Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Statistical Planning and Inference 123 (2004) 365–376

journal of statistical planning and inference

www.elsevier.com/locate/jspi

A new test for randomness and its application to some cryptographic problems☆

B.Ya. Ryabko*, V.S. Stognienko, Yu.I. Shokin

Institute of Computational Technologies of Siberian Branch of the Russian Academy of Sciences, Lavrentiev str. 6, 630090, Novosibirsk, Russia

Received 28 July 2002; accepted 26 February 2003

Abstract

We consider the problem of testing the hypothesis H_0 that the letters from some alphabet $A=\{a_1,a_2,\ldots,a_k\}$ obey the uniform distribution, when k is large. The problem is of interest for random number testing and some cryptographic applications where $k=2^{10}\sim 2^{30}$ and greater. In such a case it is difficult to use the well-known chi-square test since the sample size for it must be greater than k.

We suggest an *adaptive chi-square test* which can be successfully applied for testing some kinds of H_1 even if the sample size is much smaller than k. This statement is proved theoretically and confirmed experimentally.

MSC: 62G; 62G10

Keywords: Hypothesis testing; Chi-square test; Adaptive testing; Random number testing; Block cipher testing

1. Introduction

The chi-square test is one of the most popular hypothesis tests, which is widely applied in economics, biology, cryptography and many other fields. For example, the chi-square test is used for testing random number generators and block ciphers'

☆ Part of this work was done when B. Ryabko had visited the Centre of Excellence of European Union, Institute of Mathematics, Polish Academy of Science, Warsaw. B. Ryabko was also supported by INTAS grant no. 00-738.

* Corresponding author. Tel.: +7-383-2284938; fax: +7-383-2661039.
E-mail address: ryabko@neic.nsk.su (B.Ya. Ryabko).

0378-3758/$ - see front matter
doi:10.1016/S0378-3758(03)00149-6

suitability for random number generators (see, for example, Knuth, 1981; Rukhin et al., 2001).

In such applications the number of categories (and, consequently, the number of degrees of freedom of χ^2 distribution) is very large and, thereby, the sample size should be also large. So, in such cases, performing the chi-square test requires a lot of time. Moreover, it is often difficult to obtain such large samples and the chi-square test cannot be applied.

We suggest a new method, which we call the *adaptive chi-square test*. It is shown that the new test can be applied when the sample size is much smaller than that required for the usual chi-square test.

First, let us explain the main idea of the new test. Let there be a hypothesis H_0 which states that the letters from some alphabet $A=\{a_1,a_2,\dots,a_k\}$, $k>2$, are distributed uniformly (i.e. $p(a_1)=p(a_2)=\cdots=p(a_k)=1/k$) against the alternative hypothesis H_1 that the true distribution is not uniform. Let there be given a sample which can be used for testing. The sample is divided into two parts, which are called the *training sample* and the *testing sample*. The training sample is used for estimation of frequencies of the letter occurrences. After that the letters of the alphabet A are combined into subsets $A_1,A_2,\dots,A_s$, $s\geqslant 2$, in such a way that, first, one subset contains letters with close (or even equal) frequencies of occurrence and, second, s is much less than k (say, $k=2^{20}, s=2$). Then, the set of subsets $\{A_1,A_2,\dots,A_s\}$ is considered as a new alphabet and the new hypothesis $\hat{H}_0: p(A_1)=|A_1|/k, p(A_2)=|A_2|/k,\dots, p(A_s)=|A_s|/k$ and the alternative hypothesis, which is the negation of $\hat{H}_0$, are tested based on the second ('testing') part of the sample. Obviously, if H_0 is true, then $\hat{H}_0$ is also true and, if $\hat{H}_1$ is true, then H_1 is true. That is why this new test can be used for testing the initial H_0 and H_1. The idea of such a scheme is quite simple. If H_1 is true, then there are letters with relatively large and relatively small probabilities. Generally speaking, the high-probable letters will have relatively large frequencies of occurrence and will be accumulated in some subsets A_i whereas low-probable letters will be accumulated in the other subsets. That is why this difference can be found based on the testing sample. It should be pointed out that a decrease in the number of categories from large k to small s can essentially increase the power of the test and, therefore, can essentially decrease the required sample size. More exactly, it will be shown that the sample size can be decreased in $\sqrt{k}$ times, which can be important when k is large. We carried out some experiments in addition to a theoretical investigation of the suggested test. Namely, we tested ciphered texts in English and in Russian in order to distinguish them from random sequences. It is worth noting that the problem of recognition of ciphered texts in a natural language is of some interest for cryptology, see Schneier (1996). It turns out, that the suggested test can distinguish ciphered English and Russian texts from random bit sequences, basing on samples which are essentially smaller than it is required for usual chi-square test.

