Appeared in *Journal of Cryptology*, vol. 5, no. 2, 1992, pp. 89-105.

A Universal Statistical Test for Random Bit Generators [1]

Ueli M. Maurer

Department of Computer Science
Princeton University
Princeton, NJ 08544, USA

Abstract. A new statistical test for random bit generators is presented which, in contrast to presently used statistical tests, is universal in the sense that it can detect any significant deviation of a device's output statistics from the statistics of a truly random bit source when the device can be modeled as an ergodic stationary source with finite memory but arbitrary (unknown) state transition probabilities. The test parameter is closely related to the device's per-bit entropy which is shown to be the correct quality measure for a secret-key source in a cryptographic application. The test hence measures the cryptographic badness of a device's possible defect. The test is easy to implement and very fast and thus well-suited for practical applications. A sample program listing is provided.

Keywords. Randomness, Random bit generator, Statistical test, Entropy, Ergodic stationary source, Exhaustive key search.

1. Introduction

A random bit generator is a device that is designed to output a sequence of statistically independent and symmetrically distributed binary random variables, i.e., that is designed to be the implementation of a so-called binary symmetric source (BSS). In contrast, a pseudo-random bit generator is designed to deterministically generate a binary sequence that only appears as if it were generated by a BSS.

[1]This work was supported by Omnisec AG, Switzerland. A preliminary version of this paper was presented at CRYPTO '90, Aug. 11-15, 1990, Santa Barbara, CA, and will appear in the proceedings.

Random bit generators have many applications in cryptography, VLSI testing, probabilistic algorithms and in other fields. Their major application in cryptography is as the secret-key source of a symmetric cipher system, but random bit generators are also required for generating public-key parameters (e.g., RSA-moduli) and for generating the keystream in the well-known one-time pad system (e.g., see [10]). In these applications, the security crucially depends on the randomness of the source. In particular, a symmetric (secret-key) cipher whose security rests on the fact that an exhaustive key search is infeasible may be completely insecure when not all keys are equiprobable. Similarly, the security of the RSA public-key cryptosystem may be strongly reduced when, because of a statistical defect in the random source used in the procedure generating the primes, the two primes are with high probability chosen from a small set of primes only.

This paper is concerned primarily with the application of random bit generators as the secret-key source of a symmetric cipher system. The paper is not concerned with pseudo-random bit generators, i.e., with the security evaluation of practical keystream generators for stream ciphers. However, it is certainly a necessary (but far from sufficient) condition for security that such a keystream generator pass the test presented here.

Randomness is a property of an abstract mathematical model that is characterized by probabilities. (In the context of random number generation the term "random" is also used as a synonym for independent and uniformly distributed, i.e., for the special model of a BSS, and we will make the same use of terminology.) Whether a probabilistic model can give an exact description of reality is a philosophical question related to the question of whether the universe is deterministic or not, and seems to be impossible to answer to everyone's satisfaction. On the other hand, there exist chaotic processes in nature, such as radioactive decay and the thermal noise in a transistor, that allow the construction of a random bit generator whose behavior is for all practical applications equivalent to that of a BSS. It is a non-trivial engineering task, however, to design an electronic circuit that exploits the randomness of a physical process in such a manner that dependencies between bits or a bias in the output are avoided. In a cryptographic application it is therefore essential that such a device be tested extensively for malfunction after production, and also periodically during operation.

The new proposed statistical test for random bit generators offers two major advantages over the presently used statistical tests (including the common frequency test, serial test, poker test, autocorrelation tests and run test which are described in [1] and [7]). First, unlike these tests, the new test is able to detect any one of a very general class of possible defects (deviations from the statistics of a BSS) a generator may have, including all the defects the above mentioned tests are designed to detect. This class of defects consists of those that can be modeled by an ergodic stationary source with limited memory, which can reasonably be argued to comprise the possible defects that could occur in a practical implementation of a random bit generator. Second, the new test measures the actual cryptographic significance of a defect. More precisely, the test parameter measures the per-bit entropy of a source, which is shown to be related to the running time of the enemy's optimal key-search strategy when he exploits knowledge of the secret-key source's statistical defect. In other words, the per-bit entropy of the secret-key source measures the effective key size of a cipher system under the (for this paper natural) assumption that there exists no essentially faster way than an exhaustive key-search for breaking the cipher.

The outline of the paper is as follows. The concept of a statistical test for randomness and the theoretical and practical limitations of statistical randomness testing are discussed in Section 2. In Section 3, the model of an ergodic stationary source is introduced. An analysis of the effective key size of a cipher system with a defective secret-key source is given in Section 4. Some theoretical considerations concerning the implementation of statistical tests are given in Section 5 and some previously proposed statistical tests are reviewed. The new universal statistical test is described in Section 6 and some conclusions are drawn in the final section. A reader who is interested only in the implementation of the test but not in the theoretical and philosophical background can skip Sections 2 to 5. Section 6 is almost self-contained and provides a sample program for implementing the test.

2. The Concept of a Statistical Test

In this section the problem of deciding whether a given device outputs statistically independent and symmetrically distributed binary digits is discussed from a theoretical viewpoint. When no theoretical proof based on the device's physical structure can be given (which seems to be impossible), such a decision must be based on an observed sample output sequence of a certain length N. Let B denote the set $\{0, 1\}$. A deterministic algorithm T taking as input such a sample sequence and producing as output a binary decision is usually called a *statistical test* and can be viewed as a function

$$T : B^N \rightarrow \{\text{accept},\ \text{reject}\}$$

that divides the set B^N of binary length N sequences $s^N = s_1, \ldots, s_N$ into a (usually small) set

$$S_T \ = \ \left\{s^N : T(s^N) = \text{reject}\right\} \ \subseteq \ B^N$$

of "bad" or "non-random" sequences and the remaining set of "good" or "random" sequences. The quotation marks refer to the fact that, as is explained below, no such attribute can be given to a particular sequence. Note that although the number and positions of output bits observed by a test algorithm may depend on the sequence itself, the length N of the sample sequence can nevertheless without loss of generality be considered to be a constant equal to the maximum possible length of an observed sequence.

