where $\hat{q} \in (0,1)$.

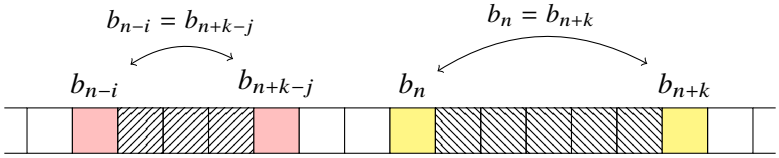

Figure 4: An example of the arrangement of blocks in the case of $1 \leq j \leq k-1$

3.2.2 Case of $j = k$

For every $b \in B^L$, we consider the event $e_2(b) = \langle A_n = i,\ A_{n+k} = j,\ b_n = b \rangle$ for $j = k$. An example for the arrangement of blocks is illustrated in Figure 5. The event $e_2(b)$ can be written as

$$\begin{aligned} e_2(b) = &\langle b_{n-i} = b, b_n = b, b_{n+k} = b \rangle \\ &\wedge \langle b_{n-i+1} \neq b, \ldots, b_{n-1} \neq b \rangle \\ &\wedge \langle b_{n+1} \neq b, \ldots, b_{n+k-1} \neq b \rangle . \end{aligned} \tag{38}$$

Since the blocks are statistically independent and uniformly distributed, we have

$$\Pr\left[e_2(b)\right] = w_r^3 \times (1 - w_r)^{i+k-2}. \tag{39}$$

We define $\mathcal{E}_2$ as the event of occurring $\langle A_n = i,\ A_{n+k} = j \rangle$ in the case of $j = k$. Then, $\mathcal{E}_2$ can be written by using Eq. (38) as

$$\mathcal{E}_2 = \bigvee_{b \in B^L} e_2(b). \tag{40}$$

Therefore, the joint distribution is derived as follows:

$$\begin{aligned} \Pr[A_n = i,\ A_{n+k} = j] &= \Pr[\mathcal{E}_2] \\ &= \Pr\left[\bigvee_{b \in B^L} e_2(b)\right] \\ &= \sum_{b \in B^L} \Pr[e_2(b)] \\ &= \sum_{b \in B_0^L \cup \cdots \cup B_L^L} \Pr[e_2(b)] \\ &= \sum_{r=0}^{L} \sum_{b \in B_r^L} \Pr\left[e_2(b)\right] \\ &= \sum_{r=0}^{L} \left(\sharp B_r^L\right) \Pr\left[e_2(b)\right] \\ &= \sum_{r=0}^{L} \binom{L}{r} w_r^3 (1 - w_r)^{i+k-2}, \end{aligned} \tag{41}$$

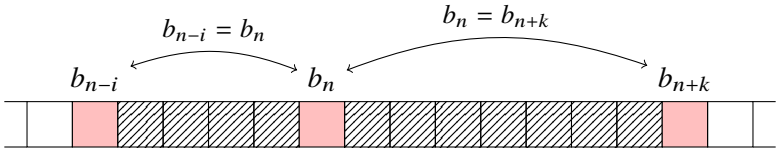

Figure 5: An example of the arrangement of blocks in the case of $j = k$

where $B_r^L := \{b \in B^L \mid \ell(b) = r\}$ and $\sharp\mathcal{S}$ be the number of elements included in a finite set $\mathcal{S}$. The sixth equality in Eq. (41) has been obtained with the fact that $\Pr[e_2(b)]$ depends only on $\ell(b)$.

3.2.3 Case of $k + 1 \leq j \leq k + i - 1$

For every $b,\ b' \in B^L$, we consider the event $e_3(b, b')\ \langle A_n = i,\ A_{n+k} = j,\ b_n = b,\ b_{n+k} = b' \rangle$ for $k + 1 \leq j \leq k + i - 1$. An example for the arrangement of blocks is illustrated in Figure 6. The event $e_3(b, b')$ can be written as

$$\begin{aligned} e_3\,(b, b') = &\left\langle b_{n-i} = b, b_n = b, b_{n+k-j} = b', b_{n+k} = b' \right\rangle \\ &\wedge \left\langle b_{n-i+1} \neq b, \ldots, b_{n+k-j-1} \neq b \right\rangle \\ &\wedge \left\langle b_{n+k-j+1} \neq b, \ldots, b_{n-1} \neq b \right\rangle \\ &\wedge \left\langle b_{n+k-j+1} \neq b', \ldots, b_{n-1} \neq b' \right\rangle \\ &\wedge \left\langle b_{n+1} \neq b', \ldots, b_{n+k-1} \neq b' \right\rangle . \end{aligned} \tag{42}$$

Since the blocks are statistically independent and uniformly distributed, we have

$$\Pr\left[e_3\,(b, b')\right] = w_{r_1}^2 w_{r_2}^2 (1 - w_{r_1})^{i-j+k-1} (1 - w_{r_1} - w_{r_2})^{j-k-1} (1 - w_{r_2})^{k-1}. \tag{43}$$

We define $\mathcal{E}_3$ as the event $\langle A_n = i,\ A_{n+k} = j \rangle$ in the case of $k + 1 \leq j \leq k + i - 1$, $\mathcal{E}_3$ can be written by using $e_3(b, b')$ as

$$\mathcal{E}_3 = \bigvee_{b \in B^L} \bigvee_{b' \in B^L \setminus \{b\}} e_3\,(b, b')\,. \tag{44}$$

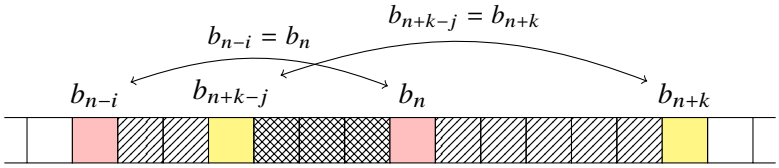

Figure 6: An example of the arrangement of blocks in the case of $k+1 \le j \le k+i-1$

Therefore, we can derive the joint distribution as

$$
\begin{aligned}
&\Pr[A_n = i,\ A_{n+k} = j] \\
&= \Pr[\mathcal{E}_3] \\
&= \Pr\left[\bigvee_{b_1 \in B^L} \bigvee_{b_2 \in B^L \setminus \{b_1\}} e_3(b, b')\right] \\
&= \sum_{b_1 \in B^L} \sum_{b_2 \in B^L \setminus \{b_1\}} \Pr\left[e_3(b, b')\right] \\
&= \sum_{r_1=0}^{L} \sum_{b_1 \in B_{r_1}^L} \sum_{r_2=0}^{L} \sum_{b_2 \in B_{r_2}^L \setminus \{b_1\}} \Pr\left[e_3(b, b')\right] \\
&= \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \sum_{b_1 \in B_{r_1}^L} \sum_{b_2 \in B_{r_2}^L} \Pr\left[e_3(b, b')\right] + \sum_{r_1=0}^{L} \sum_{r_2 \in \{r_1\}} \sum_{b_1 \in B_{r_1}^L} \sum_{b_2 \in B_{r_1}^L \setminus \{b_1\}} \Pr\left[e_3(b, b')\right] \\
&= \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1}\binom{L}{r_2} \Pr\left[e_3(b, b')\right] + \sum_{r_1=0}^{L} \sum_{r_2 \in \{r_1\}} \binom{L}{r_1} \left\{\binom{L}{r_1} - 1\right\} \Pr\left[e_3(b, b')\right],
\end{aligned}
\tag{45}
$$

where $\Pr\left[e_3(b, b')\right]$ is given in Eq. (43). The last equality in Eq. (45) holds since $\Pr[A_n = i,\ A_{n+k} = j,\ b_n = b,\ b_{n+k} = b']$ depends only on $\ell(b)$ and $\ell(b')$.

