ANALYSIS OF TESTS FOR RANDOMNESS BASED ON UNIVERSAL PREDICTORS: BERNOULLI TRIALS CASE

A.L. Kostevich, A.V. Shilkin
Research Institute for Applied Problems of Mathematics and Informatics
Belarusian State University, Minsk, BELARUS
e-mail: kostevich@bsu.by

Abstract

An approach to randomness testing for Bernoulli trials on the base of universal predictors is considered. We propose two strategies for using universal predictors and derive the power of statistical test constructed on the base of maximum-likelihood predictor for Bernoulli trials. The results are extended to CTW, SPM and Lempel-Ziv universal predictors. Comparison of test constructed on the base of Lempel-Ziv predictor with Lempel-Ziv compression test, proposed in NIST SP800-22, is performed.

1 Introduction

Randomness testing is a topical problem in cryptography [1] and simulation [2]. Traditional theoretical and empiric approaches to randomness testing [2] do not allow to construct a single test for a general alternative hypothesis. Therefore the alternative approach to randomness testing on the base of **the sequence complexity** is developed [3, 4]. Loosely speaking, a sequence is considered to be random if it can not be compressed by a data compression method. There are a lot of universal data compression methods (e.g. the Lempel-Ziv algorithm), and they are widely used for randomness testing (see [1, 3, 4]). However, data compression methods usually have complicated algorithms and the investigation of their probabilistic properties is hard, e.g. both Maurer's test [3] and Lempel-Ziv compression test [1] use estimates for unknown parameters of their statistics' distributions. Thus, an alternative approach on the base of universal predictors was proposed in [5]. It allows to construct a statistical test for randomness using predictors that are universal for general alternative hypothesis.

In this paper we refine an alternative approach [5] for the parametric case of Bernoulli trials. We derive the power of statistical test constructed on the base of the maximum-likelihood (ML) predictor. The results are extended to the Context Tree Weighting (CTW), Sampled Pattern Matching (SPM) and Lempel-Ziv (LZ) predictors, that are universal for wide variety of models, including asymmetric Bernoulli trials.

2 Randomness testing using a universal predictor

Let us introduce in short construction of test procedure from [5]. Let $X_1^t = X_1, X_2, \ldots, X_t$ be a sequence of binary random variables ($X_i \in \mathcal{A} = \{0, 1\}$) described by a set of conditional probabilities $\{\mathbf{P}\left\{X_i \mid X_1^{i-1}\right\}\}$ from a class $\mathcal{M}$ of probabilistic models. A predictor defines estimates $\{\hat{\mathbf{P}}\{X_i \mid X_1^{i-1}\}\}$ of unknown probabilities and predicts the i-th outcome according to the most probable value. The universal for a class $\mathcal{M}$ predictor has asymptotically least prediction error probability:

$$\min_{a\in\mathcal{A}} \mathbf{P}\left\{a \mid x_1^t\right\} - \mathbf{P}\{\arg\min_{a\in\mathcal{A}} \hat{\mathbf{P}}\{a \mid x_1^t\} \mid x_1^t\} \xrightarrow[t\to\infty]{\mathbf{P}} 0. \tag{1}$$

According to proposed in [5] approach, for $i = \overline{1,t}$ one predicts sequentially the values $\{X_i\}$ and one builds a sequence Y_1^t of successful prediction indicators: $Y_i = \mathbf{I}\{\hat{X}_i = X_i\}$. We consider a null hypothesis $\mathcal{H}_0$ that the sequence X_1^t is random, i.e. $\{X_i\}$ are i.i.d. symmetric Bernoulli trials. Clear, under $\mathcal{H}_0$ the indicators $\{Y_t\}$ are also i.i.d. Bernoulli trials with $\mathbf{P}\left\{Y_i = 1\right\} = 0.5$. If some $\mathcal{H}_1$ that $\max_{a\in\mathcal{A}} \mathbf{P}\left\{a \mid x_1^t\right\} = \frac{1}{2} + \varepsilon_{x_1^t} \geq \frac{1}{2}$, and $\exists \varepsilon_{x_1^*,\dots,x_t^*} > \frac{1}{2}$ is true, then application of universal for $\mathcal{H}_1$ predictor causes the following marginal probabilities: $\mathbf{P}\left\{Y_i = 1\right\} = \frac{1}{2} + \varepsilon_i$, $\varepsilon_i > 0$ as $i \to \infty$. In this case the obvious statistical test for randomness is based on frequency of successful predictions and has the form:

$$\text{decide} \begin{cases} \mathcal{H}_0, & \text{if } 2\sqrt{t}\left(S - \frac{1}{2}\right) < \Phi^{-1}(1-\alpha), \\ \mathcal{H}_1, & \text{otherwise,} \end{cases} \qquad S = \frac{1}{t}\sum_{i=1}^{t} Y_i, \tag{2}$$

where $\Phi(\cdot)$ is the standard normal c.d.f., α is a significance level. In [5] it is claimed, that the test has a significance level α and is consistent.

