ON APPROACH TO RANDOMNESS TESTING ON THE BASE OF THE UNIVERSAL PREDICTORS

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

Abstract

The proposed approach allows to construct a statistical test for randomness using the known predictors that are universal for general alternative hypotheses. The test is shown to be consistent and to have a required significance level. An example of the test based on the Lempel-Ziv algorithm is given.

1 Introduction

Randomness testing is the topical problem in cryptography and simulation [5]. There are two traditional approaches to randomness testing [2]:

– **Empiric approach.** One proposes some statistic with a known probability distribution in the case of the random sequence and constructs a statistical test for randomness testing using this statistic. But alternative hypotheses detected by the constructed test and the power of the test are usually unknown.

– **Theoretical (parametric) approach.** One chooses some meaningful (parametric) alternative hypothesis and uses the statistical technique to construct an optimal statistical test for detection of the alternative hypothesis. But, on the one hand, a set of the alternative hypotheses (and number of the tests accordingly) in applications is usually large. On the other hand, construction of the single test for a general alternative hypothesis in most cases is impossible.

Therefore the alternative approach to randomness testing on the base of **the sequence complexity** is being 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 [4, 5]). However data compression methods usually have complicated algorithms and the investigation of their probabilistic properties is hard, e.g. both Maurer's test [4] and the Lempel-Ziv compression test [5] use estimates of unknown parameters of their statistics' distributions (see also section 3). Fortunately, B.Ya. Ryabko and V.A. Monarev succeeded in construction of the mathematically well-founded test for randomness testing on the base of an *arbitrary* universal data compression algorithm [6].

In the paper we propose an alternative approach to randomness testing on the base of an *arbitrary* universal predictor. We were motivated by Yao's theorem [1] about universality of the next-bit test and close relation between universal predictors and both classical statistical methods and data compression methods.

2 Randomness testing using a universal predictor

Let $X_1, X_2, \ldots$ be a sequence of binary random variables ($X_t \in \mathcal{A} = \{0, 1\}$) described by a set of conditional probabilities $\{\mathbf{P}\{X_{t+1} \mid X_t, \ldots, X_1\}\}$. One observes the first t random variables and tries to predict the next outcome X_{t+1}.

If the probability model of the sequence is known, then the optimal forecast X^*_{t+1} is determined by the maximum of the corresponding conditional probability and the minimum prediction error probability π^*_t is attained:

$$\begin{gathered} X^*_{t+1} = \arg\max_{a\in\mathcal{A}} \mathbf{P}\{a \mid X_t, X_{t-1}, \ldots, X_1\}, \\ \pi^*_t(X_t, \ldots, X_1) = \mathbf{P}\left\{X^*_{t+1} \neq X_{t+1} \mid X_t, \ldots, X_1\right\} = \min_{a\in\mathcal{A}} \mathbf{P}\{a \mid X_t, X_{t-1}, \ldots, X_1\}. \end{gathered} \tag{1}$$

If the probability model of the sequence is unknown, then the "prediction" conditional probabilities $\{\hat{\mathbf{P}}\{X_{t+1} \mid X_t, \ldots, X_1\}\}$ should be defined. After that, the next outcome can be predicted as in (1) with the greater prediction error probability $\hat{\pi}_t$:

$$\begin{gathered} \hat{X}_{t+1} = \arg\max_{a\in\mathcal{A}} \hat{\mathbf{P}}\{a \mid X_t, X_{t-1}, \ldots, X_1\}, \\ \hat{\pi}_t(X_t, \ldots, X_1) = \mathbf{P}\left\{\hat{X}_{t+1} \neq X_{t+1} \mid X_t, \ldots, X_1\right\} \geq \pi^*_t(X_t, \ldots, X_1). \end{gathered} \tag{2}$$

Definition ([7]). *A predictor* (2) *is called* universal *for the class $\mathcal{M}$ if the prediction error tends to zero: $\hat{\pi}_t(X_t, \ldots, X_1) - \pi^*_t(X_t, \ldots, X_1) \to 0$ in probability as $t \to \infty$ w.r.t. any set of the conditional probabilities $\{\mathbf{P}\{X_{t+1} \mid X_t, \ldots, X_1\}\}$ from the class $\mathcal{M}$.*