3.2.4 Case of $j = k + i$

When $j = k + i$, the events $\langle A_n = i \rangle$ and $\langle A_{n+k} = j \rangle$ do not occur coincidentally as illustrated in Figure 7. Thus, the joint distribution is obtained as

$$
\Pr[A_n = i, A_{n+k} = j] = 0. \tag{46}
$$

3.2.5 Case of $j \ge k + i + 1$

For every $b,\ b' \in B^L$, we consider the event $\langle A_n = i,\ A_{n+k} = j, b_n = b,\ b_{n+k} = b' \rangle$ occurs when $j \ge k + i + 1$. An example for the arrangement of blocks is illustrated in Figure 8. Let $e_5(b_1, b_2)$

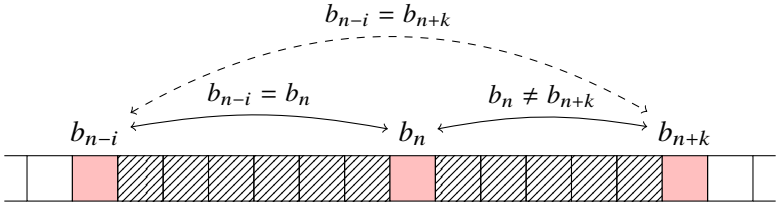

Figure 7: An example of the arrangement of blocks in the case of $j = k + i$

be the event of occurring $\langle A_n = i,\ A_{n+k} = j, b_n = b,\ b_{n+k} = b' \rangle$ which is written as

$$\begin{aligned} e_5(b, b') := &\left\langle b_{n+k-j} = b, b_{n-i} = b', b_n = b', b_{n+k} = b \right\rangle \\ &\wedge \left\langle b_{n+k-j+1} \neq b, \ldots, b_{n-i-1} \neq b \right\rangle \\ &\wedge \langle b_{n-i+1} \neq b, \ldots, b_{n-1} \neq b \rangle \\ &\wedge \langle b_{n-i+1} \neq b', \ldots, b_{n-1} \neq b' \rangle \\ &\wedge \langle b_{n+1} \neq b, \ldots, b_{n+k-1} \neq b \rangle . \end{aligned} \tag{47}$$

Since the blocks are statistically independent and uniformly distributed, we have

$$\Pr\left[e_5(b_1, b_2)\right] = w_{r_1}^2 w_{r_2}^2 (1 - w_{r_1})^{-i+j-k-1} (1 - w_{r_1} - w_{r_2})^{i-1} (1 - w_{r_2})^{k-1}. \tag{48}$$

We define $\mathcal{E}_5$ as the event $\langle A_n = i,\ A_{n+k} = j \rangle$ for $j \geq k + i + 1$. Then, $\mathcal{E}_5$ can be written by using $e_5(b_1, b_2)$ as

$$\mathcal{E}_5 = \bigvee_{b \in B^L} \bigvee_{b' \in B^L \setminus \{b_1\}} e_5(b, b'). \tag{49}$$

Therefore, we can derive the joint distribution as

$$\begin{aligned} &\Pr[A_n = i,\ A_{n+k} = j] \\ &= \Pr[\mathcal{E}_5] \\ &= \Pr\left[\bigvee_{b \in B^L} \bigvee_{b' \in B^L \setminus \{b\}} e_5(b, b') \right] \\ &= \sum_{b \in B^L} \sum_{b' \in B^L \setminus \{b\}} \Pr\left[e_5(b, b')\right] \\ &= \sum_{r_1=0}^{L} \sum_{b \in B_{r_1}^L} \sum_{r_2=0}^{L} \sum_{b' \in B_{r_2}^L \setminus \{b\}} \Pr\left[e_5(b, b')\right] \\ &= \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \sum_{b \in B_{r_1}^L} \sum_{b' \in B_{r_2}^L} \Pr\left[e_5(b_1, b_2)\right] + \sum_{r_1=0}^{L} \sum_{r_2 \in \{r_1\}} \sum_{b \in B_{r_1}^L} \sum_{b' \in B_{r_1}^L \setminus \{b\}} \Pr\left[e_5(b, b')\right] \\ &= \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1} \binom{L}{r_2} \Pr\left[e_5(b, b')\right] + \sum_{r_1=0}^{L} \sum_{r_2 \in \{r_1\}} \binom{L}{r_1} \left\{ \binom{L}{r_1} - 1 \right\} \Pr\left[e_5(b, b')\right], \end{aligned} \tag{50}$$

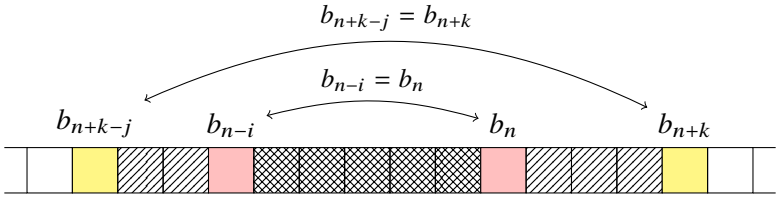

Figure 8: An example of the arrangement of blocks in the case of $j \geq k + i + 1$

where $\Pr[e_5(b_1, b_2)]$ is obtained in Eq. (48). The above relations have been obtained in the same manner in the case of $k + 1 \leq j \leq k + i - 1$.

3.2.6 Summary of the results

In the previous subsections, we have derived the marginal distribution of A_n and the joint distribution of (A_n, A_{n+k}) under the assumption of $Q \to \infty$. We give the summary of the result of the joint distribution that we have obtained in subsection 3.2. The joint distribution of (A_n, A_{n+k}) is written as follows:

$$
\begin{aligned}
&\Pr[A_n = i,\ A_{n+k} = j] \\
&= \begin{cases}
\left(\displaystyle\sum_{r=0}^{L} \binom{L}{r} w_r^2 (1 - w_r)^{i-1}\right) \times \left(\displaystyle\sum_{r=0}^{L} \binom{L}{r} w_r^2 (1 - w_r)^{j-1}\right) & (1 \leq j \leq k - 1) \\
\displaystyle\sum_{r=0}^{L} \binom{L}{r} w_r^3 (1 - w_r)^{i+k-2} & (j = k) \\
\displaystyle\sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1}\binom{L}{r_2} \Pr[e_3(b, b')] & \\
\quad + \displaystyle\sum_{r_1=0}^{L} \sum_{r_2 \in \{r_1\}} \binom{L}{r_1} \left\{\binom{L}{r_1} - 1\right\} \Pr[e_3(b, b')] & (k + 1 \leq j \leq k + i - 1) \\
0 & (j = k + i) \\
\displaystyle\sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1}\binom{L}{r_2} \Pr[e_5(b, b')] & \\
\quad + \displaystyle\sum_{r_1=0}^{L} \sum_{r_2 \in \{r_1\}} \binom{L}{r_1} \left\{\binom{L}{r_1} - 1\right\} \Pr[e_5(b, b')] & (j \geq k + i + 1)
\end{cases},
\end{aligned}
\tag{51}
$$

where $w_r = \hat{q}^r (1 - \hat{q})^{L-r}$. In Eq.(51), $\Pr[e_3(b, b')]$ and $\Pr[e_5(b, b')]$ are given in Eqs. (43) and (48), respectively.

4 The variance of references distribution

In Section 3, we have obtained the marginal distribution of A_n and the joint distribution of $(A_n,\ A_{n+k})$ for any $\hat{q} \in (0, 1)$. In this section, we provide a theoretical deviation for the variance of reference distribution of the highly sensitive test under the null hypothesis using the results given in Section 3.

4.1 Theoretical derivation of the variance

The variance of a random variable X is defined by

$$\begin{aligned} \mathrm{Var}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2, \end{aligned} \tag{52}$$

where $\mathbb{E}[X]$ is the expected value of X and $\mathrm{Var}[X]$ is the variance of X. In general, for any random variables $X_1, X_2, \ldots, X_n$, the variance of the sum of n variables is obtained by

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + 2 \sum_{1 \le i < j \le n} \mathrm{Cov}[X_i, X_j], \tag{53}$$

where $\mathrm{Cov}[X, Y]$ is the covariance defined by

$$\mathrm{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X] \times \mathbb{E}[Y]. \tag{54}$$

Let $\sigma_{C,\hat{q}}(K)^2$ be the variance for the reference distribution of the highly sensitive test with $\hat{q}$. Using Eq. (53), $\sigma_{C,\hat{q}}(K)^2$ is written as

$$\begin{aligned} \sigma_{C,\hat{q}}(K)^2 =& \mathrm{Var}[f_C(\hat{x}^n)] \\ =& \mathrm{Var}\left[\frac{1}{K}\sum_{n=1}^{K} g(A_n)\right] \\ =& \frac{1}{K^2}\left(\sum_{n=1}^{K} \mathrm{Var}[g(A_n)] + 2 \sum_{1 \le i < j \le K} \mathrm{Cov}[g(A_i), g(A_j)]\right) \\ =& \frac{1}{K^2}\left(K \times \mathrm{Var}[g(A_n)] + 2 \sum_{k=1}^{K-1} (K - k) \times \mathrm{Cov}[g(A_n), g(A_{n+k})]\right). \end{aligned} \tag{55}$$

The last equality in Eq. (55) has been obtained with the fact that the sequence of $\{A_k\}_{k=1}^{K}$ is stationary ergodic under the assumption $Q \to \infty$.

