Advances in Mathematics of Communications
Volume 00, No. 0, 2007, 1–16

Web site: http://www.aimSciences.org

ON A CLASSIFICATION OF FINITE STATISTICAL TESTS

David Lubicz

CELAR
BP 57419 35174 Bruz Cedex
France

Abstract. Statistical tests of random sequences are often used in cryptography in order to perform some routine checks for random and pseudo-random number generators. Most of the test suites available are based on the theory of hypothesis testing which allows one to decide whether a sample has been drawn following a certain distribution. In this article, we develop a theoretical foundation of statistical tests of random sequences and hypothesis testing with a focus on cryptographic applications and we draw some interesting practical consequences.

1. Introduction

Random sequences are often used in cryptography in order to diversify the behavior of algorithms, to generate secret keys, or to provide random seeds for instance. As a consequence, random number generators constitute a critical part of many cryptographic protocols and consequently of any hardware or software implementation of these protocols. It is then important to be able to assess the randomness of the generators in order to avoid any failure which would immediately impact the security. In order to do so, it is quite usual to perform statistical tests on a cryptographic device in order to check if the output behaves in a certain sense like a random sequence. This can be done in different circumstances : to validate a device during its design, to qualify a device during the production phase or to detect a failure of a device in use.

This subject is related to an important literature dealing with the definition of general test suites [7], [11], with the theoretical definition of random sequences [6], [8], [5] and more recently with the definition of a test associated to some general statistical models [9], [2].

Broadly speaking, in order to define a statistical test one must clarify several things :

- what property of randomness is involved by the test. This means giving a proper definition of a random sequence and defining what superset of the set of random sequences will pass the test.
- the link between the property of randomness and the device that we want to test. In order to do so, one must define a mathematical model for the device and discuss its validity;
- the defining of a test function which outputs `yes` or `no` given a finite output of the device.

2000 *Mathematics Subject Classification:* Primary: 94A60; Secondary: 94A15.
Key words and phrases: Random number generator, Random sequence, Statistical test, Hypothesis testing, Information theory.

In order to define a test function, it is usual to use the general framework of hypothesis testing (see [11] or [7]). Basically, hypothesis testing [10] aims at deciding whether a sample behaves like a certain distribution contained in a statistical model i.e. in a family of distributions depending of some parameters. In order to use this theory, it is necessary to give a precise definition of what a statistical model attached to a test is. Unfortunately, in the literature dealing with statistical tests and hypothesis testing, this important matter is often overlooked. For instance, in [11], the statistical models associated to the presented statistical tests are not always clearly defined. Some other authors describe some general statistical models associated to a test [9] or some general test suite justified by some empirical definition of a random sequence [5]. The aim of this paper is to give a precise and general definition of a statistical test which includes all usual tests and to explain how to attach a specific statistical model to each test. Then we give a classification of all the statistical models attached to statistical tests with respect to their capacity to distinguish certain types of deviation from a true random behavior. This classification appears to be new. One of the benefits of this general approach based on information theory is that it provides a formalism to compare the different statistical tests. For instance one would like to know if two tests are independent, i.e. that they provide different statistical information or on the contrary, if a test is more general than an other. The results of section 3.3 will be very helpful to answer that kind of question. On a theoretical side, we deduce from our classification some analogs of well-known results about equidistributed sequences. We derive also some interesting practical consequences.

Organisation of the paper. In section 2.2, we give a definition a random sequence. We use it in section 2.3 to link the classical cryptographic definition based on unpredictability [3] with the definition given in [5] used in the random sequences literature. In section 3.1, we give a definition of a finite statistical test and explain what the statistical model attached to a finite statistical test is. Next, in section 3.2, we classify the finite statistical tests with respect to their strength : we say that a finite statistical test A is *stronger* than another one B if all the sequences that pass test A pass test B. In other words, A is stronger than B if A discriminates more sequences than B does, i.e. A is more selective. In section 3.3, we state a sufficient condition for a family of finite statistical tests to ensure that a sequence that passes all the tests of the family also pass all the other finite tests. We draw some interesting consequences of this result: for instance we obtain immediately discrete analogs of well-know results. In section 4, we explain the link of the previous theory with the practical matter of testing the randomness of a device.

2. Source of randomness and unpredictability

We fix some notations to be used in the rest of the section. Let Σ be an alphabet. We denote by Σ^* the set of all sequences with terms in Σ. An element u of Σ^* is given by a sequence (u_n) of elements of Σ indexed by $n \in \mathbb{N}$. In the same way, for $k \in \mathbb{N}$, we denote by Σ^k the set of finite sequences of length k with terms in Σ. An element u of Σ^k will be written as:

$$u = u_0 \dots u_{k-1}$$

with $u_i \in \Sigma$ for $i = 0 \dots k-1$. Note that Σ^n is a finite set and then can be considered itself as an alphabet.

2.1. Statistical model of a device. It is quite usual in cryptography to think of a random number generator as a way to produce a distribution which may be described for instance by a sequence of binary random variables. In this paper, in order to clarify the link with some usual definitions of random sequences often referred to in the literature [5] about tests of randomness, we adopt a slightly different approach by saying that a random number generator outputs an infinite binary sequence. This situation corresponds to an idealisation of the reality where one can operate the device during a infinite period of time. From an infinite random sequence it is easy to recover a distribution by the computation an empirical probability as explained in section 2.2. It is also easy to produce a random sequence which behaves accordingly to a certain distribution, so that there is an easy way back and forth between the two point of views.

We suppose that a source of randomness S_T is a mapping from a parameter space T in the set Σ^*. The parameter space T may be either discrete or continuous . In the case of a physical generator T, there can be a set of continuous variables that describes the state of the RNG (temperature of the circuit, position of each of the bits). For a LFSR, T is the discrete space describing the initialization vector, the feedback polynomial and the filtering function.

In practice, the set of parameters can take into account the normal operation of the source as well as flaws. It is possible that the source produces sequences with good statistical properties for some values of the parameter in T and poor statistical properties for the other values of T. For instance, a physical random generator may be built such that it outputs bits with a bias p independent of the preceding draws. It outputs "1" with a probability of p and a "0" with a probability of $q = 1 - p$. A hard to control production process may influence the parameter p. Therefore, a method is needed to assess the generator and reject any source that has a parameter p too far from $\frac{1}{2}$. A solution to this problem is provided by the theory of hypothesis testing.

