

On the Role of the Law of Large Numbers in the Theory of Randomness¹

An. A. Muchnik and A. L. Semenov

Institute of New Technologies, Nizhnyaya Radishchevskaya 10, Moscow, 109004 Russia

Abstract—In the first part of this article, we answer Kolmogorov’s question (stated in 1963 in [1]) about exact conditions for the existence of random generators. Kolmogorov theory of complexity permits of a precise definition of the notion of randomness for an individual sequence. For infinite sequences, the property of randomness is a binary property, a sequence can be random or not. For finite sequences, we can solely speak about a continuous property, a measure of randomness. Is it possible to measure randomness of a sequence t by the extent to which the law of large numbers is satisfied in all subsequences of t obtained in an “admissible way”? The case of infinite sequences was studied in [2]. As a measure of randomness (or, more exactly, of nonrandomness) of a finite sequence, we consider the specific deficiency of randomness δ (Definition 5). In the second part of this paper, we prove that the function $\delta/\ln(1/\delta)$ characterizes the connections between randomness of a finite sequence and the extent to which the law of large numbers is satisfied.

INTRODUCTION

In 1930-s Andrei Kolmogorov founded probability theory on the base of measure theory. In [1] he writes: “The set theoretic axioms of the calculus of probability . . . had solved the majority of formal difficulties in the construction of a mathematical apparatus which is useful for a very large number of applications of probabilistic methods so successfully that the problem of finding the basis of real applications of the results of the mathematical theory of probability became of secondary importance to many investigators.”

However, Kolmogorov himself regarded the question about the basis as a principal one. In 1962, during his visit to India, he began to develop a new approach to it,² the so-called theory of descriptive complexity. Now the research area initiated by Kolmogorov has grown into a rich theory, which has important connections not only with probability theory but also with theory of algorithms, theory of coding, theory of matroids, and other fields of mathematics.

As for practical applications, the main results are to appear in future. To get them, it is required to take into account not only the descriptive complexity of a program but also an amount of resources used by it. In many cases, this task is connected with unsolved problems of computational complexity theory.

Consider a sequence of independent trials with two equiprobable outcomes, 0 and 1. The simplest and at the same time the most significant condition of randomness for a sequence of outcomes is an approximate equality of the number of zeros and the number of ones. Obviously, this requirement alone is not sufficient (for instance, the sequence 01010101 . . . does not seem to be random). But if *this condition holds for all subsequences obtained from the original one with the help of*

¹ Supported in part by the Russian Foundation for Basic Research, project nos. 01-01-00505, 02-01-10904, and 02-01-22001, and the Council on Grants for Scientific Schools.

² The first publication appeared in 1963, see [1].

“admissible” place-selection rules, then such a sequence can be considered as a random generator. Surely, the notion of an admissible rule should be made mathematically precise.³

Is the mentioned frequency criterion universal? A number of known facts about infinite sequences would rather be an evidence of the opposite. For example (see [3]),

- (i) there exists a set S of measure 0 that consists of infinite binary sequences, and for any countable family R of admissible place-selection rules there exists $s \in S$ such that the law of large numbers is satisfied in all infinite subsequences selected from s by rules belonging to R .

The construction of the set S and its property to be of measure 0 are effective, hence all its elements are intuitively nonrandom.

Kolmogorov stressed the importance of the analysis not only of limit regularities of infinite sequences but of finite sequences too. All results of our paper lie in the finite field. An analog of (i) for finite sequences would be the following:

- (ii) Assume that L is a natural number. There exists a set S of finite sequences of length L such that the cardinality of S is small enough compared with 2^L and, for any not too large family R of admissible place-selection rules, there exists $s \in S$ such that the law of large numbers is satisfied precisely enough in all not too short subsequences selected from s by rules belonging to R .

Surely, the last statement should be made more precise. What is “small enough,” “not too large,” “precisely enough,” “not too short”? Remarkably, there is a natural precise meaning of these expressions such that the negation of (ii) is true. Thus, it is possible to establish a positive connection between the notions of *frequency* and *universal* randomness (Theorem 4).

1. ANALYSIS OF KOLMOGOROV’S ARTICLE “ON TABLES OF RANDOM NUMBERS”

1.1. *Philosophical Motivation*

As Kolmogorov wrote in [1], for a long time he thought that

1. “The frequency concept based on the notion of limiting frequency as the number of trials increases to infinity does not contribute anything to substantiate the applicability of results of probability theory to real practical problems, where we always deal with a finite number of trials”;
2. “The frequency concept applied to a large but finite number of trials does not admit a rigorous formal exposition within the framework of pure mathematics.”

Kolmogorov’s opinion on the first statement had not changed (as he said in [1]).⁴ As for the second statement, Kolmogorov came to the conclusion that, using the frequency conception of randomness and the refined notion of the complexity of a program, it is possible to formulate purely mathematical conditions under which probability theory can be applied to practice. He wrote that an exact definition of the complexity of a program would be given in another paper, but in [1] the fact that the number of simple objects cannot be very large was only used. Later, Kolmogorov referred to this approach as the combinatorial one (see [4]). Definitions that correspond to the algorithmical approach were given by Kolmogorov in 1965 (see [4]). The results of [1] have natural algorithmical analogs, and we formulate and prove them. This parallel between the two approaches

³ An example of an admissible rule is “to select all digits at even places.” An example of a nonadmissible rule is “to select the digits at places such that the original sequence contains zeros at them.”

⁴ Though, in other papers, Kolmogorov emphasized that researches on infinite sequences are of great heuristic significance.

can be extended widely. Therefore, in our paper, most of the theorems are formulated in two variants, combinatorial and algorithmical, and are numbered in parallel (numbers of algorithmical theorems have the prime).

Kolmogorov considered a finite binary sequence to be a table of random numbers (a random generator) if the frequencies of zeros and ones are close to $1/2$ in all long enough subsequences obtained by not too complex place-selection rules.

The mathematical purpose of [1] was to estimate how large the complexity of place-selection rules can be to still ensure that at least one random generator exists.

1.2. Definitions

In [1], Kolmogorov defined the notion of a (nonmonotonic) place-selection rule on a finite binary sequence \mathbf{t} .

An informal description is as follows. Assume that we have a sequence of cards of the same length as a given sequence \mathbf{t} . The digits of the sequence \mathbf{t} are written on the face of the cards, the back is the same for all cards. In the beginning, cards lie face down in the order of digits in \mathbf{t} . A rule decides what card should be turned over and (before turning over) whether it should be included in the subsequence under construction. To make the next decision, the rule takes into account digits on the cards that have already been turned over. The selected digits are arranged in subsequence according to the order of selecting, not the order of the original sequence. Here is the formal definition.

Definition 1. A (*place-selection*) rule on sequences of length L is a function r that maps binary sequences of lengths from 0 to $L - 1$ to pairs from the set $\{1, \dots, L\} \times \{\text{“select”}, \text{“do not select”}\}$, where the first components of values of the function are always different on any sequence and on any of its proper extensions. Assume that \mathbf{t} is a sequence of length L . For each i from 0 to L , by induction on i we construct a binary sequence \mathbf{s}_i of length i . Put $\mathbf{s}_0 = \Lambda$. To obtain \mathbf{s}_{i+1} , we write the $\pi_1(r(\mathbf{s}_i))$ th digit of the sequence \mathbf{t} at the end of \mathbf{s}_i . The sequence selected from \mathbf{t} by the rule r is obtained from \mathbf{s}_L by eventual deleting all digits with numbers i such that $\pi_2(r(\mathbf{s}_{i-1})) = \text{“do not select”}$. (Here π_1 and π_2 denote the first and second elements of a pair.)

The selected sequence is denoted by $r[\mathbf{t}]$.

Sometimes, it is useful to consider more restricted classes of rules. *Monotonic rules* turn over the cards sequentially in the original order.⁵ *Nonadaptive rules* point out a certain set of cards at once, and the subsequence consists of digits written on these cards and arranged in the original order.

Definition 2. Assume that \mathcal{R}_L is a set of rules on sequences of length L . A sequence \mathbf{t} of length L is called an (n, ε) -random generator with respect to \mathcal{R}_L if, for every rule $r \in \mathcal{R}_L$, the selected subsequence $r[\mathbf{t}]$ has the following property:

If the length of $r[\mathbf{t}]$ is greater than or equal to n , then the difference between the fraction of zeros and $1/2$ is less than ε .

The absolute value of the difference between $1/2$ and the fraction of zeros in the sequence is called the *deviation* (of the fraction of zeros).

Let us remark that, for each $p \in [0, 1]$, we can consider (n, ε, p) -random generators, where the fraction of zeros in the selected subsequences are close to not $1/2$ but p .

How can we translate words about not too complex place-selection rules into a formal language?

⁵ Monotonic rules were first considered by Church in 1940.

Under the combinatorial approach, the *complexity* of a finite set is the binary logarithm of its cardinality.

Under the algorithmical approach, we use the *entropy* of a constructive object. The notion of entropy⁶ was defined by Kolmogorov in 1965. Actually, a class of functions was defined such that any two of them differ by a constant. Kolmogorov hoped that it would be achievable to suggest a certain “natural” programming language such that the corresponding entropy function is less than the entropy function of any other “natural” programming language plus 100. Since our results are equally applicable to any entropy function, formulations of these results may contain a constant C depending on the choice of a programming language only. We also use the notion of *conditional* entropy, which was also introduced by Kolmogorov.

The analogy between the combinatorial and algorithmical approaches is based on the following two propositions, proved by Kolmogorov.

Proposition 1. *The set of objects such that their entropy conditional to a fixed object is less than m cannot contain more than $2^m - 1$ elements.*

Proposition 2. *Assume that a set is enumerated by a program of entropy less than α and the cardinality of this set is less than 2^m . Then the entropy of its elements is less than $m + \alpha + 2 \text{lb } \alpha + C$. (Here and in the sequel, we denote by $\text{lb}(x)$ the binary logarithm of x .)*

Consider the set of all sequences such that the subsequence selected by a fixed rule has length and deviation greater than given values. The law of large numbers implies that, if these values are large enough (in other words, if we speak about long subsequences of large deviation), then the cardinality of this set is small. However, even if we do not require that the deviation is small, the cardinality can be small anyway. As an example, consider the following monotonic rule: if the digits already selected form the beginning of the sequence $01010101\dots$, then the next digit have to be selected. If the length of selection is equal to the length of the original sequence and an arbitrary deviation, the corresponding set contains only two elements.

To overcome this difficulty, we introduce the notion of a normal rule, and we will consider sets generated by normal rules only.

Definition 3. A rule r on sequences of length L is called *normal* if r selects subsequences of the same positive length from all sequences of length L .

For a normal rule r that selects subsequences of length n and for a number $\varepsilon \in [0, 1/2]$, we denote by $A_{r,\varepsilon}$ the set of all sequences such that r selects a subsequence with deviation not less than ε . Such sets are called *regular*.

Let us introduce the notation, which we will often use:

$$d(n, \varepsilon) \equiv 2n\varepsilon^2 \text{lb } \varepsilon.$$

The following facts (they will be discussed in detail later) give reasons for this notation. If a normal rule r selects subsequences of length n , then the cardinality of the regular set $A_{r,\varepsilon}$ depends on L , n , and ε only. If ε is small enough and n is large enough compared with $1/\varepsilon$, then $|A_{r,\varepsilon}| \approx 2^{L-d(n,\varepsilon)}$.

There exists an effective operation (we will call it *n -normalization*) that transforms each rule into a normal one, the length of selection become equal to n , and selections that had length exactly n do not change.

Let us describe the procedure of n -normalization of a rule. The new rule operates as the old one, but with two exceptions. If the old rule has already selected n digits, then the new rule stops selecting at this moment. If the old rule has already selected $k < n$ digits and only $n - k$ digits are not selected, then the new rule selects all of them.

⁶ A comparison of different versions of this notion can be found in [5].

1.3. Sufficient and Necessary Conditions for Existence of a Random Generator

Theorem 1 (Kolmogorov, 1963). Consider arbitrary numbers L (the length of a sequence), $\varepsilon > 0$ (the deviation), and $n \geq \varepsilon^{-4}$ (the length of selected subsequences). For any set of rules \mathcal{R}_L , if its complexity is less than

$$d(n, \varepsilon)(1 - \varepsilon),$$

then there exists an (n, ε) -random generator with respect to \mathcal{R}_L .

Proof. In [1], Kolmogorov just outlined the proof of this theorem; for this reason, we give a complete proof here.

Obviously, we can assume without loss of generality that $\varepsilon \leq 1/2$. Consider the uniform distribution on sequences of length L (all digits are independent and are equal to 0 and 1 with probability $1/2$). The probability of each sequence is equal to 2^{-L} .

Take a place-selection rule r . By r_k , denote the result of k -normalization of r . Since digits keep to be independent when their order changes, for r_k the probability of selecting a given subsequence of length k is equal to 2^{-k} .

