ISSN 0081-5438, Proceedings of the Steklov Institute of Mathematics, 2011, Vol. 274, pp. 34–89. © Pleiades Publishing, Ltd., 2011.
Published in Russian in Trudy Matematicheskogo Instituta imeni V.A. Steklova, 2011, Vol. 274, pp. 41–102.

Algorithmic Tests and Randomness with Respect to a Class of Measures

Laurent Bienvenu[a], **Peter Gács**[b], **Mathieu Hoyrup**[c], **Cristobal Rojas**[d], and **Alexander Shen**[e,f]

Received March 2011

Abstract—This paper offers some new results on randomness with respect to classes of measures, along with a didactic exposition of their context based on results that appeared elsewhere. We start with the reformulation of the Martin-Löf definition of randomness (with respect to computable measures) in terms of randomness deficiency functions. A formula that expresses the randomness deficiency in terms of prefix complexity is given (in two forms). Some approaches that go in another direction (from deficiency to complexity) are considered. The notion of Bernoulli randomness (independent coin tosses for an asymmetric coin with some probability p of head) is defined. It is shown that a sequence is Bernoulli if it is random with respect to *some* Bernoulli measure B_p. A notion of "uniform test" for Bernoulli sequences is introduced which allows a quantitative strengthening of this result. Uniform tests are then generalized to arbitrary measures. Bernoulli measures B_p have the important property that p can be recovered from each random sequence of B_p. The paper studies some important consequences of this orthogonality property (as well as most other questions mentioned above) also in the more general setting of constructive metric spaces.

DOI: 10.1134/S0081543811060058

1. INTRODUCTION

This paper, though intended to be rather self-contained, can be seen as a continuation of [11] (which itself built on earlier work of Levin) and [13].

Our enterprise is to develop the theory of randomness beyond the framework where the underlying probability distribution is the uniform distribution or a computable distribution. A randomness test $\mathbf{t}(\omega, P)$ of an object ω with respect to a measure P is defined to be a function of both the measure P and the point ω.

In some later parts of the paper, we will also go beyond the case where the underlying space is the set of finite or infinite sequences: rather, we take a constructive metric space with its algebra of Borel sets.

[a] Laboratoire d'Informatique Algorithmique: Fondements et Applications (LIAFA), CNRS UMR 7089 & Université Paris Diderot, Paris 7, Case 7014, 75205 Paris Cedex 13, France.

[b] Department of Computer Science, Boston University, 111 Cummington st., Boston, MA 02215, USA.

[c] Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA), B248, 615, rue du Jardin Botanique, BP 239, 54506 Vandœuvre-lès-Nancy, France.

[d] Department of Mathematics, University of Toronto, Bahen Centre, 40 St. George st., Toronto, Ontario, Canada M5S 2E4.

[e] Laboratoire d'Informatique Fondamentale de Marseille (LIF), Universitè Aix–Marseille, CNRS UMR 6166, 39 rue Joliot-Curie, 13453 Marseille Cedex 13, France.

[f] On leave from Institute for Information Transmission Problems (Kharkevich Institute), Russian Academy of Sciences, Bol'shoi Karetnyi per. 19, Moscow, 127994 Russia.

E-mail addresses: Laurent.Bienvenu@liafa.jussieu.fr (L. Bienvenu), gacs@bu.edu (P. Gács), Mathieu.Hoyrup@loria.fr (M. Hoyrup), crojas@math.utoronto.ca (C. Rojas), sasha.shen@gmail.com (A. Shen).

We will apply the above notion of test to define, following the ideas of [16], for a class $\mathcal{C}$ of measures having some compactness property, a "class test" $\mathbf{t}_{\mathcal{C}}(\omega)$. This is a test to decide whether an object ω is random with respect to some measure P in the class $\mathcal{C}$. We will show that in the case of the class of Bernoulli measures over binary sequences, this notion is equivalent to the class tests introduced by Martin-Löf in [20].

