Journal of Statistical Planning and Inference 217 (2022) 1–7

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Asymptotically most powerful tests for random number generators

Boris Ryabko *

Federal Research Center for Information and Computational Technologies of Siberian Branch of Russian Academy of Science, Russian Federation
Novosibirsk State University, Russian Federation

ARTICLE INFO

Article history:
Received 12 February 2020
Received in revised form 3 July 2021
Accepted 9 July 2021
Available online 17 July 2021

Keywords:
Statistical test
Randomness testing
p-value
Random number generators
Shannon entropy

ABSTRACT

The problem of constructing the most powerful test for random number generators (RNGs) is considered, where the generators are modelled by stationary ergodic processes. At present, RNGs are widely used in data protection, modelling and simulation systems, computer games, and in many other areas where the generated random numbers should look like binary numbers of a Bernoulli equiprobable sequence. Another problem considered is that of constructing effective statistical tests for random number generators (RNG). Currently, effectiveness of statistical tests for RNGs is mainly estimated based on experiments with various RNGs. We find an asymptotic estimate for the *p*-value of an optimal test in the case where the alternative hypothesis is a known stationary ergodic source, and then describe a family of tests each of which has the same asymptotic estimate of the *p*-value for any (unknown) stationary ergodic source. This model appears to be acceptable for binary sequences generated by physical devices that are used in cryptographic data protection systems.

1. Introduction

Random number generators (RNG) are widely used in many applications, including cryptographic information protection systems, modelling and simulation systems and computer games. The goal of RNG is to generate sequences of binary digits, which are distributed as a result of throwing an fair coin or, more precisely, obey the Bernoulli distribution with parameters (1/2, 1/2). As a rule, for practically used RNG this property is verified experimentally with the help of statistical tests intended for this purpose, see for a review L'Ecuyer (2017) and L'Ecuyer and Simard (2007). Nowadays, there are more than a hundred applicable statistical tests and some of them are a mandatory part of cryptographic information protection systems (Rukhin et al., 2010). Besides, there are dozens of RNGs based on different algorithms and different physical processes (Herrero-Collantes and Garcia-Escartin, 2017); see for review L'Ecuyer (2017). In such a situation, the natural question is how to compare different tests. Currently, the main method of such a comparison is numerical experiments in which different tests are practically applied to different RNGs (L'Ecuyer, 2017; L'Ecuyer and Simard, 2007, 2013; Rukhin et al., 2010).

Here we consider the problem of finding optimal tests in the case when the RNG is modelled by a stationary ergodic source. This model appears to be acceptable for binary sequences generated by physical devices that are used in

* Correspondence to: Federal Research Center for Information and Computational Technologies of Siberian Branch of Russian Academy of Science, Russian Federation.
E-mail address: boris@ryabko.net.

https://doi.org/10.1016/j.jspi.2021.07.007
0378-3758/

cryptographic data protection systems. We propose the following asymptotic solution to this problem: we first describe the asymptotic behaviour of the p-value of the optimal test for the case where the probability distribution of the RNG is a priori known, and then describe a family of statistical tests that have the same asymptotic estimates of the p-value for any distribution (which is not known in advance). More precisely, we show that in both cases, with probability 1, $\lim_{n\to\infty} -\frac{1}{n}\log \pi_\tau(x_1x_2...x_n) = 1 - h(\nu)$, where $x_1x_2...x_n$ is the sample generated by a stationary ergodic ν, τ is the test statistic, $\pi_\tau(x_1x_2...x_n)$ is the p-value, and $h(\nu)$ is the Shannon entropy of the ν.