It should be noted that the learning techniques has been applied to many statistical problems. For example, Markov Approximation and Neural Network based clustering are used to make testing more efficient, see, for example, reviews in Vapnik (1995) and Devroye et al. (1996). Besides, it is known that stratification can be used as a tool in the analysis of randomness, see L'Ecuyer and Simard (1999) and Wegenkittl and

Matsumoto (1999). In contrast to those approaches, the suggested method is intended to solve one particular statistical problem, which was not considered earlier. Namely, it is offered to increase the power of the chi-square test by learning and grouping, in case when the sample size is relatively small.

The rest of the paper is organized as follows. Section 2 contains necessary information from the mathematical statistics and some auxiliary results about the chi-square test. The description of the suggested test and its properties are given in Section 3. Section 4 contains experimental results about recognition of ciphered texts. Appendix contains some proofs.

2. Chi-square test

First we give some required information concerning the chi-square test. Let there be two following hypotheses about a probability distributions on a set (or alphabet) A:

$$\mathrm{H}_0\colon\ p(a_1)=p_1^0, p(a_2)=p_2^0,\ldots, p(a_k)=p_k^0, \qquad \mathrm{H}_1=\neg\mathrm{H}_0, \tag{1}$$

where $p=(p_1^0, p_2^0,\ldots, p_k^0)$ is a certain distribution on the A. Let $x_1,x_2,\ldots,x_N$ be a sample and v_i is the number of occurrences of $a_i\in A$ in the sample. (In statistics $a_1,\ldots,a_k$ are often called categories.) The chi-square test is applied by calculating

$$x^2=\sum_{i=1}^{k}\frac{(v_i-Np_i^0)^2}{Np_i^0}. \tag{2}$$

The less x^2, the more probable H_0. More exactly, when the chi-square test is applied, there is such a threshold constant γ, that H_0 is accepted, if $x^2<\gamma$. Otherwise, H_0 is rejected. (The constant γ depends on k and the level of significance).

It is known that x^2 asymptotically follows the chi-square distribution with $(k-1)$ degrees of freedom (χ^2_{k-1}) if H_0 is true. On the other hand, if H_1 is true, x^2 asymptotically follows a so called noncentral chi-square distribution with $(k-1)$ degrees of freedom and a parameter λ ($\hat{\chi}^2_{\lambda,k-1}$) where λ is defined by

$$\lambda=N\pi, \qquad \pi=\sum_{i=1}^{k}\frac{(p_i^0-p_i^1)^2}{p_i^0}. \tag{3}$$

Here N is the sample size and $p_k^1=p(a_k)$ when H_1 is true, see Kendall and Stuart (1961) and Lehmann (1959).

It is shown by Kendall and Stuart (1961) that

$$E_{\mathrm{H}_0}(x^2)=k-1, \qquad V_{\mathrm{H}_0}(x^2)=2(k-1), \tag{4}$$

$$E_{\mathrm{H}_1}(x^2)=(k-1)+\lambda, \qquad V_{\mathrm{H}_1}(x^2)=2(k-1)+4\lambda, \tag{5}$$

where E_{H_i} and V_{H_i} are the mean value and variance, correspondingly, when H_i is true, $i=1,2$.

If the level of significance (or a Type I error) of the chi-square test is $\alpha, \alpha\in(0,1)$, then H_0 is rejected is if $x^2\geqslant\chi^2_{k-1;(1-\alpha)}$. (Here $\chi^2_{k-1;(1-\alpha)}$ is the $(1-\alpha)$-*value* of the χ^2_{k-1} distribution.)

It is important to note that such an approximation is valid when N is quite large. Thus, many authors recommend to take such a sample size N that $Np_i^0 \geqslant 5$ for all $i = 1, \ldots, k$. (If this inequality is not true for some category i, this category should be combined with some other category and this procedure should be repeated as long as the inequality becomes true for all categories.) Obviously, if the inequality is true, then $N \geqslant 5k$.