In case there is an effective sense in which the elements of the class are mutually orthogonal, we obtain an especially simple separation of the randomness test $\mathbf{t}(\omega, P)$ into two parts: the class test and an arbitrarily simple test for "typicality" with respect to the measure P. In some natural special cases, the typicality test corresponds to a convergence property of relative frequencies, allowing one to apply the theory to any general effectively compact class of ergodic stationary processes.

Generally speaking, our tests do not necessarily possess some properties of randomness tests $\mathbf{t}(\omega, P)$ that depend on the measure P, for example, a kind of monotonicity in P. It is therefore notable that in the case of orthogonal classes, randomness is equivalent to a "blind" notion of randomness, which only considers randomness tests that do not depend on the measure P.

Here is an outline of the paper. We start with the reformulation of the Martin-Löf definition of randomness (with respect to computable measures) in terms of tests. A randomness test provides a quantitative measure of non-randomness, called "randomness deficiency"; it is finite for random sequences and infinite for nonrandom ones. There are two versions of these tests ("average-bounded" and "probability-bounded" ones); a relation between them is established.

A formula that expresses the (average-bounded) randomness deficiency in terms of prefix complexity is given (in two forms). It implies the Levin–Schnorr criterion of randomness (with prefix complexity, as in the special case first announced in Chaitin's paper [4]). Some approaches that go in another direction (from deficiency to complexity) are considered.

The notion of Bernoulli sequence (looking like the outcome of independent coin tosses for an asymmetric coin) is defined. It is shown that the set of Bernoulli sequences is the union (over all $p \in [0,1]$) of the sets of sequences that are random with respect to B_p, the Bernoulli measure with probability p; here we assume that p is given as an oracle). A notion of "uniform test" for Bernoulli sequences is introduced. Then the statement above is proved in the following quantitative form: the Bernoulli deficiency is the infimum of B_p deficiencies over all $p \in [0,1]$.

The notion of general uniform test (not restricted to the class of Bernoulli measures) is introduced. It is shown that it generalizes Martin-Löf's earlier definition of randomness (which was given only for computable measures).

Bernoulli measures B_p have the important property that p can be recovered from each random sequence of B_p. The paper studies some important consequences of this orthogonality property (as well as most other questions mentioned above) also in the more general setting of constructive metric spaces.

The following notation is useful, since inequalities hold frequently only within an additive or multiplicative constant.

Notation 1.1. We will write $f(x) \overset{*}{<} g(x)$ for an inequality between positive functions within a multiplicative constant, that is, for the relation $f(x) = O(g(x))$; more precisely, if there is a constant c with $f(x) \le cg(x)$ for all x. The relation $f \overset{*}{=} g$ means $f \overset{*}{<} g$ and $f \overset{*}{>} g$. Similarly, $f \overset{+}{<} g$ and $f \overset{\pm}{=} g$ mean inequalities within additive constants.

Let Λ denote the empty string. Logarithms are taken, by default, to base 2. We use $|x|$ to denote the length of a string x. For a finite string x and a finite or infinite string y, let $x \sqsubseteq y$ denote that x is a prefix of y. If x is a finite or infinite sequence, then its elements are written as $x(1), x(2), \dots$, and its prefix of size n will be denoted by $x(1:n)$.

Let $\overline{\mathbb{R}}_+ = [0, \infty]$ be the set of nonnegative reals, with the special value ∞ added. The binary alphabet $\{0,1\}$ will also be denoted by $\mathbb{B}$.

2. RANDOMNESS ON SEQUENCES, FOR COMPUTABLE MEASURES

In the first sections, we will study randomness over infinite binary sequences.

2.1. Lower semicomputable functions on sequences.

Definition 2.1 (binary Cantor space, Baire space)**.** We will denote by Ω the set of infinite binary sequences, and also call it the *binary Cantor space*. For a finite string x let $x\Omega$ be the set of all infinite sequences that have the finite prefix x. These sets will be called *basic open sets*, and the set of all basic open sets is called the *basis* of Ω (as a topological space). A subset of Ω is *open* if it is the union of a set of basis elements.

The set of infinite sequences of natural numbers will be called the *Baire space*. Basic open sets and open sets can be defined for it analogously.

A notion somewhat weaker than computability will play a crucial role.

