Adaptive chi-square test and its application to some cryptographic problems. *

Boris Ryabko

Abstract

We address the problem of testing the hypothesis H_0 that the letters from some alphabet $A = \{a_1, a_2, \ldots, a_k\}$, are distributed uniformly (i.e. $p(a_1) = p(a_2) = \ldots = p(a_k) = 1/k$) against the alternative hypothesis H_1 that the true distribution is not uniform, in case k is large. (It is typical for random number testing and some cryptographic problems where $k = 2^{10} \sim 2^{30}$ and more, see [2, 8, 6]). In such a case it is difficult to use the chi-square test because the sample size must be greater than k.

We suggest the *adaptive chi-square test* which can be successfully applied for testing some kinds of H_1 even in case when the sample size is much less than k. This statement is confirmed theoretically and experimentally. The theoretical proof is based on the consideration of one kind of the alternative hypothesis H_1 where the suggested test rejects the null hypothesis when the sample size is $O(\sqrt{k})$ (instead of $const \cdot k$ for the usual chi-square test). For experimental investigation of the suggested test we consider a problem of testing ciphered Russian texts. It turns out that the suggested test can distinguish the ciphered texts from random sequences basing on a sample which is much smaller than that required for the usual chi-square test.

Keywords. *hypothesis testing, chi-square test, adaptive testing, random number testing, sample size, block cipher testing .*

*The author was supported by the Russian Foundation of Basic Research under Grant No. 99-01-00586 and by INTAS under Grant no. 00-738.

1 Introduction

The chi-square test is one of the most popular hypothesis tests and it is widely applied to economics, biology, cryptography and many other fields. For example, one of cryptographic applications is the testing of random number generators and block ciphers' suitability as random number generators (see, for example, [3, 9, 6]). Moreover, it was recently shown that chi-square test of randomness can be used as a tool for breaking ciphers [2, 8].

In such cryptographic applications the number of categories (and, consequently, the number of degrees of freedom of χ^2 distribution) is very large ($2^{10} \sim 2^{30}$) and, thereby, the sample size should be also large. That is why, in such a case the implementation of the chi-square test requires a lot of time. Moreover, often it is difficult to obtain so large samples and the chi-square test cannot be employed.

We suggest a new method which is called 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.

Let us first explain the main idea of the new test. Let there be a hypothesis H_0 that the letters from some alphabet $A = \{a_1, a_2, \ldots, a_k\}, k > 2$, are distributed uniformly (i.e. $p(a_1) = p(a_2) = \ldots = p(a_k) = 1/k$) against the alternative hypothesis H_1 that the true distribution is not uniform, and 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 $A_1, A_2, \ldots, A_s, s \geq 2$, 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, $k = 2^{20}, s = 5$). Then, the set of subsets $\{A_1, A_2, \ldots, A_s\}$ is considered as a new alphabet and the new hypotheses $\hat{H}_0 : p(A_1) = |A_1|/k, p(A_2) = |A_2|/k, \ldots, p(A_s) = |A_s|/k$ and $\hat{H}_1 = \neg\hat{H}_0$ are tested basing 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, it means that 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 basing on the testing sample. It should be pointed out that a decrease in the number of categories from large k to small s (say, from 2^{20} to 5) can essentially increase the power of the test and, therefore, can essentially decrease the 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. In addition to a theoretical investigation of the suggested test we carried out some experiments. Namely, we tested ciphered texts 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 [7]. It turns out, that the suggested scheme can distinguish ciphered Russian texts from random ones, basing on samples which are essentially smaller than it is required for usual chi-square test.

The rest of the paper is organized as follows. The second section 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 the third section. The fourth section contains experimental results about recognition of ciphered texts in Russian and the 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:

$$H_0 : p(a_1) = p_1^0, p(a_2) = p_2^0, \ldots, p(a_k) = p_k^0; \quad H_1 = \neg H_0, \tag{1}$$

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 ν_i is the number of occurrences of $a_i \in A$ in the sample.(Often in statistics $a_1, \ldots, a_k$ are called categories.) The chi-square test is applied by calculating

$$x^2 = \sum_{i=1}^{k} \frac{(\nu_i - N\, p_i^0)^2}{N\, p_i^0}. \tag{2}$$

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 $p_i^1 = p(a_k)$ when H_1 is true, see ([1, 4]).