Next, we would like to have a definition of the sequences produced by the normal operation of the source. There are various definitions of a random sequence more or less adapted to cryptographic applications. The most general definition is due to Kolmogorov [6]. It is based on the length of the smallest program which outputs a certain finite binary sequence. In [8], Martin-Löf shows that a random sequence in Kolmogorov's sense passes all possible statistical tests. He also extends Kolmogorov's definition to the case of infinite sequences. Here, we use a more practical definition following Knuth's survey [5] and show how to adapt this definition in order to grasp the usual cryptographic unpredictability property.

Let W^k be the map from Σ^* to the set of sequences with terms in Σ^k, which maps $u \in \Sigma^*$ to the unique sequence $(w_n) = (W^k(u)_n) \in (\Sigma^k)^*$, such that:

$$u = w_0|w_1|\dots|w_q|\dots$$

with $|$ the concatenation.

In the following, a sequence of events is defined as a sequence $(u_n)_{n\in\mathbb{N}}$ with terms in a finite set Ω. For instance, Ω can be the alphabet Σ. For $x \in \Omega$, the *empirical probability* of the event x, denoted by

$$P_e[(u_n) = x],$$

is defined by the following limit, if it exists:

$$\lim_{k\to\infty} \frac{S_k(x)}{k}, \tag{1}$$

with $S_k = \sharp\{n < k | u_n = x\}$. It should be understood that the notation $P_e[(u_n) = x]$ assumes that this last limit is well-defined. When this limit is not well-defined, it is possible to define $P_e^{\sup}$ and $P_e^{\inf}$ as respectively $\limsup_{k\to\infty} \frac{S_k(x)}{k}$ or $\liminf_{k\to\infty} \frac{S_k(x)}{k}$. It is also possible to define conditional empirical probabilities by considering subsequences of (u_n) having a certain property. We leave the details to the reader.

To an infinite binary sequence one can associate for all $n \in \mathbb{N}^*$ a probability distribution on Σ^k given by the empirical probability of $W^k(u)_n$. In particular, a source defines a map from the set of parameters T to the set of probability distributions on Σ^k for all k that is a *statistical model* on Σ^k.

2.2. DEFINITION OF A RANDOM SEQUENCE. The definition due to Borel following [5] can now be stated.

Definition 1. Let $l \in \mathbb{N}^*$, a sequence $u \in \Sigma^*$ is l-distributed, if for all $x \in \Sigma^l$, $P_e[(W^l(u)_n) = x] = (\frac{1}{\sharp\Sigma})^l$. A sequence $u \in \Sigma^*$ is ∞-distributed if it is l-distributed for all $l \in \mathbb{N}^*$.

Following [5], it can be stated as a first approximation that a sequence is random if it is ∞-distributed. In particular, if u is a random sequence then for all $k \in \mathbb{N}^*$, $(W^k(u)_n)$ is an equidistributed sequence of words of Σ^k. If a finite subsequence of length k is picked up from a random sequence then the probability of selecting a given subsequence is the same for all words in Σ^k. This illustrates well the intuitive idea of a random phenomenon. A consequence is that it is not possible in principle to precisely decide whether a finite sequence has been generated by a truly random process.

We will see in the course of this paper that this definition due to Borel grasps a lot of important properties of randomness. However, as explained in [5], it is not tight enough to take into account all the properties that one expects from a sequence which would be truly drawn by the way of a random process.

Example 2. Let $u \in \Sigma^*$ be an $\infty-$distributed sequence and let v the sequence deduced from u by forcing to zero all the bits of rank n^2, $n \in \mathbb{N}$. One easily sees that the sequence v is also $\infty-$distributed and although not a random sequence because it is possible to predict easily the value of certain of its bits.

In order to take into account the unpredictability property, following [6], we introduce the following definition:

Definition 3. Let Σ be an alphabet. A "subsequence rule" $\mathcal{R}_{(f)}$ or simply $\mathcal{R}$ if no confusion is possible, is a computable function f from $\cup_{k=1}^{\infty}\Sigma^k$ into $\{0,1\}$.

Such a subsequence rule defines a subsequence $(u_n)\mathcal{R}$ of an infinite sequence (u_n) in the following manner : the nth term of (u_n) is in the sequence $(u_n)\mathcal{R}$ if and only if $f(u_0,\ldots,u_{n-1}) = 1$.

By computable function, we mean that there exists an algorithm f which inputs any finite number of elements of Σ with an extra termination symbol and determines the value of $f(x_0,\ldots,x_{n-1})$. For instance, we can choose a certain computational model $\mathcal{M}$ which can be that of Turing machines and say that an algorithm is an element of this computational model. It is moreover possible to suppose that this f is taken from a certain family $\mathcal{F}$ of elements of $\mathcal{M}$. This family may be obtained by adding some restriction to the time or memory consumption of the algorithm in our given computational model. In particular, we can force f to be an algorithm in polynomial time with respect to the length of the input.

Using this notion of subsequence rule, we can then assess the following definition:

Definition 4. Let $\mathcal{F}$ be a family of elements of a computational model $\mathcal{M}$. A sequence $(u_n) \in \Sigma^*$ is $R4(\mathcal{F})$ if for all subsequence rules $\mathcal{R}$ defined by a certain element f of $\mathcal{F}$, $(u_n)\mathcal{R}$ is $1-$distributed. We simply say $R4$ for $R4(\mathcal{M})$.

It is easy to see that a $R4$ sequence is $\infty-$distributed. Moreover, one can see at once that this definition excludes the pathological preceding example.

By a result of Wald [5] p. 164, it is known that there exists $R4$-sequences. In fact there are actually uncountably many of them.

In order to show the connection with the usual cryptographic notion of unpredictability we introduce a new definition. Let (s_n) be a sequence with terms in $\{0,1\}$. The density $d_{(s_n)}$ of (s_n) is the limit if it exists:

$$d_{(s_n)} = \lim_{n\to\infty} \frac{\sum_{i=0}^{n-1} s_i}{n}. \tag{2}$$

If $d_{(s_n)} = 0$ (resp. $d_{(s_n)} > 0$) we say that (s_n) is non dense (resp. dense). Again if the limit in (2) does not exist it is possible to use a lim sup or lim inf but in the course of this paper we only consider sequences with a well-defined density. Now let $(u_n) \in \Sigma^*$ and $\mathcal{R}_{(f)}$ a subsequence rule for (u_n), we say that $\mathcal{R}$ is non dense (resp. dense) with respect to (u_n) if the sequence $s_n = f(u_0, \ldots, u_{n-1})$ is non dense (resp. dense).