Definition 2.2. An open set $G \subseteq \Omega$ is called *effectively open*, or *lower semicomputable open*, or *c.e. open*, or *r.e. open* if it is the union of a computable sequence $x_i\Omega$ of basic elements. A set is *upper semicomputable closed*, or *effectively closed* if its complement is effectively open.

A set Γ is called *effectively* G_δ if there is a sequence of sets U_k, $k = 1, 2, \ldots$, effectively open uniformly in k such that $\Gamma = \bigcap_k U_k$.

A function $t\colon \Omega \to [0, \infty]$ is *lower semicomputable* if

(a) for any rational r the set $\{\omega\colon r < t(\omega)\}$ is open in Ω, that is, is a union of intervals $x\Omega$;

(b) moreover, this set is effectively open uniformly in r; that is, there exists an algorithm that gets r as an input and generates strings $x_0, x_1, \ldots$ such that the union of intervals $x_i\Omega$ is equal to $\{\omega\colon r < t(\omega)\}$.

This definition is a constructive version of the classical notion of lower semicontinuous function as in requirement (a). The same class of lower semicomputable functions has other (equivalent) definitions; here is one of them.

Definition 2.3. A function u defined on Ω and having rational values is called *basic* if the value $u(\omega)$ is determined by some finite prefix of ω.

If this prefix has length N, the function can be presented as a table with 2^N rows; each row contains N bits (the values of the first N bits of ω) and a rational number (the value of the function). Such a function is a finite object.

The proof of the following proposition is a simple exercise:

Proposition 2.4. *The* (*pointwise*) *limits of monotonic sequences of basic functions are exactly the lower semicomputable functions on* Ω*.*

Since the difference of two basic functions is a basic function, we can reformulate this criterion as follows: lower semicomputable functions are (pointwise) sums of computable series made of nonnegative basic functions.

One more way to define a lower semicomputable function goes as follows.

Definition 2.5 (generating)**.** Let T be a lower semicomputable function on the set $\{0,1\}^*$ of finite sequences of zeros and ones with nonnegative (finite or infinite) values. This means that the set of pairs (x, r) such that $r < T(x)$ is enumerable. Then the function t defined as

$$t(\omega) = \sup_{x \sqsubseteq \omega} T(x)$$

is a lower semicomputable function on Ω. We will say that the function $T(\cdot)$ *generates* the function $t(\cdot)$ if it is also monotone: $T(x) \le T(y)$ if $x \sqsubseteq y$.

The monotonicity requirement can always be satisfied by taking $T'(x) = \max_{z \sqsubseteq x} T(z)$.

Proposition 2.6. *Any lower semicomputable function t on Ω is generated by an appropriate function T on $\{0,1\}^*$ in this way.*

We may also assume that T is a computable function with rational values. Indeed, since only the supremum of T on all the prefixes is important, instead of increasing $T(x)$ for some x we may increase $T(y)$ for all $y \sqsupseteq x$ of large length; this delay allows T to be computable.

For a given lower semicomputable function t on Ω there exists a maximal monotonic function T on finite strings that generates t (in the sense just described). This maximal T can be defined as follows:

$$T(x) = \inf_{\omega \sqsupseteq x} t(\omega). \tag{1}$$

Let us now exploit the finiteness of the binary alphabet $\{0,1\}$, which implies that the space Ω is a compact topological space.

Proposition 2.7. *The function T defined by* (1) *is lower semicomputable. In the definition, we can replace* inf *by* min*.*

Proof. Evidently, the function generated by T does not exceed t. On the other hand, if $t(\omega) > r$, the semicontinuity property guarantees that the same is valid in some neighborhood of ω, so $T(x) \geq r$ for some prefix $x \sqsubseteq \omega$. Therefore, T generates t.

It remains to show that T is lower semicomputable. Indeed, $r < \inf_{\omega \sqsupseteq x} t(\omega)$ if and only if there exists some rational $r' > r$ with $r' < t(\omega)$ for all $\omega \sqsupseteq x$. The latter condition can be reformulated: the open set of all sequences ω such that $t(\omega) > r'$ is a superset of $x\Omega$. This open set is a union of an enumerable family of intervals; if these intervals cover $x\Omega$, compactness implies that this is revealed at some finite stage, so the condition is enumerable (and the existential quantifier over r' keeps it enumerable).