It is known [1] that

$$E_{H_0}(x^2) = k-1; \quad V_{H_0}(x^2) = 2(k-1) \tag{4}$$

$$E_{H_1}(x^2) = (k-1) + \lambda; \quad V_{H_1}(x^2) = 2(k-1) + 4\lambda, \tag{5}$$

where E_{H_i} and V_{H_i} are 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 $\alpha, \alpha \in (0,1)$, the hypothesis H_0 is accepted when x^2 from (2) is less than the $(1-\alpha)$ *-value* of the χ^2_{k-1} distribution. (Such a value is denoted as $\chi^2_{k-1;(1-\alpha)}$ and is defined as the solution of the equation $\int_{-\infty}^{\chi^2_{k-1;(1-\alpha)}} p_{\chi^2_{k-1}}(x)dx = (1-\alpha)$ where $p_{\chi^2_{k-1}}(x)$ is the density of χ^2_{k-1}.) If $x^2 \geq \chi^2_{k-1;(1-\alpha)}$ then H_0 is rejected.

It is important to note that such an approximation is valid when N is quite large. Thus, many authors recommend to take such the sample size N that $N\,p_i^0 > 5$ for all $i = 1, \dots, k$, see, for example, [1, 3, 6]. Obviously, it means that the following inequality should be valid

$$N > 5\,k. \tag{6}$$

It is shown in [1] that the calculation of the Type II error can be approximately carried out by

$$\int_{-\infty}^{\chi^2_{k-1;1-\alpha}} p_{\hat{\chi}^2_{k-1;\lambda}}(x)dx\,, \tag{7}$$

where $p_{\hat{\chi}^2_{k-1;\lambda}}$ is the $\hat{\chi}^2_{k-1;\lambda}$ density. We will use the following claim which summarizes mentioned above properties.

Claim 1. *The chi-square test has asymptotically the errors of the Type I α and the error of the Type II less than β, if and only if,*

$$\chi^2_{k-1;(1-\alpha)} \leq \hat{\chi}^2_{k-1;\lambda;\beta}, \tag{8}$$

where $\hat{\chi}^2_{k-1;\lambda;\beta}$ is β- value of $\hat{\chi}^2_{k-1;\lambda}$, $\chi^2_{k-1;(1-\alpha)}$ is the $(1-\alpha)$- value of χ^2_{k-1} and k is the number of categories.

The proof immediately follows from (7), if we take into account that, by definition,

$$\int_{-\infty}^{\hat{\chi}^2_{k-1;\lambda,\beta}} p_{\hat{\chi}^2_{k-1;\lambda}}(x)dx \ = \beta.$$

3 Adaptive chi- square test. Description and theoretical consideration.

First we intend to show that, in principle, the grouping of letters (or categories) can decrease the sample size of the chi-square test. For this purpose we will prove a claim about connection of the number of categories and the sample size, considering, for the sake of simplicity, only tests which have the Type II error less than $1/2$.

Claim 2. *Let chi-square test be being applied to the hypotheses H_0 and H_1 from (1).If the Type II error is less than $1/2$ and the sample size N and the number of categories k are large, then, asymptotically, the following inequality is true.*

$$N > U_{1-\alpha}\sqrt{2(k-1)}/\pi, \tag{9}$$

where π is defined in (3) and $U_{1-\alpha}$ is $(1-\alpha)$- value of the standard normal distribution. (For such a distribution the mean value is 0 and the variance is 1.)

The proof is given in Appendix.

We can see from this inequality, that it is possible to decrease the sample size N if we can decrease the number of categories k without essential decreasing of π. Moreover, this observation can be done basing on consideration of (4), (5). Indeed, from those equalities and (3) we can see that $E_{H_1}(x^2) - E_{H_0}(x^2) = N\pi\,(= \lambda)$ and does not depend on k, whereas both variances grow when k grows. So, if someone is able to decrease the number of categories k and not to decrease π, he increases the power of the test.