It turns out that asymptotically optimal tests with the required properties are known (Ryabko and Astola, 2006; Ryabko et al., 2016; Ryabko and Monarev, 2005), and are deeply connected with so-called universal codes. Note that nowadays there are many universal codes which are based on different ideas and approaches, among which we note the PPM universal code (Cleary and Witten, 1984) which is used along with the arithmetic code (Rissanen and Langdon, 1979), the Lempel–Ziv (LZ) codes (Ziv and Lempel, 1977), the Burrows–Wheeler transform (Burrows and Wheeler, 1994) which is used along with the book-stack (or MTF) code (Ryabko, 1980; Bentley et al., 1986; Ryabko et al., 1987), the class of grammar-based codes (Kieffer and Yang, 2000; Yang and Kieffer, 2000) and some others (Drmota et al., 2010; Ryabko, 1984; Louchard and Szpankowski, 1995). All these codes are universal. This means that, asymptotically, the length of the compressed file goes to the smallest possible value, i.e. the Shannon entropy ($h(\nu)$) per letter.

The main idea of randomness tests based on universal codes is rather natural: try to "compress" a test sequence by a universal code. If the sequence is significantly compressed, then it is not random, see Ryabko and Astola (2006), Ryabko et al. (2016), Ryabko and Monarev (2005) and a short description below.

The rest of the paper is organized as follows. The next section contains the necessary definitions and some basic facts used in the following. Sections 3 and 4 are devoted to the investigation of the Neyman–Pearson test and tests based on universal codes, correspondingly. We see that universal codes play an important role in testing of randomness, so we give a brief description of one of them in Appendix A; the proofs are given in Appendix B.

2. Definitions and preliminaries

2.1. The main notations

We consider an RNG which generates a sequence of letters $x = x_1x_2...x_n$, $n \geq 1$, from the alphabet $\{0, 1\}^n$. There are two statistical hypotheses: $H_0 = \{x \quad obeys\ uniform\ distribution\ (\mu_U)\ on\ \{0, 1\}^n\}$, and the alternative hypothesis $H_1 = \bar{H_0}$, that is, H_1 is negation of H_0. It is a particular case of the so-called goodness-of-fit problem, and any test for it is called a test of fit, see Kendall and Stuart (1961). Let T be a test. Then, by definition, a significance level α equals probability of the Type I error. (Recall, that Type I error occurs if H_0 is true and H_0 is rejected. Type II error occurs if H_1 is true, but H_0 is accepted.) Denote a critical region of the test T for the significance level α by $C_T(\alpha)$ and let $\bar{C}_T(\alpha) = \{0, 1\}^n \setminus C_T(\alpha)$. (Recall, that for a certain $x = x_1x_2...x_n$ the hypothesis H_0 is rejected if and only if $x \in C_T(\alpha)$.)

Let us assume that H_1 is true and the investigated sequence $x = x_1x_2...x_n$ is generated by (unknown) source ν. By definition, the test T is consistent (for ν), if for any significance level $\alpha \in (0, 1)$ the probability of Type II error goes to 0, that is

$$\lim_{n\to\infty} \nu(\bar{C}_T(\alpha)) = 0\,. \tag{1}$$

Let us give a definition of a so-called p-value, which plays an important role in the randomness testing. Let there be a statistic τ (that is, a function on $\{0, 1\}^n$) and x be a word from $\{0, 1\}^n$. A p-value ($\pi_\tau(x)$) of τ and x is defined by the equation

$$\pi_\tau(x) = \mu_U\{y : \tau(y) \geq \tau(x)\} = |\{y : \tau(y) \geq \tau(x)\}|/2^n\,. \tag{2}$$

(Here and below μ_U is a uniform distribution, $|X|$ is a number of elements X, if X is a set, and the length of X, if X is a word.)

Informally, $\pi_\tau(x)$ is the probability to meet a random point y for which $\tau(y)$ is at least as large as the observed value when considering the null hypothesis.

2.2. The consistent tests for stationary ergodic sources and universal codes

First let us give a short informal description of the universal codes. For any integer m a lossless code ϕ is defined as such a map from the set of m-letter words to the set of all binary words that for any sequence of encoded m-letter words $\phi(v_1)\phi(v_2)...$ the initial sequence $v_1v_2...$ can be found without mistakes; the formal definition can be found, for example, in Cover and Thomas (2006).

