ON THE POWER OF TEST FOR RANDOMNESS ON THE BASE OF LEMPEL-ZIV PREDICTOR

A.V. Shilkin[1], A.L. Kostevich[1,2],
[1] *Research Institute for Applied Problems of Mathematics and Informatics*
Belarusian State University, Minsk, BELARUS
[2] *JSC "Avest", Minsk, BELARUS*
e-mail: shilkinanton@gmail.com , andrew.kostevich@gmail.com

Abstract

We propose a technique on the base of universal predictors for statistical test construction for randomness testing of binary sequences. The technique allows to find the asymptotic power of the test. We use the technique to construct the test on the base of the universal Lempel-Ziv predictor and theoretically find its power for the model of i.i.d. asymmetric Bernoulli trials under two-staged procedure of test construction. We perform comparison of the proposed test with Lempel-Ziv compression test from NIST SP800-22.

1 Introduction

Randomness testing of binary sequences is the topical problem in cryptography. It arises under true random number generators testing and preliminary analysis of cryptographic algorithms. Therefore, there are many well known statistical test suites for cryptographic applications: NIST SP 800-22 [1], FIPS 140, AIS 31, and other. The suites usually include many tests in order to detect different alternative hypotheses, because construction of a single test for a single "broad" alternative hypothesis, which includes rich family of probabilistic models, in most cases is impossible.

We refine a technique [2] on the base of universal predictors for statistical test construction. This technique allows to construct a statistical test for randomness using any predictor and detect broad alternative hypothesis, for which predictor is universal.

2 Randomness testing using a universal predictor

Let $X_1^n = X_1, X_2, \ldots, X_n$ be a sequence of binary random variables ($X_i \in \mathcal{A} = \{0, 1\}$) described by a set of conditional probabilities $\{\mathbf{P}\{X_{t+1} \mid X_1^t; \theta\}\}$ from class $\mathcal{M}$ compounded by probabilistic models with parameter $\theta \in \Theta$. If θ is known then maximum-likelyhood predictor (ML-predictor) predicts the $(t+1)$-th outcome $\hat{X}_{t+1}^*$ given previous t outcomes according to the most probable value for the given θ and has the minimal prediction error $\pi_t^*(X_1^t; \theta) = \mathbf{P}\{\hat{X}_{t+1}^* \neq X_{t+1} \mid X_1^t; \theta\}$:

$$\hat{X}_{t+1}^* = \begin{cases} 0, & \mathbf{P}\{X_{t+1} = 0 \mid X_1^t; \theta\} > \mathbf{P}\{X_{t+1} = 1 \mid X_1^t; \theta\}, \\ 1, & \text{otherwise}, \end{cases}$$

$$\pi_t^*(X_1^t; \theta) = \mathbf{P}\left\{\hat{X}_{t+1}^* \neq X_{t+1} \mid X_1^t; \theta\right\} = \min_{a \in \mathcal{A}} \mathbf{P}\left\{X_{t+1} = a \mid X_1^t; \theta\right\}.$$

If θ is unknown a predictor defines estimate $\hat{\theta}_t$ on the base of X_1^t and uses it to find $\hat{X}_{t+1}$. There exist classes $\mathcal{M}$ of models with universal predictors, i.e. predictors with asymptotically same prediction error probability as ML-predictor [3]:

$$\pi_t^*(X_1^t;\theta) - \mathbf{P}\{X_{t+1} \neq \arg\max_{a\in\mathcal{A}} \mathbf{P}\{X_{t+1} = a \mid X_1^t; \hat{\theta}_t\} \mid X_1^t;\theta\} \xrightarrow[t\to\infty]{P} 0.$$

Let one sequentially predicts $\hat{X}_t$ for X_t, $t = 1, 2, \ldots$ and builds a sequence $\{Y_t\}$ of successful prediction indicators: $Y_t = \mathbf{I}\{\hat{X}_t = X_t\}$. Consider the null hypothesis $\mathcal{H}_0$ that the sequence X_1^n is random, i.e. $\{X_t\}$ are i.i.d. symmetric Bernoulli trials. Clear, under $\mathcal{H}_0$ the indicators $\{Y_t\}$ are also i.i.d. Bernoulli trials with $\mathbf{P}\{Y_t = 1\} = 0.5$. Consider an alternative

$$\begin{aligned} \mathcal{H}_1: \quad & \max_{a\in\mathcal{A}} \mathbf{P}\left\{a \mid X_1^t;\theta\right\} = \frac{1}{2} + \varepsilon_{\theta;X_1^t} \geq \frac{1}{2}, \quad \exists t^0, i_1^*, \ldots, i_{t^0}^* : \\ & 0 < \varepsilon_{\theta;i_1^*,\ldots,i_{t^0}^*} < \frac{1}{2}, \quad \mathbf{P}\left\{X_1 = i_1^*, \ldots, X_t = i_{t^0}^*;\theta\right\} > 0. \end{aligned} \tag{1}$$