The following simple example shows that there exist such alternative hypotheses H_1 that it is possible to increase power of the test grouping categories. Let the hypotheses H_0 and H_1 from (1) are defined as follows:

$$p_1^0 = p_2^0 = \ldots = p_k^0 = 1/k;$$

$$p_1^1 = p_2^1 = \ldots = p_{k/2}^1 = (1/k) + \varepsilon; \quad p_{(k/2)+1}^1 = \ldots = p_k^1 = (1/k) - \varepsilon,$$

where $0 < \varepsilon < 1/k$. From the claim 2 and the definition of π (3) we immediately obtain that $\pi = k^2\varepsilon^2$ and

$$N \geq U_{1-\alpha} \cdot \sqrt{2(k-1)} k^{-2}\varepsilon^{-2}.$$

If we combine the letters $a_1, \ldots, a_{k/2}$ and $a_{k/2+1}, \ldots, a_k$ into two subsets A_1 and A_2, correspondingly, the value of π will be the same ($\pi = k^2\varepsilon^2$) but the number of categories will be only 2. Using the claim 1, it is easy to show that in this case the sample size N will be equal to $O(k^{-2}\varepsilon^{-2})$ that is asymptotically ($const\ \sqrt{k}$) times less than the sample size of the chi-square test applied to the initial alphabet. (By the way, it can be easily seen from (4, 5). When we group the categories, the difference of mean values is not changed, whereas the variances decrease into k times.)

More generally, it is easy to see that if we combine in one subset letters for which the fraction (p_i^1/p_i^0) are equal, the required sample size can be decreased. We do not consider this method in details because, unfortunately, 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 the testing.

The more formal description of the suggested adaptive chi-square test is the following. There are hypotheses H_0 and H_1 defined by (1) and a sample. It will be convenient to denote the sample as $x_1, x_2, \ldots, x_{m+n}$, where $m + n$ is the sample size. The sample is divided into two following parts $x_1, x_2, \ldots, x_m$ and $x_{m+1}, x_2, \ldots, x_{m+n}$ which are called as 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 in one subset letters for which the fractions $(\tilde{p}_i^1/p_i^0)$ are close. After that the new following hypotheses

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

are tested basing on the testing sample $x_{m+1}, x_2, \ldots, x_{m+n}$. We do not describe an exact rule of finding the parameters m, n and s as well as do not define exactly how to construct the subsets $\{A_1, A_2, \ldots, A_s\}$, but recommend to implement some experiments for finding the parameters which make the total sample size $(m+n)$ minimal (or, at least, acceptable). The point is that there are many cryptographic 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. Such a problem will be considered in the next paragraph whereas we proceed with theoretical investigation at the rest of this paragraph.

Let us consider an example when the adaptive chi-square test can be applied when the sample size equals $O(\sqrt{k})$ whereas a usual chi-square test can be used if the sample size is more than $c\,k$, see (6). Suppose, that the number of categories k is even and let

$$p_1^0 = p_2^0 = \ldots = p_k^0 = 1/k \tag{10}$$

$$\left.\begin{array}{c} p_1^1 = p_2^1 = \ldots = p_{k/2}^1 = \frac{1}{k}(1+\delta) \\ p_{(k/2)+1}^1 = \ldots = p_k^1 = \frac{1}{k}(1-\delta), \end{array}\right\} \tag{11}$$
$$\delta \in (0,1).$$

It turns out that the adaptive chi-square test can be successfully applied when the total sample size is $O(\sqrt{k})$.

Claim 3. *Let the adaptive chi-square test be applied for testing H_0 and H_1 from (10,11). Then, for each $\delta \in (0,1)$ and $\alpha \in (0,1)$ there exist such the training sample size m and the testing sample size n that*

i) $(m+n) = O(\sqrt{k})$ and

ii) the level of significance of the test is α and the Type II error is less than 1/2.

The proof. Let A_i be the set of letters from A which occurred i times in the training sample $x_1x_2\ldots x_m, i = 0, 1, \ldots$. The proof will be based on the two following lemmas.

Lemma 1. *If k goes to ∞, $m = c\sqrt{k}$ and either H_0 or H_1 is true, then*

$$\left.\begin{array}{c} E(|A_1|) = c\sqrt{k} + 0(1) \\ E(|A_2|) = 0(1) \\ E(\sum_{r=3}^{\infty}|A_r|) = \circ(1), \end{array}\right\} \tag{12}$$

where $E(\ )$ means the expectation and c is a positive constant.

Lemma 2. *If k goes to ∞, $m = c\sqrt{k}$ and H_1 is true, then*

$$E_{H_1}(P\{a \in A_1\}) = \frac{c}{\sqrt{k}}(1+\delta^2) + \circ\left(\frac{1}{\sqrt{k}}\right), \tag{13}$$

where c is a positive constant.