We can now formulate our definition of a random sequence.

Definition 5. Let $\mathcal{F}$ be a family of elements of $\mathcal{M}$ a computational model. A sequence $(u_n) \in \Sigma^*$ is $R7(\mathcal{F})$ if for *all dense subsequence rule* $\mathcal{R}_{(f)}$ defined by a certain element f of $\mathcal{F}$, $(u_n)\mathcal{R}$ is $1-$distributed. The notation $R7$ will stand for $R7(\mathcal{M})$.

2.3. Randomness and unpredictability. In order to justify this definition, we show that it grasps the unpredictability notion usual in cryptography (see [3]). To see that, let $\mathcal{M}$ be the computational model of Turing machines.

Definition 6. Let $\mathcal{F}$ be a family of elements in $\mathcal{M}$, a $\mathcal{F}$-predictor is an algorithm in $\mathcal{F}$ which inputs a finite number $(u_0, \ldots, u_n)$ of elements of Σ and outputs an element of Σ. In order to take into account the fact that some elements of $\mathcal{M}$ may have a limited amount of memory, we moreover impose the following restriction on a $\mathcal{F}$-predictor represented by a Turing machine: the head on the tape containing the input elements $(u_0, \ldots, u_n)$ can only go in one direction.

With this last assumption we do not loose generality because an algorithm with enough memory can copy the input tape on another regular memory tape. On the other side we can allow an algorithm with finite memory to have access to a tape with a limitless number of elements of Σ stored on it.

One can define the advantage of a predictor P with respect to a sequence $(u_n) \in \Sigma^*$ as : $\mathrm{Adv}^P((u_n)) = |P_e[P(u_0, \ldots, u_{k-1}) = u_k] - 1/\sharp\Sigma|$.

Definition 7. Let $(u_n) \in \Sigma^*$, we say that (u_n) is $\mathcal{F}$-predictable (resp. $\mathcal{F}$-unpredictable) if there exists (resp. does not exists) P a $\mathcal{F}$-predictor such that $\mathrm{Adv}^P((u_n)) \neq 0$.

We have the following proposition, which should be compared to Theorem 3.3.7 of [3]

Proposition 8. Let $\mathcal{F}$ be either the family of polynomial time Turing machines or the family of finite memory Turing machines. A sequence $(u_n) \in \Sigma^*$ is in $R7(\mathcal{F})$ if and only if it is $\mathcal{F}$-unpredictable.

Proof. If P is a predictor and $s \in \Sigma$ is a symbol, we denote by P_s the subsequence rule defined by $P_s(u_0, \ldots, u_{n-1}) = 1$ if and only if $P(u_0, \ldots, u_{n-1}) = s$.

First assume that (u_n) is $\mathcal{F}-$predictable. Suppose that for all $\mathcal{F}$-predictor P and all symbol $s \in \Sigma$ such that P_s is dense and we have $P_e[P(u_0, \ldots, u_{n-1}) = u_n | P(u_0, \ldots, u_{n-1}) = s] = 1/\sharp\Sigma$. Then we have $P_e[P(u_0, \ldots, u_{n-1}) = u_n] = 1/\sharp\Sigma$ which gives a contradiction with the hypothesis. So there exists $s \in \Sigma$ and a $\mathcal{F}$-predictor P such that P_s is dense and

$$|P_e[P((u_0, \ldots, u_{n-1})) = u_n | P((u_0, \ldots, u_{n-1})) = s] - 1/\sharp\Sigma| \neq 0.$$

This means that $(u_n)\mathcal{R}_{P_s}$ is not $1-$distributed.

Now suppose that there exists a subsequence rule $\mathcal{R}_{(P)}$ with density $d > 0$ given by an algorithm P in $\mathcal{F}$. It means that $(u_n)\mathcal{R}_{(P)}$ is not 1-distributed. Then there exists $\epsilon > 0$ and $s \in \Sigma$ such that $|P_e[(u_n)\mathcal{R}_{(P)} = s] - 1/\sharp\Sigma| > \epsilon$ and we can suppose that ϵ and s are chosen such that

$$P_e[(u_n)\mathcal{R}_{(P)} = s] > 1/\sharp\Sigma + \epsilon. \tag{3}$$

Let P' be the algorithm which takes as input a finite sequence $u_0, \ldots, u_{n-1}$ of elements of Σ and outputs s if $P((u_0, \ldots, u_{n-1})) = 1$ and a random element of Σ if $P((u_0, \ldots, u_{n-1})) = 0$. Clearly, P' is in $\mathcal{F}$. We have then, using (3):

$$\begin{aligned} &P_e[P'((u_0, \ldots, u_{n-1}) = u_n)] \\ &\quad = P_e[P'((u_0, \ldots, u_{n-1}) = u_n) | P((u_0, \ldots, u_{n-1}) = 1)].d \\ &\quad + P_e[P'((u_0, \ldots, u_{n-1}) = u_n) | P((u_0, \ldots, u_{n-1}) \neq 1)](1 - d) \\ &\quad > (1/\sharp\Sigma + \epsilon).d + (1/\sharp\Sigma)(1 - d) = 1/\sharp\Sigma + \epsilon d, \end{aligned}$$

with $\epsilon d > 0$. This means that the sequence (u_n) is predictable. □

3. Theory of finite statistical tests

3.1. Definition of a statistical test. In this section, we give a definition for a statistical test. We focus on the particular case of finite tests and show how to associate a statistical model to a finite test. Then, we explain how to derive a certain class of predictors out of a statistical test and we give an entropy criterion for a sequence to be unpredictable.

We recall that the classical definition of a statistical test [13] is just an algorithm which receives as input a finite binary sequence of length n and returns a finite binary sequence of length $l(n)$ with $l(n) \leq n$. Such an algorithm transforms a distribution on the sequences of length n into a distribution on the sequences of length $l(n)$. From now on, in order to simplify the notations, we suppose that $\Sigma = \{0, 1\}$ i.e. Σ^* is the set of all binary sequences.

In our context, we have the following definition.