Since the function $t(\omega)$ is lower semicontinuous, it actually reaches its infimum on the compact set $x\Omega$, so inf can be replaced with min. $\square$

2.2. Randomness tests. We assume that the reader is familiar with the basic concepts of measure theory and integration, at least in the space Ω of infinite binary sequences. A measure P on Ω is determined by the values

$$P(x) = P(x\Omega),$$

which we will denote by the same letter P, without danger of confusion. Moreover, any function $P\colon \{0,1\}^* \to [0,1]$ with the properties

$$P(\Lambda) = 1, \qquad P(x) = P(x0) + P(x1) \tag{2}$$

uniquely defines a measure (this is a particular case of Carathéodory's theorem).

Definition 2.8 (computable measure)**.** A real number x is called *computable* if there is an algorithm that for all rational $\varepsilon > 0$ returns a rational approximation of x with error not greater than ε. Computable numbers can also be determined as limits of computable sequences of rational numbers $x_1, x_2, \dots$ such that $|x_n - x_{n+k}| \leq 2^{-n}$. An infinite sequence $s_1, s_2, \dots$ of real numbers is a *strong Cauchy* sequence if we have $|s_m - s_n| \leq 2^{-m}$ for all $m < n$.

A real-valued function defined on words (or other constructive objects) is *computable* if its values are computable uniformly from the input, that is, there is an algorithm that for each input and $\varepsilon > 0$ returns an ε-approximation of the function value on this input.

A measure P over Ω is said to be *computable* if the function $P\colon \{0,1\}^* \to [0,1]$ is computable.

Definition 2.9 (randomness test, computable measure)**.** Let P be a computable probability distribution (measure) on Ω. A lower semicomputable function t on Ω with nonnegative (possibly infinite) values is an (*average-bounded*) *randomness test* with respect to P (or a *P-test*) if the expected value of t with respect to P is at most 1, that is,

$$\int t(\omega)\, dP \le 1.$$

A sequence ω *passes* a test t if $t(\omega) < \infty$. A sequence is called *random* with respect to P if it passes all P-randomness tests (as defined above).

The intuition: when $t(\omega)$ is large, this means that the test t finds a lot of "regularities" in ω. Constructing a test, we are allowed to declare whatever we want to be a "regularity"; however, we should not find too many of them on average: if we declare too many sequences to be "regular," the average becomes too big.

This definition turns out to be equivalent to randomness as defined by Martin-Löf (see below). But let us start with the universality theorem:

Theorem 2.10. *For any computable measure P there exists a universal* (*maximal*) *P-test u: this means that for any other P-test t there exists a constant c such that*

$$t(\omega) \le cu(\omega)$$

for every $\omega \in \Omega$.

In particular, $u(\omega)$ is finite if and only if $t(\omega)$ is finite for every P-test t, so the sequences that pass the test u are exactly the random sequences.

Proof. Let us enumerate the algorithms that generate all lower semicomputable functions. Such an algorithm produces a monotone sequence of basic functions. Before letting through the next basic function of this sequence, let us check that its P-expectation is less than 2. If the algorithm considered indeed defines a P-test, this expectation does not exceed 1, so by computing the values of P with sufficient precision we are able to guarantee that the expectation is less than 2. If this checking procedure does not terminate (or gives a negative result), we just do not let the basic function through.

In this way we enumerate all tests as well as some lower semicomputable functions that are not exactly tests but are at most twice bigger than tests. It remains to sum up all these functions with positive coefficients whose sum does not exceed 1/2 (say, $1/2^{i+2}$). □

This is the definition of a randomness test in the form used in [9]. It is equivalent to the classical Martin-Löf definition of a randomness test:

Definition 2.11. Let P be a computable distribution over Ω. A sequence of open sets $U_1, U_2, \dots$ is called a *Martin-Löf test* for P if the sets U_i are effectively open in a uniform way (that is, $U_i = \bigcup_j x_{ij}\Omega$ where the double sequence x_{ij} of strings is computable) and, moreover, $P(U_k) \le 2^{-k}$ for all k.