Let us estimate the probability that a sequence \mathbf{t} is not an (n, ε) -random generator for the rule r (i.e., the length of selection is greater than or equal to n , and the deviation is ε or greater). Consider the least $k \geq n$ such that the deviation in the beginning of $r[\mathbf{t}]$ of length k is not less than ε . It is clear that

$$\begin{aligned} & \Pr\{\mathbf{t} \text{ is not an } (n, \varepsilon)\text{-random generator for } r\} \\ & \leq \Pr\{\text{the deviation in } r_n[\mathbf{t}] \geq \varepsilon\} + 2 \sum_{k=n}^L \Pr\left\{\text{the number of zeros in } r_k[\mathbf{t}] = \left\lceil k \left(\frac{1}{2} + \varepsilon\right) \right\rceil\right\} \\ & = 2 \sum_{j=\lceil n(\frac{1}{2} + \varepsilon) \rceil}^n \binom{n}{j} 2^{-n} + 2 \sum_{k=n}^L \binom{k}{\lceil k(\frac{1}{2} + \varepsilon) \rceil} 2^{-k}. \end{aligned}$$

For $j \neq k$, we use the following inequality arising from the Stirling formula:

$$\binom{k}{j} \leq \frac{e^{k \cdot h(j/k)}}{\sqrt{2\pi j(k-j)/k}},$$

where $h(x) = -x \ln x - (1-x) \ln(1-x)$ is the Shannon entropy function. The denominator of the bound is not less than $\sqrt{2\pi(1-1/k)} \geq \sqrt{2\pi \cdot 15/16}$ since $k \geq n \geq \varepsilon^{-4} \geq 16$. The sign of the derivative of $h(x)$ shows that, for $x \geq 1/2$, the function h decreases; therefore, $h(j/k) \leq h(1/2 + \varepsilon)$. Differentiating twice, we can easily show that $h(1/2 + \varepsilon) \leq \ln 2 - 2\varepsilon^2$.

Thus, the upper bound on the probability for the rule r is less than

$$2^{1-n} + \sqrt{\frac{8}{15\pi}} e^{-2n\varepsilon^2} n + \sqrt{\frac{8}{15\pi}} \frac{2e^{-2n\varepsilon^2}}{1 - e^{-2\varepsilon^2}}.$$

Simple calculations show that the last bound is strictly less than $e^{-2n\varepsilon^2(1-\varepsilon)}$ for $\varepsilon \leq 1/2$ and $n \geq \varepsilon^{-4}$ (the inequality $1 - e^{-x} \geq x/\sqrt{e}$ for $0 \leq x \leq 1/2$ is used).

To estimate the probability that a sequence is not an (n, ε) -random generator for at least one rule from \mathcal{R}_L , we multiply it by the number of rules. So the probability that a sequence is not an (n, ε) -random generator with respect to \mathcal{R}_L is strictly less than 1; therefore, there exists at least one (n, ε) -random generator. \triangle

Remark 1. Kolmogorov's theorem, as well as our results below, can be generalized to the case where 0 and 1 have frequencies close to p and $1 - p$ respectively in all long subsequences selected by simple rules.

Under the algorithmical approach, we get the following statement.

Theorem 1'. *Consider arbitrary numbers L (the length of a sequence), $\varepsilon > 0$ (the deviation), and $n \geq \varepsilon^{-4}$ (the length of selected subsequences). Then there exists an (n, ε) -random generator for the set of rules \mathcal{R}_L consisting of all rules such that their entropy conditional to L is less than*

$$d(n, \varepsilon)(1 - \varepsilon).$$

Proof. Proposition 1 implies that the set of rules with entropy less than $d(n, \varepsilon)(1 - \varepsilon)$ has complexity less than $d(n, \varepsilon)(1 - \varepsilon)$, so the algorithmical theorem trivially follows from the combinatorial result. \triangle

Theorem 2 (Kolmogorov, 1963). *Consider arbitrary numbers L (the length of a sequence), $\varepsilon \in (0, 1/20)$ (the deviation), and $n \in [\varepsilon^{-3}, L/2]$ (the length of selected subsequences). There exists a set \mathcal{R}_L of nonadaptive rules such that its complexity is less than*

$$4n\varepsilon(1 + 5\varepsilon)$$

and there does not exist an (n, ε) -random generator with respect to \mathcal{R}_L .

Proof. We should construct a set of nonadaptive place-selection rules such that, for each sequence \mathbf{t} , there is a rule r in this set that selects a long subsequence $r[\mathbf{t}]$ of large deviation.

Let

$$m = \left\lfloor \frac{1}{4\varepsilon} + \frac{1}{2} \right\rfloor, \quad L' = 2m \left\lceil \frac{n}{2m - 1} \right\rceil.$$

It can easily be checked that $L' < L$ under conditions on ε and n from the theorem assumption.

Our rules will select subsequences just from the first L' digits of a sequence. Namely, we split the beginning of length L' into m segments of equal length L'/m . A rule selects exactly one-half of digits from one of these segments and all digits from the others. Thus, a rule is determined by the number of a segment and a subset of $\{1, \dots, L'/m\}$ consisting of $L'/(2m)$ elements.

Each rule selects a subsequence of length

$$n' = (2m - 1) \left\lceil \frac{n}{2m - 1} \right\rceil \geq n.$$

Take a sequence \mathbf{t} . We prove that there is a rule r such that the deviation in $r[\mathbf{t}]$ is greater than ε . Denote the beginning of the sequence \mathbf{t} of length L' by \mathbf{t}' . Let us consider three cases.

1. Assume that there are two segments of \mathbf{t}' such that at least one-half of digits in the first one are zeros and at least one-half of digits in the second one are ones.

Consider two rules, r_1 and r_2 . To define them, we point out which digits they do *not* select.

- For r_1 , all such digits are zeros and lie in the first segment.
- For r_2 , all such digits are ones and lie in the second segment.

The number of zeros in $r_1[\mathbf{t}]$ is equal to the number of zeros in \mathbf{t}' minus $L'/(2m)$, and the number of zeros in $r_2[\mathbf{t}]$ is equal to the number of zeros in \mathbf{t}' . Thus, the numbers of zeros in these selections differ by $L'/(2m)$; hence, the difference between one of them and $n'/2$ is not less than $L'/(4m)$. The deviation in either $r_1[\mathbf{t}]$ or $r_2[\mathbf{t}]$ is not less than

$$\frac{L'/(4m)}{n'} = \frac{1}{2(2m - 1)} \geq \varepsilon.$$

2. Assume that, in all segments of \mathbf{t}' , at least one-half of all digits are zeros.

Then consider a rule that selects only zeros from (for instance) the first segment and all digits from the others. The number of zeros in the selection is not less than

$$\frac{L'}{2m} + (m - 1)\frac{L'/m}{2} = \frac{L'}{2}.$$

The deviation is not less than

$$\frac{L'/2 - n'/2}{n'} = \frac{2m - (2m - 1)}{2(2m - 1)} \geq \varepsilon.$$

3. Finally, assume that, in all segments of \mathbf{t}' , at least one-half of digits are zeros.

This case is entirely similar to the previous one (with changing zeros to ones).

Thus, we have proved that there are no (n, ε) -random generators with respect to the constructed set of rules. Now we should estimate the number of rules. It is equal to

$$m \binom{L'/m}{L'/2m} \leq m \sqrt{\frac{2}{\pi L'/m}} 2^{L'/m}.$$

It remains to note that $m\sqrt{2m/\pi L'} < 1/4$ for $n \geq \varepsilon^{-3}$ and $\varepsilon < 1/20$, and $L'/m < 2 + 4n\varepsilon(1 + 5\varepsilon)$ for $\varepsilon < 1/20$. \triangle

Theorem 1 gives us a lower bound and Theorem 2 gives an upper bound for the maximal number τ such that, for each L and each set of rules with complexity less than τ , there is at least one (n, ε) -random generator of length L . Since $d(n, \varepsilon) = 2n\varepsilon^2 \text{lb } e$ is far less than $4n\varepsilon$ for small ε , Kolmogorov tried to remove the discrepancy between the power of ε in the bounds. As is noted in [1], he had not succeeded. In the preface to the translation of [1] into Russian (see [6]), Kolmogorov reminded the reader that the problem was waiting to be solved.

The lower bound obtained by Kolmogorov turns out to be practically sharp (even for nonadaptive rules).

Theorem 3. Consider arbitrary numbers $L \geq 2$ (the length of a sequence), $\varepsilon \in (0, 1/3)$ (the deviation), and $n \in [2\varepsilon^{-3} \text{lb } L, L/2]$ (the length of selected subsequences). There exists a set \mathcal{R}_L of nonadaptive rules such that its complexity is less than

$$d(n, \varepsilon) \frac{1 + \varepsilon}{1 - n/(L - 1)}$$

and there does not exist an (n, ε) -random generator with respect to \mathcal{R}_L .

Proof. We prove the existence of such a set of rules using a probabilistic method, as in Theorem 1. But now we consider a probability distribution on rules and show that the probability of the event “there is an (n, ε) -random generator with respect to a set of rules \mathcal{R}_L ” is less than 1 .⁷

We will look for a required set of rules among nonadaptive rules that select subsequences of length exactly n ; i.e., a rule is determined by an n -element subset of the set $1, \dots, L$. The number of such rules is $\binom{L}{n}$; we define the uniform distribution on them.

⁷ There are many examples in mathematics where an implicit construction gives better estimation than all known explicit constructions. The book [7], pp. 257–261 and 273–280, contains a detailed discussion of Kolmogorov’s proof of a special case of Shannon’s theorem about noise-resistant coding, and the significance of “noneffectiveness” in reasoning is stressed.

Take a sequence \mathbf{t} of length L . Let us bound from below the probability that \mathbf{t} is not an (n, ε) -random generator with respect to a randomly chosen rule r , i.e., the deviation in $r[\mathbf{t}]$ is not less than ε . We assume that the number of zeros in \mathbf{t} is not less than the number of ones (the opposite case is considered symmetrically).

At first, we assume that the numbers $L/2$ and $n(1/2 + \varepsilon)$ are integers; in the end of the proof, we explain what should be done in other cases. Since we need a lower bound, it is sufficient to estimate the probability of the deviation such that the fraction of zeros in a selection exceeds $1/2$ by at least ε ; obviously, this probability is minimal if \mathbf{t} contains the same number of zeros and ones. Moreover, it is sufficient to estimate the probability of the deviation which is exactly equal to ε . Thus, we estimate the probability of selecting a subsequence of length n containing exactly $(1/2 + \varepsilon)n$ zeros from a sequence of $L/2$ zeros and $L/2$ ones. Evidently, this probability is equal to

$$\frac{\binom{L/2}{(\frac{1}{2} - \varepsilon)n} \binom{L/2}{(\frac{1}{2} + \varepsilon)n}}{\binom{L}{n}}$$

(we assume that $\varepsilon < 1/2$ and $n \leq L/2$).

As in Theorem 1, we use the following inequality arising from the Stirling formula:

$$\frac{e^{k \cdot h(j/k)}}{\sqrt{8j(k-j)/k}} \leq \binom{k}{j} \leq \frac{e^{k \cdot h(j/k)}}{\sqrt{2\pi j(k-j)/k}}.$$

We find that the probability is not less than

$$e^{L(\frac{1}{2}h((1-2\varepsilon)\gamma) + \frac{1}{2}h((1+2\varepsilon)\gamma) - h(\gamma))} \times \frac{\sqrt{2\pi}/4}{\sqrt{(1-4\varepsilon^2)(1-4\varepsilon^2\frac{\gamma^2}{(1-\gamma)^2})(1-\gamma)n}},$$

where $\gamma = n/L$.

It can easily be checked that the second factor in this bound is greater than $1/\sqrt{en}$.

Differentiating twice, we can verify that, for $\varepsilon \leq 1/\sqrt{8}$ and $\gamma \leq 1/2$, we have

$$\frac{1}{2}h((1-2\varepsilon)\gamma) + \frac{1}{2}h((1+2\varepsilon)\gamma) - h(\gamma) \geq -\frac{2\gamma\varepsilon^2}{1-\gamma}(1+4\varepsilon^2/3). \quad (1)$$

Finally, we get that the required probability is greater than e^{-K} , where

$$K = \frac{2n\varepsilon^2}{1-\gamma}(1+4\varepsilon^2/3) + \frac{1}{2}(1 + \ln n).$$

Thus, the probability that one rule does not select a subsequence from a given sequence such that the deviation is not less than ε turns out to be less than $(1 - e^{-K})$. Now let us independently take N random rules.⁸ The probability that a fixed sequence \mathbf{t} is an (n, ε) -random generator with respect to this set of rules is less than

$$(1 - e^{-K})^N < e^{-Ne^{-K}}$$

⁸ Some of them may coincide.

(here we use the inequality $\left(1 - \frac{1}{x}\right)^x < e^{-1}$, which is true for all $x > 1$). Multiplying by the number of sequences of length L , we get a strict upper bound for the probability that there exists at least one (n, ε) -random generator with respect to this set of rules, namely,

$$2^L e^{-Ne^{-K}} = e^{L \ln 2 - Ne^{-K}}.$$

The last expression is not greater than 1 if $N = \lceil e^K L \ln 2 \rceil < e^K L$, and the complexity of the set is not greater than

$$\frac{2n\varepsilon^2}{1 - n/L} (1 + 4\varepsilon^2/3) \text{lb } e + 2 \text{lb } L.$$

It can easily be checked that

$$d(n, \varepsilon) \frac{1 + \varepsilon}{1 - n/L} > \frac{2n\varepsilon^2}{1 - n/L} (1 + 4\varepsilon^2/3) \text{lb } e + 2 \text{lb } L \tag{2}$$

for $\varepsilon < 1/3$ and $n \geq 2\varepsilon^{-3} \text{lb } L$.