Definition 9. Let $\mathcal{F}$ be a family of elements of $\mathcal{M}$ a computational model. Let $k \in \mathbb{N}^*$, a statistical test of length k in $\mathcal{F}$ or a $(\mathcal{F}, k)$-test is an algorithm in $\mathcal{F}$ which takes as input any finite number $u_0, \ldots, u_n$ of elements of Σ and returns an elements of Σ^k. We say k-test for $(\mathcal{M}, k)$-test.

Given a k-test F and $(u_n) \in \Sigma^*$, one can define a distribution P on Σ^k by setting for $u \in \Sigma^k$, $P(u) = P_e[F(u_0, \ldots, u_n) = u]$.

Note that a $\mathcal{F}$-predictor is precisely a $(\mathcal{F}, 1)$-test.

In order to take into account the fact that a statistical test represented by a Turing machine may have a finite memory, we put on the restriction that the head on the tape which contains the input elements $u_0, \ldots, u_n$ can only go in one direction.

In the rest of this paper, we study the particular case of finite statistical tests and finite predictors. By a finite statistical test (resp. a finite predictor), we mean a statistical test (resp. predictor) given by a Turing machine with a finite amount of memory.

It should be noted that any k-bit pattern produced by a finite memory Turing machine M can be computed by another finite Turing machine M' from the global state of M which consists in the data of a dump of the memory (which does not contain the infinite input tape) of the machine M plus the position of the head on the tape and the internal state of the Turing machine coded in a suitable alphabet. So the distribution that we obtain from a finite k-test F can be computed by another finite Turing machine from the distribution obtained with the same test F which would produce a dump of its global state. We can then suppose that the output of the machine is simply its global state when the machine stops.

Now, let $s_0, \ldots, s_p$ be the global states a finite statistical test goes through between the reading of the input bits x_i and x_{i+1}. If we compute the distribution given by the frequency of occurrences of the states of the machine, the states $s_0, \ldots, s_p$ are equivalent since each of them entails the other and their number will only affect the weighting of the distribution. So we may choose to identify the states $s_0, \ldots, s_p$. As a consequence, the machine being in a state s can reach only two states s'_0 and s'_1 depending on the value of the input bit x_i being read.

Moreover, if $(u_n)_{n\in\mathbb{N}}$ is a binary sequence, and F is a finite statistical test which outputs its global state, one can compute $s' = F(u_0, \ldots, u_{n+1})$ from the data of $s = F(u_0, \ldots, u_n)$ and u_{n+1}. This property is an immediate consequence of the hypothesis that the head of F on the input tape can only go in one direction.

As a result of this discussion we have that a finite statistical test is just a finite automaton. Following the definition of a finite automaton [4], we have the

Proposition 10. There is a bijective correspondence between the set of finite statistical tests and the set of finite state automata. As a consequence, we can represent a finite statistical test F by the following triple (S, f, s_0)

- S a finite set of states $\{s_0, \ldots, s_k\}$,
- $f : S \times \Sigma \mapsto S$ a transition function,
- an initial state s_0.

The transition function f is defined by: for $s_i, s_j \in S$ and $\sigma \in \Sigma$, $f(s_i, \sigma) = s_j$ if the finite statistical test goes from state s_i to state s_j upon the reading of σ on the input tape. From this data, one can compute the map $F : \Sigma^{k+1} \to S$ for all k with the induction formula:

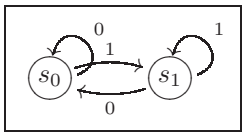
$$F(x_0, \ldots, x_k) = f(F(x_0, \ldots, x_{k-1}), x_k).$$

A finite statistical test $F = (S, f, s_0)$ can be represented by an oriented graph such that

- the nodes of the graph are the different states $s_i \in S$,
- the nodes s_i and s_j are linked by an edge labelled by $\sigma \in \Sigma$ if $f(s_i, \sigma) = s_j$.

Example 11. Let $\mathcal{F}_0$ be the finite statistical test with two states s_0 and s_1 such that $\mathcal{F}_0$ goes to state s_1 (resp. s_0) upon the reading of 1 (resp. 0). It is easy

FIGURE 1. Frequency test



to see that computing the mean value of occurrences of the states s_0 or s_1 is just another definition of the frequency (monobit) test [11]. The figure 1 gives the graph representation of this test.

More generally, for $k \in \mathbb{N}^*$, let $\mathcal{F}_k$ be the finite statistical test with 2^k states $\{s_0, \ldots, s_{2^k-1}\}$, the test being in state s_i upon the reading of the element u_l of the binary sequence (u_n) if we have $i = \sum_{\lambda=1}^{k} u_{n-k+\lambda} 2^{\lambda-1}$. The distribution that we obtain on the set of states of F_k corresponds to the distribution computed by frequency test on blocks of size k.

In the same way, one can describe the test of largest run of one [11] with a bound on the size of the run with a finite state machine. More generally, all the usual tests such as the matrix rank test, the Maurer's test, the (non)-overlapping template matching test belongs to the set of finite statistical tests ([11]).

We have an analogue definition for finite predictors.

Definition 12. A finite predictor P is given by the quadruple of (S, f, s_0, p) where
- (S, f, s_0) represents a finite statistical test,
- $p : S \mapsto \{0, 1\}$ is a function such that $p(s_i)$ gives the output of the predictor when in state s_i.

Definition 13. A finite statistical test $F = (S, f, s_0)$ (resp. $P = (S, f, s_0, p)$ a finite predictor) is unifilar if for σ, σ' two different elements of Σ and for all $s_i \in S$, we have $f(s_i, \sigma) \neq f(s_i, \sigma')$.

We remark that a finite statistical test (or predictor) is unifilar if and only if each element of the associated family of Markov chains is unifilar [1] pp. 187. It is clear that we do not lose generality if we suppose from now on that all finite statistical tests are unifilar adding if necessary some extra states to the test. The unifilar condition is used when invoking the Theorem 6.4.2 from [1] in the proposition 19 and in order to simplify the definition of the map Φ in the following paragraph.