A binary symmetric source emits every sequence of a given length N with the same probability 2^{-N} and therefore it seems to be impossible to argue that one particular sequence is "more random" than another sequence. However, an interesting approach to the problem of defining randomness for finite sequences has been taken by Kolmogorov [8] who defined the randomness of a sequence, informally, as the length of the shortest possible description of a generation rule for the sequence. A sequence can be considered "random" if one of the shortest descriptions is the sequence itself. More formally, the amount of randomness (or Kolmogorov-complexity) of a binary sequence is defined as the length of the shortest Turing-machine program for a fixed universal Turing machine that generates the sequence. Martin-Löf showed that, in an asymptotic sense, a sequence that is random according to this definition satisfies all computable statistical tests for randomness [9]. A minor problem with Kolmogorov's definition is that the length of the shortest program depends on the particular machine used. A much more severe and intrinsic problem, which is related the fact that the halting problem for Turing machines is undecidable [6], is that the Kolmogorov-complexity is not computable, even using infinite computing power. In other words, it is theoretically impossible, not only computationally infeasible, to check all possible generation rules for a given sequence and to choose the shortest one.

In view of the above it seems to be somewhat surprising that statistical randomness tests can be successfully used in practical applications, including cryptographic ones. The reason is that in many cases it may be reasonable to assume that, if a device is defective or badly designed, it behaves according to a certain probabilistic model with one or several unknown parameters, for instance a binary memoryless source or an ergodic stationary source (cf. Section 3). It is only under such an assumption, which is usually not stated explicitly, that statistical tests can be useful. As a consequence of such a restrictive assumption, however, a statistical test will not detect other types of non-randomness. For instance, the binary extension of π, the sequence 110010010000111111011010100..., can be generated deterministically and hence is not random and useless for cryptographic purposes, but it has nevertheless all commonly considered properties of a random sequence and will therefore pass every "reasonable" statistical test.

For every particular probabilistic model with specified parameters (e.g., a binary memoryless source emitting 1's with probability 0.4 and 0's with probability 0.6), the problem of deciding whether the tested device behaves according to this specified model or whether it is a BSS can be solved using the well-established framework of hypothesis testing (e.g., see [2]). For a parametrized model, however, statistical tests are generally not optimal in a hypothesis testing sense for two reasons. First, unless a probability distribution over the different models (or the parameters of a certain model) is fixed, a satisfactory overall optimality criterion cannot be defined. Second, as is often the case in hypothesis testing, the optimal strategy, even for a particular choice of parameters, may be infeasible to implement. Many statistical tests are therefore heuristic. Some tests (e.g., the frequency test and the serial test, cf. Section 5) can be interpreted as follows: the parameters of a certain statistical model are estimated from the sample sequence and a single test parameter is extracted from the differences of these estimated parameters to those of a BSS. Based on the probability distribution of the test parameter for a truly random sequence, the sample sequence is accepted or rejected. In terms of this interpretation, the advantages of the test presented in this paper can be described as follows. First, the test is based on the very general model of an ergodic stationary source (cf. Section 3) whose parameters are transition probabilities. Second, the test parameter has a cryptographic interpretation: it is very closely related to the per-bit entropy of the source, which measures the effective key size of a cipher system (cf. Section 4). Although the per-bit entropy is a function of the parameters of the model (the transition probabilities), our test does not estimate the parameters, but rather estimates the per-bit entropy directly.

3. Statistical Models for Bit Generators

The simplest probabilistic model of a bit generator is a binary memoryless source (BMS) which outputs statistically independent and identically-distributed binary random variables and is characterized by a single parameter, the probability p of emitting 1's. This model will be denoted by BMS_p. Note that a $\mathrm{BMS}_{1/2}$ is equivalent to a BSS. Another simple model, denoted by ST_p, emits 0's and 1's with equal probability, but its transition probabilities are biased: a binary digit is followed by its complement with probability p and by the same digit with probability $1-p$. This is an example of a binary stationary source with one bit of memory. In general, the probability distribution of the i-th bit of a generator's output may depend on the previous M output bits where M is the memory of the source. In many applications it is reasonable to assume that an even defective or badly designed random bit generator can well be modeled by such a source with relatively small memory.

Consider a source S that emits a sequence $U_1, U_2, U_3, \ldots$ of binary random variables. If there exists a positive integer M such that for all $n > M$, the conditional probability distribution of U_n, given $U_1, \ldots, U_{n-1}$, depends only on the most recent M output bits, i.e., such that

$$P_{U_n|U_{n-1}\ldots U_1}(u_n|u_{n-1}\ldots u_1) \;=\; P_{U_n|U_{n-1}\ldots U_{n-M}}(u_n|u_{n-1}\ldots u_{n-M}) \qquad (1)$$

for $n > M$ and for every binary sequence $(u_1, \ldots, u_n) \in B^n$, then the smallest such M is called the *memory* of the source S and $\Sigma_n = [U_{n-1}, \ldots, U_{n-M}]$ denotes its *state* at time n. Let $\Sigma_1 = [U_0, \ldots, U_{-M+1}]$ be the initial state where $U_{-M+1}, \ldots, U_0$ are dummy random variables. If in addition to (1) the source satisfies

$$P_{U_n|\Sigma_n}(u|\sigma) \;=\; P_{U_1|\Sigma_1}(u|\sigma)$$

for all $n > M$ and for all $u \in B$ and $\sigma \in B^M$, then it is called *stationary.* A stationary source with memory M is thus completely specified by the probability distribution of the initial state, P_{Σ_1}, and the state transition probability distribution $P_{\Sigma_2|\Sigma_1}$. The state sequence forms a Markov chain with the special property that each of the 2^M states has at most 2 successor states with non-zero probability. See [5], chapters XV and XVI, for a treatment of Markov chains. We will denote the 2^M possible states of the source (or the Markov chain) by the integers in the interval $[0, 2^M - 1]$. ($\Sigma_n = j$ means that $U_{n-1} \ldots U_{n-M}$ is the binary representation of j.) For the class of *ergodic* Markov chains (see [5] for a definition), which includes virtually all cases that are of practical interest, there exists an invariant state probability distribution $p_0, \ldots, p_{2^M-1}$ such that

$$\lim_{n\to\infty} P_{\Sigma_n}(j) \;=\; p_j$$

for $0 \le j \le 2^M - 1$. Moreover, the probabilities p_j's are the solution of the following system of 2^M linear equations:

$$\sum_{j=0}^{2^M-1} p_j = 1 \tag{2}$$

$$p_j \;=\; \sum_{k=0}^{2^M-1} P_{\Sigma_2|\Sigma_1}(j|k)\;p_k \qquad \text{for} \;\; 0 \le j \le 2^M - 2. \tag{3}$$

An example of an ergodic stationary source is given at the end of the next section.