It should be noted that there are other definitions of the universal predictor (see [7]) that differ from the given definition in specification of the prediction error. One can use the divergence or logarithm of the likelihood ratio for the conditional probabilities $\{\mathbf{P}\{X_{t+1} \mid X_t, \ldots, X_1\}\}$ and $\{\hat{\mathbf{P}}\{X_{t+1} \mid X_t, \ldots, X_1\}\}$ as the prediction error. In the last case and under the *parametric* description of the class $\mathcal{M}$ the prediction error coincides with the well-known likelihood ratio statistic.

We shall use $\hat{\pi}_t - \pi^*_t$ as the error probability of the predictor. This choice is more suitable for construction of a statistical test for randomness against an alternative hypothesis described by a general (*nonparametric*) class $\mathcal{M}$.

Let us define an auxiliary sequence of indicators of successful predictions. Let n binary random variables $X_1, \ldots, X_n$ be observed. For each $t = 1, 2, \ldots, n-1$ we will predict the next outcome X_{t+1} using the predictor (2), which is constructed on the base of the first t r.v. $X_1, \ldots, X_t$, and $\hat{X}_1 = 1$ for simplicity. After that, for each $t = 1, 2, \ldots, n$ we calculate the indicator the successful prediction: $Y_t = \mathbf{I}\{\hat{X}_t = X_t\}$. It is easily seen that the sequence $\{Y_t\}$ has the following properties:

P1. If the predictor (2) is deterministic, then the sequence $\{X_t\}$ can be restored from the sequence $\{Y_t\}$ (i.e. there is no loss of information when considering only $\{Y_t\}$), so $\mathbf{P}\{Y_t = 1 \mid Y_{t-1}, \ldots, Y_1\} = \mathbf{P}\{Y_t = 1 \mid X_{t-1}, \ldots, X_1\}$.

P2. If the predictor (2) is universal, then as $t \to \infty$

$$\mathbf{P}\{Y_t = 1 \mid X_{t-1}, \ldots, X_1\} - \max_{a\in\mathcal{A}} \mathbf{P}\{a \mid X_{t-1}, \ldots, X_1\} \to 0 \text{ in probability.}$$

Let us consider a null hypothesis $\mathcal{H}_0$ that the sequence $\{X_t\}$ is random:

$$\mathcal{H}_0: \quad \{X_t\} \text{ are i.i.d. Bernoulli r.v.}, \quad \mathbf{P}\{X_t = 0\} = \mathbf{P}\{X_t = 1\} = \frac{1}{2}.$$

Clear, under $\mathcal{H}_0$ the indicators $\{Y_t\}$ are also i.i.d. Bernoulli r.v. with $\mathbf{P}\{Y_t = 1\} = 0.5$.

Let us consider an alternative hypothesis $\mathcal{H}_1$ that the sequence $\{X_t\}$ is described by the following conditional probabilities:

$$\mathcal{H}_1: \quad \max_{a \in \mathcal{A}} \mathbf{P}\{a \mid X_{t-1}, \ldots, X_1\} = \frac{1}{2} + \varepsilon_{t, X_{t-1}, \ldots, X_1}, \quad \varepsilon_{t, x_{t-1}, \ldots, x_1} \geq 0. \tag{3}$$

If the predictor (2) is universal for $\mathcal{H}_1$, then the probability distribution (3) induces an unknown multivariate probability distribution of $\{Y_t\}$ with the marginal probabilities:

$$\mathcal{H}_1^{(Y)}: \quad P\{Y_t = 1\} = \frac{1}{2} + \varepsilon_t, \quad \varepsilon_t \geq 0 \quad \text{as} \quad t \to \infty.$$

In this case the natural statistical test for randomness is based on the frequency of the successful predictions and has the form:

$$\text{decide} \begin{cases} \mathcal{H}_0, & \text{if } 2\sqrt{n}\left(S - \frac{1}{2}\right) < \Phi^{-1}(1-\alpha), \\ \mathcal{H}_1, & \text{otherwise}, \end{cases} \quad S = \frac{1}{n}\sum_{t=1}^{n} Y_t, \tag{4}$$

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

Claim 1. *If the test* (4) *is based on an arbitrary predictor, then the type I error probability of the test tends to* α *as* $n \to \infty$.