There exists a map Φ from the set of finite statistical tests to the set of finite Markov chains parametrised by $\Lambda = (\lambda_{s_0}, \ldots, \lambda_{s_k}) \in [0, 1]^{k+1}$ that we denote by $\mathcal{M}_F(\Lambda)$. Φ maps $F = (S, f, s_0)$ with $S = \{s_0, \ldots, s_k\}$ onto $\mathcal{M}_F(\Lambda) = \{(S, M(\Lambda), Z_0), \Lambda = (\lambda_{s_0}, \ldots, \lambda_{s_k}) \in [0, 1]^{k+1}\}$ given by
- S a set of states.
- a transition matrix $M(\Lambda) = [m_{ij}(\Lambda)]$, $0 \leq i, j \leq k$ with $m_{ij} = \lambda_{s_i}$ if $f(s_i, 1) = s_j$, $m_{ij} = 1 - \lambda_{s_i}$ if $f(s_i, 0) = s_j$ and $m_{ij} = 0$ otherwise. One can check immediately that $M(\Lambda)$ is a stochastic matrix i.e. we have $\sum_{j=0}^{k} m_{ij} = 1$.
- an initial distribution Z_0 which is the distribution with total weight 1 on the state s_0.

Definition 14. Let $F = (S, f, s_0)$ be a finite statistical test. Let $s_i \in S$, s_i is said to be aperiodic if the greatest common divisor of the length of all the paths in the

graph representation of F starting in s_i and coming back to s_i is 1. We say that F is aperiodic if all its states are aperiodic.

The statistical test F is said to be indecomposable if there does not exist a proper subset $W \subset S$ such that $\cup_{\sigma\in\Sigma} f(W,\sigma) \subset W$.

We say that F is ergodic if it is aperiodic and indecomposable.

Remark 15. By definition, if a statistical test $F = (S, f, s_0)$ is not indecomposable, then the set of states S can be decomposed into a disjoint union of proper subsets W_i such that $\cup_{\sigma\in\Sigma} f(W_i,\sigma) \subset W_i$ plus some other states T. Moreover, starting from the state $s_0 \in S$, the statistical test will ultimately go into one of the W_i upon the reading of a random sequence. This means that the test F is equivalent to one of its restriction to the states W_i, the choice being made depending only on the first few bits of the random sequence. As a consequence, we do not loose generality by supposing that all the finite statistical tests are indecomposable. We make this assumption for the rest of this paper.

Lemma 16. A finite statistical test $F = (S, f, s_0)$ is indecomposable (resp.) ergodic if and only if all elements of the family of Markov chains $\Phi(F)$ are indecomposable (resp. ergodic) apart from a finite subset of the parameter set.

Proof. A Markov chain is ergodic if and only if it is aperiodic and indecomposable. The result follows immediately from the definitions, when we consider the subset of the parameter set of all non null parameters. □

Definition 17. Let Λ be a set of parameters, and Ω be a probability space. A *statistical model* on Ω is a probability distribution on Ω depending on the parameters Λ.

Let $F = (S, f, s_0)$ be a finite statistical test with set of states $\{s_0, \ldots, s_k\}$ that we consider as an equidistributed probability space. Let $\Phi(F) = \{(S, M(\Lambda), s_0), \Lambda \in]0, 1[^{k+1}\}$ be the family of Markov chains associated to F. These Markov chains have a unique stationary state of probability by [1] theorem 6.3.3 $W = [w_0, \ldots, w_k]$ on S which can be computed using the Chapman-Kolmogorov relation:

$$WM(\Lambda) = W.$$

In this way, we have provided S with a statistical model with set of parameters Λ.

Example 18. The statistical model associated to the frequency (monobit) test as described in example 11 has two parameters λ_{s_0} and λ_{s_1} linked by the relation $\lambda_{s_0} + \lambda_{s_1} = 1$ (see figure 1). This is nothing but the statistical model of a binary memoriless source (BMS).

Next, we introduce a well-known fundamental invariant which measures the uncertainty of a sequence relatively to a statistical test. Let $F = (S, f, s_0)$ be a finite statistical test with $S = \{s_0, \ldots, s_k\}$ and $T = (S, M(\Lambda), Z_0) \in \Phi(F)$. By definition, T gives a sequences of random variables $Z_0, Z_1, \ldots, Z_n, \ldots$ over the set of states S. We define the entropy of T as the limit:

$$H(T) = \lim_{n\to\infty} \frac{H(Z_0, \ldots, Z_n)}{n+1}, \tag{4}$$

where H, the usual entropy function is defined by

$$H(Z_0, \ldots, Z_n) = -\sum_{0\le i_0,\ldots,i_n\le k} p(z_{0,i_0}, \ldots, z_{n,i_n}) \log(p(z_{0,i_0}, \ldots, z_{n,i_n}))$$

with $p(z_{0,i_0},\dots,z_{n,i_n}) = P[Z_0 = s_{i_0},\dots,Z_n = s_{i_n}]$.

In the rest of this section, we let $\sharp$ be the index function from S into $\{0,\dots,k\}$ defined by $i = \sharp(s_i)$.

Proposition 19. Let F be a finite statistical test with set of states S. Let $T = (S, M(\Lambda), s_0) \in \Phi(F)$. Then $H(T) = 1$ if and only if $\Lambda = (1/2,\dots,1/2)$.

Proof. Let $[w_0,\dots,w_k]$ be the stationary state of probability of T. As T is unifilar, by [1] Theorem 6.4.2, we have

$$H(T) = \sum_{i=0}^{k} w_i H(m_{i\sharp(f(s_i,1))}, m_{i\sharp(f(s_i,0))}).$$

As $\sum_{i=0}^{k} w_i = 1$, $H(T) = 1$ if and only if for all i, $H(m_{i\sharp(f(s_i,1))}, m_{i\sharp(f(s_i,0))}) = 1$ but this is only the case when $m_{i\sharp(f(s_i,1))} = 1/2$ for all i. □

If (u_n) is an element of Σ^* then it defines a map that we also denote by (u_n) from the set of finite statistical tests $F = (S, f, s_0)$ to their associated family of Markov chains $(S, M(\Lambda), s_0)$. For this we have to explain how to compute $\Lambda = (\lambda_{s_0},\dots,\lambda_{s_k}) \in [0,1]^k$. We just set $\lambda_{s_i} = P_e[u_n = 1 | F((u_0,\dots,u_{n-1})) = s_i]$.

Corollary 20. Let $F = (S, f, s_0)$ be a finite ergodic statistical test and let $(u_n) \in \Sigma^*$. We have $H((u_n)(F)) < 1$ if and only if there exists a function $p : S \to \Sigma$ such that if we consider the predictor $P = (S, f, s_0, p)$ then $\mathrm{Adv}^P((u_n)) > 0$.