4. The Effective Key Size of a Cipher with a Defective Key Source

A good practical cipher is designed such that no essentially faster attack is known than an exhaustive key search. The size of the key space is chosen large enough to ensure that to succeed in such an exhaustive search, even with only very small probability of success, requires an infeasible searching effort. If not all possible values of the secret key have equal *a priori* probability, then the enemy's optimal strategy in an exhaustive key search is to start with the most likely key and to continue testing keys in order of decreasing probabilities. Let Z denote the secret key, let n be its length in bits and let $z_1, z_2, \ldots, z_{2^n}$ be a list of the key values satisfying

$$P_Z(z_1) \;\ge\; P_Z(z_2) \;\ge\; \cdots \;\ge\; P_Z(z_{2^n}).$$

For a given source S and for δ satisfying $0 \le \delta \le 1$ let $\mu_S(n, \delta)$ denote the minimum number of key values an enemy must test (using the optimal key-searching strategy) in order to find the correct key with probability at least δ when S is used to generate the n-bit key Z, i.e.,

$$\mu_S(n, \delta) \;=\; \min \left\{ k : \sum_{i=1}^{k} P_Z(z_i) \ge \delta \right\}. \tag{4}$$

We define the *effective key size* of a cipher system with key source S to be $\log_2 \mu_S(n, \frac{1}{2})$, i.e., the logarithm of the minimum number of keys an enemy must try in order to find the correct key with probability at least 50%. The choice $\delta = 1/2$ in this definition is somewhat arbitrary, but in general, for large enough n, $\log_2 \mu_S(n,\delta)/n$ is almost independent of δ when δ is not extremely close to 0 or 1. Note that when the key is truly random, i.e., when S is a binary symmetric source, then $\log_2 \mu_S(n, \frac{1}{2}) = n - 1$.

We now determine the effective key size of a cipher system whose key source is BMS_p. Without loss of generality assume that $0 < p \leq 1/2$. Note that the source ST_p described in the previous section can be modeled by the source BMS_p with a summator at the output (integrating modulo 2 the output bits of the BMS_p). Therefore the set of probabilities of keys and hence also the effective key size is identical for both sources. The probability distribution of Z is given by

$$P_Z(z) \;=\; p^{w(z)}(1-p)^{n-w(z)},$$

where $w(z)$ denotes the Hamming weight of z. In order to succeed with probability approximately 1/2 the enemy must examine all keys z with Hamming weight $w(z) \leq pn$. The effective key size is thus well approximated by

$$\log_2 \mu_{\mathrm{BMS}_p}(n, \tfrac{1}{2}) \;\approx\; \log_2 \sum_{i=0}^{pn} \binom{n}{i}. \tag{5}$$

From eq. A.21 in [13] one can derive the inequalities

$$\frac{1}{\sqrt{8t(n-t)/n}}\, 2^{nH(t/n)} \;\leq\; \binom{n}{t} \;\leq\; \sum_{i=0}^{t} \binom{n}{i} \;\leq\; 2^{nH(t/n)} \tag{6}$$

for $t \leq n/2$, where $H(x)$ is the binary entropy function defined by

$$H(x) \;=\; -x \log_2 x - (1-x)\log_2(1-x) \tag{7}$$

for $0 < x < 1$ and by $H(0) = H(1) = 0$. Note that $H(x) = H(1-x)$ for $0 \leq x \leq 1$. Inequalities (6) suggest the following accurate approximation:

$$\log_2 \sum_{i=0}^{t} \binom{n}{i} \;\approx\; nH(t/n),$$

which together with (5) gives

$$\log_2 \mu_{\mathrm{BMS}_p}(n, \tfrac{1}{2}) \;\approx\; nH(p).$$

Using (6) one can prove that this approximation is asymptotically precise, i.e., that

$$\lim_{n\to\infty} \frac{\log_2 \mu_{\mathrm{BMS}_p}(n,\delta)}{n} \;=\; H(p)$$

for $0 < \delta < 1$.

Note that the entropy per output bit of the source BMS_p, $H(p)$, is hence equal to the factor by which the effective key size is reduced. Shannon proved (see [11], theorem 4) that for a general ergodic stationary source S,

$$\lim_{n\to\infty} \frac{\log_2 \mu_S(n,\delta)}{n} \;=\; H_S,$$

for $0 < \delta < 1$, where H_S is the per-bit entropy of S defined as

$$H_S = - \sum_{j=0}^{2^M-1} p_j \sum_{k=0}^{2^M-1} P_{\Sigma_2|\Sigma_1}(k|j) \log_2 P_{\Sigma_2|\Sigma_1}(k|j), \tag{8}$$

and where the stationary state probabilities p_j are for $0 \leq j \leq 2^M - 1$ defined by (3). In other words, for the general class of ergodic stationary sources, the per-bit entropy H_S is the correct measure of their cryptographic quality when they are used as the secret-key source of a cipher system. Conversely, the per-bit redundancy, $1 - H_S$, is the correct measure of the cryptographic badness of a key source. Because every state j can have at most two successor states with non-zero probability, namely $j^* = (2j) \bmod 2^M$ and $j^{**} = (2j+1) \bmod 2^M$, the expression (8) can be simplified:

$$H_S = \sum_{j=0}^{2^M-1} p_j \, H(P_{\Sigma_2|\Sigma_1}(j^*|j)). \tag{9}$$

Example: Consider a source that emits independent and symmetrically distributed bits except when two consecutive bits are identical, in which case the next bit is different with probability 0.8. For instance, when two 0's have occurred, the next bit is 1 with probability 0.8 and 0 with probability 0.2, but when the pair 01 occurred, the next bit is 0 or 1 both with probability 0.5. This source is an ergodic stationary source with memory $M = 2$, and it is easy to verify that the state transition probabilities are given by $P_{\Sigma_2|\Sigma_1}(0|0) = 0.2$, $P_{\Sigma_2|\Sigma_1}(1|0) = 0.8$, $P_{\Sigma_2|\Sigma_1}(2|1) = 0.5$, $P_{\Sigma_2|\Sigma_1}(3|1) = 0.5$, $P_{\Sigma_2|\Sigma_1}(1|2) = 0.5$, $P_{\Sigma_2|\Sigma_1}(3|2) = 0.5$, $P_{\Sigma_2|\Sigma_1}(1|3) = 0.8$ and $P_{\Sigma_2|\Sigma_1}(3|3) = 0.2$. The stationary state probabilities can be obtained as a solution of the system (2),(3): $p_0 = p_3 = 5/26$ and $p_1 = p_2 = 4/13$. The per-bit entropy is according to (9) equal to $2(5/26)H(0.2)+2(4/13)H(0.5) = (5/13)\cdot 0.7219+(8/13)\cdot 1 = 0.893$. The output of this source is thus 10.7% redundant.

5. Review of Some Previous Statistical Tests

As mentioned in Section 2, a statistical test T for sequences of length N is a function $T : B^N \to \{\text{accept, reject}\}$ which divides the set B^N of binary length N sequences $s^N = s_1, \ldots, s_N$ into a (small) set

$$S_T = \left\{ s^N : T(s^N) = \text{reject} \right\} \subseteq B^N$$

of "bad" sequences and the remaining set of "good" sequences. The probability that a sequence generated by a BSS is rejected is

$$\rho = \frac{|S_T|}{2^N}$$

and will be called the *rejection rate*. In a practical test, ρ should be small, for example $\rho \approx 0.001 \dots 0.01$.

A statistical test T for a reasonable sample length N cannot feasibly be implemented by checking a list of the set S_T. Instead, a statistical test T is typically implemented by specifying an efficiently computable test function f_T that maps the binary length N sequences to the real numbers $\mathcal{R}$:

$$f_T : B^N \to \mathcal{R} : \; s^N \mapsto f_T(s^N) \; .$$

The probability distribution of the real-valued random variable $f_T(R^N)$ is determined, where R^N denotes a sequence of N statistically independent and symmetrically distributed binary random variables, and a lower and an upper threshold t_1 and t_2, respectively, are specified such that

$$\Pr[f_T(R^N) \leq t_1] \; + \; \Pr[f_T(R^N) \geq t_2] \; = \; \rho \, .$$

Usually $\Pr[f_T(R^N) \leq t_1] \approx \Pr[f_T(R^N) \geq t_2] \approx \rho/2$. The set S_T of "bad" sequences with cardinality $|S_T| = \rho 2^N$ is defined by

$$S_T \; = \; \left\{ s^N \in B^N : \; f_T(s^N) \leq t_1 \;\; \text{or} \;\; f_T(s^N) \geq t_2 \right\} . \tag{10}$$

Usually, f_T is chosen such that $f_T(R^N)$ is distributed (approximately) according to a well-known probability distribution, most often the normal distribution or the χ^2 distribution with d degrees of freedom for some positive integer d. Since extensive numerical tables of these distributions are available, such a choice strongly simplifies the specification of t_1 and t_2 for given ρ and N. The normal distribution results when a large number of independent and identically distributed random variables are summed. The χ^2 distribution with d degrees of freedom results when the squares of d independent and normally distributed random variables with zero mean and variance 1 are summed.

In the sequel we briefly review the most popular statistical tests for random bit generators. The simplest test is the *frequency test* T_F which is used to determine whether a generator is biased and is based on the model BMS_p with one parameter. For a sample sequence $s^N = s_1, \dots, s_N$, the test parameter $f_{T_F}(s^N)$ is defined as

$$f_{T_F}(s^N) = \frac{2}{\sqrt{N}} \left(\sum_{i=1}^{N} s_i - N/2 \right) .$$

The number of 1's in a random sequence $R^N = R_1, \dots, R_N$ is distributed according to a binomial distribution which is very well approximated by the normal distribution with mean $N/2$ and variance $N/4$ since $E[R_i] = 1/2$ and $\text{Var}[R_i] = 1/4$ for $1 \leq i \leq N$. Thus the probability distribution of $f_{T_F}(R^N)$ is for large enough N well approximated by the normal distribution with zero mean and variance 1, and reasonable values for the rejection thresholds in (10) are $t_2 = -t_1 \approx 2.5 \dots 3$.

In the so-called *serial test* T_S with parameter L, the sample sequence s^N is cut into N/L consecutive blocks of length L (e.g., $L=8$), and the number $n_i(s^N)$ of occurrences of the binary representation of the integer i is determined for $0 \leq i \leq 2^L - 1$. f_{T_S} is defined as

$$f_{T_S}(s^N) \;=\; \frac{L2^L}{N} \sum_{i=0}^{2^L-1} \left(n_i(s^N) - \frac{N}{L2^L}\right)^2 .$$

A slightly simplified explanation of this formula is that the term $N/(L2^L)$ is the expected value of $n_i(s^N)$, and the purpose of the term $L2^L/N$ is to normalize the (unsquared) terms in the sum, which have zero mean, to have variance 1. The probability distribution of $f_{T_S}(R^N)$ is for large N very well approximated by the χ^2 distribution with $2^L - 1$ degrees of freedom. The serial test is based on the difficult to motivate statistical model of a source that emits statistically independent blocks of length L.