We will consider so-called universal codes which have the following property: for any stationary ergodic ν defined on the set of all infinite binary words $x = x_1x_2...$, with probability one

$$\lim_{n\to\infty} \frac{1}{n}|\phi(x_1x_2...x_n)| = h(\nu)\,, \tag{3}$$

where $h(\nu)$ is the Shannon entropy of ν (see for definition Cover and Thomas (2006)). Such codes exist, see, for example, Ryabko and Astola (2006), Ryabko et al. (2016) and Appendix A. Note, that a goal of codes is to "compress" sequences, i.e. make a length of the codeword $\phi(x_1x_2...x_n)$ as small as possible. The property (3) shows that the universal codes are asymptotically optimal, because the Shannon entropy is a lower bound of the length of the compressed sequence (per letter), see Cover and Thomas (2006).

Let us back to considered problem of hypothesis testing. Suppose, it is known that a sample sequence $x = x_1x_2...$ was generated by stationary ergodic source and, as before, we consider the same H_0 against the same H_1. Let ϕ be a universal code. The following test is a particular case of a goodness-of-fit test suggested in Ryabko and Astola (2006), Ryabko et al. (2016):

If $n - |\phi(x_1...x_n)| \geq -\log_2 \alpha$ then H_0 is rejected, otherwise accepted. Here, as before, α is the significance level, $|\phi(x_1...x_n)|$ is the length of encoded ("compressed") sequence.

We denote this test by T_ϕ and its statistic by τ_ϕ, i.e.

$$\tau_\phi(x_1...x_n) = n - |\phi(x_1...x_n)| \,. \tag{4}$$

It turns out that the test T_ϕ is consistent for any stationary ergodic source. More precisely, the following theorem is proven in Ryabko and Monarev (2005)

Let ν be a probability process, $\alpha \in (0, 1)$ and ϕ be a code. Then the Type I error of the described test is not larger than α and, if ν is stationary ergodic and ϕ is universal, then the Type II error goes to 0, when $n \to \infty$.

It is worth noting that the power of such tests was estimated experimentally (see Ryabko et al. (2016)), where real data compressors were used instead of universal codes. The point is that real data compressors are implementations of universal codes, and we can use them in (4) as code ϕ. In such a situation, the asymptotic behaviour of a specific data compressor cannot be guaranteed (because it is a specific computer program), but, anyway, the error of Type I is not greater than α. Such an experiment was carried out with two data compressors RAR and ARJ and, for comparison, with tests from Rukhin et al. (2010). It turns out that the data compression tests are superior to the test from Rukhin et al. (2010) (see Ryabko et al. (2016), part 2.4 "The Experiments").

3. Asymptotic behaviour of a *p*-value of the Neyman–Pearson test

Suppose, that H_1 is true and sequences $x \in \{0, 1\}^n$ obey a certain distribution ν. It is well-known in mathematical statistics that the optimal test (*NP*-test or likelihood-ratio test) is described by Neyman—Pearson lemma and the critical region of this test is defined as follows:

$$C_{NP}(\alpha) = \{x : \ \mu_U(x)/\nu(x) \leq \lambda_\alpha\} \,,$$

where $\alpha \in (0, 1)$ is the significance level and the constant λ_α is chosen in such a way that $\mu_U(C_{NP}(\alpha)) = \alpha$, see Kendall and Stuart (1961). (We did not take into account that the set $\{0, 1\}^n$ is finite. Strictly speaking, in such a case a randomized test should be used, but in what follows we will consider asymptotic behaviour of the tests for large n and this effect will be negligible.) The p-value for the *NP*-test can be derived from the definition (2), if we put $\tau(x) = \nu(x)$ and take into account that by definition, $\mu_U(x) = 2^{-n}$ for any $x \in \{0, 1\}^n$. So,

$$\pi_{NP}(x) = \mu_U\{y : \nu(y) \geq \nu(x)\} = |\{y : \nu(y) \geq \nu(x)\}|/2^n \,. \tag{5}$$

The following theorem describes an asymptotic behaviour of p-values for stationary ergodic sources for *NP* test.