In the next place, we derive $\mathrm{Var}[g(A_n)]$ and $\mathrm{Cov}[g(A_n), g(A_{n+k})]$ in Eq. (55). First, $\mathrm{Var}[g(A_n)]$ can be written as

$$\mathrm{Var}[g(A_n)] = \mathbb{E}[\{g(A_n)\}^2] - (\mathbb{E}[g(A_n)])^2 \tag{56}$$

by Eq. (52). From the definition of expected value, the first term of the right hand side of Eq. (56) can be calculated as

$$\begin{aligned}\mathbb{E}[\{g(A_n)\}^2] &= \sum_{i=1}^{\infty}\{g(i)\}^2\Pr[A_n = i] \\ &= \sum_{i=2}^{\infty}\left[\left\{(\log_2 \mathrm{e})\sum_{k=1}^{i-1}\frac{1}{k}\right\}^2 \times \sum_{r=0}^{L}\binom{L}{r}w_r^2(1-w_r)^{i-1}\right].\end{aligned} \tag{57}$$

The second term of the right hand side of Eq. (56) is equal to $\{L \times H(\hat{q})\}^2$ from Eq. (26). Secondly, $\mathrm{Cov}[g(A_n), g(A_{n+k})]$ in Eq. (55) can be calculated as

$$\begin{aligned}\mathrm{Cov}[g(A_n), g(A_{n+k})] &= \mathbb{E}[g(A_n)g(A_{n+k})] - \mathbb{E}[g(A_n)] \times \mathbb{E}[g(A_{n+k})] \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g(i)g(j)\Pr[A_n = i,\ A_{n+k} = j] - \{L \times H(\hat{q})\}^2\end{aligned} \tag{58}$$

where $\Pr[A_n = i,\ A_{n+k} = j]$ has been obtained in Eq. (51). The first equality in Eq. (58) has been obtained from Eq. (54). The second equality in Eq. (58) has been obtained from the definition of the expected value and Eq. (26).

By combining Eqs. (55)–(58), $\sigma_{C,q}(K)^2$ is rewritten as

$$\begin{aligned}\sigma_{C,\hat{q}}(K)^2 =& \frac{1}{K}\left(\sum_{i=2}^{\infty}\left[\left\{(\log_2 \mathrm{e})\sum_{k=1}^{i-1}\frac{1}{k}\right\}^2 \times \sum_{r=0}^{L}\binom{L}{r}w_r^2(1-w_r)^{i-1}\right] - \{L \times H(\hat{q})\}^2\right) \\ &+ \frac{2}{K^2}\sum_{k=1}^{K-1}(K-k)\left\{\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g(i)g(j)\Pr[A_n = i,\ A_{n+k} = j] - \{L \times H(\hat{q})\}^2\right\}.\end{aligned} \tag{59}$$

4.2 Numerical results

In this subsection, we show some results of experiments. In the following, we approximate the infinite double sum as finite sum when we compute $\sigma_{C,\hat{q}}(K)^2$ by Eq. (59) since it is unable to compute an infinite summation by computer. Even if computing infinite sum is inevitable, we explore the covariance given in (58) to calculate the value more efficiently by computational experiment in Appendix B.

4.2.1 Experiment 1

We confirm that $\sigma_{C,\hat{q}}(K)^2$ can be computed accurately by Eq. (59) when $L = 4$. In the numerical computation, we set $Q = 10 \times 2^L$. Figure 9 shows the result of computed variance of reference distribution when $\hat{q} = 0.33$, 0.4 and 0.5. Note that we set $\hat{q} = 0.33$ since it is said to be optimal for detecting the deviation of a binary sequence in [23]. As seen in the Figure 9, the variance decreases by $O(\frac{1}{K})$. We can express the variance as $\frac{D_K(\hat{q})}{K}$. Table 1 shows the coefficient $D_K(\hat{q})$ when we approximate variance by $\frac{D_K(\hat{q})}{K}$.

In the next place, we show the result of the numerical experiment for computing an unbiased variance of binary sequences generated by a pseudo random generator. The procedure of the experiment is as follows.

Step1: Set L, Q, K, $\hat{q}$, M and N.

Step2: Generate M pieces of binary sequences $x^{n,1}, \ldots x^{n,M}$ by pseudo random number generator, where $x^{n,i}$ for $i = 1, 2, \ldots, M$ is a binary sequence of length $n = L \times (Q + K)$.

Step3: Convert each binary sequence $x^{n,i}$ into $\hat{x}^{n,i}$ with $\hat{q}$ from Eq. (28) by using pseudo random number generator.

Step4: For each converted binary sequence $\hat{x}^{n,i}$, compute the test statistical value $f_i = f_C(\hat{x}^{n,i})$ from Eq. (19).

Step5: Compute an unbiased variance defined by

$$u^2 = \frac{1}{M-1} \sum_{i=1}^{M} (f_i - \overline{f})^2, \tag{60}$$

where $\overline{f}$ is the arithmetic mean of $f_1, f_2, \ldots, f_M$.

Step6: Repeat Step2 to Step4 in N times, and obtain N unbiased variances $u_1^2, u_2^2, \ldots, u_N^2$. Then, compute the arithmetic mean value of unbiased variances by

$$\overline{u}^2 = \frac{1}{N} \sum_{i=1}^{N} u_i^2. \tag{61}$$

In the numerical simulation, we set $L = 4$, $Q = 10 \times 2^L$, $\hat{q} = 0.33$, $M = 1000$ and $N = 30$. We also set K as 10^3, 2×10^3, 4×10^3, 6×10^3, 8×10^3, 10×10^3, 12×10^3, 14×10^3 and 16×10^3.

We used Mersenne Twister as the pseudo random number generator in Step2 and Step3. Figure 10 shows the result of the experiment, and we can confirm that the simulated unbiased variance coincides with the result obtained by Eq. (59) in precisely.

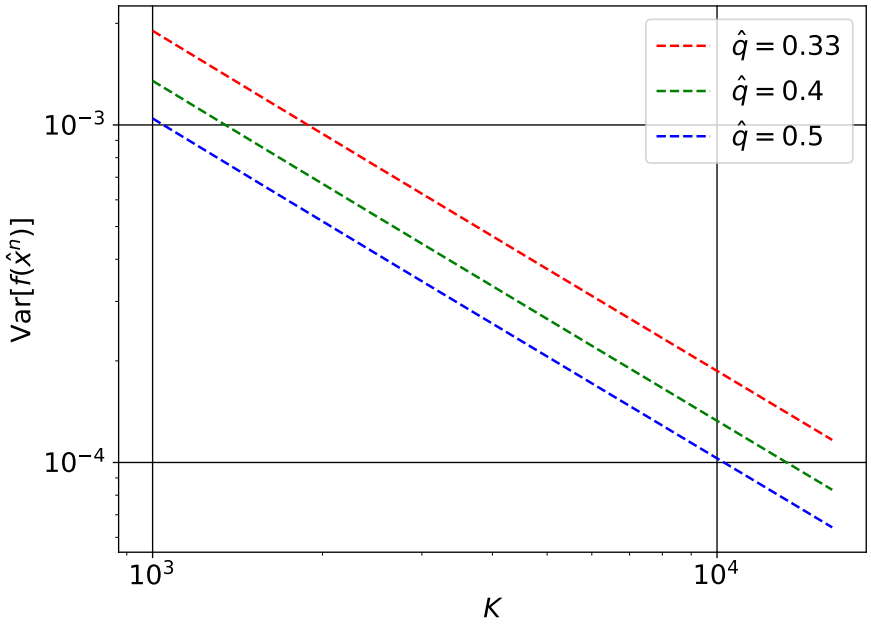

Figure 9: The variance of the reference distribution computed based on Eq. (59) with $\hat{q} = 0.33,\ 0.4$ and 0.5 (Copyright(C)2020 IEICE, [33] Figure 1)

Table 1: $D_K(\hat{q})$ for different values of $\hat{q}$ and K

$\hat{q}$	$D_{10000}(\hat{q})$	$D_{20000}(\hat{q})$	$D_{30000}(\hat{q})$	$D_{40000}(\hat{q})$
0.33	1.867364	1.865492	1.864868	1.864556
0.4	1.328692	1.327430	1.327009	1.326799
0.5	1.028395	1.027449	1.027134	1.026976