3 The power of the test for Bernoulli trials

Consider as $\mathcal{H}_1$ the parametric model of independent Bernoulli trials for the sequence X_1^t with $p = \mathbf{P}\left\{X_i = 1\right\}$. The optimal predictor for this model is based on maximum-likelihood function. Therefore the counts $n_0(i)$, $n_1(i)$ of zeroes and ones are used to predict X_{i+1}:

$$\hat{X}_{i+1}^{ML} = \begin{cases} 1, & \text{if } n_1(i) \geq n_0(i), \\ 0, & \text{otherwise,} \end{cases} \qquad n_1(i) = \sum_{j=1}^{i} X_j, \quad n_0(i) = i - n_1(i). \tag{3}$$

By the law of large numbers, $n_1(i)/i \to p$ in probability as $i \to \infty$.

Now let us introduce two strategies for prediction.

The *first strategy* consists in training predictor on the first part of the sequence and predicting the second part. Let X_1^t be observed and m, n be positive numbers $(m + n = t)$. The first part X_1^m is used to estimate the parameter p, i.e. to count $n_0(m)$, $n_1(m)$. The second part X_{m+1}^n is used to build S-statistic of test (2). It should be noted that X_{m+1}^n is not used to update counts of zeroes and ones, i.e. $n_k(i) = n_k(m)$, $i \in \{m+1, \dots, n\}$, $k = 0, 1$. One can see that in the first strategy we try to avoid making wrong predictions in the beginning but S-statistic of the test (2) is built on $n < t$ observations.

The *second strategy* is a sequential prediction and re-training of the predictor.

Theorem 1. *Let $\mathcal{H}_1$ be true and the first strategy be used. Under $m, n \to \infty$ the power of the test* (2) *constructed on the base of ML-predictor* (3) *has the following form:*

$$W_{m,n} \approx \Big(1 - \Phi\big(c_1(\frac{1}{2} - p)\sqrt{m}\big)\Big)\Big(1 - \Phi\big(\frac{c_1}{2}\Phi^{-1}(1-\alpha) + (\frac{1}{2} - p)\sqrt{n}\big)\Big) +$$
$$+ \Big(\Phi\big(c_1(\frac{1}{2} - p)\sqrt{m}\big)\Big)\Big(1 - \Phi\big(\frac{c_1}{2}\Phi^{-1}(1-\alpha) - c_1(\frac{1}{2} - p)\sqrt{n}\big)\Big), \quad c_1 = \big(p(1-p)\big)^{-\frac{1}{2}}.$$

Proof follows from independence of X_1^m and X_{m+1}^n and the asymptotic normality of statistic $\hat{p} = \sum_{i=1}^k X_i$ for the parameter $p : \sqrt{k}(\hat{p} - p) \sim \mathcal{N}(0, p(1-p)),$ as $k \to \infty$.

Theorem 2. *Let $\mathcal{H}_1$ be true and the second strategy be used. Under $t \to \infty$ the power of the test* (2) *constructed on the base of ML-predictor* (3) *has the following form:*

$$W_t \approx 1 - \Phi\Big(\frac{\Phi^{-1}(1-\alpha)}{2\sqrt{t}\sigma} + \frac{\frac{1}{2} - \mu}{\sigma}\Big),$$
$$\sigma^2 \approx \frac{1}{t}p(1-p) + \frac{1}{t}(1-2p)^2\Phi(c_2\sqrt{t})(1 - \Phi(c_2\sqrt{t})) +$$
$$+ \frac{2}{t^2}(2p-1)^2 \sum_{u=1}^{t-1}\sum_{v=u+1}^{t}\big(\Phi(c_2\sqrt{u}, c_2\sqrt{v}) - \Phi(c_2\sqrt{u})\Phi(c_2\sqrt{v})\big) + O(t^{-\frac{3}{2}}) < \infty,$$
$$\mu \approx p + (1-2p)\Phi(c_2\sqrt{t}) + O(t^{-\frac{1}{2}}), \quad c_2 = \big(\frac{1}{2} - p\big)/\sqrt{p(1-p)}.$$

Proof. In the second strategy prediction at time t is made on $t-1$ observations, so $\{Y_t\}$ are dependent r.v. To find mean and variance one must find $\sum_{i=1}^{t-1}\Phi(c\sqrt{t})$, $\sum_{i=1}^{t-1}\Phi^2(c\sqrt{t})$. This was done by using integration instead of summation. Using joint distribution $\mathbf{P}\{Y_u, Y_v\}$ allows to find $\mathbf{Cov}\{Y_u, Y_v\}$.