A set N is called a *constructive* (*effective*) *null set* for the measure P if there is a Martin-Löf test $U_1, U_2, \dots$ with the property $N = \bigcap_k U_k$. Note that effective null sets are constructive G_δ sets.

A sequence $\omega \in \Omega$ is said to *pass* the test $U_1, U_2, \dots$ if it is not in N. It is *Martin-Löf random* with respect to the measure P if it is not contained in any constructive null set for P.

The following theorem is not new (see, for example, [19]).

Theorem 2.12. *A sequence ω passes all average-bounded P-tests* (= *passes the universal P-test*) *if and only if it is Martin-Löf random with respect to P.*

Proof. If t is a test, then the set of all ω such that $t(\omega) > N$ is an effectively open set that can be found effectively given N. This set has P-measure at most $1/N$ (by Chebyshev's inequality), so the set of sequences ω that do not pass t (that is, $t(\omega)$ is infinite) is an effectively P-null set.

On the other hand, let us show that for every effectively null set Z there exists an average-bounded test that is infinite at all its elements. Indeed, for every effectively open set U the function 1_U that is equal to 1 inside U and to 0 outside U is lower semicomputable. Then we can get a test $\sum_i 1_{U_i}$. The average of this test does not exceed $\sum_i 2^{-i} = 1$, while the sum is infinite for all elements of $\bigcap_i U_i$. □

When talking about randomness for a computable measure, we will write *randomness* from now on, understanding Martin-Löf randomness, since no other kind will be considered.

Sometimes it is useful to switch to the logarithmic scale.

Definition 2.13. For every computable measure P, we will fix a universal P-test and denote it by $\mathbf{t}_P(\omega)$. Let $\mathbf{d}_P(\omega)$ be the logarithm of the universal test $\mathbf{t}_P(\omega)$:

$$\mathbf{t}_P(\omega) = 2^{\mathbf{d}_P(\omega)}.$$

With other kinds of test, it will also be our convention to use $\mathbf{t}$ (boldface) for the universal test and $\mathbf{d}$ (boldface) for its logarithm.

In a sense, the function $\mathbf{d}_P$ measures the randomness deficiency in bits.

The logarithm, along with the requirement $\int \mathbf{t}_P(\omega)\,dP \le 1$, implies that $\mathbf{d}_P(\omega)$ may have some negative values, and even values $-\infty$. By just choosing a different universal test we can always make $\mathbf{d}_P(\omega)$ bounded below by, say, -1, and also integer-valued. On the other hand, if we want to make it nonnegative, we will have to lose the property $\int 2^{\mathbf{d}_P(\omega)} dP \le 1$, though we may still have $\int 2^{\mathbf{d}_P(\omega)}\,dP \le 2$. It will still have the following property:

Proposition 2.14. *The function $\mathbf{d}_P(\cdot)$ is lower semicomputable and is the largest (up to an additive constant) among all lower semicomputable functions such that the P-expectation of $2^{\mathbf{d}_P(\cdot)}$ is finite.*

As we have shown, for any fixed computable measure P the value $\mathbf{d}_P(\omega)$ (and $\mathbf{t}_P(\omega)$) is finite if and only if the sequence ω is Martin-Löf random with respect to P.

Remarks 2.15. 1. Each Martin-Löf's test $(U_1, U_2, \dots)$ is more directly related to a lower semicomputable function $F(\omega) = \sup_{\omega \in U_i} i$. This function has the property $P[F(\omega) \ge k] \le 2^{-k}$. Such functions will be called *probability-bounded* tests; they were used in [30]. We will return to such functions later (Subsection 2.3).

2. We have defined $\mathbf{d}_P(\omega)$ separately for each computable measure P (up to a constant). We will later give a more general definition of randomness deficiency $\mathbf{d}(\omega, P)$ as a function of two variables P and ω that coincides with $\mathbf{d}_P(\omega)$ for every computable P up to a constant depending on P.