Now let us return to our assumption that the numbers $L/2$ and $n(1/2 + \varepsilon)$ are integers. If the number L is odd, we can make the same reasoning for the beginnings of original sequences of length $L - 1$, and in the final formula n/L is replaced by $n/(L - 1)$. The number $n(1/2 + \varepsilon)$ can always be converted into an integer by changing ε to $\varepsilon' \geq \varepsilon$ such that $\varepsilon' - \varepsilon < 1/n$. The theorem assumption easily implies that, if $\varepsilon < 1/3$, then $\varepsilon' < 1/\sqrt{8}$. Thus, inequality (1) holds if we replace ε by ε' , and inequality (2) holds if we do not change its left-hand side and replace ε on the right-hand side by ε' . \triangle

Remark 2. In the proof of Theorem 2, the set \mathcal{R}_L is polynomially computable, i.e., an algorithm have been constructed that, given the number of a rule from \mathcal{R}_L and an argument of the rule, computes (within polynomial time) the result of applying the rule to the argument. As for the set \mathcal{R}_L from Theorem 3, this is likely to be false since the proof is probabilistic.

Open problem. Is it possible to improve Theorem 3 in order to make the set \mathcal{R}_L polynomially computable (informally speaking, so that \mathcal{R}_L would become an explicitly defined set)?

Theorem 3'. Consider a natural $L \geq 2$ (the length of a sequence), a rational $\varepsilon \in (0, 1/3)$ (the deviation), and a natural $n \in [2\varepsilon^{-3} \text{lb } L, L/2]$ (the length of selected subsequences). There does not exist an (n, ε) -random generator with respect to the set \mathcal{R}_L consisting of all nonadaptive rules such that their conditional entropy under known L is less than

$$d(n, \varepsilon) \frac{1 + \varepsilon}{1 - n/L} + C.$$

Proof. By the previous theorem, there is no (n, ε) -random generator of length L with respect to a certain set of nonadaptive rules of complexity less than $d(n, \varepsilon) \frac{1 + \varepsilon}{1 - n/(L - 1)}$. Let us show that, given L , n , and ε , a set of rules with this property can be constructed algorithmically.

Indeed, given a set \mathcal{R}_L of nonadaptive rules and a sequence \mathbf{t} of length L , we can effectively determine whether \mathbf{t} is an (n, ε) -random generator with respect to \mathcal{R}_L (it is sufficient to apply each rule from \mathcal{R}_L to \mathbf{t} and to calculate the deviation). Examining each sequence of length L , we can verify whether there exists an (n, ε) -random generator with respect to \mathcal{R}_L . Examining all sets of nonadaptive rules with a given cardinality, we find the required set (if there are several sets with this property, we take the first set in our enumeration).

Given L , n , and ε , this set can be enumerated by a program whose entropy depends on the programming language only. Proposition 2 (relativized with respect to L , n , and ε) implies that

the conditional entropy of each of these rules does not exceed the complexity of the set plus a constant. If we add rules (all the remaining with entropy less than $d(n, \varepsilon) \frac{1 + \varepsilon}{1 - n/(L - 1)} + C$), the property (nonexistence of an (n, ε) -random generator) cannot be violated. Finally, note that the difference

$$d(n, \varepsilon) \frac{1 + \varepsilon}{1 - n/(L - 1)} - d(n, \varepsilon) \frac{1 + \varepsilon}{1 - n/L}$$

is less than 1. \triangle

2. COMPARISON OF UNIVERSAL AND FREQUENCY RANDOMNESS

2.1. Philosophical Motivation

It is traditional in mathematical statistics to use tests in order to clarify ideas of randomness. Assume that we have a finite sequence of cards (see the paragraph before Definition 1), their faces down, each containing a binary digit. In this formulation, to each sequence of cards we assign a certain (unknown to us) sequence of zeros and ones. Informally, nonrandomness (with respect to the uniform distribution) of this sequence means that it is possible to make a nontrivial prediction about its behavior. If one offers a sequence and, after the cards are opened, the sequence on them coincides with the offered one, then this sequence should be considered as maximally nonrandom. If one offers a set containing a relatively small number of elements and, after the cards are opened, the sequence on them belongs to the offered set, then this sequence should be considered as nonrandom too (having a bigger “measure of nonrandomness” for the smaller cardinality of the set). Such sets are usually called tests.

Practically, sometimes it is convenient to consider tests of a certain special kind. Frequency tests are especially important. To each rule r selecting a subsequence of length n from any sequence of length L and to each deviation ε , we assign a set⁹ (a frequency test) containing sequences \mathbf{t} such that $\ell(\mathbf{t}) = L$, $\ell(r[\mathbf{t}]) = n$, and the fraction of zeros in $r[\mathbf{t}]$ differs from $1/2$ by at least ε . We could say that the universal concept of randomness is reduced to the frequency one if any test U could be covered by a small collection of frequency tests F_1, \dots, F_m . In this case, it is important that m is very small compared with the cardinality of U , and the cardinality of each F_i is not much greater than the cardinality of U . Theorem 4 gives particular estimations of this kind. Theorem 5 shows that the estimations from Theorem 4 cannot be improved substantially. Theorem 6 implies that frequency tests F_i cannot be defined with the help of a more restricted (than in Theorem 4) class.

It is worth saying that the combinatorial results from the first part of this article can also be interpreted as an analysis of the possibility to cover a set U by frequency tests F_1, \dots, F_m . But in this case, as U , we consider the set of all sequences.

Under the combinatorial approach, as in classical probability theory, we cannot precisely define what is “one (who does not know a sequence) offers a set of small cardinality that contains this sequence.” Under the algorithmical approach, in place of such a set we take the set of all sequences of small (compared with their length) entropy. In other words, one (“universal”) test is studied. An advantage is that we can speak about a complexity measure of an individual sequence. A disadvantage is that the universal test is not defined explicitly (it is enumerable but not decidable).

As in the first part of the article, algorithmical theorems have the numbers of parallel combinatorial theorems with the prime.

⁹ In Definition 3, such sets are referred to as regular.

2.2. The Combinatorial Approach

Definition 4. The *deficiency* $\Delta(S)$ of a set S consisting of binary sequences of a certain length L is the difference between L and the complexity of the set S , i.e., $\Delta(S) = L - \text{lb } |S|$. The *specific deficiency* $\delta(S)$ is the value $\delta(S) = \Delta(S)/L$.

The specific deficiency characterizes a measure of nontriviality of a set if it is used as a test.

Theorem 4. Let $\delta \in (0, e^{-e^8})$ and consider a natural $L \geq (1/\delta)^5$. Consider sets of binary sequences of length L . For an arbitrary set S of specific deficiency $\geq \delta$, there is a family of regular sets of specific deficiency greater than

$$\delta' = \frac{\delta}{\ln(1/\delta)}(1 - \beta)$$

that covers S and contains at most $Le^{1/\delta'}$ sets (as β , we can take $\frac{2 \ln \ln(1/\delta)}{\ln(1/\delta)}$).

Proof. For a set S of sequences of length L , let us consider the following game. Mathematician and Nature make L moves in turn. At the i th move, Mathematician places an $x_i \in [0, 1]$ bet on a digit 0 or 1, and Nature chooses an element $t_i \in \{0, 1\}$ so that, after L moves, the sequence \mathbf{t} would belong to S . At first, the capital of Mathematician is equal to zero; then, at each move, it increases by the value of the bet if Mathematician have guessed the next digit t_i , and decreases by the same value otherwise. The capital may be negative, and for this reason the game is called the *game "on credit."* Let us show that, for any set S , there is a strategy of Mathematician that allows him to win not less than $\Delta(S) \ln 2$.

Let us introduce some notation. If \mathbf{s} is an extension of a sequence \mathbf{t} (in other words, \mathbf{t} is a prefix of \mathbf{s}), we write $\mathbf{s} \sqsupset \mathbf{t}$ (it is possible that $\mathbf{s} = \mathbf{t}$). The set $S_{\mathbf{t}} = \{\mathbf{s} \mid \mathbf{s} \sqsupset \mathbf{t}, \mathbf{s} \in S\}$ is the set of all extensions of \mathbf{t} belonging to S . Let $\mathbf{t}_{1:i}$ be the beginning of a sequence \mathbf{t} obtained before the $(i + 1)$ st move.

Mathematician's strategy is as follows. If $|S_{\mathbf{t}_{1:i}0}| \geq |S_{\mathbf{t}_{1:i}1}|$ (i.e., if most of the extensions of $\mathbf{t}_{1:i}$ belonging to S begin with 0), then Mathematician bets on 0; otherwise, he bets on 1. The value x of the bet is determined by the equation $\frac{1+x}{2} = \frac{|S_{\mathbf{t}_{1:i}0}|}{|S_{\mathbf{t}_{1:i}}|}$ or $\frac{1+x}{2} = \frac{|S_{\mathbf{t}_{1:i}1}|}{|S_{\mathbf{t}_{1:i}}|}$ respectively.

Let us show that the value $K_i + \Delta_i \ln 2$ does not decrease for this strategy, where K_i is Mathematician's capital after the i th move and $\Delta_i = \Delta(S_{\mathbf{t}_{1:i}}) - i$ is the deficiency of the set of all possible finishes of the game. This implies that Mathematician's gain K_L is not less than $\Delta(S) \ln 2$, since at the beginning $K_0 = 0$, $\Delta_0 = \Delta(S)$, and after the last move $\Delta(S_{\mathbf{t}_{1:L}}) - L = L - \text{lb } 1 - L = 0$.

We should prove that the sum $(K_{i+1} - K_i) + (\Delta_{i+1} - \Delta_i) \ln 2$ is always nonnegative. The second term is equal to $-\ln 2 - \ln \frac{|S_{\mathbf{t}_{1:(i+1)}}|}{|S_{\mathbf{t}_{1:i}}|}$. The first term is equal to the bet x or $(-x)$. In the first case, we have $x - \ln 2 - \ln \frac{1+x}{2} = x - \ln(1+x) \geq 0$ (by the well-known inequality for the logarithm). In the second case, we have $-x - \ln 2 - \ln \frac{1-x}{2} = -x - \ln(1-x) \geq 0$. Thus, we obtain that the gain is not less than $\delta L \ln 2$, where $\delta = \delta(S)$.

We will use this strategy for constructing a set of rules that give a cover of S by regular sets. The idea is the following: given S , we construct regular sets of large deficiency that cover "almost all" of S ; for the rest of S , we repeat the same construction. We estimate the number of covered sequences using our winning strategy (it will be slightly modified in order to decrease the number of possible bet values).

First of all, we modify the strategy. To this end, let us take the number B nearest to $\ln(1/\delta)$ from above such that $M = \frac{B}{\delta \ln 2}$ is natural, and round off all bet values to the nearest number of

the form k/M (where $k \in \mathbb{N}$). Note that the decrease in the gain at each move is less than $1/M$, and the modified strategy wins more than $\delta L \ln 2(1 - 1/B)$. Now nonzero bets can run over M different values only.

To each set of admissible values of bets (subset of $\{1/M, 2/M, \dots, 1\}$) and to each binary digit (0 or 1), assign the following rule. It reads a sequence digit by digit and applies the winning strategy to the part read previously; the next digit is selected to the subsequence if the strategy prescribes to bet this value on this digit. The number of rules of this form is not greater than $2 \cdot 2^M$. For each n , apply n -normalization to each rule. After that, the number of normal rules is not greater than $2L \cdot 2^{B/(\delta \ln 2)} < 2L \cdot 2^{1+\text{lb}(1/\delta)/\delta} = 4L(1/\delta)^{1/\delta}$.

For almost all sequences $\mathbf{t} \in S$, we will find a normal rule r of the form described above and a number ε such that the regular set $A_{r,\varepsilon}$ contains \mathbf{t} and the deficiency of $A_{r,\varepsilon}$ is large enough. The inequality $\varepsilon_1 > \varepsilon_2$ implies that $A_{r,\varepsilon_1} \subseteq A_{r,\varepsilon_2}$; therefore, for each r , we can take the least ε such that the deficiency of $A_{r,\varepsilon}$ is large enough. Most part of S turns out to be covered by not more than $4L(1/\delta)^{1/\delta}$ sets, and for a small exceptional set $\tilde{S} \subset S$ we repeat the whole construction; a recursive calculation shows that the total number of sets in the cover is not too large.

Let us take a sequence $\mathbf{t}_0 \in S$. The strategy wins more than $\delta L \ln 2(1 - 1/B)$ on it, betting on zero and on one. If we allow betting on 0 or on 1 only, then at least in one case the gain is greater than $\frac{1}{2}\delta L \ln 2(1 - \frac{1}{B})$. Let us choose the corresponding digit (assume that it is 0) and in the sequel consider bets on 0 only (we will ignore bets on 1).

For any beginning \mathbf{t} of a sequence from S , denote by $d_i(\mathbf{t})$ the difference between the numbers of wins and losses on bets of value i/M (bets on 0 only) and denote by $n_i(\mathbf{t})$ the total number of such bets (when Nature moves in accordance with \mathbf{t}).