3. Adaptive chi-square test: description and theoretical consideration

First we give some informal explanation. Let us assume that there exist such letters a_i from the alphabet A for which the fractions (p_i^1/p_i^0) are equal. If we combine the letters from A into subsets in such a way that one subset contains letters for which the fraction (p_i^1/p_i^0) are equal, then the required sample size can be decreased. Indeed, if we denote the number of such groups as $\check{k}$, we obtain from (4) and (5) the equalities

$$E_{\mathrm{H}_0}(x^2) = \check{k} - 1, \quad V_{\mathrm{H}_0}(x^2) = 2(\check{k} - 1),$$

$$E_{\mathrm{H}_1}(x^2) = (\check{k} - 1) + \lambda, \quad V_{\mathrm{H}_1}(x^2) = 2(\check{k} - 1) + 4\lambda.$$

Obviously, the number of groups $\check{k}$ is less than k and we can see from the last equalities, (4) and (5) that the difference $E_{\mathrm{H}_1}(x^2) - E_{\mathrm{H}_0}(x^2)$ is the same for the initial and grouped alphabet, whereas both variancies $V_{\mathrm{H}_0}(x^2)$ and $V_{\mathrm{H}_1}(x^2)$ of the grouped alphabet are less than for the initial one. That is why the required sample size can be decreased, if the chi-square test is applied to the grouped alphabet.

We do not consider this method in details because the alternative hypothesis H_1 is not known beforehand. In order to overcome obstacles, we suggest, first, to estimate the frequencies of occurrence of letters from the alphabet A using a part of the sample and, then, to implement the grouping using frequencies instead of probabilities $p_1^1, p_2^1, \ldots, p_k^1$. After that the independent second part of the sample is used for testing.

The more formal description of the suggested adaptive chi-square test is as follows. There are hypotheses H_0 and H_1 defined by (1) and the sample $x_1, x_2, \ldots, x_N$. The sample is divided into two following parts $x_1, x_2, \ldots, x_m$ and $x_{m+1}, x_2, \ldots, x_N$, which are called the training sample and the testing sample, correspondingly. The training part is used for finding the frequencies of occurrence of letters from the alphabet A which will be denoted by $\tilde{p}_1^1, \tilde{p}_2^1, \ldots, \tilde{p}_k^1$. Then, we divide the alphabet A into subsets $\{A_1, A_2, \ldots, A_s\}$, $s > 1$, combining the letters for which the fractions $(\tilde{p}_i^1/p_i^0)$ are close, in one subset. After that the new following hypotheses

$$\hat{\mathrm{H}}_0\colon\ p(A_1) = \sum_{a_i \in A_1} p_i^0,\ p(A_2) = \sum_{a_i \in A_2} p_i^0, \ldots, p(A_s) = \sum_{a_i \in A_s} p_i^0, \quad \hat{\mathrm{H}}_1 = \neg \hat{\mathrm{H}}_0$$

are tested based on the testing sample $x_{m+1}, x_2, \ldots, x_N$. We do not describe the exact rule of finding the parameters m, N and s and do not also define exactly how to construct the subsets $\{A_1, A_2, \ldots, A_s\}$, but we recommend to implement some experiments for finding the parameters, which make the total sample size N minimal (or, at least, acceptable). The point is that there are many problems in cryptography and other applications where it is possible to implement some experiments for optimizing the

parameter values and, then, to test hypothesis basing on independent data. For example, there are some problems in cryptography where it is possible to carry out experiments with known "secret" keys for finding suitable m and N and, then, use them for real problems. Such a problem will be considered in the next paragraph while we proceed with theoretical investigation at the rest of this paragraph.

Let us consider an example where the required sample size of the adaptive chi-square test is equal to $O(\sqrt{k})$, whereas a usual chi-square test can be used if the sample size is more than k. Let us assume, that the number of categories k is even and let

$$p_1^0 = p_2^0 = \cdots = p_k^0 = \tfrac{1}{k}, \tag{6}$$

$$p_{i_1}^1 = p_{i_2}^1 = \cdots = p_{i_{k/2}}^1 = \tfrac{1}{k}(1 + 1/2),$$

$$p_{i_{(k/2)+1}}^1 = \cdots = p_{i_k}^1 = \tfrac{1}{k}(1 - 1/2), \tag{7}$$

where $\{i_1, \ldots, i_{k/2}\} \cup \{i_{(k/2)+1}, \ldots, i_k\} = \{1, \ldots, k\}$. It turns out that the adaptive chi-square test can be successfully applied when the total sample size is $O(\sqrt{k})$.