Note. *It follows from theorems 1,2, that for any $m = \varepsilon t, 0 < \varepsilon < 1$ using strategy 2 for prediction is asymptotically more effective, whereas for fixed t one can find the value m that maximizes the power of test that uses strategy 1 for prediction.*

The ML-predictor for Bernoulli trials uses optimal sufficient statistic to estimate the parameter p. Universal predictors, due to universality for wide variety of models, use different statistics to estimate unknown probability. To predict the next outcome the SPM, LZ, CTW predictors use frequencies of s-grams $\{n_{i_1,\dots,i_s}\}$. In the case of the SPM and LZ predictors, the length s is unbounded as $t \to \infty$ random variable and depends on the sequence X_1^t. The CTW predictor uses weighted values of $\{n_{i_1,\dots,i_s}\}$, $s \in \{1, 2, \dots, D\}$, where D is fixed parameter. Since the law of large numbers holds in the sense that $\frac{n_{i_1,\dots,i_s,1}}{n_{i_1,\dots,i_s}} \to p$ as $t \to \infty$ in probability, SPM, LZ, CTW predictors are asymptotically equivalent to ML-predictor for the model of Bernoulli trials.

We perform Monte-Carlo simulation experiments to estimate the power of the test (2) based on ML-predictor (3) for Bernoulli trials with $p = 0.6, t = 800$. Figure 1 presents the theoretical power (denoted by line) and the Monte-Carlo estimates for the power (dashed line) of test (2), (3) using strategy 2. The theoretical power of test (2), (3) using strategy 1 with m observations for studying is denoted by $\circ$ for

$m = 50$, by • for $m = 200$, and by ◇ for $m = 550$. The Monte-Carlo estimates lied too close to corresponding theoretical power, thus we removed them from Figure 1 to avoid overload. One can see, that Monte-Carlo estimates agreed with theoretical results.

The LZ compression test is widely known test based on universal compression methods. But it is noted in [1] that the conditions of application of theoretical mean 50171.7 and variance 33.59 of test statistic for sequence of fixed length $t = 10^6$ remain ambiguous, and the statistical estimates 69586.25 and 70.44 of that parameters are taken.

Figure 2 presents estimated power of the test (2) constructed on the base of LZ-predictor using strategy 1 for prediction ($m = 2 \cdot 10^5$, denoted by •), the estimated power of LZ compression test [1] (denoted by ∘), and the estimated power of test on the base of ML-predictor (denoted by line) w.r.t the parameter p of Bernoulli trials of length $t = 10^6$. One can see that the test on the base of LZ-predictor is more powerful then LZ compression test [1] and its power lies closer to the power of the test on the base of optimal ML-predictor (3).

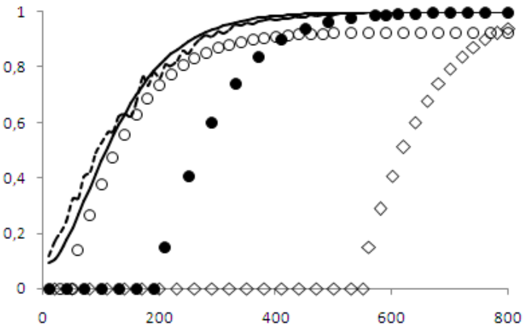

Figure 1: Performance of tests

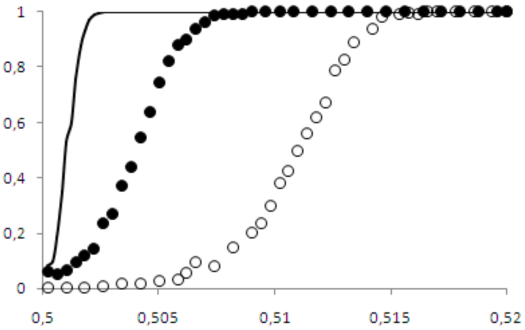

Figure 2: Comparison of the powers

References

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

[2] Knuth E.E. (1981). *The art of computer programming*, vol. 2. Addison-Wesley.

[3] Maurer U. (1992). A universal statistical test for random bit generators. *J. of Cryptology*. Vol. **5** (2), pp. 89-105.

[4] Ryabko B.Ya., Monarev V.A. (2005). Using information theory approach to randomness testing. *J. of Statistical Planning and Inference*. Vol. **133** (1), pp. 95-110.

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