The gain on \mathbf{t}_0 is equal to

$$\sum_{i=1}^M \frac{i}{M} d_i(\mathbf{t}_0).$$

Let us exclude all bets (of value i/M) such that $n_i(\mathbf{t}_0) \leq L/M^2$. The gain on one move is not greater than the value of the bet, that is, 1; hence, after the bets have excluded, the total gain decreases by at most $1 \cdot \frac{L}{M^2} \cdot M = L/M = \delta L \ln 2/B$. Thus,

$$\sum_{i: n_i(\mathbf{t}_0) > \frac{L}{M^2}} \frac{i}{M} d_i(\mathbf{t}_0) > \frac{1}{2} \delta L \ln 2 \left(1 - \frac{3}{B}\right). \quad (3)$$

For each $\mathbf{t}_1 \in S$, consider the set

$$S(\mathbf{t}_1, i) = \left\{ \mathbf{t} \in S \mid n_i(\mathbf{t}) = n_i(\mathbf{t}_1), d_i(\mathbf{t}) \leq \frac{i}{M} n_i(\mathbf{t}_1) \left(1 - \frac{1}{B}\right) \right\}.$$

Claim 1. *If $n_i(\mathbf{t}_1) > L/M^2$, then $|S(\mathbf{t}_1, i)| < |S|e^{-L/2B^2M^4}$.*

We will prove this claim later, and now we continue the proof of the theorem.

Put

$$\tilde{S} = \bigcup_{\substack{\mathbf{t}_1 \in S \\ i: n_i(\mathbf{t}_1) > L/M^2}} S(\mathbf{t}_1, i).$$

Let us note that a set $S(\mathbf{t}_1, i)$ is uniquely determined by two parameters, i and $n_i(\mathbf{t}_1)$, which can range over M and L different value respectively. Therefore,

$$|\tilde{S}| < LM|S|e^{-L/2B^2M^4}.$$

Let us consider two cases:

1. $\mathbf{t}_0 \in \tilde{S}$;
2. $\mathbf{t}_0 \notin \tilde{S}$.

In the first case, \mathbf{t}_0 belongs to the exceptional set, and we do not try to cover this set by regular sets of the form described above. Instead, we repeat our construction for the set \tilde{S} with the same bound on δ (the deficiency of \tilde{S} is greater than the deficiency of S). As a result, we obtain the following chain of decreasing sets: $S^{(0)} = S$, $S^{(1)} = \tilde{S}^{(0)}$, $S^{(2)} = \tilde{S}^{(1)}$, \dots . This chain is finite since, at each step, the cardinality (a natural number) of the corresponding set becomes at least $e^{L/2B^2M^4}/LM$ times as small. The length of the chain is not greater than

$$1 + \frac{\ln |S|}{\ln(e^{L/2B^2M^4}/LM)} < 1 + \frac{L \ln 2}{L/2B^2M^4 - \ln L - \ln M}.$$

Using bounds for L and δ from the theorem assumption, we obtain that the chain length is less than $\frac{25 \ln^6(1/\delta)}{\delta^4}$. For each element of the chain, the number of rules is not greater than $4L(1/\delta)^{1/\delta}$; therefore, the total number of rules is less than $Le^{1/\delta'}$.

Consider the second case now. Let us construct a regular set covering \mathbf{t}_0 , that is, point out a rule (a collection of bets values) and a deviation.

For brevity, put $n_i = n_i(\mathbf{t}_0)$ and $d_i = d_i(\mathbf{t}_0)$. The condition $\mathbf{t}_0 \notin \tilde{S}$ implies in particular that, if $n_i > L/M^2$, then $\mathbf{t}_0 \notin S(\mathbf{t}_0, i)$, or, equivalently, $d_i > \frac{i}{M}n_i(1 - \frac{1}{B})$. From this and by (3), we get

$$\sum_{i: n_i > \frac{L}{M^2}} \frac{d_i^2}{n_i} > \sum_{i: n_i > \frac{L}{M^2}} d_i \frac{i}{M} \left(1 - \frac{1}{B}\right) > \frac{1}{2} \delta L \ln 2 \left(1 - \frac{4}{B}\right).$$

Consider a regular set that is determined by a collection of bets $J \subseteq \{i \mid n_i > L/M^2\}$, the length of a selected subsequence $n = \sum_{i \in J} n_i$, and the deviation $\varepsilon = \frac{1}{2n} \sum_{i \in J} d_i$. This set contains \mathbf{t}_0 (recall that d_i is the difference between the numbers of wins and losses). By the Chernoff bound, we have that the specific deficiency of the set is not less than $2n\varepsilon^2 \ln e/L$. The proof will be completed if we show that this value is greater than $\frac{\delta}{\ln(1/\delta)}(1 - \beta)$ for some J .

Assume the contrary, i.e., for any $J \subseteq \{i \mid n_i > L/M^2\}$, we have

$$\frac{\left(\sum_{i \in J} d_i\right)^2}{\left(\sum_{i \in J} n_i\right) 2L \ln 2} \leq \frac{\delta}{\ln(1/\delta)}(1 - \beta).$$

In other notation,

$$\left(\sum_{i \in J} \frac{d_i}{n_i} n_i\right)^2 \leq Z \left(\sum_{i \in J} n_i\right),$$

where $Z = \frac{\delta}{\ln(1/\delta)}(1 - \beta) 2L \ln 2$.

Claim 2. If $\sum_{i \in J} \frac{d_i}{n_i} n_i \leq \sqrt{Z \sum_{i \in J} n_i}$ for all J , then

$$\sum \frac{d_i^2}{n_i} \leq Z + \frac{Z}{4} \left(\ln \sum n_i - \ln Z\right),$$

where the sum is taken over all i such that $n_i > L/M^2$.

This claim will also be proved later.

Now, if we recall the bound for $\sum d_i^2/n_i$ obtained previously, we get

$$\frac{1}{2}\delta L \ln 2 \left(1 - \frac{4}{B}\right) < Z + \frac{Z}{4}(\ln L - \ln Z)$$

(here we take into account that $\sum n_i \leq L$). After simple transformations, we get

$$\left(1 - \frac{4}{\ln(1/\delta)}\right) < (1 - \beta) \left(1 + \frac{\ln \ln(1/\delta) + 4 - \ln \ln 4 - \ln(1 - \beta)}{\ln(1/\delta)}\right),$$

which is wrong for $\beta = \frac{2 \ln \ln(1/\delta)}{\ln(1/\delta)}$ and $\delta < e^{-e^8}$. Δ

It remains to prove the two claims used.

Proof of Claim 2. We should obtain an estimation of the sum $\sum \frac{d_i^2}{n_i} = \sum \left(\frac{d_i}{n_i}\right)^2 n_i$ from a given estimation of the sum $\sum \frac{d_i}{n_i} n_i$. Let us change these sums to integrals of step functions. Since the order of indices i is not important, assume that $i = 1, \dots, I$ and $\frac{d_i}{n_i}$ monotonically nonincreases. Consider a nonincreasing function f on the half-open interval $[0, \sum_{i=1}^I n_i)$ defined by the equality

$$f(x) = \frac{d_k}{n_k} \quad \text{if} \quad N_{k-1} \leq x < N_k,$$

where $N_k = \sum_{i=1}^k n_i$. Obviously, these sums over $i = 1, \dots, k$ are equal to integrals from 0 to N_k of f and f^2 respectively. By the claim condition, for every k we have

$$\int_0^{N_k} f(x) dx \leq \sqrt{ZN_k}.$$

Using the convexity of the square root, we can extend this inequality to arbitrary values of the upper limit of the integral. Indeed, for $N_k \leq u \leq N_{k+1}$, we get

$$\int_0^u f(x) dx = \int_0^{N_k} f(x) dx + \frac{d_{k+1}}{n_{k+1}}(u - N_k).$$

For $u = N_{k+1}$, the inequality implies that for all u we have

$$\frac{d_{k+1}}{n_{k+1}}(u - N_k) \leq \left(\sqrt{ZN_{k+1}} - \int_0^{N_k} f(x) dx \right) \frac{u - N_k}{N_{k+1} - N_k}.$$

Therefore,

$$\int_0^u f(x) dx \leq \sqrt{ZN_{k+1}} \frac{u - N_k}{N_{k+1} - N_k} + \sqrt{ZN_k} \left(1 - \frac{u - N_k}{N_{k+1} - N_k}\right) \leq \sqrt{Zu}.$$

We want to represent the right-hand side as an integral of a bounded nonincreasing function g . Note that $f(x) \leq 1$ since $d_k \leq n_k$. Put

$$g(x) = \begin{cases} 1 & \text{if } x < Z, \\ \frac{\sqrt{Z}}{2\sqrt{x}} & \text{if } x \geq Z. \end{cases}$$

Let us remark that, if $u \geq Z$, then $\int_0^u g(x) dx = \sqrt{Zu}$. Hence,

$$\int_0^u (g(x) - f(x)) dx \geq 0$$

for all $u \in [0, N_I]$. It remains to show that $\int_0^u (g^2(x) - f^2(x)) dx \geq 0$.

The second mean-value theorem for integrals asserts that, for any integrable function $\psi(x)$ and any nonnegative nonincreasing function $\varphi(x)$, there exists a point $\xi \in [a, b]$ such that

$$\int_a^b \varphi(x)\psi(x) dx = \varphi(a) \int_a^\xi \psi(x) dx.$$

The functions $\psi(x) = g(x) - f(x)$ and $\varphi(x) = g(x) + f(x)$ satisfy conditions of this theorem; hence,

$$\int_0^u (g^2(x) - f^2(x)) dx = (g(0) + f(0)) \int_0^\xi (g(x) - f(x)) dx \geq 0.$$

Let us verify that $Z < N_I$. In fact,

$$\sum n_i = \sum \frac{n_i^2}{n_i} \geq \sum \frac{d_i^2}{n_i} > \frac{1}{2} \delta L \ln 2 \left(1 - \frac{4}{B}\right) > \frac{\delta}{\ln(1/\delta)} (1 - \beta) 2L \ln 2 = Z$$

(recall that δ is small).

Thus,

$$\sum_i \frac{d_i^2}{n_i} = \int_0^{N_I} f^2(x) dx \leq \int_0^{N_I} g^2(x) dx = Z + \frac{Z}{4} (\ln N_I - \ln Z). \quad \triangle$$

Proof of Claim 1. Consider the set $S(\mathbf{t}_1, i)$. Note that the condition

$$d_i(\mathbf{t}) \leq \frac{i}{M} n_i(\mathbf{t}_1) \left(1 - \frac{1}{B}\right)$$

may be rewritten in the following form:

$$\left(\frac{1}{2} + \frac{i}{2M}\right) n_i - \frac{in_i}{2BM} \geq \frac{n_i + d_i}{2}.$$

To estimate the cardinality of $S(\mathbf{t}_1, i)$, we introduce two auxiliary probability measures, Pr_S and Pr . The measure Pr_S is the uniform distribution on S , that is, $\text{Pr}_S(X) = |X \cap S|/|S|$. The measure Pr is the Bernoulli measure with the probability of zero equal to $\frac{1}{2} + \frac{i}{2M}$.

Let $X_{\nu,\sigma}$ be the set of all sequences of length ν that contain at most σ zeros. Assume that $\mathbf{t} \in \{0, 1\}^{\leq L}$. Let $z_i(\mathbf{t})$ be the number of wins on bets on zero of value i/M when Nature moves in accordance with the sequence \mathbf{t} and Mathematician moves in accordance with the (modified) strategy. Note that $z_i(\mathbf{t}) = (n_i(\mathbf{t}) + d_i(\mathbf{t}))/2$. Assuming that $S_{\mathbf{s}} \neq \emptyset$, let us prove the inequality¹⁰

$$\forall \nu \forall \sigma \quad \Pr_S\{n_i(\mathbf{t}) = \nu + n_i(\mathbf{s}), z_i(\mathbf{t}) \leq \sigma + z_i(\mathbf{s}) \mid \mathbf{t} \sqsupset \mathbf{s}\} \leq \Pr(X_{\nu,\sigma})$$

by descent induction on the length of \mathbf{s} from L to 0.

Assume that $\nu < 0$ or $\sigma < 0$; then the probability on the left-hand side equals 0. Now assume that $\nu \geq 0$ and $\sigma \geq 0$.

Let $\ell(\mathbf{s}) = L$, then

if $\nu = 0$, then the probability on the right-hand side equals 1;

if $\nu \neq 0$, then the probability on the left-hand side equals 0.

The induction base is proved.