Theorem. *Let α and β be in the interval $(0,1)$. Then, we can find such $k_{\alpha,\beta}$ that for each even $k > k_{\alpha,\beta}$ there exists an adaptive chi-square test with the training sample size $m(k)$ and testing sample size $n(k)$, such that for every partition $\{\{a_{i_1}, \ldots, a_{i_{k/2}}\}, \{a_{i_{(k/2)+1}}, \ldots, a_{i_k}\}\}$, and H_0, H_1 complying with (6), (7), the following is true*:

(i) $(m(k) + n(k)) \leqslant \dot{c}\sqrt{k}$, *where* $\dot{c} > 0$ *and does not depend on* k,
(ii) *the Type I error is less than* α *and the Type II error is less than* β.

Proof. Let $\hat{A}_i$ be the set of letters from A which occurred i times in the training sample $x_1 x_2 \ldots x_m, i = 0, 1, \ldots$. The proof will be based on the following lemmas.

Lemma 1. *If k goes to infinity, C is a positive constant, H_0 is true and $m = \lceil C\sqrt{k} \rceil$, then $E(\sum_{r=2}^{k}(r-1)|\hat{A}_r| \leqslant \dot{C}$, where $E()$ means the expectation and $\dot{C}$ does not depend on k.*

Lemma 2. *If k goes to infinity, C is a positive constant, $m = \lceil C\sqrt{k} \rceil$ and H_1 is true, then*

$$E\left(\sum_{a \in \hat{A}_1} p(a)\right) = 5C/(4\sqrt{k}) + O(1/k),$$

$$E\left(\sum_{a \in \hat{A}_2} p(a)\right) = 7C^2/(4k) + o(1/k),$$

$$E\left(\sum_{r=3}^{k}\left(\sum_{a \in \hat{A}_r} p(a)\right)\right) = o(1/k).$$

The proofs of the lemmas are given in appendix. Let us proceed with the proof of the theorem.

Let the training sample size $m(k)$ and the tasting sample size $n(k)$ be defined by

$$m(k) = \lfloor \sqrt{k} \rfloor, \quad n(k) = \lceil c\sqrt{k} \rceil, \tag{8}$$

where the positive constant c will be defined later.

By definition, the test uses two following subsets (superletters):

$$A_0 = \hat{A}_0, A_1 = \bigcup_{r=1}^{k} \hat{A}_r. \tag{9}$$

Obviously, $m(k) = \sum_{r=1}^{k} r|\hat{A}_r|$ and, consequently,
$|A_1| = m(k) - \sum_{r=1}^{k} (r-1)|\hat{A}_r|$. From this equality and (8) we obtain

$$\sqrt{k} - \sum_{r=2}^{k} (r-1)|\hat{A}_r| < |A_1| < \sqrt{k} + 1. \tag{10}$$

Let us define

$$P(A_i \backslash \mathrm{H}_j) = \sum_{a \in A_i} p(a), \quad i, j = 0, 1. \tag{11}$$

$P(A_i \backslash \mathrm{H}_j)$ are random variables for $i, j = 0, 1$, because they depend on the training sample $x_1, \ldots, x_m$. Their estimations will be based on Lemmas 1, 2 and the following well-known Markov inequality: For any non-negative random variable ξ and any $\Delta > 0$

$$\Pr\{\xi \geqslant \Delta\} \leqslant E(\xi)/\Delta.$$

From Lemma 1 and (8) we can see that $E(\sum_{r=2}^{k}(r-1)|\hat{A}_r|) < \tilde{C}$, where $\tilde{C}$ does not depend on the alphabet size k. If we define $\Delta = \sqrt[3]{k}$ and apply Markov inequality to $|\bigcup_{r=2}^{k} \hat{A}_r|$ we obtain the inequality

$$\Pr\left\{ \sum_{r=2}^{k} (r-1)|\hat{A}_r| \geqslant \sqrt[3]{k} \right\} \leqslant \tilde{C}/\sqrt[3]{k}.$$