Theorem 1. *If ν is a stationary ergodic measure, then, with probability 1,*

$$\lim_{n\to\infty} -\frac{1}{n} \log \pi_{NP}(x) = 1 - h(\nu) \,, \tag{6}$$

where $h(\nu)$ is the Shannon entropy of ν, see for definition Cover and Thomas (2006).

The *NP*-test is optimal in the sense that its probability of a Type II error is minimal, but when testing RNG the alternative distribution is unknown, and, hence, some other tests should be used. It turns out that the above described test T_ϕ has the same asymptotic behaviour as *NP*-test.

4. Asymptotically optimal tests for randomness

The following theorem describes an asymptotic behaviour of p-values for stationary ergodic sources for tests which are based on universal codes.

Theorem 2. *Let ϕ be a universal code and the test T_ϕ with statistic τ_ϕ(4) is applied. Then for any stationary ergodic measure ν, with probability 1,*

$$\lim_{n\to\infty} -\frac{1}{n} \log \pi_{\tau_\phi}(x) = 1 - h(\nu) \,, \tag{7}$$

where π_{τ_ϕ} is the p-value.

Note that this theorem gives some idea of the relation between the Shannon entropy of the (unknown) process ν and the required sample size. Indeed, suppose that a *NP* test is used and the desired significance level is α. Then, we can see that (asymptotically) α should be less than $\pi_{NP}(x)$ and from (6) we obtain $n > -\log\alpha/(1-h(\nu))$ (for the most powerful test). It is known that the Shannon entropy is 1 if and only if ν is the uniform measure μ_u. Therefore, in a certain sense, the difference $1-h(\nu)$ estimates the distance between the distributions, and the last inequality shows that the required sample size goes to infinity if ν approaches the uniform distribution. It is also worth noting that the value $1-h(\nu)$ equals the so-called Kullback–Leibler (KL) distance between ν and the uniform distribution. More precisely, by definition

$$\lim_{n\to\infty}\frac{1}{n}\sum_{v\in\{0,1\}^n}\nu(v)\log\frac{\nu(v)}{m_U(v)}=1-h(\nu),$$

where the left part is the limit of the KL distance (Cover and Thomas, 2006).

The next simple example illustrates the theorems. Let there be a test κ and a generator (a measure ν) that generates sequences of independent binary digits with, say, $\nu(0)=0.501$, $\nu(1)=0.499$. Suppose that $\lim_{n\to\infty}-\frac{1}{n}\log\pi_\kappa(x)=c$, where c is a positive constant. Let us consider the following "decimation test" $\kappa^{1/2}$: an input sequence $x_1x_2....x_n$ is transformed into $x_1x_3x_5...x_{2\lfloor n/2\rfloor-1}$ and then κ is applied to this transformed sequence. Obviously, for this test $\lim_{n\to\infty}-\frac{1}{n/2}\log\pi_{\kappa^{1/2}}(x)=c$, and, hence, $\lim_{n\to\infty}-\frac{1}{n}\log\pi_{\kappa^{1/2}}(x)=c/2$. Thus, the value $-\frac{1}{n}\log\pi_\kappa(x_1...x_n)$ seems to be a reasonable estimate of the power of the test for large n.

5. Conclusion

It is known that mathematical statistics and information theory are closely related, see for review Cover and Thomas (2006). For example, KL-distance is deeply connected with Fisher information and Sanov's theorem (Cover and Thomas, 2006). In this paper, it is shown that Information Theory and, in particular, universal coding, allows one to find asymptotically optimal statistics for the rather important problem of testing RNGs.