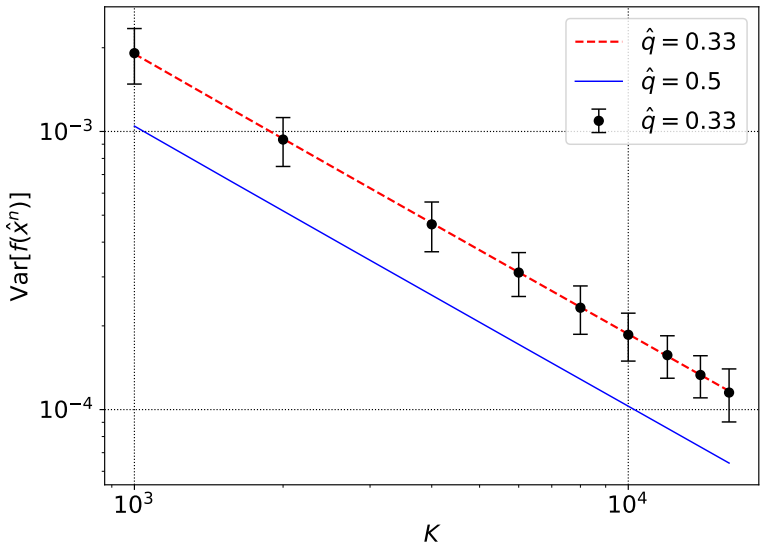

Figure 10: The variance of the reference distribution computed in the experiment. The solid blue line and the broken red line are the variances computed based on Eq. (59) with $\hat{q} = 0.5$ and $\hat{q} = 0.33$, respectively. The block points show the arithmetic mean of the unbiased variance with $\hat{q} = 0.33$ using Mersenne Twister. (Copyright(C)2020 IEICE, [33] Figure 2)

4.2.2 Experiment 2

We have seen that the variance of reference distribution of the highly sensitive test can be computed when $L = 4$. We consider the case of $L = 8$ and $\hat{q} = 0.33$ which are recommended in [23]. However, the computational cost for $L = 8$ is too high to compute directly since the recommendation value for K is $1000 \times 2^8 = 256000$. To overcome the obstacle, we derive the fitted curve. We approximate the variance of the reference distribution as

$$\sigma_{C,\hat{q}}(K)^2 = \frac{1}{K}\left(a + \frac{b}{K}\right), \tag{62}$$

where a and b are real valued constants. These constants can be obtained with any two points $(K_1, \sigma_{C,\hat{q}}(K_1)^2)$ and $(K_2, \sigma_{C,\hat{q}}(K_2)^2)$ by

$$\begin{aligned} a &= \frac{1}{K_1 - K_2}\left(K_1^2\sigma_{C,\hat{q}}(K_1)^2 - K_2^2\sigma_{C,\hat{q}}(K_2)^2\right), \\ b &= \frac{K_1K_2}{K_2 - K_1}\left(K_1\sigma_{C,\hat{q}}(K_1)^2 - K_2\sigma_{C,\hat{q}}(K_2)^2\right). \end{aligned} \tag{63}$$

Table 2 shows the pairs of (a, b) obtained from $(K_1, K_2) = (40000, 45000)$ and the standard deviation $\tilde{\sigma}_{C,\hat{q}}(K)$ for $K = 1000 \times 2^L$ obtained from Eq. (62). Figure 62 show the fitted curve. Using the fitted curve, we obtained the standard deviation for $K = 1000 \times 2^8$ as

$$\sigma_{C,0.33}(1000 \times 2^8) = 0.003488600339. \tag{64}$$

To confirm the accuracy of Eq. (64), we computed $\sigma_{C,0.33}(1000 \times 2^8)$ using MT in 10 times. For each trial, we used 4×10^6 pieces of binary sequences and set $Q = 10 \times 2^8$. Table 2 and Figure 12 show the results of this experiment. These results support that the value represented in Eq. (64) is more accurate than the value in previous study.

Table 2: The pairs of (a, b) when $(K_1, K_2) = (40000, 45000)$ and the standard deviation obtained from Eq. (62)

(K_1, K_2)	(a, b)	$\tilde{\sigma}_{C,q}(K)$
$(40000, 45000)$	$(3.112098237555, 897.7504381251)$	0.003488600339

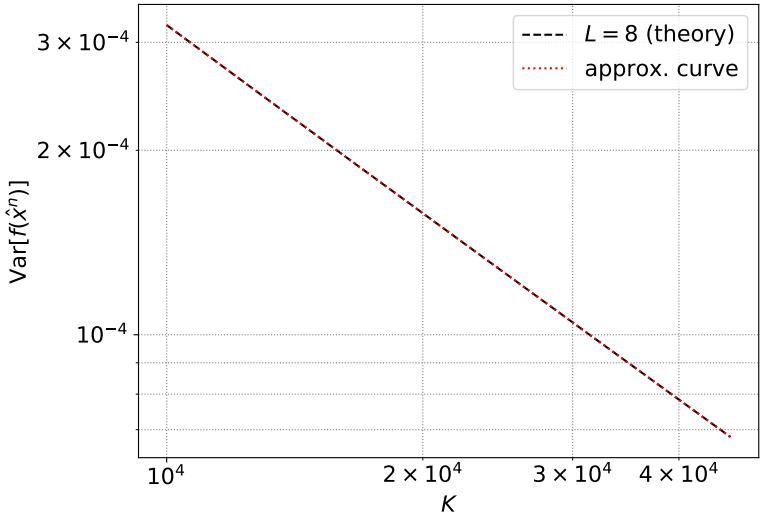

Figure 11: The fitted curve

Table 3: Value of $\sigma_{C,0.33}(1000 \times 2^8)$ computed using the Mersenne Twister

Trial No.	$\sigma_{C,0.33}(1000 \times 2^8)$
1	0.00348911
2	0.00348889
3	0.00348837
4	0.00349002
5	0.00348612
6	0.00348572
7	0.00348889
8	0.00348767
9	0.00348672
10	0.00349002
Total	$0.00348816 \pm 2.44 \times 10^{-6}$

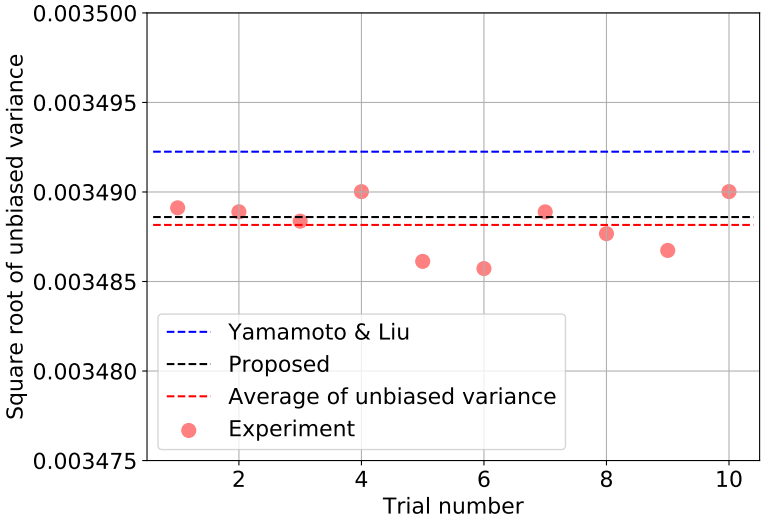

Figure 12: The variance $\sigma_{C,0.33}(1000 \times 2^8)$. Each point shows an unbiased variance derived empirically and the red broken line shows the arithmetic mean. The blue broken line shows the value given in [23]. The black broken line represented in Eq. (64).

4.2.3 Experiment 3

We investigated the difference between the value derived in Experiment 2 and the value given in [23]. In the follows, we refer to the value represented in Eq. (64) as proposed value and the value given in [23] as Yamamoto's value. We used MT [34] and AES-128 CTR [20] as pseudo random number generators.

Each test was performed for 10^5 sequences of length $n = 2068480$-bit. We divided them into 100 sets of 1000 sequences. We assigned pass or rejected for each set based on two-level tests, "proportion test" and "uniformity test", described in subsection 1.2. Recall that the significance level for uniformity test is 0.0001. Since the significance level for uniformity test is so small that we cannot observe the difference, we executed uniformity test with the significance levels 0.01 and 0.05. By the same reason, we additionally performed proportion test with $\xi = 2$ as well as $\xi = 3$.

The results of Experiment 2 imply that more sequences should be used to observe the differences between the highly sensitive test with the proposed value and the test with the Yamamoto's value. We cannot expect to get any meaningful results because there are some approximation errors. For instance, we assume that the p-value takes any real value in $[0, 1]$, but practically it can take only discrete values. When we use an enormous number of sequences, we cannot avoid the effect of such errors. Thus, the purpose of the experiments is only to confirm that the proposed value does not cause any problem in practical situation.