From the last inequality and (10) we obtain the estimation

$$\Pr\{||A_1| - \sqrt{k}| \geqslant \sqrt[3]{k}\} \leqslant \tilde{C}/\sqrt[3]{k}.$$

If the hypothesis H_0 is true, then $p(a) = 1/k$ for all $a \in A$ and we obtain from the last inequality and (9), (11) that

$$\Pr\left\{ \left| P(A_1 \backslash \mathrm{H}_0) - \frac{1}{\sqrt{k}} \right| \geqslant (1/\sqrt[3]{k^2}) \right\} \leqslant \tilde{C}/\sqrt[3]{k}.$$

In the same way we can derive from the Lemma 2 that

$$\Pr\left\{ \left| P(A_1 \backslash \mathrm{H}_1) - \frac{5}{4\sqrt{k}} \right| \geqslant (1/\sqrt[3]{k^2}) \right\} \leqslant \tilde{C}_1/\sqrt[3]{k}.$$

Taking into account that the testing sample size $n(k) = \lceil c\sqrt{k} \rceil$ (see (8)), we can see from two last inequalities that, with the probability 1,

$$P(A_1 \backslash \mathrm{H}_0) n(k) = c + \mathrm{o}(1), \quad P(A_1 \backslash \mathrm{H}_1) n(k) = 5c/4 + \mathrm{o}(1),$$

where $k \to \infty$. In torn, from those equalities and (2) we can see that (with probability 1) $x^2 = c/16 + \mathrm{o}(1)$, if H_1 is true and $x^2 = \mathrm{o}(1)$, if H_0 is true. Taking into account that chi-square test can be applied if the sample size is not less than $5\tilde{k}$, where $\tilde{k}$ is number of categories (or superletters) we define the constant $\ddot{c}$ by $\ddot{c} = \max\{5, 16\chi^2_{1,1-\alpha}\}$. Now we can see that the test can be applied, if c in (8) is not less than $\ddot{c}$. From the definition (2) and two last equalities we can see that H_0 will be accepted with probability 1, if H_0 is true, and rejected (with probability 1), if H_1 is true, when $k \to \infty$. It means that there exists such k^* that H_0 will be accepted with probability larger than $1 - \alpha$, if H_0 is true and $k > k^*$, and H_0 will be rejected with probability larger than $(1 - \beta)$, if H_1 is true and $k > k^*$. We obtain from the definition $\ddot{c}$ and (8) that the total sample size $(m(k) + n(k))$ is less than $(\ddot{c} + 1)\sqrt{k}$. So, we can see that the theorem is proved for $k_{\alpha,\beta} = k^*$.

Comment 1. The theorem can be extended. Namely, if we consider the following more general hypothesis H_1

$$p^1_{i_1} = p^1_{i_2} = \cdots = p^1_{i_{k/2}} = \tfrac{1}{k}(1 + \delta),$$

$$p^1_{i_{(k/2)+1}} = \cdots = p^1_{i_k} = \tfrac{1}{k}(1 - \delta),$$

$\delta \in (0, 1)$ instead of (7), we can prove the claim 2, but the constant c will be depended on δ. The way of proving is completely the same.

Comment 2. It can be easily seen that the test power is maximal if the training sample size and the testing sample size are equal, but we do not focus an attention on this fact, because, generally speaking, there exist alternative hypotheses H_1 for which it is not true.

4. The experiments

As it was shown by Maurer (1992), the problem of testing randomness is very important for many cryptographic applications. In this section we consider one of such applications concerned with the block ciphers.

Block ciphers have been widely used in practice and attracted attention of many researches. Recently, National Institute of Standards and Technology (USA) carried out a competition "Advanced Encryption Standard (AES)", whose purpose was to find a new block cipher which could be used as a standard. The cipher has to meet many requirements and, in particular, its output should look like completely random even if the input is not; see Nechvatal et al. (2000) and Soto and Bassham (2001). (Here the completely random output means a bit sequence generated by the Bernoulli source with equal probabilities of 0's and 1's.) For example, even if the input is a natural

language text (English, Russian, etc.), the ciphertext has to be indistinguishable from a completely random sequence.