To prove the induction step, let us consider two cases. For brevity, here we use the following notation:

$$P(\mathbf{s}, \nu, \sigma) = \Pr_S\{n_i(\mathbf{t}) = \nu + n_i(\mathbf{s}), z_i(\mathbf{t}) \leq \sigma + z_i(\mathbf{s}) \mid \mathbf{t} \sqsupset \mathbf{s}\},$$

$$\Pr_S\{\mathbf{t} \sqsupset \mathbf{s}0 \mid \mathbf{t} \sqsupset \mathbf{s}\} = \frac{|S_{\mathbf{s}0}|}{|S_{\mathbf{s}}|} = \frac{1+x}{2}.$$

1. In position \mathbf{s} , the strategy does not place a bet of value i/M on zero. Then

$$P(\mathbf{s}, \nu, \sigma) = \frac{1+x}{2}P(\mathbf{s}0, \nu, \sigma) + \frac{1-x}{2}P(\mathbf{s}1, \nu, \sigma)$$

$$\leq \left(\frac{1+x}{2} + \frac{1-x}{2}\right) \Pr(X_{\nu,\sigma}) = \Pr(X_{\nu,\sigma}).$$

2. In position \mathbf{s} , the strategy places a bet of value i/M on zero. Recall that in this case $i/M \leq x$. Then

$$P(\mathbf{s}, \nu, \sigma) = \frac{1+x}{2}P(\mathbf{s}0, \nu-1, \sigma-1) + \frac{1-x}{2}P(\mathbf{s}1, \nu-1, \sigma)$$

$$\leq \frac{1+x}{2} \Pr(X_{\nu-1, \sigma-1}) + \frac{1-x}{2} \Pr(X_{\nu-1, \sigma}).$$

On the other hand,

$$\Pr(X_{\nu,\sigma}) = \frac{1+i/M}{2} \Pr(X_{\nu-1, \sigma-1}) + \frac{1-i/M}{2} \Pr(X_{\nu-1, \sigma}).$$

It is obvious that $\Pr(X_{\nu-1, \sigma}) \geq \Pr(X_{\nu-1, \sigma-1})$. So, reducing the weight of the lesser term (from $(1+x)/2$ to $(1+i/M)/2$) and appropriately enlarging the weight of the greater one, we enlarge the sum. Hence we get the desired inequality $P(\mathbf{s}, \nu, \sigma) \leq \Pr(X_{\nu,\sigma})$.

Thus,

$$|S(\mathbf{t}_1, i)|/|S| = \Pr_S\{n_i(\mathbf{t}) = \nu, z_i(\mathbf{t}) \leq \sigma\} \leq \Pr(X_{\nu,\sigma}),$$

where $\sigma = (1/2 + i/2M)\nu - i\nu/2BM$ and $\nu = n_i(\mathbf{t}_1)$.

¹⁰ Here $\Pr_S\{U \mid V\}$ denotes the probability that U is true under the condition that V is true, that is, $\Pr_S(\{\mathbf{t} \mid U, V\})/\Pr_S(\{\mathbf{t} \mid V\})$.

It follows from the Chernoff bound that

$$\Pr(X_{\nu,\sigma}) \leq e^{-2(\frac{i}{2BM})^2\nu}.$$

Taking into account that $i \geq 1$ and $n_i(\mathbf{t}_1) > L/M^2$, we obtain the final bound

$$|S(\mathbf{t}_1, i)| < |S|e^{-\frac{L}{2B^2M^4}}. \quad \Delta$$

Remark 3. It can be proved that in Theorem 4 (in contrast to Theorem 3) we cannot restrict ourselves to families of regular sets with the same deviation ε .

Remark 4. The proof of Theorem 4 uses regular sets generated by monotonic rules only. Theorem 6 shows that using regular sets generated by nonadaptive rules is not sufficient.

Theorem 4 gives a bound for the specific deficiency of covering regular sets. The following theorem shows that this bound is sharp.

Theorem 5. *Let $\delta \in (0, e^{-e^8})$ and consider a natural $L \geq (1/\delta)^5$. Consider sets of binary sequences of length L . There exists a set S of specific deficiency greater than δ that cannot be covered by less than $e^{L\delta^4/70}$ regular sets of specific deficiency not less than*

$$\frac{2\delta}{\ln(1/\delta)}.$$

Proof. Let

$$m = \left\lfloor \left(2 - 4\sqrt{\delta \ln(1/\delta)} \right) / \delta \ln 2 \right\rfloor, \quad k = \lfloor L/2m \rfloor.$$

A sequence of length L is represented as a concatenation of m pairs of sequences having length k and a “remainder” having length $L - 2km < 2m$. In each pair, one sequence corresponds (a more precise explanation will be given later) to the digit 0, and the other corresponds to the digit 1; for the i th pair, these sequences are called the i th 0- and 1-segments of an original sequence \mathbf{t} of length L and are denoted by $\mathbf{t}_{i,0}$ and $\mathbf{t}_{i,1}$ respectively.

For each i from 1 to m , let us define sets $S_{i,0}$ and $S_{i,1}$ of length- k sequences. Let $p_i = \frac{1}{2} + d\frac{1}{\sqrt{i \ln m}}$. The set $S_{i,0}$ consists of sequences with not less than $k(p_i - 1/m)$ zeros and $S_{i,1}$ consists of sequences with not less than $k(p_i - 1/m)$ ones. The set S contains all sequences \mathbf{t} such that $\mathbf{t}_{i,0} \in S_{i,0}$ and $\mathbf{t}_{i,1} \in S_{i,1}$ for each i .

First let us estimate the specific deficiency of S . The cardinality of S is equal to the product of $|S_{i,0}| \cdot |S_{i,1}|$ over all i multiplied by 2^{L-2km} , the number of possible “remainders.” Therefore, from the Chernoff bound (for the uniform distribution on all binary sequences) we obtain

$$|S| \leq 2^{L-2km} \prod_{i=1}^m \left(2^k e^{-2k \left(\frac{1}{\sqrt{i \ln m}} - \frac{1}{m} \right)^2} \right)^2 = 2^{L-2km} 2^{2km} e^{-4k \sum_{i=1}^m \left(\frac{1}{\sqrt{i \ln m}} - \frac{1}{m} \right)^2}.$$

Since $1 - \frac{\sqrt{i \ln m}}{m} \geq 1 - \sqrt{\frac{\ln m}{m}}$ and $\sum_{i=1}^m \frac{1}{i} > \ln m$, we have

$$\sum_{i=1}^m \left(\frac{1}{\sqrt{i \ln m}} - \frac{1}{m} \right)^2 = \sum_{i=1}^m \frac{1}{i \ln m} \left(1 - \frac{\sqrt{i \ln m}}{m} \right)^2 > \frac{\ln m}{\ln m} \left(1 - \sqrt{\frac{\ln m}{m}} \right)^2.$$

Hence we get $|S| < 2^L e^{-4k\left(1-2\sqrt{\frac{\ln m}{m}}\right)}$, and so the specific deficiency of S is greater than

$$\frac{4k}{L \ln 2} \left(1 - 2\sqrt{\frac{\ln m}{m}}\right) > \left(\frac{2}{m \ln 2} - \frac{4}{L \ln 2}\right) \left(1 - 2\sqrt{\frac{\ln m}{m}}\right) > \delta.$$

Take a cover of S consisting of less than $e^{L\delta^4/70}$ regular sets. Let us prove that at least one set in this cover has specific deficiency less than $\frac{2\delta}{\ln(1/\delta)}$.

The plan of the proof is as follows. Consider the set of rules that generate regular sets from the cover. For each rule r , the sequence is called r -typical if the deviation in the selected subsequence is small compared with its length. We will prove that the set S contains a sequence that is typical for all rules under consideration (since the probability with respect to a certain distribution of the opposite event is strictly less than 1). Therefore, one set from the cover (namely, the set containing this typical sequence) has small deficiency.

Let \mathbf{t} be a sequence of length L and r be a normal place-selection rule. To define r -typicalness, assume that the number of ones in the subsequence $r[\mathbf{t}]$ is greater than the number of zeros; in the opposite case, it is necessary to change all ones to zeros and vice versa in the definition below.

Denote by n the length of a selected subsequence $r[\mathbf{t}]$, denote by n_i the number of digits selected by r from the i th 1-segment of the sequence, and denote by n' the total number of digits selected by r from 0-segments of the sequence (i.e., $n \geq \sum_{i=1}^m n_i + n'$ and $n_i \leq k$). The sequence \mathbf{t} is called r -typical if the following conditions hold:

- The number of ones selected from the i th 1-segment is less than $n_i p_i + \sqrt{\frac{kn_i}{m^3}}$;
- Either $n' < k/m$ or, among digits selected from 0-segments, the number of ones is less than the number of zeros.

Let us introduce an auxiliary probabilistic measure on sequences of length L : digits are independent, digits from the i th 0-segment (1-segment) are equal to zero (or, respectively, to one) with probability p_i , and digits from the “remainder” are equal to zero with probability $1/2$.

Let us estimate the probability (with respect to this measure) that a sequence \mathbf{t} (of length L) selected by guess does not belong to the set S :

$$\Pr\{\mathbf{t} \notin S\} \leq \sum_{i=1}^m (\Pr\{\mathbf{t}_{i,0} \notin S_{i,0}\} + \Pr\{\mathbf{t}_{i,1} \notin S_{i,1}\}) \leq 2me^{-2k\left(\frac{1}{m}\right)^2} = e^{-\frac{2k}{m^2}\left(1 - \frac{m^2 \ln 2m}{2k}\right)}.$$

Now let us fix a place-selection rule r and estimate the probability that a sequence \mathbf{t} (of length L) selected by guess is not r -typical (as in the definition of r -typicalness, we assume that the number of ones in the subsequence $r[\mathbf{t}]$ is greater than the number of zeros; the opposite case is considered similarly).

First, let us estimate the probability that, at least for one i from the i th 1-segment, at least $n_i p_i + \sqrt{\frac{kn_i}{m^3}}$ ones are selected. It is not greater than the sum of probabilities of the corresponding events for each i . For a fixed $n_i > 0$, using the Chernoff bound, we get

$$e^{-2\frac{1}{n_i}\left(\sqrt{\frac{kn_i}{m^3}}\right)^2}.$$

Since n_i is not given, we sum up the probabilities over all possible values of $n_i \leq k$. The final upper bound is

$$mke^{-\frac{2k}{m^3}}.$$

Now let us estimate the probability that, from 0-segments, $n' \geq k/m$ digits are selected and the number of ones among these digits is not less than the number of zeros. For a fixed n' , the probability does not exceed

$$e^{-2n'(p_m - \frac{1}{2})^2} \leq e^{-2\frac{k}{m} \frac{1}{m \ln m}}.$$

Summing up over all possible $n' \leq mk$, we finally get the bound

$$mke^{-\frac{2k}{m^2 \ln m}}.$$

We see that the bound on the probability of violating the first condition of r -typicalness is much greater than the bound on that of violating the second condition. Taking into account the symmetric case (change 0 to 1), we obtain that, for a sequence selected by guess, the probability of being non- r -typical is less than $2mke^{-\frac{L}{m^4} + 1} \leq e^{-\frac{L}{m^4} + \ln L + 1}$. Multiplying by the number of regular sets from the cover of S , we find that the probability of “being not typical for at least one rule” is considerably less than 1 when $\delta < e^{-e^8}$ and $L \geq (1/\delta)^5$. If we add a small probability that a sequence does not belong to S , this bound remains to be less than 1. Thus, there exists a sequence $\mathbf{t} \in S$ that is typical for all rules from the cover.

Take the set A from the cover that contains this \mathbf{t} and let r be the corresponding rule. Let us majorize a possible value of the specific deficiency of A , using the r -typicalness of \mathbf{t} . Consider two cases.

1. $n \leq 8k/(\ln m \ln 2)$. The deficiency of A is not less than n ; therefore, the specific deficiency is not less than

$$\frac{n}{L} \leq \frac{4}{m \ln m \ln 2}.$$

2. $n > 8k/(\ln m \ln 2)$. Let us majorize the difference between the number of ones in the subsequence $r[\mathbf{t}]$ and the half of the subsequence length (the argument for zeros is parallel). This difference consists of the corresponding differences for 1-segments, for 0-segments, and for the “remainder.” We bound the difference for the “remainder” by the half of its length, which is less than m . The typicalness implies that, for 0-segments, either the total length of the selected subsequence (and hence the doubled difference) is less than k/m , or the number of zeros in the subsequence is greater than the number of ones. For the same reason, on 1-segments, the difference is less than

$$\sum_{i=1}^m \left(\frac{n_i}{\sqrt{i \ln m}} + \sqrt{\frac{kn_i}{m^3}} \right) = \frac{1}{\sqrt{\ln m}} \sum_{i=1}^m \left(\frac{n_i}{\sqrt{i}} \right) + \sqrt{\frac{k}{m^3}} \sum_{i=1}^m \sqrt{n_i}.$$

Since the square root is a convex function, we get

$$\sum_{i=1}^m \sqrt{n_i} \leq m \sqrt{\sum_{i=1}^m n_i/m} \leq \sqrt{nm}.$$

On one hand, $\sum_{i=1}^m (n_i/\sqrt{i}) \leq n$. On the other hand, to estimate the sum $\sum_{i=1}^m (n_i/\sqrt{i})$ for $n > 4k$, note that, if n_i increases by one and n_{i+1} decrease by one, then the sum will increase. We know

that $n_i \leq k$ and $\sum n_i \leq n$; therefore, increasing the values of the first n_i until they attain k , we obtain

$$\sum_{i=1}^m \left(\frac{n_i}{\sqrt{i}} \right) \leq \sum_{i=1}^{\lceil n/k \rceil} \left(\frac{k}{\sqrt{i}} \right) < k \cdot 2\sqrt{n/k} = 2\sqrt{nk}.$$

Finally, we obtain that the difference between the number of ones in $r[\mathbf{t}]$ and $n/2$ is less than

$$\begin{aligned} m + \frac{k}{2m} + \frac{\sqrt{nk}}{m} + \frac{n}{\sqrt{\ln m}} &< \frac{n}{\sqrt{\ln m}} \left(1 + \frac{\sqrt{\ln^3 m}}{m} \right) && \text{if } n \leq 4k, \\ m + \frac{k}{2m} + \frac{\sqrt{nk}}{m} + \frac{2\sqrt{nk}}{\sqrt{\ln m}} &< \frac{2\sqrt{nk}}{\sqrt{\ln m}} \left(1 + \frac{\sqrt{\ln m}}{m} \right) && \text{if } n > 4k. \end{aligned}$$

Since the regular set A contains the sequence \mathbf{t} , its specific deficiency is not greater than

$$\begin{aligned} \frac{1}{L} \frac{2}{n \ln 2} \frac{n^2}{\ln m} (1 + o(1)) &\leq \frac{4}{m \ln m \ln 2} (1 + o(1)) && \text{if } n \leq 4k, \\ \frac{1}{L} \frac{2}{n \ln 2} \frac{4nk}{\ln m} (1 + o(1)) &\leq \frac{4}{m \ln m \ln 2} (1 + o(1)) && \text{if } n > 4k. \end{aligned}$$

The lower bounds for the binomial coefficients,

$$\binom{n}{j} \geq \frac{e^{n \cdot h(j/n)}}{\sqrt{8j(n-j)/n}}$$

and for the Shannon entropy function (if $\varepsilon \leq 1/\sqrt{12}$),

$$h\left(\frac{1}{2} + \varepsilon\right) \geq \ln 2 - 2\varepsilon^2(1 + \varepsilon^2),$$

show that the terms $o(1)$ in the previous expressions are less than $1/\ln(1/\delta)$. On the other hand, $\ln m > \ln(1/\delta)(1 + 1.05/\ln(1/\delta))$.