Tables 4 and 5 show the results of proportion test with $\xi = 3$ and with $\xi = 2$, respectively. Tables 6, 7 and 8 show the results of uniformity test with the significance levels 0.0001, 0.01 and 0.05, respectively. Note that "MT/AES" means that a tested sequence is generated by Mersenne Twister and AES-128 CTR is used for flipping each bit. "AES/MT" and "AES/AES" are defined in the same manner. The results in Tables 4, 5, 6, 7 and 8 support that the proposed value is robust.

Table 4: Number of sets rejected by proportion test with $\xi = 3$

	Yamamoto [23]	Proposed
MT/AES	0	0
AES/MT	0	0
AES/AES	0	0

Table 5: Number of sets rejected by proportion test with $\xi = 2$

	Yamamoto [23]	Proposed
MT/AES	2	3
AES/MT	4	4
AES/AES	6	6

Table 6: Number of sets rejected by uniformity test with the significance level 0.0001

	Yamamoto [23]	Proposed
MT/AES	0	0
AES/MT	0	0
AES/AES	0	0

Table 7: Number of sets rejected by uniformity test with the significance level 0.01

	Yamamoto [23]	Proposed
MT/AES	0	0
AES/MT	0	1
AES/AES	0	0

Table 8: Number of sets rejected by uniformity test with the significance level 0.05

	Yamamoto [23]	Proposed
MT/AES	2	2
AES/MT	3	6
AES/AES	4	5

5 Conclusion

In this thesis, we have theoretically derived the variance for the reference distribution of highly sensitive universal statistical test. In deriving process, the marginal distribution of A_n and the joint distribution of (A_n, A_{n+k}) have been derived theoretically with the fact that the sequence of $\{A_k\}_{k=1}^{K}$ can be considered as stationary ergodic under the assumption of $Q \to \infty$. We have shown that the value of the variance can be numerically computed for the block size $L = 4$. Because of computational cost, we used an fitted curve to get the numerical value of the variance for $L = 8$. Since the value obtained by the fitted curve may have some error, we have compared with the unbiased variance computed from binary sequences generated by a pseudo random number generator. By this experiment, we have confirmed that the value obtained from the fitted curve is more consistent with the experimental result than the existing value which has been obtained by a numerical experiment. We can state that the value of the variance in this thesis is superior subject to the existing one. Thus, we recommend that the value obtained in this thesis should be used when the highly sensitive universal statistical test is performed so that randomness of binary sequences can be tested more precisely.

Acknowledgments

The author would like to express his sincere gratitude to Professor Ken Umeno for his guidance and helpful advice. I would also like to express my appreciation for Assistant Professor Atsushi Iwasaki for his support and encouragement in keeping my progress.

Furthermore, I would like to extend my thanks to the member of Physical Statistical Laboratory for their support during my study. I would especially grateful to Dr. Hirofumi Tsuda and Mr. Shinji Kakinaka for their extended discussions and valuable suggestions which have greatly contributed to the improvement of this thesis.

Finally, I wish to thank my family for their support and encouragement throughout my study.

References

[1] A. N. Kolmogorov, "Three approaches to the quantitative definition of information," *International Journal of Computer Mathematics*, vol. 2, no. 1-4, pp. 157–168, 1968.

[2] G. J. Chaitin, "On the length of programs for computing finite binary sequences," *Journal of the ACM (JACM)*, vol. 13, no. 4, pp. 547–569, 1966.

[3] G. J. Chaitin, "On the length of programs for computing finite binary sequences: statistical considerations," *Journal of the ACM (JACM)*, vol. 16, no. 1, pp. 145–159, 1969.

[4] G. J. Chaitin, "A theory of program size formally identical to information theory," *Journal of the ACM (JACM)*, vol. 22, no. 3, pp. 329–340, 1975.

[5] O. Demuth, "Remarks on the structure of tt-degrees based on constructive measure theory," *Commentationes Mathematicae Universitatis Carolinae*, vol. 29, no. 2, pp. 233–247, 1988.

[6] P. Martin-Löf, "The definition of random sequences," *Information and Control*, vol. 9, no. 6, pp. 602–619, 1966.

[7] P. Martin-Löf, "Complexity oscillations in infinite binary sequences," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 19, no. 3, pp. 225–230, 1971.

[8] C. P. Schnorr, "A unified approach to the definition of random sequences," *Mathematical systems theory*, vol. 5, no. 3, pp. 246–258, 1971.

[9] C. P. Schnorr, "Process complexity and effective random tests," *Journal of Computer and System Sciences*, vol. 7, no. 4, pp. 376–388, 1973.

[10] C. P. Schnorr and P. Fuchs, "General random sequences and learnable sequences," *The Journal of Symbolic Logic*, vol. 42, no. 3, pp. 329–340, 1977.

[11] L. Blum, M. Blum, and M. Shub, "A simple unpredictable pseudo-random number generator," *SIAM Journal on Computing*, vol. 15, no. 2, pp. 364–383, 1986.

[12] M. Blum and S. Micali, "How to generate cryptographically strong sequences of pseudo-random bits," *SIAM journal on Computing*, vol. 13, no. 4, pp. 850–864, 1984.

[13] P. L'Ecuyer and R. Simard, "Testu01: Ac library for empirical testing of random number generators," *ACM Transactions on Mathematical Software (TOMS)*, vol. 33, no. 4, p. 22, 2007.

[14] W. Schindler, "Functionality classes and evaluation methodology for deterministic random number generators," *Anwendungshinweise and Interpretation (AIS)*, pp. 5–11, 1999.

[15] W. Killmann and W. Schindler, "A proposal for: Functionality classes and evaluation methodology for true (physical) random number generators," *T-Systems debis Systemhaus Information Security Services and Bundesamt für Sicherheit in der Informationstechnik (BSI), Tech. Rep*, 2001.

[16] W. Caelli, E. Dawson, L. Nielsen, and H. Gustafson, "Crypt–x statistical package manual," *Measuring the Strength of Stream and Block ciphers*, 1992.

[17] P. FIPS, "140-2," *Security requirements for cryptographic modules*, vol. 25, 2001.

[18] A. Rukhin, J. Soto, J. Nechvatal, M. Smid, and E. Barker, "A statistical test suite for random and pseudorandom number generators for cryptographic applications," Tech. Rep., Booz-Allen and Hamilton Inc Mclean Va, 2001.

[19] L. E. Bassham III, A. L. Rukhin, J. Soto, J. R. Nechvatal, M. E. Smid, E. B. Barker, S. D. Leigh, M. Levenson, M. Vangel, D. L. Banks, *et al.*, "Sp 800-22 rev. 1a. a statistical test suite for random and pseudorandom number generators for cryptographic applications," 2010.

[20] V. Rijmen and J. Daemen, "Advanced encryption standard," *Proceedings of Federal Information Processing Standards Publications, National Institute of Standards and Technology*, pp. 19–22, 2001.

[21] U. M. Maurer, "A universal statistical test for random bit generators," *Journal of cryptology*, vol. 5, no. 2, pp. 89–105, 1992.

[22] J. S. Coron, "On the security of random sources," in *International Workshop on Public Key Cryptography*, pp. 29–42, Springer, 1999.

[23] H. Yamamoto and Q. Liu, "Highly sensitive universal statistical test," in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 700–704, IEEE, 2016.

[24] P. Elias, "Interval and recency rank source coding: Two on-line adaptive variable-length schemes," *IEEE Transactions on Information Theory*, vol. 33, no. 1, pp. 3–10, 1987.

[25] F. M. Willems, "Universal data compression and repetition times," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 54–58, 1989.

[26] J. S. Coron and D. Naccache, "An accurate evaluation of maurer's universal test," in *International Workshop on Selected Areas in Cryptography*, pp. 57–71, Springer, 1998.

[27] G. H. Hardy, E. M. Wright, *et al.*, *An introduction to the theory of numbers*. Oxford university press, 1979.

[28] S. Wegenkittl, "Entropy estimators and serial tests for ergodic chains," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2480–2489, 2001.

[29] G. Choe and D. Kim, "Average convergence rate of the first return time," in *Colloquium Mathematicae*, vol. 84, pp. 159–171, 2000.

[30] M. Abadi and A. Galves, "A version of maurer's conjecture for stationary ψ-mixing processes," *Nonlinearity*, vol. 17, no. 4, p. 1357, 2004.

[31] Y. S. Kim, "Estimation of rényi entropy of order α based on the nearest neighbor distance," in *2014 International Symposium on Information Theory and its Applications*, pp. 125–129, IEEE, 2014.

[32] Y. S. Kim, "Low complexity estimation method of Rényi entropy for ergodic sources," *Entropy*, vol. 20, no. 9, p. 657, 2018.