Of course, theoretically, this cannot be achieved. The point is that a usual representation of texts in computers is very redundant. More exactly, one letter of the text is often represented as an 8-bit word (or byte). So, if we take into account the well-known estimation of the Shannon entropy of English (see, e.g., Cover and Thomas, 1991), we can see that the Shannon entropy per bit is much less than 1. It can be easily seen that the Shannon entropy of the ciphered text is not greater than the sum of the key entropy and the entropy of the input. Hence, if, for example, someone uses such a cipher for which the lengths of input and output files are equal and applies this cipher to an English language text using one key for the large text, the Shannon entropy of the ciphered text will be approximately the same as for the original text and, consequently, will be less than 1 per bit. It is well known that the Shannon entropy of the completely random sequence is 1 (per bit), that is why the ciphered text in English (and other languages) is not completely random, if a large file is ciphered using one key.

So, apparently, the problem of constructing tests which can distinguish ciphered texts from random sequences can be considered as a good example for estimation of a power of statistical tests, because, on the one hand, it is known that ciphered texts cannot be completely random in principle and, on the other hand, the ciphers are constructed in such a way that the ciphered sequences should look random. (As much as possible). Besides, the problem of testing of the ciphered texts is of some interest for cryptography, see Schneier (1996). That is why the problem of distinguishing a ciphered text from a random bit sequence was chosen for experimental investigation of the adaptive chi-square test.

Let us describe the experiments in more details. We considered the block ciphers Rijndael and RC6. The first cipher has been proposed by NIST as Advanced Encryption Standard (AES) and the second one was selected as a finalist for AES and is widely used in practice. We applied adaptive chi-square test to ciphered English and Russian texts using as source of texts "Moshkov Library" (http://lib.ru/). Texts were combined in large files and each file was ciphered by either Rijndael or RC6 with 128-bit block length in such a way that one key was used for ciphering of all blocks from one file. (Such mode of encryption is called Electronic Code Book). Then we took 40 files of texts in English and 40 such files in Russian. Each file was ciphered with a randomly chosen key and tested for randomness using new algorithms and the usual chi-square test.

When the new algorithm was applied, each ciphered file was divided into 24-bit words and the obtained sequence was considered as a text over an alphabet of 2^{24} letters. (By definition, the alphabet letters are all 24-bit words.) The adaptive chi-square test was applied to testing the hypothesis about randomness (i.e. $H_0 = \{p(a_1) = p(a_2) = \cdots = p(a_{2^{24}}) = 2^{-24}\}$, $H_1 = \neg H_0$). The ciphered files were divided into two equal parts in such a way that the first part was used as a training sample and the second part as a testing sample. The training sample was used for estimation of the number of the letter occurrences and the alphabet was divided into three subsets $\{A_0, A_1, A_2\}$ in such a way that the set A_0 contained letters which were not met in the training sample,

Table 1
Results of experiments

Length	102,400	204,800	512,000	1,024,000	2,048,000
	Usual/new	Usual/new	Usual/new	Usual/new	Usual/new
RC6 Russian	4/11	4/13	4/28	23/32	27/35
RC6 English	7/19	6/24	5/31	27/31	27/35
AES Russian	3/10	4/18	5/24	18/31	25/34
AES English	4/17	8/26	8/28	25/30	30/33

A_1 contained letters which were met once and A_2 contained all other letters. Then, according to description of the adaptive chi-square test, the hypotheses

$$\hat{\mathrm{H}}_0 = \{p(A_i) = |A_i|/2^{24},\ i = 0, 1, 2\}, \quad \hat{\mathrm{H}}_1 = \neg\hat{\mathrm{H}}_0$$

were tested based on the testing sample. The level of significance was 0.05.

It is worth noting that a lot of experiments were carried out for finding the value of the parameter s, the lengths of training and testing samples and the number of subsets. The aim of the experiments was to find the parameters, which maximize the test power. It is also important to note that the new (independent) data were used for described below experiments.

The results are given in Table 1. For example, the second column contains results for 102,400-files. The number 11 from the first column "new" means that the hypothesis H_0 about randomness was rejected 11 times, when 40 Russian files ciphered by RC6 were tested by the adaptive chi-square test. Analogously, the last number 33 in the last column "new" means that H_0 was rejected 33 times, when 40 English files ciphered by AES were tested by the adaptive chi-square test.