The proofs of the lemmas are given in Appendix. Let us proceed with the proof of Claim 3. Let the training sample size m and the tasting sample size n be defined by

$$m = c\,\lceil \sqrt{k}\,\rceil, \quad n = \lceil \sqrt{k}\,\rceil \tag{14}$$

where c is a positive constant which will be fixed later. As it follows from the lemma 1,with the probability 1 there exist only three nonempty subsets A_0, A_1, A_2 , when k goes to ∞. Therefor, from the definition (3) we obtain

$$\pi = \sum_{i=0}^{2} \frac{(P(A_i/H_0) - P(A_i/H_1)^2}{P(A_i/H_0)},$$

where $P(A_i/H_j) = \sum_{a \in A_i} p(a),\ j = 0,1$. Obviously,

$$\pi > \frac{(P(A_1/H_0) - P(A_1/H_1))^2}{P(A_1/H_0)}.$$

If we take into account that $P(A_1/H_0) = \frac{1}{k}|A_1|$ then, from the (12,13) and the last inequality we obtain that

$$\pi > \frac{\left[\left(\frac{c}{\sqrt{k}}(1+\delta^2) + \circ\left(\frac{1}{\sqrt{k}}\right) - \left(\frac{c}{\sqrt{k}} + \circ\left(\frac{1}{\sqrt{k}}\right)\right)\right]^2}{\frac{c}{\sqrt{k}} + \circ\left(\frac{1}{\sqrt{k}}\right)} = \left(\frac{c}{\sqrt{k}} + \circ\left(\frac{1}{\sqrt{k}}\right)\right)\left(\delta^4 + \circ\left(\frac{1}{\sqrt{k}}\right)\right).$$

So,

$$\pi > \frac{c}{\sqrt{k}}\ \delta^4 + \circ\left(\frac{1}{\sqrt{k}}\right),$$

when k goes to ∞. From this inequality, (3) and (14) we can see that asymptotically $\lambda = c\delta^4 + O(1)$ since $N = n = \sqrt{k}$. If we take into account that the number of the nonempty subsets (and, consequently, the number of categories) is 3 and a desirable value of the Type II error (β) should be not more than 1/2, we obtain from the last equality and Claim 1 that the following inequality should be valid

$$\chi^2_{2;(1-\alpha)} \le \hat{\chi}^2_{2;\,c\delta^4;\,1/2}. \tag{15}$$

When c grows, the parameter $\lambda = c\delta^4$ and the mean value of the noncentral chi-square distribution increase whereas $\chi^2_{2;(1-\alpha)}$ is not changed. Therefore, there exists such a constant $\tilde{c}$, that

$$\chi^2_{2;(1-\alpha)} = \hat{\chi}^2_{2;\,\tilde{c}\delta^4;\,1/2}$$

and (15) is true if $c > \tilde{c}$. On the other words, if the training sample size m and the testing sample size n are defined by (14) and $c > \tilde{c}$, the adaptive chi-square test has the level of significance α and its Type II error is less than 1/2. The claim is proved.

4 The experiments.

The block ciphers have been widely used in practise and have attracted attention of many researches. Thus, 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. Such a block cipher has to meet many requirements and, in particular, an output bit sequence of such a cipher should look like random even in case when an input sequence is not random (see [5, 9]). For example, if the input is a text in a human language (English, Russian, French, etc.), the ciphered text has to look like a random bit sequence (As much, as it is possible. Of course, such a ciphered sequence cannot be random simply because its limiting entropy is less than 1 per bit.) Apparently, the problem of constructing tests which can distinguish ciphered text and random sequences can be considered as a good example for estimation of a power of such tests. Moreover, this problem is of some interest for cryptology, see [7]. That is why the problem of distinguishing ciphered text and random bit sequences was chosen for experimental investigation of the adaptive chi-square test.

Let us describe the experiments more exactly. We considered the block ciphers Rijndael and RC6. The first of them has been proposed by NIST as Advanced Encryption Standard as well as the second was selected as a finalist for AES and widely used in practise, see [9, 5]. We applied adaptive chi-square test to ciphered Russian text using as source of Russian texts "Moshkov Library" (http://www.moshkov.ru/) which is one of the largest INTERNET libraries of Russian books. A lot of experiments with different values of the lengths of the training and the testing samples, the rule of grouping and values of other parameters were carried out. Below we describe one of them.