2.3. Average-bounded and probability-bounded deficiencies. Let us refer, for example, to [19, 26] for the definition and basic properties of plain and prefix (Kolmogorov) complexity. We will define prefix complexity in Definition 2.18 below, though. We will not use complexities explicitly in the present section, but just refer to some of their properties by analogy.

The definition of a test given above resembles the definition of prefix complexity; we can give another one which is closer to plain complexity. These tests are called *probability-bounded* tests. For them we use a weaker requirement: we require that the P-measure of the set of all sequences ω such that $t(\omega) > c$ does not exceed $1/c$. (This property is true if the integral does not exceed 1, due to Chebyshev's inequality.)

In the logarithmic scale this requirement can be restated as follows: the P-measure of the set of all sequences whose deficiency is greater than n does not exceed 2^{-n}. If we restrict tests to integer values, we arrive at the classical Martin-Löf tests (see also Remark 2.15.1).

While constructing a universal test in this sense, it is convenient to use the logarithmic scale and consider only integer values of n. As before, we enumerate all tests and "almost-tests" d_i (where the measure is bounded by a twice bigger bound) and then take the weighted maximum in the following way:

$$\mathbf{d}(\omega) = \max_i [d_i(\omega) - i] - c.$$

Then $\mathbf{d}$ is less than d_i only by $i + c$, and the set of all ω such that $\mathbf{d}(\omega) > k$ is the union of sets where $d_i(\omega) > k + i + c$. Their measures are bounded by $O(2^{-k-i-c})$, and for a suitable c the sum of measures is at most 2^{-k}, as required.

In this way we get two measures of non-randomness that can be called "average-bounded deficiency" $\mathbf{d}^{\text{aver}}$ (the first one, related to the tests called "integral tests" in [19]) and "probability-bounded deficiency" $\mathbf{d}^{\text{prob}}$ (the second one). It is easy to see that they define the same set of nonrandom sequences (= sequences that have infinite deficiency). Moreover, the finite values of these two functions are also rather close to each other:

Proposition 2.16. $\mathbf{d}^{\text{aver}}(\omega) \overset{+}{<} \mathbf{d}^{\text{prob}}(\omega) \overset{+}{<} \mathbf{d}^{\text{aver}}(\omega) + 2\log \mathbf{d}^{\text{aver}}(\omega)$.

Proof. Any average-bounded test is also a probability-bounded test, so $\mathbf{d}^{\text{aver}}(\omega) \overset{+}{<} \mathbf{d}^{\text{prob}}(\omega)$.

For the other direction, let d be a probability-bounded test (in the logarithmic scale). Let us show that $d - 2\log d$ is an average-bounded test. Indeed, the probability of the event "$d(\omega)$ is between $i-1$ and i" does not exceed $1/2^{i-1}$, the integral of $2^{d-2\log d}$ over this set is bounded by $2^{-i+1}2^{i-2\log i} = 2/i^2$, and therefore the integral over the entire space converges.

It remains to note that the inequality $a \overset{+}{<} b + 2\log b$ follows from $b \overset{+}{>} a - 2\log a$. Indeed, we have $b \geq a/2$ (for large enough a), hence $\log a \leq \log b + 1$, and then $a \overset{+}{<} b + 2\log a \overset{+}{<} b + 2\log b$. □

In the general case the question of the connection between boundedness in average and boundedness in probability is addressed in paper [24]. It is shown there (and this is not difficult) that if $u\colon [1,\infty] \to [0,\infty]$ is a monotonic continuous function with $\int_1^\infty (u(t)/t^2)\,dt \leq 1$, then $u(t(\omega))$ is an average-bounded test for every probability-bounded test t, and that this condition for u cannot be improved. (Our estimate is obtained by choosing $u(t) = t/\log^2 t$.)

Remark 2.17. This statement resembles the relation between the prefix and plain description complexity. However, now the difference is bounded by the logarithm of the *deficiency* (that is, bounded independently of the length for the sequences that are close to random), not of the *complexity* (as usual), which would be normally growing with the length.