As it was mentioned above, the usual chi-square test was also applied for testing H_0. For this purpose each file was considered as a bit stream and divided into s-bit subwords (blocks) and the hypothesis H_0 that each word $u \in B_s = \{0, 1\}^s$ has the probability 2^{-s} was tested against $\mathrm{H}_1 = \neg\mathrm{H}_0$, where s took values $1, 2, \ldots$. As it has been noted, the chi-square test can be applied for testing H_0 if the sample size is not less than $5|B|$. That is why H_0 was tested for such s that $8L/s \geqslant 52^s$, where L is the length of the file in bytes. (Indeed, the sample size is equal to $\lfloor 8L/s \rfloor$ as well as the alphabet size is 2^s.) Thus, for the files of the length of 102,400 bytes H_0 was tested for $s = 1, 2, 3, \ldots, 13$, for the length 204,800 H_0 was tested for $s = 1, 2, 3, \ldots, 14$, etc. As before, the level of significance was 0.05.

The largest number of cases when H_0 was rejected is written down in the table. (So, we write down the best result for the usual chi-square test over all possible values of s.) For example, 40 Russian texts ciphered by RC6 were tested by the usual chi-square test for the block lengths $s = 1, 2, \ldots, 14$. When $s = 1$, H_0 was rejected 2 times, when $s = 2$–3 times, etc. The maximal value of rejections was 4 and it was obtained when $s = 8, 10$ and 12. So, 4 is written down in the corresponding column "usual". Similarly, when 40 2,048,000-byte English files ciphered by AES were tested by the usual chi-square test with $s = 16$, H_0 was rejected 30 times and this number of rejections was larger

than for all other values of the block length s. That is why 30 is written down at the bottom of the last column "usual".

We can see from the table that the new test can detect non-randomness more efficiently than the usual chi-square test. In other words, the power of the adaptive chi-square test is larger than that of the usual one, when the sample size is relatively small.

Acknowledgements

The authors wish to thank Prof. Subir Ghosh for valuable advice and Prof. George Marsaglia for providing them with "The Marsaglia Random Number CDROM".

Appendix A

Proof of Lemma 1. From (6) we obtain

$$E(|\hat{A}_2|) = k\binom{m}{2}\left(\frac{1}{k}\right)^2\left(1-\frac{1}{k}\right)^{m-2} = \frac{1}{k}\,\frac{m(m-1)}{2}\left[\left(1-\frac{1}{k}\right)^k\right]^{(m-2)/k}.$$

Obviously, $E(|\hat{A}_2|) \leqslant m^2/(2k)$. From the equality $m = \lceil C\sqrt{k}\rceil$, we can see, that $E(|\hat{A}_2|) = \mathrm{O}(1)$. For $\hat{A}_r, r > 2$, analogously

$$E(|\hat{A}_r|) = k\binom{m}{r}\left(\frac{1}{k}\right)^r\left(1-\frac{1}{k}\right)^{m-r} < \frac{1}{r!}\,\frac{1}{k^{\frac{r}{2}-1}} + \mathrm{o}\left(\frac{1}{k^{\frac{r}{2}-1}}\right).$$

If we upper bound the last value by $1/k^{(r/2-1)}$ and calculate the sum of the geometrical progression, we obtain that $\sum_{r=2}^{k}(r-1)E(|\hat{A}_r|) = \mathrm{o}(1)$.

The statement of the lemma follows from three last equalities and the last inequality. The lemma is proved.

Proof of the lemma 2. Let us define $A_i^+ = \{a_{i_1},\ldots,a_{i_{k/2}}\}\cap\hat{A}_i$ and $A_i^- = \{a_{i_{k/2+1}},\ldots,a_{i_k}\}\cap\hat{A}_i, i = 0,1,\ldots$, and let the number of occurrences of letters from $\{a_{i_1},\ldots,a_{i_{k/2}}\}$ and $\{a_{i_{(k/2)+1}},\ldots,a_{i_k}\}$ in the training sample $x_1\ldots x_m$ be m_1 and m_2, correspondingly. Obviously,

$$E(m_1) = \frac{m}{2}(1+1/2), \quad E(m_2) = \frac{m}{2}(1-1/2). \tag{A.1}$$

From the definition (7) we can see that

$$E\left(\sum_{a\in A_1} p(a)\right) = E\left(\sum_{a\in A_1^+} p(a)\right) + E\left(\sum_{a\in A_1^-} p(a)\right)$$
$$= E(|A_1^+|)(3/(2k)) + E(|A_1^-|)(1/(2k)). \tag{A.2}$$