Acknowledgement

Research was supported by Russian Foundation for Basic Research (grant no. 18-29-03005).

Appendix A. The universal code

In this section we describe a so-called twice-universal code which is universal for stationary ergodic sources. It is based on optimal universal codes for Markov chains, developed by Krichevsky (1968, 1993) and the twice-universal code (Ryabko, 1984). Denote by M_i, $i=1,2,\dots$ the set of Markov chains with memory (connectivity) i, and let M_0 be the set of Bernoulli sources. For stationary ergodic μ and an integer r we denote by $h_r(\mu)$ the r-order entropy (per letter) and let $h_\infty(\mu)$ be the limit entropy; see for definitions Cover and Thomas (2006).

Krichevsky (1968, 1993) described the codes $\psi_0,\psi_1,\dots$ which are asymptotically optimal for $M_0, M_1, \dots$, correspondingly. If the sequence $x_1x_2...x_n$, $x_i\in A$, is generated by a source $\mu\in M_i$, the following inequalities are valid almost surely (a.s.):

$$h_i(\mu)\le|\psi_i(x_1...x_t)|/t\ \le h_i(\mu)+((|A|-1)|A|^i+C)/t, \tag{8}$$

where t grows. (Here C is a constant.) The length of a codeword of the twice-universal code ρ is defined as the following "mixture":

$$|\rho(x_1...x_t)|=-\log\sum_{i=0}^{\infty}\omega_{i+1}\,2^{-|\psi_i(x_1...x_t)|} \tag{9}$$

(It is well-known in Information Theory (Cover and Thomas, 2006) that there exists a code with such codeword lengths, because $\sum_{x_1...x_t\in A^t}2^{-|\rho(x_1...x_t)|}=1$.) This code is called twice-universal because for any M_i, $i=0,1,\dots$, and $\mu\in M_i$ the equality (8) is valid (with different C). Besides, for any stationary ergodic source μ a.s.

$$\lim_{t\to\infty}|\rho_i(x_1...x_t)|/t\ =h_\infty(\mu). \tag{10}$$

Appendix B

Proof of Theorem 1. The well-known Shannon–McMillan–Breiman (SMB) theorem states that for the stationary ergodic source ν and any $\epsilon>0,\delta>0$ there exists such $n'(\epsilon,\delta)$ that

$$\nu\{x\in\{0,1\}^n:\ h(\nu)-\epsilon<-\frac{1}{n}\log\nu(x)<\\ h(\nu)+\epsilon\ \}>1-\delta\quad for\ \ n>n'(\epsilon,\delta), \tag{11}$$

see Cover and Thomas (2006). From this we obtain

$$\nu\{x \in \{0,1\}^n : \ 2^{-n(h(\nu)-\epsilon)} > \ \nu(x) > 2^{-n(h(\nu)+\epsilon)}\} > 1-\delta \tag{12}$$

for $n > n'(\epsilon, \delta)$. It will be convenient to define

$$\Phi_{\epsilon,n} = \{x \in \{0,1\}^n : \ h(\nu) - \epsilon < -\frac{1}{n}\log \nu(x) < h(\nu) + \epsilon \ \} \tag{13}$$

From this definition and (12) we obtain

$$(1-\delta)2^{n(h(\nu)-\epsilon)} \le |\Phi_{\epsilon,n}| \le 2^{n(h(\nu)+\epsilon)}. \tag{14}$$

For any $x \in \Phi_{\epsilon,n}$ define

$$\Lambda_x = \{y : \nu(y) \ge \nu(x) \} \bigcap \Phi_{\epsilon,n}. \tag{15}$$

Note that, by definition, $|\Lambda_x| \le |\Phi_{\epsilon,n}|$ and from (14) we obtain

$$|\Lambda_x| \le 2^{n(h(\nu)+\epsilon)}. \tag{16}$$

For any $\rho \in (0,1)$ we define $\Psi_\rho \subset \Phi_{\epsilon,n}$ such that

$$\nu(\Psi_\rho) = \rho \ \& \ \forall u \in \Psi_\rho, \ \forall v \in (\Phi_{\epsilon,n} \setminus \Psi_\rho) : \nu(u) \ge \nu(v). \tag{17}$$