Thus, in all cases, at least one set from the cover of S has specific deficiency less than

$$\frac{2\delta}{\ln(1/\delta)}. \quad \triangle$$

Theorem 6. *For any $\sigma > 0$ and any $L \geq 12 + 6\text{lb}(1/\sigma)$, there exists a set of binary sequences of length L and of specific deficiency $\geq 1/3$ that cannot be covered by less than $2^{\sigma L/2}$ nonadaptive regular sets of specific deficiency $\geq \sigma$.*

Proof. Consider an ordered collection of $2^{\lfloor 2L/3 \rfloor}$ sequences of length L (possibly, with repetitions). The specific deficiency of the set of these sequences (without repetitions) is not less than $1/3$. Therefore, to prove the theorem, it is sufficient to find a collection that cannot be covered by any family consisting of less than $2^{\sigma L/2}$ nonadaptive regular sets of specific deficiency not less than σ (we may assume that $\sigma \leq 1$).

The number of collections consisting of $2^{\lfloor 2L/3 \rfloor}$ sequences is equal to

$$\binom{2^L}{2^{\lfloor 2L/3 \rfloor}}.$$

Let us estimate the number of collections that cannot be covered by families of the kind described.

Take any of such families, \mathcal{A}_L . The number of collections that are covered by \mathcal{A}_L is equal to

$$\left| \bigcup \mathcal{A}_L \right|^{2^{\lfloor 2L/3 \rfloor}}.$$

The definition of the family \mathcal{A}_L implies that

$$\left| \bigcup \mathcal{A}_L \right| \leq |\mathcal{A}_L| \max\{|A| : A \in \mathcal{A}_L\} < 2^{\sigma L/2} \cdot 2^{L-\sigma L} = 2^{L-\sigma L/2}.$$

Thus, the number of collections covered by one family is less than

$$2^{(L-\sigma L/2) \cdot 2^{\lfloor 2L/3 \rfloor}}.$$

The family \mathcal{A}_L corresponds to a family of nonadaptive rules \mathcal{R}_L . Choosing various deviation, we get different (generally speaking) families of regular sets corresponding to the same set of rules. However, as the deviation increases, the class of covered collections can only decrease. Thus, one family of $< 2^{\sigma L/2}$ rules can cover less than

$$2^{(L-\sigma L/2) \cdot 2^{\lfloor 2L/3 \rfloor}}$$

collections.

Note that, in essence, a nonadaptive rule is a subset of $\{1, \dots, L\}$; hence, the number of nonadaptive rules is equal to 2^L . The number of families consisting of less than $2^{\sigma L/2}$ nonadaptive rules is less than

$$2^{L \cdot 2^{\sigma L/2}}.$$

Therefore, such families can cover less than

$$2^{L \cdot 2^{\sigma L/2}} \cdot 2^{(L-\sigma L/2) \cdot 2^{\lfloor 2L/3 \rfloor}} = \left(2^L\right)^{2^{\sigma L/2} + (1-\frac{\sigma}{2})2^{\lfloor 2L/3 \rfloor}}$$

collections. To show that there are collections that are not covered, it is sufficient to verify that

$$\left(2^L\right)^{2^{\sigma L/2} + (1-\frac{\sigma}{2})2^{\lfloor 2L/3 \rfloor}} \leq \left(2^L\right)^{2^{\lfloor 2L/3 \rfloor}}.$$

This inequality arises from the inequality $L(2/3 - \sigma/2) \geq 2 - \text{lb } \sigma$, which is provided by $\sigma \leq 1$ and $L \geq 12 + 6 \text{lb}(1/\sigma)$. \triangle

2.3. The Algorithmical Approach

Definition 5. The *deficiency* of a nonempty binary sequence \mathbf{t} of length L is the value $L - K(\mathbf{t} | L)$, where $K(\mathbf{t} | L)$ is the entropy of \mathbf{t} conditional to L . The *specific deficiency* is the deficiency divided by the length. (An additive term from the definition of the entropy divided by the length of a sequence tends to zero as the length tends to infinity. So, in the limit, the notion of specific deficiency is invariant.)

In the sequel, to estimate the deficiency of sequences, we will use the deficiency of auxiliary regular sets. Let us introduce the following notation. Assume that a normal rule r selects a subsequence of length n ; denote the deficiency of a regular set $A_{r,\varepsilon}$ by $D(n, \varepsilon)$. An important property of the deficiency of a regular set is that the deficiency does not depend on L (the length of sequences to which place-selection rules are applied). For a given n , D is a monotone increasing step function of $\varepsilon \in [0, 1/2]$; boundaries of steps are rationals with the denominator $2n$, values of the function are binary logarithms of rationals, and all these rationals computably depend on n .

When ε is small enough and n is large enough compared with $1/\varepsilon$, we have $D(n, \varepsilon) \sim d(n, \varepsilon)$.

At first sight, a natural analog of Theorem 4 would be the following:

Assertion 1. Assume that $\delta > 0$ is small enough and a natural L is large enough. Assume that \mathbf{t} is a binary sequence of length L and of specific deficiency $\geq \delta$. Then there exists a rule r such that

$$K(r|L) \leq \alpha(\delta) \quad \text{and} \quad D(n, \varepsilon)/L \geq \delta'(\delta),$$

where n is the length of $r[\mathbf{t}]$, ε is the deviation in $r[\mathbf{t}]$, α and δ' are positive functions.

But the following theorem trivially implies the negation of Assertion 1.

Theorem 4''. Let α and δ' be positive numbers. For any natural L large enough, there exists a binary sequence \mathbf{t} of length L and of specific deficiency greater than $1/2$ such that, for any rule r ,

$$K(r|L) \leq \alpha \quad \Rightarrow \quad D(n, \varepsilon)/L < \delta' \vee n = 0,$$

where n is the length of $r[\mathbf{t}]$ and ε is the deviation in $r[\mathbf{t}]$.

Proof. Denote by \mathcal{R}_L the family of all rules r such that $K(r|L) \leq \alpha$. Given numbers $[\alpha]$, L , and $|\mathcal{R}_L|$, one can effectively find a list of elements of \mathcal{R}_L . Denote by r_n the n -normalization of a rule r .

Consider the uniform distribution on binary sequences of length L . It follows from the definition of the function D that, for any rule r , the probability of the set $A_{r_n, \varepsilon}$ is equal to $2^{-D(n, \varepsilon)}$.

For each n , take the least step bound ε such that $D(n, \varepsilon) \geq \delta' L$. If such ε exists, it is a rational from $[0, 1/2]$ with the denominator $2n$. Denote by E this collection of pairs $\langle n, \varepsilon \rangle$. Since $|\mathcal{R}_L| < 2^{\alpha+1}$ and $|E| \leq L$, the probability of the union of sets $A_{r_n, \varepsilon}$ for $r \in \mathcal{R}_L$ and $\langle n, \varepsilon \rangle \in E$ does not exceed

$$\sum_{\langle n, \varepsilon \rangle \in E} 2^{-D(n, \varepsilon) + \alpha + 1} \leq 2^{-\delta' L + \alpha + 1 + \text{lb } L}.$$

For L large enough, the last expression is strictly less than 1. Given lists of elements of \mathcal{R}_L and E , using exhaustive search one can find a sequence \mathbf{t} of length L that does not belong to $A_{r_n, \varepsilon}$ for all $r \in \mathcal{R}_L$ and $\langle n, \varepsilon \rangle \in E$.

Now let us try to briefly encode information sufficient for constructing a list of elements of E . Clearly, one can construct E if he knows L and the pair $\langle n, \varepsilon \rangle$ such that the function D achieves its minimum on E at the point $\langle n, \varepsilon \rangle$. Thus, the entropy of the list conditional to L does not exceed $2 \text{lb } L + O(1)$.

The sequence \mathbf{t} is constructed from numbers L , $[\alpha]$, $|\mathcal{R}_L|$, and the collection E ; therefore, $K(\mathbf{t}|L) \leq 2 \text{lb } L + C(\alpha)$. The specific deficiency of \mathbf{t} is equal to $1 - K(\mathbf{t}|L)/L$, which is greater than $1/2$ for large enough L .

Assume that $r \in \mathcal{R}_L$, $n = \ell(r[\mathbf{t}])$, and ε is the deviation in $r[\mathbf{t}]$. Assume that $D(n, \varepsilon)/L \geq \delta'$. Then, by the construction, \mathbf{t} does not belong to $A_{r_n, \varepsilon}$; hence, the deviation in $r[\mathbf{t}]$ is strictly less than ε . This contradiction concludes the proof. \triangle

To repair the analogy with the combinatorial approach, a place-selection rule under the algorithmical approach should be interpreted as not an explicit function but a program that gets the length of a sequence L as an additional input and computes the function r from Definition 1; now, the function r may be defined *not everywhere*.

Definition 6. The *frequency α -deficiency* of a binary sequence \mathbf{t} of length L is the maximum of the value $D(n, \varepsilon)/L$ over rules r that are determined by programs such that $r[\mathbf{t}]$ is defined and the entropy of the program conditional to L is less than α (where n is the length of $r[\mathbf{t}]$ and ε is the deviation in $r[\mathbf{t}]$).

Let us remark that, as α increases, the frequency α -deficiency also increases.

Theorem 4'. Assume that $\delta_1 > 0$ is small enough. Then, for each rational $\delta_0 > 0$, one can effectively find L_0 such that, for any $\delta \in [\delta_0, \delta_1]$ and natural $L \geq L_0$, the following condition holds:

If \mathbf{t} is a binary sequence of length L and of specific deficiency δ , then the frequency α -deficiency of \mathbf{t} is greater than

$$\delta' = \frac{\delta}{\ln(1/\delta)}(1 - B(\delta)),$$

where $\alpha = \frac{4 \ln(1/\delta)}{\delta}$ and $B(\delta) = O\left(\frac{\ln \ln(1/\delta)}{\ln(1/\delta)}\right)$.

Remark 5. We see that this theorem corresponds to Assertion 1.

Proof. Let us construct a rule that selects a subsequence of large deviation from \mathbf{t} . Split the sequence \mathbf{t} into two parts of approximately equal lengths (i.e., their lengths are equal to $\lfloor L/2 \rfloor$ and $\lceil L/2 \rceil$). Split each part again into two parts of approximately equal lengths and so on. Before the $(i + 1)$ st stage of the recursive construction of our rule, we choose one part of \mathbf{t} obtained after i splits (before the first stage, this part is the sequence \mathbf{t} itself). Denote this part by \mathbf{u}_i and denote its complement¹¹ by \mathbf{v}_i . We assume that, before the $(i + 1)$ st stage, exactly the cards from \mathbf{v}_i were turned over, and neither of them was selected. We assume also that the specific deficiency of \mathbf{u}_i conditional to \mathbf{v}_i is greater than $\delta + \lambda(i - 1)$, where λ is a number from the interval $\left(\frac{\delta}{\ln(1/\delta)}, \frac{2\delta}{\ln(1/\delta)}\right)$ such that the number $(1 - \delta + 3\lambda)$ is rational. For large L , the last assumption is obviously true for $i = 0$.

Now let us describe the $(i + 1)$ st stage. Split the sequence \mathbf{u}_i into two parts of approximately equal lengths and denote them by \mathbf{w}_1 and \mathbf{w}_2 . Let $\gamma = \lfloor (1 - \delta + 3\lambda)\ell(\mathbf{w}_1) \rfloor$. Consider two possible cases.