Lemma 1. *Let there exists the universal predictor for broad class $\mathcal{M}$ compounded by models described by $\mathcal{H}_1$ (1). If the universal for $\mathcal{M}$ predictor is used to construct $\{Y_t\}$ then $\mathbf{P}\{Y_t = 1\} = 0.5 + \varepsilon_{\theta;t}$ and there exists t^* such that $\varepsilon_{\theta;t} > 0$ for all $t \geq t^*$.*

Thus, the natural statistical test for $\mathcal{H}_0$ has the form:

$$\text{accept} \begin{cases} \mathcal{H}_0, & \text{if } 2\sqrt{n}\left(S_n - \frac{1}{2}\right) < \Delta, \\ \mathcal{H}_1, & \text{otherwise,} \end{cases} \quad S_n = \frac{1}{n}\sum_{t=1}^{n} Y_t, \ \Delta = \Phi^{-1}(1-\alpha), \tag{2}$$

where $\Phi(\cdot)$ is the standard normal c.d.f., α is a significance level.

3 Test based on the universal Lempel-Ziv predictor

Let X_1^n be the sequence of i.i.d. Bernoulli trials with unknown success probability $\theta \in (0;1)$. Consider simple hypothesis $\mathcal{H}_0$: $\theta = 0.5$ against complex alternative $\mathcal{H}_1$:

$$\mathcal{H}_1: \quad \mathbf{P}\left\{X_t = 1\right\} = \theta, \quad \theta = 0.5 + \epsilon, \quad 0 < |\epsilon| < 0.5. \tag{3}$$

Let us recall briefly the universal for stationary ergodic Markov chains of finite order Lempel-Ziv predictor [4] (LZ-predictor). The testing sequence is parsed into words according to Lempel-Ziv algorithm: new word is added in vocabulary, if it is the shortest one that has not yet been added. The vocabulary is organized into a binary tree: adding a word corresponds to adding leaves required to “read” the word from the tree. “Reading” a word implies parsing from a root to internal node or leaf: if the value of next letter is “0” one goes to the left from the current node, otherwise one goes to the right. After adding all words the weight of each leaf is set to 1. The weight of internal node is defined recursively as the sum of weights of its offsprings. Let *path* X_1^t be a traverse from the root “in the direction” of X_1, then X_2 and so on. If at moment $l < t$ the traversing reaches a leaf, then traversing resets and starts again from the root “in the direction” of X_{l+1}.

Let a *context* be defined as the path X_j^t where no reset happened. Let $N_{X_j,X_{j+1},\dots,X_t}$ be a weight of node where path ended. Now the Lempel-Ziv estimators of conditional probabilities are defined as:

$$\hat{\mathbf{P}}^{LZ}\{X_{t+1}=a|X_1^t\} = N_{X_j,X_{j+1},\dots,X_t,a}/N_{X_j,X_{j+1},\dots,X_t}. \tag{4}$$

Note that the estimator (4) is based on frequencies of the words and it may be biased under insufficient number of observations. We define L as the maximal length of contexts with frequency $N_{X_1^L} \geq K$, $\forall X_1^L \in \mathcal{A}^L$ for some given K and truncate the Lempel-Ziv tree to height L. Note that the weights of nodes in truncated tree remain unchanged.

We use two-staged procedure for test construction: let X_1^n be partitioned into two parts of size m and k respectively. Let us denote the first part by X_{-m+1}^0 and the second part by X_1^k. The first part X_{-m+1}^0 is used to build estimates (4). The second part X_1^k is used to build the test statistic in the following way. The first prediction is made from root and corresponds to value on the first level, which weight in the tree is maximal, then Y_1 is built. That is “memoryless” prediction, as the length of context is equal 0, and it is equivalent to the ML-predictor. Then we move from root in the direction of X_1. Now prediction for $\hat{X}_2$ is made from node, correspondent to X_1. It may be treated as ML-predictor for the first order Markov chain. After predicting X_L we reset to root and build a prediction for X_{L+1} again using “memoryless” predictor.

We will continue under condition that the Lempel-Ziv tree is calculated on X_{-m+1}^0 and it is **fixed**. Let $k \bmod (L+1) = 0$ for simplicity. Then all indicators of successful predictions can be divided into groups:

$$S_{k,m} = \tfrac{1}{k}\textstyle\sum_{t=1}^{k} Y_t = \tfrac{1}{k}\sum_{i=0}^{L} S^{(i)}, \quad S^{(i)} = \sum_{t=0}^{k/(L+1)-1} Y_{1+i+(L+1)t}\ ,$$

where $S^{(i)}$ is the sum of indicators of successful predictions using context of length i. Each context $s = s_1^l$ of length $l(s) = l$ specifies the node in the tree, and the prediction $\hat{X}^{(s)}$ from context s is known for given the weighted Lempel-Ziv tree. Let $\tilde{p}(s)$ be the probability of successful prediction after context s under hypothesis (3): $\tilde{p}(s) = ((1-\theta)\cdot \mathbf{I}\{\hat{X}^{(s)} = 0\} + \theta \cdot \mathbf{I}\{\hat{X}^{(s)} = 1\})$, and C be the set of contexts that will be used for predictions.