Proof. Let $T = (u_n)(F) \in \Phi(F)$ and let $M = [m_{ij}]$ be the transition matrix of T. Let $m_i = m_{ij}$ with j such that $f(s_i, 1) = s_j$. We define the function p such that $p(s_i) = 1$ if $m_i > 1/2$ and $p(s_i) = 0$ if $m_i \leq 1/2$ and consider the predictor $P = (S, f, s_0, p)$. Let $I \subset \{0,\dots,k\}$ be such that $\forall i \in I$, $m_i > 1/2$ and let $J = \{0,\dots,k\}\backslash I$. We set $w_i = P_e[F((u_0,\dots,u_{n-1})) = s_i]$. By definition

$$\mathrm{Adv}^P((u_n)) \geq \left(\sum_{i\in I} w_i m_i + \sum_{i\in J} w_i 1/2\right) - 1/2 = \sum_{i\in I} w_i(m_i - 1/2).$$

The result follows from this last formula and proposition 19. □

The meaning of the following corollary is that a sequence is unpredictable by all finite predictors if and only if its entropy is maximal.

Corollary 21. Let $\mathcal{M}_0$ be the set of finite Turing machines, then a sequence (u_n) is in $R7(\mathcal{M}_0)$ if and only if for all finite statistical test F, $H((u_n)(F)) = 1$.

Proof. The result follows from corollary 20 and proposition 8. □

3.2. CLASSIFICATION OF FINITE STATISTICAL TESTS. In this section, we define an order relation on the set of finite ergodic statistical tests. Basically, let F and F' be finite statistical tests, if F is stronger than F' then every sequence that passes the test F also passes F'.

Definition 22. Let $F = (S, f, s_0)$ and $F' = (S', f', s'_0)$ be two finite ergodic statistical tests. A morphism between F and F' is a map $\chi : S \to S'$ such that $\chi(s_0) = s'_0$ and compatible with the transition functions i.e. for all $s \in S$ and $\sigma \in \Sigma$, we have

$$\chi(f(s,\sigma)) = f'(\chi(s),\sigma).$$

A morphism $\chi : F \to F'$ of finite ergodic statistical tests induces a map $\chi_M : \Phi(F) \to \Phi(F')$ on the associated family of Markov chains defined as follows: if $\Phi(F) = (S, M(\Lambda), s_0)$ and $\Phi(F') = (S', M'(\Lambda'), t_0)$, then we set

$$\chi_M((S, M(\lambda_{s_0}, \dots, \lambda_{s_k}), s_0)) = (S', M'(\lambda'_{t_0}, \dots, \lambda'_{t_l}), t_0)$$

with for $i = 0, \dots, l$,

$$\Big(\sum_{s_j \in \chi^{-1}(t_i)} w_{s_j}\Big)\lambda'_{t_i} = \sum_{s_j \in \chi^{-1}(t_i)} w_{s_j}\lambda_{s_j}, \tag{5}$$

where $[w_{s_0}, \dots, w_{s_k}]$ is the stationary state of probability of $(S, M((\lambda_{s_0}, \dots, \lambda_{s_k})), s_0)$.

Proposition 23. Let $\chi : F \to F'$, be a morphism of statistical tests and χ_M the associated morphism on Markov chains. Let $(S, M(\Lambda), s_0)$ be an element of the family $\Phi(F)$ then

$$H((S, M(\Lambda), s_0)) \leq H(\chi_M(S, M(\Lambda), s_0))$$

Proof. Denote by $\zeta : [0, \dots, k] \to [0, \dots, l]$ the map such that $\chi(s_i) = t_{\zeta(i)}$. Let $[w_0, \dots, w_k]$ and $[w'_0, \dots, w'_l]$ be the stationary states of probability of F and F'. From the Chapman-Kolmogorov relation and (5), we deduce that

$$w'_j = \sum_{i \in \zeta^{-1}(j)} w_i. \tag{6}$$

Next, set $\lambda_{i0} = m_{i\sharp(f(s_i,0))}$, $\lambda_{i1} = m_{i\sharp(f(s_i,1))}$ for $i = 0, \dots, k$ and in the same manner, $\lambda'_{i0} = m'_{i\sharp(f'(t_i,0))}$, $\lambda'_{i1} = m'_{i\sharp(f'(t_i,1))}$ for $i = 0, \dots, l$. We have

$$H(T) = \sum_{i=1}^{k} w_i H(\lambda_{i0}, \lambda_{i1}) = \sum_{j=0}^{l} \sum_{i \in \zeta^{-1}(j)} w_i H(\lambda_{i0}, \lambda_{i1}), \tag{7}$$

On the other hand, write

$$H(T') = \sum_{j=0}^{l} w'_j H(\lambda'_{j0}, \lambda'_{j1}). \tag{8}$$

So the proposition will be achieved if we have the following inequality

$$w'_j H(\lambda'_{j0}, \lambda'_{j1}) \geq \sum_{i \in \zeta^{-1}(j)} w_i H(\lambda_{i0}, \lambda_{i1}),$$

which can be rewritten following equations (5) and (6), as

$$H(\frac{1}{W}\sum_{i \in \zeta^{-1}(j)} w_i\lambda_{i0}, \frac{1}{W}\sum_{i \in \zeta^{-1}(j)} w_i\lambda_{i1}) \geq \frac{1}{W}\sum_{i \in \zeta^{-1}(j)} w_i H(\lambda_{i0}, \lambda_{i1}),$$

where $W = \sum_{i \in \zeta^{-1}(j)} w_i$. But this inequality is an immediate consequence of the convexity of the function $x \mapsto x \ln x$ on $\mathbb{R}_+^*$. □

Definition 24. Let F and F' be two finite statistical tests. If there exists a morphism $\chi : F \to F'$, we say that F is stronger than F' and we denote $F \geq F'$. If $F \geq F'$ and $F' \geq F$, we say that F and F' are isomorphic.

Proposition 25. If F and F' are isomorphic statistical tests then $\Phi(F)$ and $\Phi(F')$ are the same family of Markov chains.

Proof. It suffices to remark that if F and F' are isomorphic then the graph representation of F and F' are isomorphic. □

Proposition 26. The relation $\geq$ is an order relation in the set $\mathcal{F}$ of all finite statistical tests modulo isomorphism.

Proof. The relation $\geq$ is clearly transitive and reflexive. Moreover by definition if we have $F \geq F'$ and $F' \geq F$ then $F \simeq F'$. □

Proposition 27. If $\chi : F \to F'$ is a morphism between two finite statistical tests then we have $\chi_M((u_n)(F)) = (u_n)(F')$.