[33] Y. Hikima, A. Iwasaki, and K. Umeno, "The variance of the reference distribution of highly sensitive universal test constructed on the basis of maurer's test," in *Proceedings of Symposium on Cryptography and Information Security*, pp. 2A3–4 (in Japanese), IEICE, 2020.

[34] M. Matsumoto and T. Nishimura, "Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator," *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 8, no. 1, pp. 3–30, 1998.

A Proof of $C = -\frac{\ln 2}{\gamma}$

In this appendix, we provide the proof of the following theorem.

Theorem 1 *For any $x \in (0,1)$, we have the following relation*

$$\lim_{x\to 0+}\left[v(x)+\log_2 x\right] = -\frac{\gamma}{\ln 2} = -0.832746\cdots, \tag{65}$$

where

$$v(x) = x\sum_{i=1}^{\infty}(1-x)^{i-1}\log_2 i, \tag{66}$$

and γ is Euler's constant defined by

$$\gamma := \lim_{n\to\infty}\left(\sum_{i=1}^{n}\frac{1}{i} - \ln n\right). \tag{67}$$

Proof. Let $s = 1-x$. We have

$$\begin{aligned} v(1-s)+\log_2(1-s) &= (1-s)\sum_{i=1}^{\infty}s^{i-1}\log_2 i + \log_2(1-s) \\ &= \frac{1}{\ln 2}\left\{(1-s)\sum_{i=1}^{\infty}s^{i-1}\ln i + \ln(1-s)\right\} \\ &= \frac{1}{\ln 2}\left\{(1-s)\times\frac{1}{1-s}\sum_{i=1}^{\infty}s^i\times\ln\frac{i+1}{i} - \sum_{i=1}^{\infty}\frac{s^i}{i}\right\} \\ &= \frac{1}{\ln 2}\sum_{i=1}^{\infty}s^i\left\{\ln\left(1+\frac{1}{i}\right)-\frac{1}{i}\right\}. \end{aligned} \tag{68}$$

From the above equations, we have

$$\begin{aligned} \lim_{s\to 1-}\left[v(1-s)+\log_2(1-s)\right] &= \frac{1}{\ln 2}\lim_{s\to 1-}\left[\sum_{i=1}^{\infty}s^i\left\{\ln\left(1+\frac{1}{i}\right)-\frac{1}{i}\right\}\right] \\ &= \frac{1}{\ln 2}\sum_{i=1}^{\infty}\lim_{s\to 1-}\left[s^i\left\{\ln\left(1+\frac{1}{i}\right)-\frac{1}{i}\right\}\right] \\ &= \frac{1}{\ln 2}\sum_{i=1}^{\infty}\left\{\ln\left(1+\frac{1}{i}\right)-\frac{1}{i}\right\}. \end{aligned} \tag{69}$$

For the derivation of the second equation, we have used the fact that the infinite series converges absolutely which allows us to exchange the limits. Since $\ln n$ can be written as

$$\ln n = \ln\left(\frac{2}{1}\cdot\frac{3}{2}\cdot\frac{4}{3}\cdot\cdots\cdot\frac{n}{n-1}\right) = \sum_{k=1}^{n-1}\ln\left(1+\frac{1}{k}\right), \tag{70}$$

Euler's constant γ defined by Eq. (67) can also be represented as

$$\begin{aligned}\gamma &= \lim_{n\to\infty}\left[\sum_{k=1}^{n-1}\left\{\frac{1}{k}-\ln\left(1+\frac{1}{k}\right)\right\}+\frac{1}{n}\right] \\ &= \sum_{n=1}^{\infty}\left\{\frac{1}{n}-\ln\left(1+\frac{1}{n}\right)\right\}.\end{aligned} \tag{71}$$

Therefore, we arrive at the following result

$$\lim_{s\to 1-}\left[\nu(1-s)+\log_2(1-s)\right] = -\frac{\gamma}{\ln 2}. \tag{72}$$

The theorem is obtained if we substitute s into $1-x$. □

B Exploration of the covariance given in Eq. (58)

In this appendix, we explore the covariance given in Eq. (58) to calculate the value more efficiently by computational experiments. Notice that the covariance is written as

$$\mathrm{Cov}[g(A_n),\, g(A_{n+k})] = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g(i)g(j)\Pr\left[A_n = i,\, A_{n+k} = j\right] - \{L \times H(\hat{q})\}^2, \tag{73}$$

where g is given in Eq. (20), $\Pr\left[A_n = i,\, A_{n+k} = j\right]$ is given in Eq. (51), and H is a binary entropy function. Since the term $\{L \times H(\hat{q})\}^2$ is irrelevant to an infinite series, we write the first term of the right hand side (r.h.s.) of Eq. (73) as

$$\overline{S}_k := \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g(i)g(j)\Pr\left[A_n = i,\, A_{n+k} = j\right]. \tag{74}$$

Firstly, we show the following Lemma concerning an infinite series to calculate the value expressed Eq. (74).

Lemma 2 *For any $z \in (0, 1)$, the following relation holds*

$$\sum_{i=1}^{\infty} g(i) \times (1-z)^i = -\frac{1}{\ln 2} \times \frac{1-z}{z} \times \ln z, \tag{75}$$

where

$$g(m) = (\log_2 \mathrm{e}) \sum_{k=1}^{m-1} \frac{1}{k}. \tag{76}$$

Proof. Let $t = 1 - z$, and $\overline{S}$ be the left hand side of Eq. (75), that is,

$$\begin{aligned} \overline{S} &= \sum_{i=1}^{\infty} g(i) \times t^i \\ &= g(1) \times t + \sum_{i=2}^{\infty} g(i) \times t^i. \end{aligned} \tag{77}$$

Multiplying the both sides of Eq. (77) by $t(\neq 0)$, we obtain the following relation

$$\begin{aligned} t \times \overline{S} &= \sum_{i=1}^{\infty} g(i) \times t^{i+1} \\ &= \sum_{i=2}^{\infty} g(i-1) \times t^i. \end{aligned} \tag{78}$$

Subtracting Eq. (78) from Eq. (77), we obtain the following relation

$$\begin{aligned}(1-t)\overline{S} &= g(1)\times t^1 + \sum_{i=2}^{\infty}\{g(i)-g(i-1)\}\times t^i \\ &= 0 + \sum_{i=2}^{\infty}(\log_2 \mathrm{e})\times\frac{1}{i-1}\times t^i \\ &= (\log_2 \mathrm{e})\times t\times\sum_{i=1}^{\infty}\frac{t^i}{i} \\ &= (\log_2 \mathrm{e})\times t\times\{-\ln(1-t)\}\,.\end{aligned} \tag{79}$$

To obtain the last equality in the above equations, we have used the Taylor series for $|t| < 1$. Dividing both sides of Eq. (79) by $1-t(\neq 0)$, we arrive at the following result

$$\overline{S} = -\frac{1}{\ln 2}\times\frac{t}{1-t}\times\ln(1-t). \tag{80}$$

The lemma is obtained if we substitute t into $1-z$. □

In the next place, we calculate the value given in Eq. (74) more in details.

B.1 Case of $1 \le j \le k-1$

For $1 \le j \le k-1$, Eq. (74) is written as

$$\begin{aligned}\overline{S}_k &:= \sum_{i=1}^{\infty}\sum_{j=1}^{k-1} g(i)g(j)\Pr\left[A_n = i,\ A_{n+k} = j\right] \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{k-1} g(i)g(j)\left(\sum_{r=0}^{L}\binom{L}{r}w_r^2(1-w_r)^{i-1}\right)\times\left(\sum_{r=0}^{L}\binom{L}{r}w_r^2(1-w_r)^{j-1}\right) \\ &= \sum_{r=0}^{L}\left\{\binom{L}{r}\frac{w_r^2}{1-w_r}\sum_{i=1}^{\infty} g(i)(1-w_r)^i\right\}\times\sum_{r=0}^{L}\left\{\binom{L}{r}\frac{w_r^2}{1-w_r}\sum_{j=1}^{k-1} g(j)(1-w_r)^j\right\}.\end{aligned} \tag{81}$$

From Lemma 2, the infinite series of i in the above equations is written as

$$\sum_{i=1}^{\infty} g(i)(1-w_r)^i = -\frac{1}{\ln 2}\times\frac{1-w_r}{w_r}\times\ln w_r. \tag{82}$$