Lemma 2. *Under hypothesis* (3) *statistic* $S^{(i)}$*, calculated with the use of the given Lempel-Ziv tree built on* X_{-m+1}^0*, has the following properties:*

$$\mathbf{E}\left\{S^{(i)}\right\} = k\mu_{(i)}/(L+1), \quad \mathbf{D}\left\{S^{(i)}\right\} = k\mu_{(i)}(1-\mu_{(i)})/(L+1),$$
$$\mu_{(i)} = \sum_{s\in C:l(s)=i} \mathbf{P}\{s\}\tilde{p}(s), \quad \mathbf{P}\{s\} = \theta^{\sum_j s_j}(1-\theta)^{(i-\sum_j s_j)}.$$

Theorem 1. *Under hypothesis* $\mathcal{H}_1$ (3) *and* $k \to \infty$ *the power* $W_k(X_{-m+1}^0, \theta)$ *of test* (2)*, based on LZ-predictor built on given* X_{-m+1}^0*, has the following asymptotic expression:*

$$\left|W_k(X_{-m+1}^0,\theta) - \left(1 - \Phi\left(\frac{\Delta}{2\sqrt{n}\sqrt{\sigma}} + \frac{0.5-\mu}{\sqrt{\sigma}}\right)\right)\right| \to 0, \qquad \mu = \frac{1}{(L+1)}\sum_{i=0}^{L}\mu_{(i)},$$

$$\sigma = \frac{1}{k(L+1)}\Bigg(\sum_{i=0}^{L}\mu_{(i)}(1-\mu_{(i)}) + 2\sum_{j>i}\sum_{s\in C:l(s)=j}\mathbf{P}\{s\}\tilde{p}(s)\Big(\mathbf{I}\{\hat{X}(s_1^i) = s_{i+1}\} - \mu_{(i)}\Big)\Bigg).$$

4 Computation experiment

Let us perform comparison of the proposed test with Lempel-Ziv compression test from NIST SP800-22. Lempel-Ziv compression test was widely used in cryptographic applications as a part of NIST SP800-22 test suite. But it is noted in [1] that Lempel-Ziv compression test lacks theoretical foundation: theoretical mean and variance are equal to 50171.7 and 33.59 correspondingly, whereas their statistical estimates are equal to 69586.25 and 70.44 for a sequence of fixed length $n = 10^6$.

In order to confirm theoretical results of theorem 1 we perform Monte-Carlo experiments to estimate the power of the test (2) based on LZ-predictor (4). Figure 1 presents the theoretical power (denoted by line) and the Monte-Carlo estimates for the power (denoted by •) of test (2) for the given length $n = 10^6$ with $\theta = 0.503$, $m = 2 \cdot 10^5$, $k = 8 \cdot 10^5$, $K = 10^3$. One can see, that Monte-Carlo estimates agreed with theoretical results.

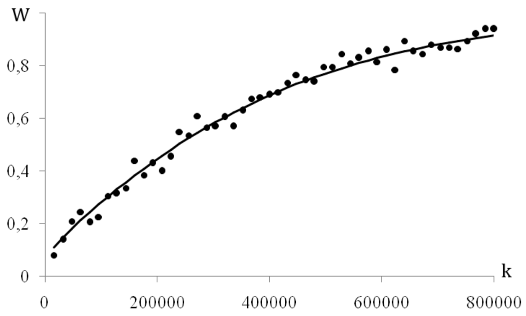

Figure 1: Performance of test

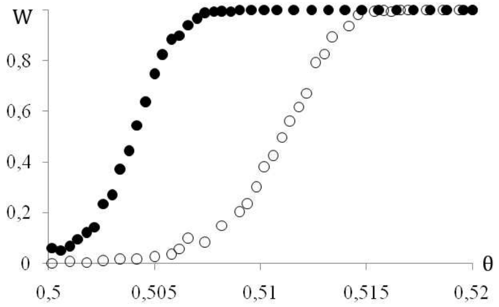

Figure 2: Comparison of the powers

Figure 2 presents estimated power of Lempel-Ziv compression test from NIST SP800-22 (denoted by ○) and the estimated power of the proposed test (2), constructed on the base of LZ-predictor (denoted by •) w.r.t. the parameter θ of Bernoulli trials with $n = 10^6$, $m = 2 \cdot 10^5$, $k = 8 \cdot 10^5$, $K = 8 \cdot 10^3$. One can see that the proposed test is more powerful than Lempel Ziv compression test.

References

[1] NIST Special Publication 800-22. (2001). *A statistical test suite for random and pseudorandom number generators for cryptographic applications.*

[2] Kostevich A.L., Shilkin A. V. (2007). On Approach to Randomness Testing on the base of the Universal Predictors. *Proc. of the 8 Int. Conf. "Computer Data Analysis and Modeling: Complex Stochastic Data and Systems")*, Vol. **1**, pp. 256-259.

[3] Suzuki J. (2003). Universal prediction and universal coding. *Systems and Computers*, Vol. **34** (6), pp. 1-11.

[4] Feder, M. (1992) Universal prediction of individual sequences / M. Feder, N. Merhav, M. Gutman // IEEE Trans. on Inf. Theory. — Vol. 38(4). — P. 1258–1270.