In the *run test* T_R with parameter L, the number $n_i^0(s^N)$ of 0-runs of length i and similarly the number $n_i^1(s^N)$ of 1-runs of length i in the sample sequence s^N are determined for $1 \leq i \leq L$ (e.g., $L = 15$). f_{T_R} is defined as

$$f_{T_R}(s^N) \;=\; \sum_{b\in\{0,1\}} \sum_{i=1}^{L} \frac{(n_i^b(s^N) - N/2^{i+2})^2}{N/2^{i+2}}$$

and the probability distribution of $f_{T_R}(R^N)$ is for large N very well approximated by the χ^2 distribution with $2L$ degrees of freedom because the terms in the sum are the squares of independent random variables that are virtually normally distributed with zero mean and variance 1.

An *autocorrelation test* with delay τ for the sequence $s^N = s_1, \ldots, s_N$ is a frequency test for the sequence $s_1 \oplus s_{1+\tau}, s_2 \oplus s_{2+\tau}, \ldots, s_{N-\tau} \oplus s_N$, where $\oplus$ denotes addition modulo 2. This test is used to detect a possible correlation between bits at distance τ and is for $\tau = 1$ based on the model ST_p (cf. Section 3).

In many practical applications a combination of several of these tests is used which corresponds to a single test T for which the set S_T is defined as the set of sequences that pass all these tests. Note that in general it is difficult to determine the rejection rate for such a combined test because the tests are not independent.

6. The New Universal Statistical Test T_U

The new statistical test T_U proposed in this section offers two main advantages over the statistical tests discussed in the previous section:

(1) Rather than being tailored to detecting a specific type of statistical defect, the new test is able to detect any one of the very general class of statistical defects that can be

modeled by an ergodic stationary source with finite memory, which includes all those detected by the tests discussed in the previous section and can reasonably be argued to comprise the possible defects that could realistically occur in a practical implementation of a random bit generator.

(2) The test measures the actual amount by which the security of a cipher system would be reduced if the tested generator G were used as the key source, i.e., it measures the effective key size $\mu_G(n, \frac{1}{2})$ of a cipher system with key source G (cf. Section 4). Therefore, statistical defects are weighted according to the potential damage they would cause in a cryptographic application.

These two advantages are due to the fact that for the general class of binary ergodic stationary sources with finite memory $M \leq L$ (cf. Section 3), where L is a parameter of the test, and for an arbitrary (unknown) choice of the conditional probabilities of the model, the resulting test parameter f_{T_U} is closely related to the per-bit entropy H_S of the source (cf. Section 4). This claim will be justified after the following description of the test. (In another context, a completely different use of entropy in a statistical test has previously been proposed in [3].)

The test T_U is specified by the three positive integer-valued parameters L, Q and K. To perform the test T_U, the output sequence of the generator is partitioned into adjacent non-overlapping blocks of length L. The total length of the sample sequence s^N is $N = (Q+K)L$, where K is the number of steps of the test and Q is the number of initialization steps. Let

$$b_n(s^N) \;=\; [s_{L(n-1)+1}, \ldots, s_{Ln}]$$

for $1 \leq n \leq Q+K$ denote the n-th block of length L of the sample sequence $s^N = s_1, \ldots, s_N$. For $n = Q+1, \ldots, Q+K$, the sequence is scanned for the most recent occurrence of the block $b_n(s^N)$, i.e., the least positive integer $i \leq n$ is determined such that $b_n(s^N) = b_{n-i}(s^N)$. Let the integer-valued quantity $A_n(s^N)$ be defined as taking on the value i if the block $b_n(s^N)$ has previously occurred and otherwise let $A_n(s^N) = n$. The test function $f_{T_U}(s^N)$ is defined as the average of the logarithm (to the base 2) of the K terms $A_{Q+1}(s^N), A_{Q+2}(s^N), \ldots, A_{Q+K}(s^N)$. More formally, the test function $f_{T_U} : B^N \to \mathcal{R} :\; s^N \mapsto f_{T_U}(s^N)$ is defined by

$$f_{T_U}(s^N) \;=\; \frac{1}{K} \sum_{n=Q+1}^{Q+K} \log_2 A_n(s^N) \tag{11}$$

where for $Q+1 \leq n \leq Q+K$, $A_n(s^N)$ is defined by

$$A_n(s^N) \;=\; \begin{cases} n & \text{if there exists no positive } i \leq n \text{ such that } b_n(s^N) = b_{n-i}(s^N), \\ \min\{i :\; i \geq 1,\; b_n(s^N) = b_{n-i}(s^N)\} & \text{otherwise.} \end{cases} \tag{12}$$

```
program UniversalTest(input,output);
const L=8; V=256; Q=2000; K=20000;
var i,n:  integer;  sum,fTU: real;
   tab:  array [0..V-1] of integer;
   block:  array [1..max] of integer;
begin
   for i:=0 to V-1 do tab[i]:=0;          (* initialization *)
   for n:=1 to Q do tab[block[n]]:=n;    (*        ''        *)
   sum:=0.0;
   for n:=Q+1 to Q+K do begin
      sum:=sum+ln(n-tab[block[n]]);
      tab[block[n]]:=n;
   end;
   fTU:=(sum/K)/ln(2.0); writeln(fTU);
end.
```

Figure 1: Listing of a PASCAL program for computing the test parameter $f_{T_U}(s^N)$ for a given sequence $s^N = s_1, \ldots, s_N$ that is assumed to be stored blockwise in the array `block` ($b_n(s^N) = [s_{L(n-1)+1}, \ldots, s_{Ln}]$ is stored in `block[n]`).