1. Let $K(\mathbf{w}_1 | \mathbf{v}_i) < \gamma$ and $K(\mathbf{w}_2 | \mathbf{v}_i) < \gamma$.

Since the entropy function is enumerable from above, given \mathbf{v}_i and γ , these facts can be discovered within, say, T_1 and T_2 steps of the enumeration. We do not know the numbers T_1 and T_2 . Our aim is to know the greater of them. Without loss of generality, assume that $T_1 \leq T_2$. Then the rule should turn over all cards from \mathbf{w}_2 , selecting nothing. Knowing \mathbf{w}_2 , we can wait for its appearance in the enumeration, and hence we can know T_2 . Now consider the set S of sequences \mathbf{s} of length $\ell(\mathbf{w}_1)$ such that we can discover within less than T_2 steps that the inequality $K(\mathbf{s} | \mathbf{v}_i) < \gamma$ holds. Obviously, $\mathbf{w}_1 \in S$. The set S contains less than 2^γ elements; therefore, its specific deficiency is greater than $1 - \gamma/\ell(\mathbf{w}_1) \geq \delta - 3\lambda$. (If $T_1 \geq T_2$, then we obtain a similar inequality since $\ell(\mathbf{w}_1) \leq \ell(\mathbf{w}_2)$.)

Now apply Theorem 4 to the set S with this bound for the specific deficiency. The proof of Theorem 4 shows that there exists a family of rules \mathcal{R} such that

$$|\mathcal{R}| \leq \left(\frac{1}{\delta - 3\lambda}\right)^{1/(\delta - 3\lambda) + O(1)},$$

and regular sets generated by normalizations of rules from \mathcal{R} cover S . One of these regular sets (denote it by A) contains \mathbf{w}_1 , and its specific deficiency is $\geq \frac{\delta - 3\lambda}{\ln(1/(\delta - 3\lambda))}(1 - \beta)$, where $\beta = O\left(\frac{\ln \ln(1/(\delta - 3\lambda))}{\ln(1/(\delta - 3\lambda))}\right) = O\left(\frac{\ln \ln(1/\delta)}{\ln(1/\delta)}\right)$. Let the normal rule generating A selects from \mathbf{w}_1 a subsequence of length n and deviation ε . Since the deficiency of A is not greater than $D(n, \varepsilon)$, we

¹¹ Here, the complement is the binary encoding of the sequence \mathbf{t} , where digits of \mathbf{u}_i are replaced by a special new symbol.

have

$$\begin{aligned} \frac{D(n, \varepsilon)}{L} &\geq \frac{\delta - 3\lambda}{\ln(1/(\delta - 3\lambda))} (1 - \beta) = \frac{\delta}{\ln(1/\delta)} \left(1 - O\left(\frac{1}{\ln(1/\delta)}\right) \right) (1 - \beta) \\ &= \frac{\delta}{\ln(1/\delta)} \left(1 - O\left(\frac{\ln \ln(1/\delta)}{\ln(1/\delta)}\right) \right). \end{aligned}$$

The place-selection rule that we are constructing stops when Case 1 is realized.

2. Let $K(\mathbf{w}_1 | \mathbf{v}_i) \geq \gamma$ or $K(\mathbf{w}_2 | \mathbf{v}_i) \geq \gamma$.

To be definite, assume that $K(\mathbf{w}_1 | \mathbf{v}_i) \geq \gamma$. By the theorem about the entropy of a pair, we have $K(\mathbf{u}_i | \mathbf{v}_i) = K(\mathbf{w}_1 | \mathbf{v}_i) + K(\mathbf{w}_2 | \mathbf{w}_1, \mathbf{v}_i) - O(\log K(\mathbf{u}_i | \mathbf{v}_i))$ (clearly, the pair $\langle \mathbf{w}_1, \mathbf{v}_i \rangle$ can be identified with the complement of \mathbf{w}_2). Hence it follows that

$$\begin{aligned} K(\mathbf{w}_2 | \mathbf{w}_1, \mathbf{v}_i) &= K(\mathbf{u}_i | \mathbf{v}_i) - K(\mathbf{w}_1 | \mathbf{v}_i) + O(\log K(\mathbf{u}_i | \mathbf{v}_i)) \\ &\leq (1 - \delta - \lambda(i - 1))\ell(\mathbf{u}_i) - (1 - \delta + 3\lambda)\ell(\mathbf{w}_1) + O(\log \ell(\mathbf{u}_i)) \\ &\leq (1 - \delta - 2\lambda i - \lambda)\ell(\mathbf{w}_2) + O(\log \ell(\mathbf{u}_i)). \end{aligned}$$

For large $\ell(\mathbf{u}_i)$, it is easy to deduce the following inequality for the specific deficiency of \mathbf{w}_2 conditional to \mathbf{w}_1 and \mathbf{v}_i :

$$1 - \frac{K(\mathbf{w}_2 | \mathbf{w}_1, \mathbf{v}_i)}{\ell(\mathbf{w}_2)} > \delta + \lambda i.$$

After that, the place-selection rule that we are constructing passes to the $(i + 2)$ nd stage with $\mathbf{u}_{i+1} = \mathbf{w}_2$.

The specific deficiency is always not greater than 1; hence, the second case can be realized less than $(1 - \delta)/\lambda$ times. Therefore, the number of stages in the rule constructed does not exceed a number independent of L . This yields that, for large enough L , the assumption that $\ell(\mathbf{u}_i)$ is large holds for all i .

It remains to estimate the entropy (conditional to L) of the rule constructed. To construct it, we used the following information:

- the value of $1 - \delta + 3\lambda$;
- for each stage: whether the first or second case holds;
- for the second case at each stage: which inequality holds, $K(\mathbf{w}_1 | \mathbf{v}_i) \geq \lfloor (1 - \delta + 3\lambda)\ell(\mathbf{w}_1) \rfloor$ or $K(\mathbf{w}_2 | \mathbf{v}_i) \geq \lfloor (1 - \delta + 3\lambda)\ell(\mathbf{w}_1) \rfloor$;
- for the first case at the last stage: which inequality holds, $T_1 \leq T_2$ or $T_1 \geq T_2$, and also the number of the set A in the family \mathcal{R} .

To find the rational $(1 - \delta + 3\lambda)$, it is sufficient to find $(\delta - 3\lambda)$; for this, it is sufficient to find a natural from the interval $\left(\left(\delta - \frac{3\delta}{\ln(1/\delta)} \right)^{-1}, \left(\delta - \frac{6\delta}{\ln(1/\delta)} \right)^{-1} \right)$. Since the length of the interval is greater than 1, it contains a natural, and the entropy of this natural is not greater than $\text{lb}(1/\delta) + O(1)$.

Thus, an upper bound for the entropy of the rule can be obtained from the following sum by adding several logarithms of this sum (logarithms appear because of encoding tuples):

$$\text{lb} \frac{1}{\delta} + O(1) + 2 \left(\frac{1 - \delta}{\lambda} + O(1) \right) + \text{lb} |\mathcal{R}| + O(1).$$

For small δ , this bound is less than $\frac{4 \ln(1/\delta)}{\delta}$.

Finally, let us remark that, every time we wrote “ L is large enough,” the corresponding lower bound for L could be effectively derived from δ_0 . The value L_0 is the maximum of these bounds. \triangle

Remark 6. In Theorem 4, we could restrict ourselves to regular sets generated by monotonic rules (however, using only nonadaptive rules was not sufficient). Conversely, in Theorem 4', non-monotonic rules are necessary, as Theorem 6' shows.

Theorem 5'. *Assume that $\delta_0 > 0$ is small enough and consider a natural $L > (1/\delta_0)^5$. There exists a binary sequence \mathbf{t} of length L and of specific deficiency greater than δ_0 such that its frequency α -deficiency is less than*

$$\frac{2\delta_0}{\ln(1/\delta_0)}(1 + 3\delta_0)$$

for $\alpha = L\delta_0^4/70$.

Proof. Take the set of rules such that their entropy conditional to L is less than α . Some of these rules are defined not everywhere; let us extend them to all sequences (of length L) in an arbitrary way. Take n -normalization of these rules for all $n = 1, \dots, L$. The number of rules obtained is less than $L \cdot 2^\alpha < e^{L\delta_0^4/70}$. Consider the set of sequences S that was (effectively) constructed in Theorem 5 for the numbers $\delta = \frac{1}{\lfloor 1/\delta_0 \rfloor - 1}$ and L . Since $\delta > \delta_0$, we have $L > (1/\delta)^5$, $e^{L\delta^4/70} > e^{L\delta_0^4/70}$, and by Theorem 5 there exists $\mathbf{t} \in S$ such that

$$\max \{D(n, \varepsilon)/L \mid n = \ell(r[\mathbf{t}]), K(r|L) < \alpha\} < \frac{2\delta}{\ln(1/\delta)} < \frac{2\delta_0}{\ln(1/\delta_0)}(1 + 3\delta_0).$$

The entropy of any $\mathbf{t} \in S$ conditional to L does not exceed $2 \text{lb}(1/\delta) + \text{lb} |S|$. Therefore, the specific deficiency of \mathbf{t} is not less than $1 - \frac{\text{lb} |S| + 2 \text{lb}(1/\delta)}{L}$. By Theorem 5, the specific deficiency of S (which equals $1 - \text{lb} |S|/L$) is greater than δ . Hence, the specific deficiency of \mathbf{t} is greater than $\delta - 2 \text{lb}(1/\delta)/L > \delta_0$. \triangle

Definition 7. The *monotonic α -frequency deficiency* of a binary sequence \mathbf{t} of length L is the maximum of the value $D(n, \varepsilon)/L$ over monotonic rules r generated by programs of entropy less than α , where n is the length of $r[\mathbf{t}]$, ε is the deviation in $r[\mathbf{t}]$, and the entropy is conditional to L .

Theorem 6'. *Assume that $\delta > 0$ is small enough and consider a natural $L \geq (1/\delta)^2$. Then there exists a binary sequence \mathbf{t} of length L and of specific deficiency greater than $1/2$ such that its monotonic $(\delta L/4)$ -frequency deficiency is less than δ .*

Proof. To construct the sequence \mathbf{t} , we will consider a game similar to that used in the proof of Theorem 4. But in this case the value of a bet may be any number less than the current capital (and the game is called the *game "on cash"*). The initial capital is equal to 1. It is convenient to say that a bet is a fraction σ of the current capital; then the current capital is multiplied by $(1 + \sigma)$ in the case of winning (if the next digit has been guessed right), and by $(1 - \sigma)$ in the other case.

To each monotonic rule r , binary digit b , and number $\sigma \in [0, 1]$, assign the following strategy in the game: at step i , the rule r is applied to the segment $\mathbf{t}_{1:i-1}$; if the rule prescribes to select the next digit, the bet σ is placed on b , otherwise, the bet is not placed. If r is undefined on a segment $\mathbf{t}_{1:i-1}$ (the program does not stop), then the strategy is undefined at all subsequent steps and the current capital becomes undefined too. Clearly, if the digit b occurs n_b times in the subsequence $r[\mathbf{t}]$ of length n , then the final capital of the strategy is equal to $(1 + \sigma)^{n_b}(1 - \sigma)^{n - n_b}$ (and the deviation is equal to $\left| \frac{1}{2} - \frac{n_b}{n} \right|$, hence an estimation of the deviation can be obtained from an estimation of the capital).

Assume that $1/\delta$ is natural; in the end of the proof, we explain what should be done with an arbitrary δ .

Let \mathcal{R}_L be the set of all monotonically selecting programs of entropy less than $\delta L/4$, and R be the number of them. Without loss of generality, we can assume that a monotonically selecting program

(possibly, undefined everywhere) is assigned to each codeword; moreover, if codewords are different, the assigned programs are different too. Therefore, $R = 2^{\lceil \delta L/4 \rceil} - 1$. For $n = \left\lceil \frac{\ln(4LR)}{\delta} \right\rceil, \dots, L$,

define the bet $\sigma_n = \sqrt{\frac{\ln(4LR)}{n}}$. We will consider the set of strategies assigned to all possible σ_n , $r \in \mathcal{R}_L$, and $b \in \{0, 1\}$.

Let us introduce several auxiliary values, which will be useful in the construction of \mathbf{t} . Let S_i be the total capital of all strategies that are defined at step i (this capital depends on $\mathbf{t}_{1:i-1}$ only). The initial capital S_0 is equal to the number of strategies and hence is less than $2LR$.

The idea of the construction is as follows: at each step, we choose t_i so that the total capital does not increase; then the final capital of a strategy is not greater than the total initial capital S_0 , and we obtain an estimation of the deviation in the sequence. However, we cannot perform this directly since we cannot verify algorithmically whether a rule is defined on a given segment; therefore, we cannot compute S_{i+1} for both possible values of t_i . Thus, at each step of our algorithm, we need additional information.

Let Q_i be the total capital at step i of strategies that will be defined at the next step, $i+1$, too (i.e., defined on the segment $\mathbf{t}_{1:i}$). It is obvious that $Q_i \leq S_i$.