Let us estimate $E(|A_1^+|)$ and $E(|A_1^-|)$. For each m_1 we obtain the following:

$$E(|A_2^+|) = k \binom{m_1}{2} \left(\frac{3}{2k}\right)^2 \left(1 - \frac{3}{2k}\right)^{m_1-2} \leqslant \frac{9}{4k} \frac{m_1(m_1-1)}{2}.$$

If we take into account that $m_1 \leqslant m$ and $m = C\sqrt{k} + \mathrm{O}(1)$, we can see that

$$E_{\mathrm{H}_1}(|A_2^+|) < 9C^2/8 + \mathrm{o}(1). \tag{A.3}$$

Analogously, for each m_1

$$E(|A_r^+|) \leqslant k \binom{m}{r} \left(\frac{3}{2k}\right)^r \left(1 - \frac{3}{2k}\right)^{m_1-r} < \frac{3^r}{2^r r!} \frac{1}{k^{(r/2)-1}} + \mathrm{o}\left(\frac{1}{k^{(r/2)-1}}\right).$$

If we upper bound the last value by $3^r/(2^r k^{(r/2-1)})$ and calculate the sum of the geometrical progression, we obtain that $\sum_{r=2}^{k} E(|A_r^+|) = \mathrm{o}(1)$. Analogously, we can easily obtain that

$$E_{\mathrm{H}_1}(|A_2^-|) < C^2/8 + \mathrm{o}(1), \quad \sum_{r=2}^{k} E(|A_r^-|) = \mathrm{o}(1).$$

By definition, $m_1 = \sum_{r=1}^{k} r|A_r^+|, m_2 = \sum_{r=1}^{k} r|A_r^-|$ and we obtain from (A.3) and the last inequalities that

$$E(|A_1^+|) = E(m_1) + \mathrm{O}(1), E(|A_1^-|) = E(m_2) + \mathrm{O}(1).$$

From this equalities, (A.2) and (A.1) we can see that

$$E_{\mathrm{H}_1}(P\{a \in A_1\}) = \frac{1}{2}\frac{m}{k}(3/2)^2 + \frac{1}{2}\frac{m}{k}(1/2)^2 + \mathrm{O}\left(\frac{1}{k}\right).$$

If we take into account that $m = \lceil C\sqrt{k} \rceil$ we derive the first statement of the lemma from the last equality. The proof of other statements is completely analogous. The lemma is proved.

References

Cover, T.M., Thomas, J.A., 1991. Elements of Information Theory. Wiley, New York.

Devroye, L., Györfi, L., Lugosi, G., 1996. A Probabilistic Theory of Pattern Recognition. Springer, New York.

Kendall, M.G., Stuart, A., 1961. The Advanced Theory of Statistics, Vol. 2. Inference and Relationship. Charles Griffin, London.

Knuth, D.E., 1981. The Art of Computer Programming, Vol. 2. Addison-Wesley, Reading, MA.

L'Ecuyer, P., Simard, R., 1999. Beware of linear congruential generators with multipliers of the form $a = \pm 2^q \pm 2^r$. ACM Trans. Model. Comput. Simulation 25 (3), 367–374.

Lehmann, E.L., 1959. Testing Statistical Hypotheses. Wiley, New York.

Maurer, U., 1992. A universal statistical test for random bit generators. J. Cryptology 5 (2), 89–105.

Nechvatal, J., et al., 2000. Report on the Development of the Advanced Encryption Standard (AES). http://csrc.nist.gov/encryption/aes/round2/r2report.pdf.

Rukhin, A., et al., 2001. A statistical test suite for random and pseudorandom number generators for cryptographic applications. NIST Special Publication 800-22 (with revision dated May 15). http://csrc.nist.gov/rng/SP800-22b.pdf.

Schneier, B., 1996. Applied Cryptography. Wiley, New York.

Soto, J., Bassham, L., 2001. Randomness testing of the advanced encryption standard finalist candidates. Proceedings AES3, New York. http://csrc.nist.gov/encryption/aes/round2/conf3/papers/30-jsoto.pdf.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer, New York.

Wegenkittl, S., Matsumoto, M., 1999. Getting rid of correlations among pseudorandom numbers: discarding versus tempering. ACM Trans. Model. Comput. Simulation 9 (3), 282–294.