Rather than by scanning the previous blocks $b_{n-1}(s^N), b_{n-2}(s^N), \ldots$ for the most recent occurrence of the block $b_n(s^N)$, for every n, the test T_U can be implemented much more efficiently by using a table (denoted in Figure 1 as `tab`) of size $V = 2^L$ that stores for each L-bit block the time index of its most recent occurrence. For each block $b_n(s^N)$ the procedure consists of two simple steps: (1) $A_n(s^N)$ is easily computed as $n - \texttt{tab}(b_n(s^N))$ and the term $\log_2 A_n(s^N)$ is added to an accumulator, and (2) $\texttt{tab}(b_n(s^N))$ is updated to the new most recent time index n of the block $b_n(s^N)$. A sample PASCAL program for implementing the test is listed in Figure 1. The sequence s^N is for illustration purposes assumed to be stored blockwise in the array `block`, i.e., `block[n]` contains the integer whose binary representation is $b_n(s^N)$. Clearly, in a realistic implementation, the sequence s^N may be too long to be stored completely. In such a case there will for example be a function which, when called, increments the index n and returns the n-th block $b_n(s^N)$ of s^N. The function `ln` computes the natural logarithm. Note that $\log_2(x) = \ln(x)/\ln(2)$.

For performing a statistical randomness test one needs to know the distribution of the test parameter for a truly random sequence in order to specify the acceptance and rejection regions for the test parameter of a sample sequence. The mean and variance of a single term $\log_2 A_n(R^N)$ of the sum defining $f_{T_U}(R^N)$ can be computed for $Q \to \infty$ according to (16) and (17) below. Because the expected value of the average of several random variables is equal to the average of the expected values, the expected value $E[f_{T_U}(R^N)]$ of the test parameter f_{T_U} for a random sequence R^N is equal to $E[\log_2 A_n(R^N)]$. The variance of the sum of statistically independent random variables is equal to the sum of the variances. However, the quantities $A_n(R^N)$ are not completely independent, and as a consequence, the variance of $f_{T_U}(R^N)$ is somewhat smaller than expected. Let $c(L,K)$ denote the factor by which the standard deviation of $f_{T_U}(R^N)$ is reduced compared to what it would be if the terms $A_n(R^N)$ were independent, i.e., let

$$\mathrm{Var}[f_{T_U}(R^N)] \;=\; c(L,K)^2 \cdot \mathrm{Var}[\log_2 A_n(R^N)]/K.$$

For $L \geq 3$, $c(L, 2^L)$ is very close to 0.8, and for $K \gg 2^L$, $c(L,K)$ is close to 0.5, 0.6 and 0.65 for $L = 4$, $L = 8$ and $L = 12$, respectively. Extensive simulations have suggested that for $K \geq 2^L$,

$$c(L,K) \;\approx\; 0.7 - 0.8/L + (4 + 32/L)K^{-3/L}/15 \tag{13}$$

is a good approximation for the constant $c(L,K)$. In summary, the distribution of the test parameter $f_{T_U}(R^N)$ for a truly random sequence has a mean value of precisely $E[f_{T_U}(R^N)]$ and is very well approximated by the normal distribution with standard deviation

$$\sigma \;=\; c(L,K)\sqrt{\mathrm{Var}[\log_2 A_n(R^N)]/K}, \tag{14}$$

where $E[f_{T_U}(R^N)]$ and $\mathrm{Var}[\log_2 A_n(R^N)]$ are listed in Table I for $1 \leq L \leq 16$.

To implement the test T_U we recommend to choose the parameters L between 6 and 16, inclusive, $Q \geq 10 \cdot 2^L$ and K as large as possible (e.g., $K = 1000 \cdot 2^L$). This choice for Q guarantees that with high probability, every L-bit pattern occurs at least once in the first Q blocks of a random sequence. We also recommend to choose a rejection rate of $\rho \approx 0.001 \ldots 0.01$, depending on the application. A device should be rejected if and only if either $f_{T_U}(s^N) < t_1$ or $f_{T_U}(s^N) > t_2$, where the thresholds t_1 and t_2 are defined by

$$t_1 = E[f_{T_U}(R^N)] - y\sigma \quad \text{ and } \quad t_2 = E[f_{T_U}(R^N)] + y\sigma,$$

where the standard deviation σ is given by (14) and where y, the number of standard deviations that $f_{T_U}(s^N)$ is allowed to be away from the mean value, must be chosen such that $\mathcal{N}(-y) = \rho/2$. $\mathcal{N}(x)$ is the integral of the normal density function and is defined as

$$\mathcal{N}(x) \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\xi^2/2} d\xi.$$

A table of $\mathcal{N}(x)$ can be found in almost every book on statistics or probability theory (e.g., see [5], p. 176). For example, to obtain a rejection rate of $\rho = 0.01$ or $\rho = 0.001$ one must choose $y = 2.58$ or $y = 3.30$ respectively. Note that σ decreases like $1/\sqrt{K}$ when K increases. Like for any other statistical test, increasing the length of the sample sequence reduces the standard deviation and therefore allows to detect smaller deviations from the statistics of a BSS. Note that the fact that $c(L,K)$ is known only approximately can lead to a rejection rate

L	$E[f_{T_U}(R^N)]$	$\text{Var}[\log_2 A_n(R^N)]$	L	$E[f_{T_U}(R^N)]$	$\text{Var}[\log_2 A_n(R^N)]$
1	0.7326495	0.690	9	8.1764248	3.311
2	1.5374383	1.338	10	9.1723243	3.356
3	2.4016068	1.901	11	10.170032	3.384
4	3.3112247	2.358	12	11.168765	3.401
5	4.2534266	2.705	13	12.168070	3.410
6	5.2177052	2.954	14	13.167693	3.416
7	6.1962507	3.125	15	14.167488	3.419
8	7.1836656	3.238	16	15.167379	3.421

Table I. Expected value of $f_{T_U}(R^N)$ and variance of $\log_2 A_n(R^N)$ for the test T_U with parameters L, $Q \to \infty$ and K, where R^N is a truly random sequence. $\text{Var}[f_{T_U}(R^N)]$ is equal to $c(L,K)^2 \cdot \text{Var}[\log_2 A_n(R^N)]$ where $c(L,K)$ is well approximated by (13).

that is slightly different from ρ, but has no other effect on the test. The precise computation of the constants $c(L,K)$ would require a considerable if not prohibitive computing effort.