Proof. Let $[s_0, \ldots, s_k]$ and $[t_0, \ldots, t_l]$ be respectively the set of states of F and F'. Let $W = [w_0, \ldots, w_k]$ and $W' = [w'_0, \ldots, w'_l]$ be the stationary states of probability of $(u_n)(F)$ and $(u_n)(F')$. Then, we have

$$P[W' = t_i] = \sum_{j \in \zeta^{-1}(i)} P[W = s_j] = \sum_{j \in \zeta^{-1}(i)} w_i.$$

This last formula defines uniquely the stationary state of probability of $(u_n)(F')$. From this and the Chapman-Kolmogorov equalities, we deduce that $(u_n)(F') = \chi_M((u_n)(F))$. □

The following corollary means that if a test F is stronger than a test F' then every sequence that passes the test F also passes the test F'.

Corollary 28. If $\chi : F \to F'$ is a map between two finite statistical tests then for any sequence $(u_n) \in \Sigma^*$, $H((u_n)(F)) \leq H((u_n)(F'))$.

Proof. This is an immediate consequence of propositions 23 and 27. □

3.3. COMPLETE FAMILIES OF TESTS. At this point, we have introduced a certain order relation on the set of all finite statistical tests. We have seen that if a test F is stronger and F' then every sequence that pass the test F also pass the test F'. A natural question is the following: is there any finite statistical test in the set of all finite statistical tests which is stronger than all the others. The answer to this question is clearly no since if such a test would exist its number of states would be greater than the number of states of all other finite statistical test. This is not possible. But then, one would like to describe a certain infinite sequence of finite statistical tests F_i, that we call complete family of tests, such that if a sequence passes all the F_i then it will pass all the finite statistical tests. We see in the following that an answer to this question can be deduced from the seminal work of Shannon on information theory [12]. We give here some precise statements.

Definition 29. Let $F = (S_i, f_i, s_{0i})\ i \in I$, be a family of finite statistical tests. Consider the transition function $\times_{i\in I} f_i$ on the Cartesian product of the set of states $\times_{i\in I} S_i$ defined by $(\times_{i\in I} f_i)((s_i)_{i\in I}, \sigma) = (f_i(s_i, \sigma))_{i\in I}$ with $(s_i)_{i\in I} \in \times_{i\in I} S_i$ and $\sigma \in \Sigma$. Let W in $\times_{i\in I} S_i$ be the smallest set among the one which contains $(s_{0i})_{i\in I}$ and stable by the action of $\times_{i\in I} f_i$ i.e. we have $(\times_{i\in I} f_i)(W) \subset W$.

If I is a finite set $(W, \times_i f_i, (s_{0i})_{i\in I})$ is a finite statistical state called the product of $(F_i)_{i\in I}$ and denoted $\times_{i\in I} F_i$.

Lemma 30. Let F and F' be two finite statistical tests. Then there exist two natural morphisms of statistical tests $p : F \times F' \to F$ and $p' : F \times F' \to F'$ given by the first (resp. the second) projection of $S \times S'$ onto S (resp. onto S').

Proof. This is a simple verification. □

Proposition 31. The product of finite statistical tests is associative and commutative. Moreover, if F and F' are two statistical tests we have $F \times F' \simeq F$ if and only if there exists a morphism of finite statistical tests $\chi : F \to F'$.

Proof. The first assertion is a direct consequence of the definitions. For the second, if $F \times F' \simeq F$ then by the preceding lemma there exists a morphism from $F \simeq F \times F'$ to F'. Reciprocally, if there exists a morphism $\chi : F \to F'$ then we have the following sequence of morphisms :

$$F \simeq F \times F \to F \times F' \to F$$

from which we deduce that $F \simeq F \times F'$. □

One question which often arises when dealing with statistical tests is to be able to decide whether a certain set of tests are independent i.e. do they cover a great variety of unrelated statistical properties? Our definition of a finite statistical test cast some new light on this problem. We have embedded all the finite tests in a big statistical model with an infinite set of parameters and it is now possible to decide if two statistical tests are independent. Let F and F' be two finite statistical tests then the bigger their product is, the more independent they are. Their are two extreme interesting cases:

- when $F \times F' = F$ then there is a morphism from F to F' and the set of parameters of $\Phi(F)$ is contained in the set of parameters of $\Phi(F')$.
- when the set of states of $F \times F'$ is the Cartesian product of the set of states of F and F', then F and F' are completely independent.

Let $F = (S_i, f_i, s_{0i})$, $i \in I$ be a family of finite statistical test. If we do not suppose I finite then $\times_{i\in I} F_i$ is not anymore a finite statistical test since we obtain an automaton with an infinite number of states. Nevertheless we can consider $\times_{i\in I} F_i$ as an infinite directed graph. We have the

Theorem 32. Let $(F_i), i \in I$ be a family of finite statistical tests. If $\times_{i\in I} F_i$ is a tree, i.e. a graph without cycle then for all binary sequences (u_n), and all finite statistical tests F, we have

$$H((u_n)(F)) \geq \inf\{H((u_n)(F_i)), i \in I\}.$$

Proof. Let (F_i), $i \in I$ be a family of finite statistical tests such that $\times_{i\in I} F_i$ is a tree. Let F be any finite statistical test and (u_n) be a binary sequence. We know that $(u_n)(F)$ (resp. for all $i \in I$, $(u_n)(F_i)$) is a Markov chain which gives in particular a sequence X_n^F (resp $X_n^{F_i}$) of random variables. As $(u_n)(F)$ and $(u_n)(F_i)$ are unifilar, we can moreover assume that X_n^F and $X_n^{F_i}$ have value in Σ (see [1] pp. 188).

The hypothesis that $\times_{i\in I} F_i$ is a tree implies that for all n there exists a $l(n)$ such that for all $k \geq l(n)$ the subgraph of the graph representation of F_k given by all the nodes which are at distance less than n from s_0 is a tree. As a consequence, for all n there exists $l(n)$ such that for all $k \geq l(n)$, the knowledge of the value $X_0^{F_k}, \ldots, X_n^{F_k}$ is equivalent to the knowledge of $u_0, \ldots, u_n$. We have then $H(X_{n+1}^{F_k} | X_0^{F_k}, \ldots, X_n^{F_k}) \leq H(X_{n+1}^F | X_0^F, \ldots, X_n^F)$.