Russian texts were combined in large files and each such a file was ciphered by either Rijndael or RC6 with 128- bit key size. One randomly chosen key was used for the ciphering of one file. Then, the ciphered file was divided into 24- bit words and the obtained sequence was considered as a text under alphabet from 2^{24} letters. (By definition, the alphabet letters are all 24- bit words.) The adaptive chi-square test was applied to testing the H_0 hypothesis about randomness (i.e. $H_0 = \{p(a_1) = p(a_2) = \ldots = p(a_{2^{24}}) = 2^{-24}\}$, $H_1 = \neg H_0$). For the final experiments there were taken ten 4916- kilobyte files which did not used during the previous experiments. Each of them was divided into two equal parts. The first part was used as a training sample as well as the second as a testing sample. The training sample was used for estimating of the number of the letter occurrences and the alphabet was divided into 4 subsets $\{A_0, \ldots, A_3\}$ as follows. A_0 contained letters which were not met in the training sample, A_1 and A_2 contained letters which were met 1 and 2 times, correspondingly, as well as A_4 contained all other letters. Then, according to description of the adaptive chi -square test, the hypotheses

$$\hat{H}_0 = \{p(A_0) = |A_0|/2^{24}, \ldots, p(A_3) = |A_3|/2^{24}\}\,, \hat{H}_1 = \neg \hat{H}_0$$

were tasted basing on the testing sample. Each of the mentioned 10 files was ciphered by Rijndael with a randomly chosen key and tested by the described adaptive chi-square test. In the all cases the calculated x^2 was more than 99.9% - value of chi-square distribution with 3 degrees of freedom and, therefor, H_0 was rejected. The same calculation was repeated with the RC6 cipher and again in all ten cases H_0 was rejected. So, the test stably detects nonrandomness of the described ciphered files.

It is important to note that the usual chi-square test can be applied if the length of tested files is not less that $5\,2^{24}$ letters (6), where each letter is a 24-bit word. So, the file length should be around 251658 kilobytes in total, that exceeds the used file length more than 200 times.

5 Appendix.

Proof of the claim 2. The Type II error should be less than 1/2, therefor, from the claim 1 we obtain that

$$\chi^2_{k-1;(1-\alpha)} \leq \hat{\chi}^2_{k-1;\lambda;0.5},$$

where α is the Type I error. It is known (see [1, 4]) that chi-square distribution can be approximated by the normal distribution with parameters (4),(5), when k is large. If we take into account that the mean value of the normal distribution is equal to its median, we obtain from the last inequality and (4),(5) that

$$(k-1) + U_{1-\alpha} \cdot \sqrt{2(k-1)} \leq (k-1) + \lambda,$$

where $U_{1-\alpha}$ is the $(1-\alpha)$- value of the standard normal distribution. Therefore,

$$\lambda \geq U_{1-\alpha}\sqrt{2(k-1)}.$$

From this inequality and (3) we obtain (9). The claim is proved.

Proof of Lemma 1. Let us first consider the case when H_0 is true and, consequently, (10) is valid. Then,

$$E(|A_2|) = k \cdot \left(\begin{array}{c} m \\ 2 \end{array} \right) \left(\frac{1}{k}\right)^2 \left(1-\frac{1}{k}\right)^{m-2} = \frac{1}{k}\frac{m(m-1)}{2}\left[\left(1-\frac{1}{k}\right)^k\right]^{\frac{m-2}{k}}.$$

From the well known inequality $(1-\frac{1}{k})^k < e^{-k}$ and the definition of m in (14) we can see, that

$$E(|A_2|) \leq \frac{1}{2}e^{-\frac{c}{\sqrt{k}}} + 0\left(\frac{1}{\sqrt{k}}\right).$$

Hence,

$$E(|A_2|) = 0(1). \tag{16}$$

For A_r, $r > 2$, analogously

$$E(|A_r|) = k\left(\begin{array}{c} m \\ r \end{array} \right)\left(\frac{1}{k}\right)^r \left(1-\frac{1}{k}\right)^{m-r} < \frac{1}{r!}\ \frac{1}{k^{\frac{r}{2}-1}}e^{-\frac{c}{\sqrt{k}}} + \circ\left(\frac{1}{k^{\frac{r}{2}-1}}\right).$$