Therefore, we obtain the following relation

$$\begin{aligned}\overline{S}_k =& \sum_{r=0}^{L}\left\{\binom{L}{r}\frac{w_r^2}{1-w_r}\times\left(-\frac{1}{\ln 2}\times\frac{1-w_r}{w_r}\times\ln w_r\right)\right\}\\ &\times\sum_{r=0}^{L}\left\{\binom{L}{r}\frac{w_r^2}{1-w_r}\times\sum_{j=1}^{k-1}(1-w_r)^j\right\}\\ =&-\frac{1}{\ln 2}\sum_{r=0}^{L}\left\{\binom{L}{r}w_r\ln w_r\right\}\times\sum_{r=0}^{L}\left\{\binom{L}{r}\frac{w_r^2}{1-w_r}\sum_{j=1}^{k-1}(1-w_r)^j\right\}.\end{aligned} \tag{83}$$

B.2 Case of $j = k$

In the case of $j = k$, Eq. (74) is written as

$$\begin{aligned}\overline{S}_k &= \sum_{i=1}^{\infty} g(i)\sum_{r=0}^{L}\binom{L}{r}w_r^3(1-w_r)^{k+i-2}\sum_{j\in\{k\}} g(j)\\ &= \sum_{r=0}^{L}\binom{L}{r}w_r^3(1-w_r)^{k-2}\sum_{i=1}^{\infty} g(i)(1-w_r)^i\times g(k)\\ &= g(k)\times\sum_{r=0}^{L}\binom{L}{r}w_r^3(1-w_r)^{k-2}\left(-\frac{1}{\ln 2}\times\frac{1-w_r}{w_r}\times\ln w_r\right)\\ &= -\frac{g(k)}{\ln 2}\times\sum_{r=0}^{L}\binom{L}{r}w_r^2(1-w_r)^{k-1}\ln w_r.\end{aligned} \tag{84}$$

The third equation in Eq. (84) has been obtained from Lemma 2.

B.3 Case of $k + 1 \le j \le k + i - 1$

Recall that the joint distribution for $k + 1 \le j \le k + i - 1$ is written as

$$\begin{aligned}\Pr[A_n = i,\ A_{n+k} = j] =& \sum_{r_1=0}^{L}\sum_{r_2\neq r_1}\binom{L}{r_1}\binom{L}{r_2}\Pr[e_3(b_1,b_2)]\\ &+\sum_{r_1=0}^{L}\sum_{r_2\in\{r_1\}}\binom{L}{r_1}\left\{\binom{L}{r_1}-1\right\}\Pr[e_3(b_1,b_2)],\end{aligned} \tag{85}$$

where $\Pr[e_3(b_1, b_2)]$ is expressed as

$$\begin{aligned}\Pr[e_3(b_1,b_2)] =& w_{r_1}^2\times w_{r_2}^2\times(1-w_{r_1})^{i-j+k-1}\times(1-w_{r_1}-w_{r_2})^{j-k-1}\times(1-w_{r_2})^{k-1}\\ =&\phi_k(r_1,r_2)\times(1-w_{r_1})^i\times\left(1-\frac{w_{r_2}}{1-w_{r_1}}\right)^j,\end{aligned} \tag{86}$$

where

$$\phi_k(r_1, r_2) = w_{r_1}^2 w_{r_2}^2 (1 - w_{r_1})^{k-1} (1 - w_{r_1} - w_{r_2})^{-k-1} (1 - w_{r_2})^{k-1}. \tag{87}$$

Then, Eq. (74) is expressed as

$$\begin{aligned} \overline{S}_k = &\sum_{i=1}^{\infty} \sum_{j=k+1}^{k+i-1} g(i)g(j) \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1}\binom{L}{r_2} \Pr[e_3(b_1, b_2)] \\ &+ \sum_{i=1}^{\infty} \sum_{j=k+1}^{k+i-1} g(i)g(j) \sum_{r_1=0}^{L} \sum_{r_2 \in \{r_1\}} \binom{L}{r_1} \left\{ \binom{L}{r_1} - 1 \right\} \Pr[e_3(b_1, b_2)]. \end{aligned} \tag{88}$$

Now, we consider the first term in r.h.s. of Eq. (88). Let $\overline{A}_1$ be this term of Eq. (88). We have

$$\begin{aligned} \overline{A}_1 &= \sum_{i=1}^{\infty} \sum_{j=k+1}^{k+i-1} g(i)g(j) \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1}\binom{L}{r_2} \Pr[e_3(b_1, b_2)] \\ &= \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1}\binom{L}{r_2} \phi_k(r_1, r_2) \sum_{i=1}^{\infty} \sum_{j=1}^{k+i-1} g(i)g(j)(1 - w_{r_1})^i \times \left(1 - \frac{w_{r_2}}{1 - w_{r_1}}\right)^j \\ &= \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1}\binom{L}{r_2} \phi_k(r_1, r_2) \sum_{j=k+1}^{\infty} \sum_{i=k+1}^{k+i-1} g(i)g(j)(1 - w_{r_1})^i \times \left(1 - \frac{w_{r_2}}{1 - w_{r_1}}\right)^j \\ &= \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1}\binom{L}{r_2} \phi_k(r_1, r_2) \sum_{j=k+1}^{\infty} \left\{ g(j) \left(1 - \frac{w_{r_2}}{1 - w_{r_1}}\right)^j \times \sum_{i=j-k+1}^{\infty} g(i)(1 - w_{r_1})^i \right\}. \end{aligned} \tag{89}$$

In the course of the derivation of the above relations, the second equation has been obtained from Eq. (87). The third equation has been obtained by exchanging the summation over i and j. Then, the infinite series with respect to i in the last equation of Eq. (89) can be calculated as

$$\begin{aligned} \sum_{i=j-k+1}^{\infty} g(i)(1 - w_{r_1})^i &= \sum_{i=1}^{\infty} g(i)(1 - w_{r_1})^i - \sum_{i=1}^{j-k} g(i)(1 - w_{r_1})^i \\ &= -\frac{1}{\ln 2} \times \frac{1 - w_{r_1}}{w_{r_1}} \times \ln w_{r_1} - \sum_{i=1}^{j-k} g(i)(1 - w_{r_1})^i. \end{aligned} \tag{90}$$

In the above relations, we have used the result of Lemma 2. Hence, the first term in r.h.s. of Eq. (88) can be expressed as

$$\overline{A}_1 = \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1}\binom{L}{r_2} \phi_k(r_1, r_2) \tag{91}$$

$$\times \sum_{j=k+1}^{\infty} \left[g(j) \left(1 - \frac{w_{r_2}}{1 - w_{r_1}}\right)^j \times \left\{ -\frac{1}{\ln 2} \times \frac{1 - w_{r_1}}{w_{r_1}} \times \ln w_{r_1} - \sum_{i=1}^{j-k} g(i)(1 - w_{r_1})^i \right\} \right]. \tag{92}$$

We can derive the second term in r.h.s. of Eq. (88) in the same way as the first term. Let $\overline{A}_2$ be the second term of Eq. (88). Then, we have

$$\begin{aligned}\overline{A}_2 = &\sum_{r_1=0}^{L} \sum_{r_2 \in \{r_1\}} \binom{L}{r_1} \left\{ \binom{L}{r_1} - 1 \right\} \phi(r_1, r_2) \\ &\times \sum_{j=k+1}^{\infty} \left[g(j) \left(1 - \frac{w_{r_2}}{1 - w_{r_1}} \right)^j \times \left\{ -\frac{1}{\ln 2} \times \frac{1 - w_{r_1}}{w_{r_1}} \times \ln w_{r_1} - \sum_{i=1}^{j-k} g(i)(1 - w_{r_1})^i \right\} \right].\end{aligned} \tag{93}$$

Therefore, Eq. (88) can be written as

$$\begin{aligned}\overline{S}_k =&\overline{A}_1 + \overline{A}_2 \\ =&\sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1} \binom{L}{r_2} \phi_k(r_1, r_2) \\ &\times \sum_{j=k+1}^{\infty} \left[g(j) \left(1 - \frac{w_{r_2}}{1 - w_{r_1}} \right)^j \times \left\{ -\frac{1}{\ln 2} \times \frac{1 - w_{r_1}}{w_{r_1}} \times \ln w_{r_1} - \sum_{i=1}^{j-k} g(i)(1 - w_{r_1})^i \right\} \right] \\ &+ \sum_{r_1=0}^{L} \sum_{r_2 \in \{r_1\}} \binom{L}{r_1} \left\{ \binom{L}{r_1} - 1 \right\} \phi(r_1, r_2) \\ &\times \sum_{j=k+1}^{\infty} \left[g(j) \left(1 - \frac{w_{r_2}}{1 - w_{r_1}} \right)^j \times \left\{ -\frac{1}{\ln 2} \times \frac{1 - w_{r_1}}{w_{r_1}} \times \ln w_{r_1} - \sum_{i=1}^{j-k} g(i)(1 - w_{r_1})^i \right\} \right],\end{aligned} \tag{94}$$

where $\phi_k(r_1, r_2)$ is given in Eq. (87).