By definition,

$$H(X^F) = \lim_{n\to\infty} H(X_{n+1}^F | X_0^F, \ldots, X_n^F),$$

and we know by [1] pp.186 that the sequence $H(X_{n+1}^{F_{l(n)}} | X_0^{F_{l(n)}}, \ldots, X_n^{F_{l(n)}})$ is non-increasing so

$$H(X^{F_{l(n)}}) \leq H(X_{n+1}^{F_{l(n)}}|X_0^{F_{l(n)}}, \ldots, X_n^{F_{l(n)}}) \leq H(X_{n+1}^F|X_0^F, \ldots, X_n^F).$$

For a $\epsilon > 0$, taking n such that $H(X_{n+1}^F|X_0^F, \ldots, X_n^F) - H(X^F) < \epsilon$, we obtain that for all $\epsilon > 0$ there exists an n_0 such that $H(X^{F_{n_0}}) \leq H(X^F) + \epsilon$ and this complete the proof. □

A consequence of theorem 32 is the following: if we have found a family $(F_i)_{i\in I}$ of finite statistical tests such that $\times_{i\in I} F_i$ is a tree then every sequence (u_n) that pass all the tests $(F_i)_{i\in I}$ also pass all the finite statistical tests.

Example 33. Let $(u_n) \in \Sigma^*$. Let $(G_i)_{i\in\mathbb{N}^*}$ be the family of finite ergodic statistical tests such that the state s of G_i upon the reading of the k^{th} element of the sequence (u_n) is given by the vector of size i $(u_{\lfloor k/i\rfloor i}, u_{\lfloor k/i\rfloor i+1}, \ldots, u_{\lfloor k/i\rfloor i+r}, 0, \ldots, 0)$ where $r = k \mod i$. It is clear that a sequence passes the tests G_i if and only if it is $i-$distributed. Moreover the family of tests $(G_i)_{i\in\mathbb{N}^*}$ verifies the condition of theorem 32.

As a consequence of the preceding example, we have the

Corollary 34. Let $(u_n) \in \Sigma^*$, we have the following equivalent propositions

1. (u_n) is ∞-distributed,
2. (u_n) passes all the finite statistical tests,
3. (u_n) is unpredictable for all finite predictors.

This corollary may be viewed as an analog the theorem B of [5] pp. 153 which states that all distributions computed in a certain way from a $\infty-$distributed sequence are the same.

4. Testing a device

We remark that all statistical models associated to a finite statistical test describe for certain values of their parameters a distribution which is verified when the input random variables are Bernoulli with parameter 1/2. The connection between the theory of hypothesis testing and random test is then the following: one chooses a statistical model adapted to the device to be tested, then takes for H_0 the value of the parameter of the model which corresponds to the distribution given by Bernoulli law with parameter 1/2 and as alternative hypothesis the contrary case.

For example, if we know that the statistical model of the device is a binary memoriless source (BMS) then one can use the only monobit frequency test [5]: one can prove that this test is the best possible for this model.

We have seen that some statistical models are more general than others. For instance the (BMS) model is contained in the general model of an ergodic stationary source with memory. The advantage of the more specific test is that it will be in general more sensitive to a given bias. But it will be useless to detect some other non random behavior.

In a general manner, the practice of statistical test in order to check the random behavior of a device is characterized by

- the choice of a statistical model adapted to the device;
- the choice of a restricted set of statistical tests associated to this statistical model.

In practice, the choice of a statistical model used to describe the device is based on the some heuristic hypothesis. The experience shows that most of the time, the problems which arise in the design of a random device will affect its behavior in some very precise manner : for instance, it may be the duplication of some bits or a kind of stuttering which correspond to the repetition of the same pattern at periodic intervals or also the production of unbalanced patterns of bits. This kind of non random behavior may be easily identified using the frequency test, Maurer's test or the collision tests. These tests should always be part of a routine random test suite. However if one of these tests detects a deviation it is very difficult to decide if this is due to a failure of the device in the absence of a systematic explanation.

In the process of selecting a test suite for a device, one should also keep in mind its cryptographic purpose. For instance, in order to generate some initialisation vector, one can use a linear feedback shift register because it will produce a long sequence of non repeating blocks, but by definition this random generator will behave poorly with respect to the linear complexity test.

Last but not least, one should also take care of the fact that in some circumstances random tests may jeopardise the quality of a random generator : for instance, if a statistical test with a none well adjusted rejection rate is used to select blocks of data at the output of a generator then it will induce a bias easy to detect using the same test on a sample of the output. This may be of some consequences in some applications where the output of the random number generator has to be indistinguishable from a truly random sequence.

Acknowledgements

The author would like to thank Didier Alquié and Franck Landelle for their useful comments during the preparation of this article.

`<david.lubicz@univ-rennes1.fr>`

References

[1] Robert B. Ash. *Information theory*. Dover Publications Inc., New York, 1990. Corrected reprint of the 1965 original.

[2] Jean-Sébastien Coron and David Naccache. An accurate evaluation of Maurer's universal test. In *Selected areas in cryptography (Kingston, ON, 1998)*, volume 1556 of *Lecture Notes in Comput. Sci.*, pages 57–71. Springer, Berlin, 1999.

[3] Oded Goldreich. *Foundations of cryptography*. Cambridge University Press, Cambridge, 2001. Basic tools.

[4] John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages, And Computation*. Addison Wesley Longman, 2006.

[5] Donald E. Knuth. *The art of computer programming. Vol. 2*. Addison-Wesley Publishing Co., Reading, Mass., second edition, 1981. Seminumerical algorithms, Addison-Wesley Series in Computer Science and Information Processing.

[6] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Internat. J. Comput. Math.*, 2:157–168, 1968.

[7] George Marsaglia. Diehard. Available at `http://www.csis.hku.hk/ diehard/`.

[8] Per Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966.

[9] Ueli M. Maurer. A universal statistical test for random bit generators. *J. Cryptology*, 5(2):89–105, 1992.

[10] William Mendenhall and Richard L. Scheaffer. *Mathematical statistics with applications*. Duxbury Press, North Scituate, Mass., 1973.

[11] NIST. A statistical test suite for random and pseudorandom number generators for cryptographic applications. Available at `http://csrc.nist.gov/CryptoToolkit/tkrng.html`.

[12] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948.
[13] Andrew C. Yao. Theory and applications of trapdoor functions. *IEEE*, pages 80–91, 1982.