Claim 2. *If the test* (4) *is based on the universal for* $\mathcal{H}_1$ *predictor,* $\{Y_t\}$ *satisfy the central limit theorem and* $\mathbf{D}\{S\} \asymp n^{-1}$ *as* $n \to \infty$, *then the power of the test is:*

$$W_n \approx \Phi\left(c_1 \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t - c_2\right) \qquad \textit{and} \qquad W_n \to 1 \quad \textit{as} \quad \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \to +\infty,$$

where c_1, c_2 *are positive constants.*

3 Test based on the Lempel-Ziv predictor

It is known that the Lempel-Ziv predictor is universal for the class $\mathcal{M}$ of stationary and ergodic Markov sources of finite order. The Lempel-Ziv algorithm partitions a sequence into phrases of variable size such that a new phrase is the shortest subsequence not seen in the past as a phrase. The Lempel-Ziv algorithm predicts the next outcome (see (2)) as the most frequent suffix given a prefix of a phrase.

The Lempel-Ziv test of NIST [5] is based on the number of phrases in a sequence of length $n = 10^6$. But it is noted in [5] that the conditions of application of theoretical mean 50171.7 and variance 33.59 of this statistic remain ambiguous, and the statistical estimates 69586.25 and 70.44 of that parameters are taken.

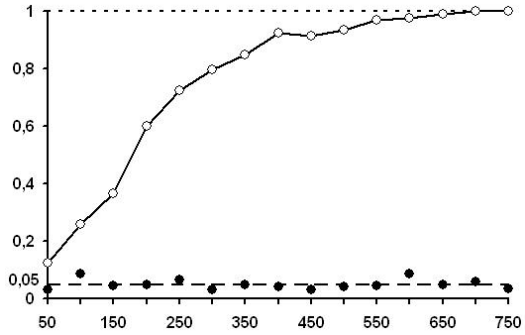

Figure 1: Performance of the test

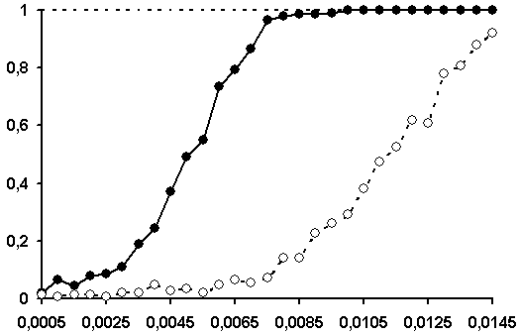

Figure 2: Comparison of the powers

In contrast to [5] and other approaches, we proposed the well-founded way to construct the test on the base of the Lempel-Ziv algorithm or any other predictor.

We conducted simulation experiments to estimate the performance of the test (4) based on the Lempel-Ziv predictor. As the hypothesis $\mathcal{H}_1$ we used the Markov chain model with $\mathbf{P}\{X_{t+1} = 1 \mid X_t = 0\} = \mathbf{P}\{X_{t+1} = 0 \mid X_t = 1\} = 0.5 + \varepsilon$. Figure 1 presents calculated estimates of the type I error probability (denoted by $\bullet$) and the power (denoted by $\circ$) of the proposed test w.r.t. length n, where $\alpha = 0.05$, $\varepsilon = 0.15$. One can see that the proposed test is consistent and the type I error probability of the test is approximately α. Figure 2 presents calculated estimates of the power of the proposed test (denoted by $\bullet$) and the Lempel-Ziv test of NIST [5] (denoted by $\circ$) w.r.t the parameter ε of the Markov chain of length $n = 10^6$. One can see that the power of the proposed test is greater then the power of NIST's test.

References

[1] Goldreich O., Goldwasser S., Micali S. (1986). How to construct random functions. *J. of the Association for Computing Machinery*. Vol. **33**, pp. 792-807.

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

[3] Kolmogorov A.N. (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission*. Vol. **1**, pp. 3-11.

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

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

[6] 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.

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