(That is, Ψ_ρ contains the most probable words whose total probability equals ρ. If there are several such sets we can take any of them.) Let us consider any $x \in (\Phi_{\epsilon,n} \setminus \Psi_\rho)$. Taking into account the definition (14) and (17) we can see that for this x

$$|\Lambda_x| \ge \rho|\Phi_{\epsilon,n}| \ge \rho(1-\delta)2^{n(h(\nu)-\epsilon)}. \tag{18}$$

So, from this inequality and (16) we obtain

$$\rho(1-\delta)2^{n(h(\nu)-\epsilon)} \le |\Lambda_x| \le \ 2^{n(h(\nu)+\epsilon)}. \tag{19}$$

From Eqs. (12), (13) and (17) we can see that $\nu(\Phi_{\epsilon,n} \setminus \Psi_\rho) \ge (1-\delta)(1-\rho)$. Taking into account (19) and this inequality, we can see that

$$\nu\{x : \ h(\nu) - \epsilon + \log(\rho(1-\delta))/n \le \log|\Lambda_x|/n \le h(\nu) + \epsilon\} \ge (1-\delta)(1-\rho). \tag{20}$$

From the definition (5) of $\pi_{NP}(x)$ and the definition (15) of Λ_x, we can see that $\pi_{NP}(x) = |\Lambda_x|/2^n$. Taking into account this equation and (20) we obtain the following:

$$\nu\{x : \ 1 - (h(\nu) - \epsilon + \log(\rho(1-\delta))/n) \ge -\log \pi_{NP}(x)/n \ge 1 - (h(\nu)+\epsilon)\} \ge (1-\delta)(1-\rho). \tag{21}$$

Clearly, there exists such $n^*(\rho)$ that for $n > n^*(\rho)$ $\quad -\log(\rho(1-\delta))/n < \epsilon$. Taking into account (11) we can see that

$$\nu\{x : \ 1 - (h(\nu) - 2\epsilon) \ge -\log \pi_{NP}(x)/n \ge 1 - (h(\nu)+\epsilon)\} \ge (1-\delta)(1-\rho) \tag{22}$$

for $n > \max(n'(\epsilon,\delta), n^*(\rho))$. This inequality is valid for any $\rho \in (0,1)$ and, in particular, for $\rho = \delta$. So, from (22) we obtain

$$\nu\{x : \ 1 - (h(\nu) - 2\epsilon) \ge -\log \pi_{NP}(x)/n \ge 1 - (h(\nu)+\epsilon)\} \ge (1-2\delta).$$

for $n > \max(n'(\epsilon,\delta), n^*(\delta))$.

Having taken into account that this inequality is valid for all positive ϵ and δ, we obtain the statement of the theorem.

Proof of Theorem 2. This is similar to the previous one. First, for any $\epsilon > 0, \delta > 0$ we define

$$\hat{\Phi}_{\epsilon,n} = \{x : h(\nu) - \epsilon < |\phi(x_1 \ldots x_n)|/n < h(\nu) + \epsilon \}. \tag{23}$$

Note that from (3) we can see that there exists such $n''(\epsilon,\delta)$ that, for $n > n''(\epsilon,\delta)$,

$$\nu(\hat{\Phi}_{\epsilon,n}) > 1-\delta. \tag{24}$$

We will use the set $\Phi_{\epsilon,n}$ (see (13)). Having taken into account the SMB theorem (11) and (24), we can see that

$$\nu(\hat{\Phi}_{\epsilon,n} \cap \Phi_{\epsilon,n}) > 1 - 2\delta\,, \tag{25}$$

if $n > \max(n'(\epsilon, \delta), n''(\epsilon, \delta))$.