The definition of T_U is based on the idea, which was independently suggested by Ziv [14], that a universal statistical test can be based on a universal source coding algorithm. A generator should pass the test if and only if its output sequence cannot be compressed significantly. However, instead of actually compressing the sample sequence we only need to compute a quantity that is related to the length of the compressed sequence. The formulation of our test was motivated by considering the universal source coding algorithms of Elias [4] and of Willems [12], which partition the data sequence into adjacent non-overlapping blocks of length L. For $L \to \infty$, these algorithms can be shown to compress the output of every discrete stationary source to its entropy. The universal source coding algorithm due to Ziv and Lempel [15] seems to be less suited for application as a statistical test because it seems to be difficult to define a test function f_T such that the expected value of $f_T(R^N)$ can be computed. No indication of the suitability of the Ziv-Lempel algorithm for a practical implementation of a statistical test is given in [14].

In the sequel we derive expressions and numerical values for the quantities $E[f_{T_U}(R^N)]$ and $\text{Var}[\log_2 A_n(R^N)]$ under the admissible assumption that $Q \to \infty$. For a source emitting the sequence of binary random variables $U^N = U_1, U_2, \ldots, U_N$ we have

$$\Pr[A_n(U^N)=i] \;=\; \sum_{b \in B^N} \Pr\left[b_n(U^N) = b,\ b_{n-1}(U^N) \neq b, \ldots, b_{n-i+1}(U^N) \neq b, b_{n-i}(U^N) = b\right]$$

for $i \geq 1$. When the blocks $b_n(U^N)$ are statistically independent and identically distributed, then the above probability factors:

$$\Pr[A_n(U^N) = i] \;=\; \sum_{b \in B^N} (\Pr[b_n(U^N) = b])^2 \cdot (1 - \Pr[b_n(U^N) = b])^{i-1} \tag{15}$$

for $i \geq 1$. For a binary symmetric source we thus have

$$\Pr[A_n(R^N) = i] \;=\; 2^{-L}(1 - 2^{-L})^{i-1}$$

for $i \geq 1$. Hence

$$E[f_{T_U}(R^N)] \;=\; E[\log_2 A_n(R^N)] \;=\; 2^{-L} \sum_{i=1}^{\infty} (1 - 2^{-L})^{i-1} \log_2 i. \tag{16}$$

The variance of $\log_2 A_n(R^N)$ is

$$\begin{aligned} \mathrm{Var}[\log_2 A_n(R^N)] \;&=\; E[(\log_2 A_n(R^N))^2] - (E[\log_2 A_n(R^N)])^2 \\ &=\; 2^{-L} \sum_{i=1}^{\infty} (1 - 2^{-L})^{i-1} (\log_2 i)^2 - (E[f_{T_U}(R^N)])^2. \end{aligned} \tag{17}$$

Table I was compiled using (16) and (17) and summarizes $E[f_{T_U}(R^N)]$ and $\mathrm{Var}[\log_2 A_n(R^N)]$ for $1 \leq L \leq 16$. Note that $E[f_{T_U}(R^N)]$ is closely related to the entropy of a block, which is L bits. In fact, it is shown below that $E[f_{T_U}(R^N)] - L$ converges to the constant -0.8327 as $L \to \infty$.

In order to show that for $L \to \infty$, $E[f_{T_U}(R^N)] - L$ and $\mathrm{Var}[\log_2 A_n(R^N)]$ converge (exponentially fast) to constants, let $v(r)$ and $w(r)$ be defined as

$$v(r) \;\triangleq\; r \sum_{i=1}^{\infty} (1 - r)^{i-1} \log_2 i \tag{18}$$

and

$$w(r) \;\triangleq\; r \sum_{i=1}^{\infty} (1 - r)^{i-1} (\log_2 i)^2. \tag{19}$$

One can show that

$$\lim_{r \to 0} [v(r) + \log_2 r] \;=\; \lim_{r \to 0} \int_r^{\infty} e^{-\xi} \log_2 \xi \, d\xi \;=\; -0.832746 \;\triangleq\; C \tag{20}$$

and

$$\lim_{r \to 0} [w(r) - (\log_2 r)^2 + 2C \log_2 r] \;=\; \lim_{r \to 0} \int_r^{\infty} e^{-\xi} (\log_2 \xi)^2 d\xi \;=\; 4.117181 \;\triangleq\; D. \tag{21}$$

Note that $E[f_{T_U}(R^N)] = v(2^{-L})$ and hence it follows from (20) that

$$\lim_{L \to \infty} \Big(E[f_{T_U}(R^N)] - L \Big) \;=\; C.$$

L	p	$E[f_{T_U}(U^N_{\mathrm{BMS}_p})]$	$Lh(p)+C$	$\mathrm{Var}[\log_2 A_n(U^N_{\mathrm{BMS}_p})]$
8	0.50	7.18367	7.16725	3.239
8	0.45	7.12687	7.10945	3.393
8	0.40	6.95559	6.93486	3.844
8	0.35	6.66713	6.63980	4.561
8	0.30	6.25683	6.21758	5.482
16	0.50	15.16738	15.16725	3.421
16	0.45	15.05179	15.05165	3.753
16	0.40	14.70268	14.70246	4.733
16	0.35	14.11275	14.11234	6.319
16	0.30	13.26886	13.26791	8.425

Table II. Relation between the per-bit entropy of a biased binary memoryless source BMS_p and the expected value $E[f_{T_U}(U^N_{\mathrm{BMS}_p})]$ of the test parameter for the output of such a source.