B.4 Case of $j = k + i$

Equation (74) in the case of $j = k + i$ is equal to 0 from Eq. (51).

B.5 Case of $j \geq k + i + 1$

Recall that the joint distribution for $j \geq k + i + 1$ is written as

$$\begin{aligned}\Pr[A_n = i,\ A_{n+k} = j] = &\sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1} \binom{L}{r_2} \Pr[e_5(b_1, b_2)] \\ &+ \sum_{r_1=0}^{L} \sum_{r_2 \in \{r_1\}} \binom{L}{r_1} \left\{ \binom{L}{r_1} - 1 \right\} \Pr[e_5(b_1, b_2)],\end{aligned} \tag{95}$$

where $\Pr[e_5(b_1, b_2)]$ is expressed as

$$\begin{aligned}\Pr[e_5(b_1, b_2)] =&w_{r_1}^2 \times w_{r_2}^2 \times (1 - w_{r_1})^{-i+j-k-1} \times (1 - w_{r_1} - w_{r_2})^{i-1} \times (1 - w_{r_1})^{k-1} \\ =&\psi_k(r_1, r_2) \times \left(1 - \frac{w_{r_2}}{1 - w_{r_1}} \right)^i \times (1 - w_{r_1})^j,\end{aligned} \tag{96}$$

where

$$\begin{aligned}\psi_k(r_1, r_2) &= w_{r_1}^2 w_{r_2}^2 (1 - w_{r_1})^{-k-1} (1 - w_{r_1} - w_{r_2})^{-1} (1 - w_{r_1})^{k-1} \\ &= w_{r_1}^2 w_{r_2}^2 (1 - w_{r_1})^{-2} (1 - w_{r_1} - w_{r_2})^{-1}.\end{aligned} \tag{97}$$

Then, Eq. (74) is expressed as

$$\begin{aligned}\overline{S}_k = &\sum_{i=1}^{\infty} \sum_{j=k+i+1}^{\infty} g(i)g(j) \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1}\binom{L}{r_2} \Pr[e_5(b_1, b_2)] \\ &+ \sum_{i=1}^{\infty} \sum_{j=k+i+1}^{\infty} g(i)g(j) \sum_{r_1=0}^{L} \sum_{r_2 \in \{r_1\}} \binom{L}{r_1} \left\{ \binom{L}{r_1} - 1 \right\} \Pr[e_5(b_1, b_2)].\end{aligned} \tag{98}$$

Firstly, we consider the first term in the r.h.s. of Eq. (88). Let $\overline{B}_1$ be the first term of Eq. (88). Then, $\overline{B}_1$ is written as

$$\begin{aligned}\overline{B}_1 &= \sum_{i=1}^{\infty} \sum_{j=k+i+1}^{\infty} g(i)g(j) \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1}\binom{L}{r_2} \Pr[e_5(b_1, b_2)] \\ &= \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1}\binom{L}{r_2} \psi_k(r_1, r_2) \sum_{i=1}^{\infty} \sum_{j=k+i+1}^{\infty} g(i)g(j) \left(1 - \frac{w_{r_2}}{1 - w_{r_1}}\right)^i \times (1 - w_{r_1})^j \\ &= \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1}\binom{L}{r_2} \psi_k(r_1, r_2) \sum_{i=1}^{\infty} g(i) \left(1 - \frac{w_{r_2}}{1 - w_{r_1}}\right)^i \times \sum_{j=k+i+1}^{\infty} g(j)(1 - w_{r_1})^j.\end{aligned} \tag{99}$$

In the course of the derivation of the above relations, the second equation has been obtained from Eq. (87). The third equation has been obtained by exchange of infinite series. Then, the infinite sum with respect to j in the last equation of Eq. (99) can be calculated as

$$\begin{aligned}\sum_{j=k+i+1}^{\infty} g(j)(1 - w_{r_1})^j &= \sum_{j=1}^{\infty} g(j)(1 - w_{r_1})^j - \sum_{j=1}^{k+i} g(j)(1 - w_{r_1})^j \\ &= -\frac{1}{\ln 2} \times \frac{1 - w_{r_1}}{w_{r_1}} \times \ln w_{r_1} - \sum_{j=1}^{k+i} g(j)(1 - w_{r_1})^j.\end{aligned} \tag{100}$$

In the above relations, we use the result of Lemma 2. Hence, the first term of Eq. (98) can be expressed as

$$\overline{B}_1 = \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1}\binom{L}{r_2} \psi_k(r_1, r_2) \tag{101}$$

$$\times \sum_{i=1}^{\infty} \left[g(i) \left(1 - \frac{w_{r_2}}{1 - w_{r_1}}\right)^i \times \left\{ -\frac{1}{\ln 2} \times \frac{1 - w_{r_1}}{w_{r_1}} \times \ln w_{r_1} - \sum_{j=1}^{k+i} g(j)(1 - w_{r_1})^j \right\} \right]. \tag{102}$$

We can derive the second term of Eq. (98) in the same way as the first term. Let $\overline{B}_2$ be the second term of Eq. (98). Then, we have

$$\begin{aligned}\overline{B}_2 = &\sum_{r_1=0}^{L} \sum_{r_2 \in \{r_1\}} \binom{L}{r_1} \left\{ \binom{L}{r_1} - 1 \right\} \psi(r_1, r_2) \\ &\times \sum_{i=1}^{\infty} \left[g(i) \left(1 - \frac{w_{r_2}}{1 - w_{r_1}} \right)^i \times \left\{ -\frac{1}{\ln 2} \times \frac{1 - w_{r_1}}{w_{r_1}} \times \ln w_{r_1} - \sum_{j=1}^{k+i} g(j)(1 - w_{r_1})^j \right\} \right].\end{aligned} \tag{103}$$

Therefore, Eq. (98) can be expressed as

$$\begin{aligned}\overline{S}_k =& \overline{B}_1 + \overline{B}_2 \\ =& \sum_{r_1=0}^{L} \sum_{r_2 \neq r_1} \binom{L}{r_1} \binom{L}{r_2} \psi_k(r_1, r_2) \\ &\times \sum_{i=1}^{\infty} \left[g(i) \left(1 - \frac{w_{r_2}}{1 - w_{r_1}} \right)^i \times \left\{ -\frac{1}{\ln 2} \times \frac{1 - w_{r_1}}{w_{r_1}} \times \ln w_{r_1} - \sum_{j=1}^{k+i} g(j)(1 - w_{r_1})^j \right\} \right] \\ &+ \sum_{r_1=0}^{L} \sum_{r_2 \in \{r_1\}} \binom{L}{r_1} \left\{ \binom{L}{r_1} - 1 \right\} \psi(r_1, r_2) \\ &\times \sum_{i=1}^{\infty} \left[g(i) \left(1 - \frac{w_{r_2}}{1 - w_{r_1}} \right)^i \times \left\{ -\frac{1}{\ln 2} \times \frac{1 - w_{r_1}}{w_{r_1}} \times \ln w_{r_1} - \sum_{j=1}^{k+i} g(j)(1 - w_{r_1})^j \right\} \right],\end{aligned} \tag{104}$$

where $\psi_k(r_1, r_2)$ is given in Eq. (97).

Master's Thesis

Study on a further improvement of Maurer's universal statistical test

Guidance

Professor Ken UMENO
Assistant Professor Atsushi IWASAKI

Yasunari HIKIMA

Department of Applied Mathematics and Physics

Graduate School of Informatics

Kyoto University

February 2020

Study on a further improvement of
Maurer's universal statistical test

Yasunari HIKIMA

February 2020

Study on a further improvement of Maurer's universal statistical test

Yasunari HIKIMA

Abstract

Maurer's universal statistical test is a hypothesis test for evaluating the randomness of a binary sequence and it is included in NIST SP 800-22 which is one of the most famous test suites. The test statistic of Maurer's test relates to the entropy of a tested sequence and hence the test can detect various defects of the sequence about randomness. It has been reported that flipping a part of bits in a sequence makes Maurer's test more sensitive. The test with flipping is called highly sensitive universal statistical test. To perform the highly sensitive test, the variance for the reference distribution is necessary, however, the theoretical value has not been derived. In this thesis, we theoretically derive the variance for the reference distribution of the highly sensitive test and investigate the validity for testing randomness.