From this moment, the proof begins to repeat the proof of the first theorem, if we use the set $(\hat{\Phi}_{\epsilon,n} \cap \Phi_{\epsilon,n})$ instead of $\Phi_{\epsilon,n}$. Namely, define

$$\hat{\Lambda}_x = \{y : |\phi(y)| \le |\phi(x)|\ \} \cap (\hat{\Phi}_{\epsilon,n} \cap \Phi_{\epsilon,n}) \tag{26}$$

and $\hat{\Psi}_\rho$ is such a subset of $(\hat{\Phi}_{\epsilon,n} \cap \Phi_{\epsilon,n})$ that

$$\nu(\hat{\Psi}_\rho) = \rho \ \&\ \forall u \in \Psi_\rho\,, \forall v \in ((\hat{\Phi}_{\epsilon,n} \cap \Phi_{\epsilon,n}) \setminus \Psi_\rho) :$$

$$|\phi(u)| \le |\phi(v)|\,. \tag{27}$$

Let us consider any $x \in ((\hat{\Phi}_{\epsilon,n} \cap \Phi_{\epsilon,n}) \setminus \hat{\Psi}_\rho)$. Taking into account the definition (25) and (26), we obtain

$$\rho(1-2\delta)2^{n(h(\nu)-\epsilon)} \le |\hat{\Lambda}_x| \le\ 2^{n(h(\nu)+\epsilon)}\,. \tag{28}$$

From Eqs. (25) and (27) we can see that $\nu((\hat{\Phi}_{\epsilon,n} \cap \Phi_{\epsilon,n}) \setminus \hat{\Psi}_\rho) \ge (1-2\delta)(1-\rho)$. Taking into account (28) and this inequality, we can see that

$$\begin{aligned}\nu\{x :\ & h(\nu) - \epsilon + \log(\rho(1-2\delta))/n \\ & \le \log|\hat{\Lambda}_x|/n \le h(\nu) + \epsilon\} \ge (1-2\delta)(1-\rho).\end{aligned} \tag{29}$$

From the definition of p-value (2) and the definition (26), we can see that $\pi_{\tau_\phi}(x) = |\hat{\Lambda}_x|/2^n$. Taking into account this equation and (29) we obtain the following:

$$\begin{aligned}\nu\{x :\ & 1 - (h(\nu) - \epsilon + \log(\rho(1-\delta))/n) \ge \\ & - \log \pi_{\tau_\phi}(x)/n \ge 1 - (h(\nu) + \epsilon)\} \ge (1-2\delta)(1-\rho).\end{aligned} \tag{30}$$

Clearly, there exists such $n^{**}(\rho)$ that for $n > n^{**}(\rho)$ $\quad -\log(\rho(1-2\delta))/n < \epsilon$. Taking into account we can see from (30) that

$$\begin{aligned}\nu\{x :\ & 1 - (h(\nu) - 2\epsilon) \ge \\ & - \log \pi_{\tau_\phi}(x)/n \ge 1 - (h(\nu) + \epsilon)\} \ge (1-2\delta)(1-\delta)\end{aligned} \tag{31}$$

for $n > \max(n'(\epsilon, \delta), n''(\epsilon, \delta), n^{**}(\rho))$. So, from (31) we obtain

$$\begin{aligned}\nu\{x :\ & 1 - (h(\nu) - 2\epsilon) \ge \\ & - \log \pi_{\tau_\phi}(x)/n \ \ge\ 1 - (h(\nu) + \epsilon)\} \ge (1-3\delta).\end{aligned}$$

for $n > \max(n'(\epsilon, \delta), n''(\epsilon, \delta), n^{**}(\delta))$.

Having taken into account that this inequality is valid for all positive ϵ and δ, we obtain the statement of the theorem.