From (17) it follows that $\mathrm{Var}[\log_2 A_n(R^N)] = w(2^{-L}) - v(2^{-L})^2$ which together with $\lim_{L\to\infty}[w(r) - v(r)^2] = \lim_{L\to\infty}[w(r) - (C - \log_2 r)^2]$ and (21) gives

$$\lim_{L\to\infty} \mathrm{Var}[\log_2 A_n(R^N)] \;=\; D - C^2 \;=\; 3.423715.$$

We now analyze the performance of the test for a biased binary memoryless source BMS_p with output sequence $U^N_{\mathrm{BMS}_p}$. The blocks $b_n(U^N_{\mathrm{BMS}_p})$ are statistically independent and thus using (15), (20) and the fact that for $L \to \infty$, $\Pr[b_n(U^N_{\mathrm{BMS}_p}) = b] \to 0$ for all $b \in B^L$ one can show that

$$\lim_{L\to\infty} \Big(E[f_{T_U}(U^N_{\mathrm{BMS}_p})] - Lh(p) \Big) \;=\; C$$

for $0 < p < 1$. This demonstrates that the test T_U measures the entropy of any binary memoryless source up to a constant. Table II summarizes $E[f_{T_U}(U^N_{\mathrm{BMS}_p})]$, $Lh(p) + C$ and $\mathrm{Var}[\log_2 A_n(U^N_{\mathrm{BMS}_p})]$ for $L = 8$ and $L = 16$ and for several values of p and demonstrates the close relationship between the expected value of the test parameter and the entropy of the source BMS_p. Some entries of Table II were computed by Maarten van der Ham on a CRAY Y/MP computer at CWI, Amsterdam.

By arguments similar to those used in [12] one can prove that for every binary ergodic stationary source S with output sequence U^N_S,

$$\lim_{L\to\infty} \frac{E[f_{T_U}(U^N_S)]}{L} \;=\; H_S.$$

We conjecture that this asymptotic relation between $E[f_{T_U}(U^N_S)]$ and H_S can be made even more precise, namely that

$$\lim_{L\to\infty}\Big(E[f_{T_U}(U_S^N)] - Lh(p)\Big) \;=\; C.$$

7. Conclusions

The new statistical test described in this paper is based on a more general statistical model than those previously considered in the context of statistical tests, namely an ergodic stationary source with memory $M \leq L$, where L is a parameter of the test. This model can reasonably be argued to comprise most defects that can realistically be expected in a practical implementation of a random bit generator based on a chaotic physical process such as the thermal noise in a transistor. Another novel feature of the test is that it measures the actual cryptographic significance of a possible defect, namely the per-bit redundancy.

The performance of a statistical test depends in a crucial manner on the statistical model on which the test is based. The more general the model, the wider is the class of possible defects that can be detected. On the other hand, the more restricted the model, the better a test based on this model is generally suited for detecting a defect that can be described by the model, i.e., a shorter sample sequence is needed to detect a defect. When designing a statistical test for testing the randomness of a device's output sequence it is therefore very important that an appropriate model is used. To illustrate this, consider the performances of the frequency test and of our new test on a device that can be modeled as a binary memoryless source emitting 1's with probability 0.45 and 0's with probability 0.55. Because the per-bit entropy $H(0.45) = 0.9928$ of this source is very close to 1, the universal test will need a much longer sample sequence to detect the non-randomness of this source with the same detection probability as a frequency test. For this example and for $L = 8$ one can show that the sequence must be 29 times longer for the universal test. On the other hand, the frequency test is unable to detect any dependencies between consecutive bits. Therefore, if for a certain application a bias in the distribution of 0's and 1's is the only defect that can reasonably be expected, a frequency test is optimal. Note also that because the per-bit entropy measures the effective key size, using the above biased source would only slightly reduce the security of a cipher system. Of course, we do not suggest that a source with such a bias be used in practice because any deviation from the statistics of a BSS may indicate that there exists a possibly much stronger hidden defect.

Acknowledgements

Major parts of this research were performed while the author was with the Institute for Signal and Information Processing, Swiss Federal Institute of Technology, Zürich, Switzerland. The problem of designing efficient statistical tests for random bit generators was suggested to the author by Omnisec AG, Switzerland. In particular, it is a pleasure to thank Dr. P. Schmid and Martin Benninger for stimulating discussions and for their generous support. I am also grateful to Andi Loeliger and Jim Massey for helpful discussions and to Maarten van der Ham for correcting some previously inaccurate entries in Table II.

References

[1] H. Beker and F. Piper, *Cipher Systems*, London: Northwood Books, 1982.

[2] R.E. Blahut, *Principles and Practice of Information Theory*, Reading, MA: Addison-Wesley, 1987.

[3] E.J. Dudewicz and E.C. van der Meulen, Entropy-based tests of uniformity, *J. American Statistical Association*, vol. 76, no. 376, Dec. 1981, pp. 967-974.

[4] P. Elias, Interval and recency rank source coding: Two on-line adaptive variable-length schemes, *IEEE Transactions on Information Theory*, vol. 33, Jan. 1987, pp. 3-10.

[5] W. Feller, *An Introduction to Probability Theory and its Applications*, third ed., vol. 1, New York, NY: Wiley, 1968.

[6] J.E. Hopcroft and J.D. Ullman, *Introduction to Automata Theory, Languages, and Computation*, Reading, MA: Addison-Wesley, 1979.

[7] D.E. Knuth, *The Art of Computer Programming*, vol. 2, 2nd edition, Reading, MA: Addison-Wesley, 1981.

[8] A.N. Kolmogorov, Three approaches to the quantitative definition of information, *Problemy Peredachi Informatsii*, vol. 1, no. 1, pp. 3-11, 1965.

[9] P. Martin-Löf, The definition of random sequences, *Information and Control*, vol. 9, 1966, pp. 602-619.

[10] J.L. Massey, An introduction to contemporary cryptology, *Proceedings of the IEEE*, vol. 76, no. 5, 1988, pp. 533-549.

[11] C.E. Shannon, A mathematical theory of communication, *Bell System Technical Journal*, vol. 27, Oct. 1948, pp. 379-423 and 623-656.

[12] F.M.J. Willems, Universal data compression and repetition times, *IEEE Transactions on Information Theory*, vol. 35, Jan. 1989, May 1977, pp. 54-58.

[13] J.M. Wozencraft and B. Reiffen, *Sequential Decoding*, Cambridge, MA: Techn. Press of the M.I.T., 1960.

[14] J. Ziv, Compression, tests for randomness and estimating the statistical model of an individual sequence, in: *Sequences* (Ed. R.M. Capocelli), Berlin: Springer Verlag, 1990.

[15] J. Ziv and A. Lempel, A universal algorithm for sequential data compression, *IEEE Transactions on Information Theory*, vol. 23, pp. 337-343.