Define the sequences P_i and m_i :

$$\begin{aligned} P_0 &= S_0, \\ m_i &= \left\lfloor \frac{P_i - Q_i}{2R} \sqrt{\delta} \right\rfloor, \\ P_{i+1} &= P_i + 2R - \frac{2Rm_i}{\sqrt{\delta}}. \end{aligned}$$

It easily follows that

$$P_i - \frac{2Rm_i}{\sqrt{\delta}} - \frac{2R}{\sqrt{\delta}} < Q_i \leq P_i - \frac{2Rm_i}{\sqrt{\delta}}.$$

Now let us describe the procedure of constructing the next digit t_{i+1} given a segment $\mathbf{t}_{1:i}$ and the values of P_i and m_i . The procedure consists of two stages:

1. All strategies are applied to the segment $\mathbf{t}_{1:i}$ in parallel; the current capital of strategies already defined on this segment is computed until their total capital exceeds $P_i - \frac{2Rm_i}{\sqrt{\delta}} - \frac{2R}{\sqrt{\delta}}$.
2. Then the value of t_{i+1} is chosen so that the total gain of strategies that were defined at Stage 1 is nonpositive.

First, let us prove that the algorithm is well defined. If the segment $\mathbf{t}_{1:i}$ has already been constructed, then the values S_i , Q_i , P_i , and m_i are defined. Assume that the algorithm was given correct P_i and m_i . During Stage 1, there will be a moment when every strategy that is defined on this segment has already been defined. Their total capital is equal to $Q_i > P_i - \frac{2Rm_i}{\sqrt{\delta}} - \frac{2R}{\sqrt{\delta}}$; hence, at a certain moment, Stage 1 will stop. Let us note that, for each (defined) strategy, the sum of the gains for $t_{i+1} = 0$ and $t_{i+1} = 1$ is zero. The same is true for any set of (defined) strategies; therefore, the total gain at Stage 2 cannot be positive for both values of t_{i+1} .

In the sequel, we need several properties of the values used.

1. $S_i \leq P_i$.

By the definition, $S_0 = P_0$. Let us estimate S_{i+1} . The total capital of the strategies that were defined during Stage 1 (denote it by Q'_i) does not decrease. The capital of other strategies is

equal to $Q_i - Q'_i$, and it becomes not more than $(1 + \sqrt{\delta})$ times as large since the values of bets do not exceed $\sqrt{\delta}$. It is obvious that $Q_i - Q'_i < Q_i - \left(P_i - \frac{2Rm_i}{\sqrt{\delta}} - \frac{2R}{\sqrt{\delta}}\right) \leq \frac{2R}{\sqrt{\delta}}$. Consequently,

$$S_{i+1} \leq Q'_i + (Q_i - Q'_i)(1 + \sqrt{\delta}) = Q_i + (Q_i - Q'_i)\sqrt{\delta} < \left(P_i - \frac{2Rm_i}{\sqrt{\delta}}\right) + \frac{2R}{\sqrt{\delta}}\sqrt{\delta} = P_{i+1}.$$

2. $m_i \geq 0$.

The inequality follows from the fact that $Q_i \leq S_i \leq P_i$.

3. $\sum m_i < 2L\sqrt{\delta}$.

Let us sum up the equalities $P_{i+1} = P_i + 2R - \frac{2Rm_i}{\sqrt{\delta}}$ over all i . We get $P_L = P_0 + 2RL - 2R\sum m_i/\sqrt{\delta}$ and, taking into account that $P_0 < 2RL$ and $P_L > S_L \geq 0$, we have

$$\sum m_i = (2RL + P_0 - P_L)\frac{\sqrt{\delta}}{2R} < 2L\sqrt{\delta}.$$

Now we can estimate the entropy of the constructed sequence \mathbf{t} . The algorithm should know the set of strategies and the sequences P_i and m_i . The set of strategies can be generated from L and δ , and P_i can be generated from m_i , L , and δ . Thus,

$$K(\mathbf{t}) \leq K(\langle m_0, \dots, m_{L-1} \rangle | L, 1/\delta) + 2\text{lb } L + 2\text{lb}(1/\delta) = K(\langle m_0, \dots, m_{L-1} \rangle | L, 1/\delta) + o(L).$$

The entropy of $\langle m_0, \dots, m_{L-1} \rangle$ conditional to L and $1/\delta$ can be majorized by the binary logarithm of the number of tuples such that a tuple consists of L naturals and the sum of these naturals is less than $N = \lceil 2L\sqrt{\delta} \rceil$. It can easily be shown that the tuple $\langle m_0, \dots, m_{L-1} \rangle$ is uniquely determined by the multiset of its partial sums $\{m_0, m_0 + m_1, \dots, m_0 + \dots + m_{L-1}\}$ (repetitions are allowed). The partial sums range over the set $\{0, \dots, N - 1\}$. Thus, the number of possible sequences $\{m_i\}$ does not exceed the number of unordered samples with replacement of N things L at a time, which is equal to

$$\binom{N + L}{L} < \frac{(N + L)^N}{N!} < \left(\frac{(N + L)e}{N}\right)^N.$$

As a result, we have

$$\begin{aligned} K(\langle m_0, \dots, m_{L-1} \rangle | L, 1/\delta) &\leq \text{lb} \binom{N + L}{L} + O(1) \\ &\leq N \text{lb} \frac{(N + L)e}{N} + O(1) \leq \lceil 2L\sqrt{\delta} \rceil \text{lb} \left(1 + \frac{1}{2\sqrt{\delta}}\right) e + O(1) < L/3. \end{aligned}$$

Since $K(\mathbf{t}) < L/2$, we get that the specific deficiency of \mathbf{t} is greater than $1/2$.

It remains to estimate the monotonic frequency deficiency of the sequence \mathbf{t} . To this end, let us fix a rule r and consider the subsequence $r[\mathbf{t}]$ of length n with deviation ε . Let us majorize $D(n, \varepsilon)$. Consider two possible cases.

1. $n \leq 7\delta L/8$. Since $D(n, \varepsilon) \leq n$, we obtain $D(n, \varepsilon)/L \leq \frac{7}{8}\delta$.
2. $n > 7\delta L/8$. Assume that the number of zeros in $r[\mathbf{t}]$ is greater than the number of ones. Take the strategy assigned to the rule r that places a bet σ_n on 0. The final capital of this strategy does not exceed the total capital $S_L < P_L \leq P_0 + 2RL < 4RL$ (this follows from the equality

$P_L = P_0 + 2RL - 2R \sum m_i / \sqrt{\delta}$). On the other hand, the capital of the strategy is equal to $(1 + \sigma_n)^{n(\frac{1}{2} + \varepsilon)} (1 - \sigma_n)^{n(\frac{1}{2} - \varepsilon)}$. We have the inequality

$$\frac{n}{2} ((1 + 2\varepsilon) \ln(1 + \sigma_n) + (1 - 2\varepsilon) \ln(1 - \sigma_n)) < \ln(4RL);$$

hence,

$$2\varepsilon \ln \frac{1 + \sigma_n}{1 - \sigma_n} + \ln(1 - \sigma_n^2) < \frac{2}{n} \ln(4RL).$$

Using the fact that $\ln \frac{1+x}{1-x} \geq 2x$ if $0 \leq x < 1$ and $\ln(1-x) \geq -2x$ if $0 \leq x \leq 1/2$, we obtain

$$\varepsilon < \frac{\frac{2}{n} \ln(4LR) + 2\sigma_n^2}{4\sigma_n} = \frac{\ln(4LR)}{2n\sigma_n} + \frac{\sigma_n}{2} = \sqrt{\frac{\ln(4LR)}{n}}.$$

The Stirling formula implies that

$$\binom{n}{j} \geq \frac{e^{n \cdot h(j/n)}}{\sqrt{8j(n-j)/n}}.$$

Differentiating twice, we get (if $\varepsilon \leq 1/\sqrt{5}$)

$$h\left(\frac{1}{2} + \varepsilon\right) \geq \ln 2 - 2\varepsilon^2(1 + 10\varepsilon^2/3).$$

Using these inequalities, we can easily check that

$$\begin{aligned} D(n, \varepsilon) &< 2n\varepsilon^2 \ln e(1 + 10\varepsilon^2/3) + \ln n < 2 \ln(4LR) \ln e \left(1 + \frac{10 \ln(4LR)}{3n}\right) + \ln n \\ &< 2 \ln(4LR) \left(1 + \frac{10 \ln(4LR)}{3(7\delta L/8)}\right) + \ln n \leq \frac{5}{6} \delta L + O(\ln L) \end{aligned}$$

and therefore $D(n, \varepsilon)/L < \frac{7}{8} \delta$ (recall that δ is small, $L \geq (1/\delta)^2$, and $\varepsilon \leq \frac{1}{\sqrt{5}}$).

Finally, if the number $1/\delta$ is not a natural, then we can carry out the same reasoning for the number $\delta_0 = \frac{1}{\lceil 1/\delta \rceil} \geq \delta$, and obviously $\frac{7}{8} \delta_0 < \delta$. \triangle

2.4. On Optimality of Bounds

All bounds that are proved in this paper seem to be close to optimal ones, with the only exception, and now we discuss it. Let δ and δ' range over reals from $(0, 1)$, L range over naturals, and S range over the set of binary sequences of length L . Denote by $\Phi_{L, \delta}(\delta')$ the maximum (over sets S with deficiency $\geq \delta$) of the least number of regular sets with deficiency $\geq \delta'$ such that their union covers S . In essence, Theorems 4 and 5 give an estimation of this function Φ . Namely, for some functions ρ_1 and ρ_2 , we have

$$\begin{aligned} \delta' \leq c_1 \frac{\delta}{\ln(1/\delta)} &\Rightarrow \Phi_{L, \delta}(\delta') < L \rho_1(1/\delta), \\ \delta' \geq c_2 \frac{\delta}{\ln(1/\delta)} &\Rightarrow \Phi_{L, \delta}(\delta') > e^{L/\rho_2(1/\delta)}, \end{aligned}$$

where $c_1 \approx 1$, $c_2 = 2$, and $L \geq (1/\delta)^5$. In other words, when δ' changes from $c_1 \frac{\delta}{\ln(1/\delta)}$ to $c_2 \frac{\delta}{\ln(1/\delta)}$, the function Φ makes a jump from an expression linear in L to an expression exponential in L . If we are interested in the behavior of Φ for δ fixed and $L \rightarrow \infty$, then the bound looks optimal (though it is possible to sharpen the constants c_1 and c_2). But now consider the behavior of Φ as $\delta \rightarrow 0$. Then, in the obtained bounds, the function ρ_2 increases polynomially and the function ρ_1 increases exponentially.

Open problem. Is it possible to strengthen the upper bound on Φ so that ρ_1 becomes polynomial?

An. Muchnik proved (see [2]) that, for a smaller upper bound on δ' , a linear function can be taken as ρ_1 .¹² More exactly, we have

$$\delta' \leq c_3 \delta^2 \quad \Rightarrow \quad \Phi_{L,\delta}(\delta') < L(c_4/\delta).$$

Thus, a positive decision of the problem seems to be likely enough.

The authors are grateful to N.K. Vereshchagin for interesting discussions about Kolmogorov theory, which were one of the incentives to begin the present work. The authors are grateful to A.V. Chernov for a great help in preparing the text for publication. The main contents of the paper was reported at Kolmogorov seminar of Moscow State University in the spring of 2002, and we thank its participants for attention.

REFERENCES

1. Kolmogorov, A.N., On Tables of Random Numbers, *Sankhyā, Indian J. Statist., Ser. A*, 1963, vol. 25, no. 4, pp. 369–376. Reprinted in *Theor. Comp. Sci.*, 1998, vol. 207, pp. 387–395.
2. Muchnik, An.A., Semenov, A.L., and Uspensky, V.A., Mathematical Metaphysics of Randomness, *Theor. Comp. Sci.*, 1998, vol. 207, no. 1–2, pp. 263–317.
3. van Lambalgen, M., Von Mises's Definition of Random Sequences Reconsidered, *J. Symb. Logic*, 1987, vol. 52, pp. 725–755.
4. Kolmogorov, A.N., Three Approaches to the Quantitative Definition of Information, *Probl. Peredachi Inf.*, 1965, vol. 1, no. 1, pp. 3–11 [*Probl. Inf. Trans. (Engl. Transl.)*, 1965, vol. 1, no. 1, pp. 1–7].
5. Uspensky, V.A. and Shen, A.Kh., Relations between Varieties of Kolmogorov Complexities, *Math. Syst. Theory*, 1996, vol. 29, pp. 271–292.
6. Kolmogorov, A.N., On Tables of Random Numbers, in *Semiotika i informatika*, Moscow: VINITI, 1982, vol. 18, pp. 3–13. Reprinted in Kolmogorov, A.N., *Teoriya informatsii i teoriya algoritmov (Information Theory and Theory of Algorithms)*, Moscow: Nauka, 1987, pp. 204–213.
7. Yaglom, A.M. and Yaglom, I.M., *Veroyatnost' i informatsiya*, Moscow: Fizmatgiz, 1960, 2nd ed. Translated under the title *Probability and Information*, Boston: Reidel, 1983.
8. Durand, B. and Vereshchagin, N., Kolmogorov–Loveland Stochasticity for Finite Strings, 2002. Available from <http://markov.math.msu.ru/~ver/papers/kolm-love.ps>.

¹² An algorithmical analog of this result was recently obtained by N. Vereshchagin and B. Durand, see [8]. We used some elements of their argument in the proof of Theorem 4'.