References

Bentley, J., Sleator, D., Tarjan, R., Wei, V., 1986. A locally adaptive data compression scheme. Commun. ACM 29 (4), 320–330.

Burrows, M., Wheeler, D.J., 1994. A block-sorting lossless data compression algorithm.

Cleary, J., Witten, I., 1984. Data compression using adaptive coding and partial string matching. IEEE Trans. Commun. 32 (4), 396–402.

Cover, T.M., Thomas, J.A., 2006. Elements of Information Theory. Wiley-Interscience, New York, NY, USA.

Drmota, M., Yu. Reznik, Szpankowski, W., 2010. Tunstall code, Khodak variations, and random walks. IEEE Trans. Inform. Theory 56 (6), 2928–2937.

Herrero-Collantes, M., Garcia-Escartin, J.C., 2017. Quantum random number generators. Rev. Modern Phys. 89, 015004.

Kendall, M., Stuart, A., 1961. The advanced theory of statistics, Vol. 2: Inference and Relationship. Hafner Publishing Company, New York, NY, USA.

Kieffer, J.C., Yang, E.-H., 2000. Grammar-based codes: a new class of universal lossless source codes. IEEE Trans. Inform. Theory 46 (3), 737–754.

Krichevsky, R., 1968. A relation between the plausibility of information about a source and encoding redundancy. Probl. Inf. Transm. 4 (3), 48–57.

Krichevsky, R., 1993. Universal Compression and Retrival. Kluwer Academic Publishers.

L'Ecuyer, P., 2017. History of uniform random number generation. In: Proceedings of the WSC 2017-Winter Simulation Conference. Las Vegas, NV, USA, pp. 3–6.

L'Ecuyer, P., Simard, R., 2007. TestU01: A C library for empirical testing of random numbergenerators. ACM Trans. Math. Software 33 (4), Article 22.

L'Ecuyer, P., Simard, R., TestU01: A Software Library in ANSI C for Empirical Testing of Random Number Generators; User's Guide, 2013. Software user's guide, version of May 16, 2013, Available at http://simul.iro.umontreal.ca/testu01/tu01.html.

Louchard, G., Szpankowski, W., 1995. Average profile and limiting distribution for a phrase size in the Lempel–Ziv parsing algorithm. IEEE Trans. Inform. Theory 41 (2), 478–488.

Rissanen, J., Langdon, G.G., 1979. Arithmetic coding. IBM J. Res. Dev. 23 (2), 149–162.

Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., Heckert, A., Dray, J., Vo, S., 2010. A Statistical Test Suite for RandOm and PseudorandOm Number Generators for Cryptographic Applications. National Institute of Standards and Technology.

Ryabko, B.Y., 1980. Data compression by means of a book stack. Probl. Inf. Transm. 16 (4), 265–269.

Ryabko, B., 1984. Twice-universal coding. Probl. Inf. Transm. 3, 173–177.
Ryabko, B., Astola, J., 2006. Universal codes as a basis for time series testing. Stat. Methodol. 3, 375–397.
Ryabko, B., Astola, J., Malyutov, M., 2016. Compression-Based Methods of Statistical Analysis and Prediction of Time Series. Springer International Publishing Switzerland.
Ryabko, B., Horspool, N.R., Cormack, G.V., Sekar, S., Ahuja, S.B., 1987. Technical correspondence. Commun. ACM 30 (9), 792–797.
Ryabko, B.Y., Monarev, V.A., 2005. Using information theory approach to randomness testing. J. Stat. Plan. Inference 133 (1), 95–110.
Yang, E.-H., Kieffer, J.C., 2000. Efficient universal lossless data compression algorithms based on a greedy sequential grammar transform. I. without context models. IEEE Trans. Inform. Theory 46 (3), 755–777.
Ziv, J., Lempel, A., 1977. A universal algorithm for sequential data compression. IEEE Trans. Inform. Theory 23 (3), 337–343.