If we upper bound the last value by $e^{-\frac{c}{\sqrt{k}}}/k^{(r/2\,-1)}$ and calculate the sum of the geometrical progression, we obtain that

$$\sum_{r=2}^{\infty} E(|A_2|) = \circ(1). \tag{17}$$

If we recall that, by definition, A_i is the set of letters from A which occurred i times in the training sample $x_1x_2\ldots x_m, i = 0, 1, \ldots$ and take into account that $m = c\,\sqrt{k}$ (see (14)) , we obtain from (16) and (17) that $E(A_1) = c\sqrt{k} + 0(1)$. So, the lemma is proved in case H_0 is true.

If H_1 is true , the same scheme of the proof can be applied. Namely, let us define $A_i^+ = \{a_1, \ldots, a_{\frac{k}{2}}\} \bigcap A_i$ and $A_i^- = \{a_{\frac{k}{2}+1}, \ldots, a_k\} \bigcap A_i, i = 0, 1, \ldots.$ Repeating the previous proof for A_i^+ and A_i^- we obtain

$$\left.\begin{array}{c} E(|A_2^+|) \leq const, \\ E(|A_2^-|) \leq const, \\ \sum_{r=3}^{\infty} E(|A_r^+|) = \sum_{r=3}^{\infty} E(|A_r^-|) = \circ(1). \end{array}\right\} \tag{18}$$

Taking into account that $A_i = A_i^+ \bigcup A_i^-$, we obtain (16) and (17). The lemma is proved.

Proof of the lemma 2. Let us denote as m_1 and m_2 the number of occurrences of letters from $\{a_1, \ldots, a_{\frac{k}{2}}\}$ and $\{a_{\frac{k}{2}+1}, \ldots, a_k\}$, correspondingly, in the training sample $x_1 \ldots x_m$. Obviously,

$$E_{H_1}(m_1) = m \cdot \frac{1}{2}(1+\delta), \quad E_{H_1}(m_2) = m \cdot \frac{1}{2}(1-\delta). \tag{19}$$

By definition

$$E_{H_1}(P\{a \in A_1\}) = E_{H_1}\left(\sum_{a \in A_1} p(a)\right)$$

and, obviously,

$$E_{H_1}\left(\sum_{a \in A_1} p(a)\right) = E_{H_1}\left(\sum_{a \in A_1^+} p(a)\right) + E_{H_1}\left(\sum_{a \in A_1^-} p(a)\right) \tag{20}$$

From (18) we obtain that

$$E_{H_1}(A_1^+) = E_{H_1}(m_1) + 0(1), \quad E_{H_1}(A_1^-) = E_{H_1}(m_2) + 0(1).$$

¿From those equalities, (20) and (19) we can see that

$$E_{H_1}(P\{a \in A_1\} = \frac{1}{2}\ \frac{m}{k}(1+\delta)^2 + \frac{1}{2}\ \frac{m}{k}(1-\delta)^2 + 0\Big(\frac{m}{k}\Big).$$

If we take into account that $m = c\sqrt{k}$ (see (14)) we derive from the last equality (13). The lemma is proved.

References

[1] Kendall M.G., Stuart A. *The advanced theory of statistics; Vol.2: Inference and relationship* . London, 1961.

[2] Knudsen R.L., Meier W. *Correlation in RC6.* Private communication. (Available at http://www.ii.uib.no/ larsr/papers/rc6.ps).

[3] Knuth D.E. *The art of computer programming.* Vol.2. Addison Wesley, 1981.

[4] Lehmann E.L. *Testing Statistical Hypotheses.* Wiley, New York,1959.

[5] J.Nechvatal and others. *Report on the Development of the Advanced Encryption Standart (AES)* , 2000, in: http://csrc.nist.gov/encryption/aes/round2/r2report.pdf

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

[7] Schneier B. *Applied Cryptography.* Wiley, 1996.

[8] Shimoyama T., Takeuchi K., Hayakawa Ju. Correlation attack to the block cipher RC5 and the simplified variants of RC6. in: Proceedings AES3, 2001, New York, (AES3 Paper Submissions) http://csrc.nist.gov/encryption/aes/round2/conf3/aes3papers.html .

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