Question 1. It would be interesting to understand whether the two tests differ only by a shift of scale or in some more substantial way. Such a more substantial difference could be confirmed by two families of sequences ω_i and ω_i' for which

$$\mathbf{d}^{\text{aver}}(\omega_i) - \mathbf{d}^{\text{aver}}(\omega_i') \to +\infty$$

as $i \to \infty$, while

$$\mathbf{d}^{\text{prob}}(\omega_i) - \mathbf{d}^{\text{prob}}(\omega_i') \to -\infty.$$

The authors do not know whether such families exists.

2.4. A formula for average-bounded deficiency. Let us recall some concepts connected with the prefix description complexity. For reference, consult, for example, [19, 26].

Definition 2.18. A set of strings is called *prefix-free* if no element of it is a prefix of another element. A computable partial function $T\colon \{0,1\}^* \to \{0,1\}^*$ is called a *self-delimiting interpreter* if its domain of definition is a prefix-free set. We define the complexity $\mathrm{Kp}_T(x)$ of a string x with respect to T as the length of a shortest string p with $T(p) = x$. It is known that there is an *optimal* (self-delimiting) interpreter, that is, a (self-delimiting) interpreter U with the property that for every self-delimiting interpreter T there is a constant c such that for every string x we have $\mathrm{Kp}_U(x) \le \mathrm{Kp}_T(x) + c$. We fix an optimal self-delimiting interpreter U and denote $\mathrm{Kp}(x) = \mathrm{Kp}_U(x)$.

We also denote $\mathbf{m}(x) = 2^{-\mathrm{Kp}(x)}$ and call it the *discrete a priori probability.*

The "a priori" name comes from some interpretations of a property that distinguishes the function $\mathbf{m}(x)$ among certain "weight distributions" called semimeasures.

Definition 2.19. A function $f\colon \{0,1\}^* \to [0,\infty)$ is called a *discrete semimeasure* if $\sum_x f(x) \le 1$.

Lower semicomputable semimeasures arise as the output distribution of a randomized algorithm using a source of random numbers and outputting some word (provided the algorithm halts; with some probability, it may not halt).

It is easy to check that $\mathbf{m}(x)$ is a lower semicomputable discrete semimeasure.

Recall the following fact.

Proposition 2.20 (coding theorem)**.** *Among lower semicomputable discrete semimeasures, the function* $\mathbf{m}(x)$ *is maximal within a multiplicative constant*; *that is, for every lower semicomputable discrete semimeasure* $f(x)$, *there is a constant* c *such that* $c\,\mathbf{m}(x) \ge f(x)$ *for all* x.

The universal average-bounded randomness test $\mathbf{t}_P$ (the largest lower semicomputable function with bounded expectation) can be expressed in terms of the a priori probability (and therefore prefix complexity):

Proposition 2.21. *Let* P *be a computable measure and let* $\mathbf{t}_P$ *be the universal average-bounded randomness test with respect to* P. *Then*

$$\mathbf{t}_P(\omega) \stackrel{*}{=} \sum_{x \sqsubseteq \omega} \frac{\mathbf{m}(x)}{P(x)}.$$

(*If* $P(x) = 0$, *then the ratio* $\mathbf{m}(x)/P(x)$ *is considered to be infinite.*)

Proof. A lower semicomputable function on sequences is a limit of an increasing sequence of basic functions.

Without loss of generality, we may assume that each increase is made on some cylinder $x\Omega$. In other terms, we increase the "weight" $w(x)$ of x and let our basic function on ω be the sum of the weights of all prefixes of ω. The weights increase gradually: at any moment, only finitely many weights differ from zero. In terms of weights, the average-boundedness condition translates into

$$\sum_x P(x)w(x) \le 1,$$

so, after multiplying the weights by $P(x)$, this condition corresponds exactly to the semimeasure requirement. Let us note that due to the computability of P, the lower semicomputability is conserved in both directions (when multiplying or dividing by $P(x)$). More formally, the function

$$\sum_{x \sqsubseteq \omega} \frac{\mathbf{m}(x)}{P(x)}$$

is a lower semicomputable average-bounded test: its integral is exactly $\sum_x \mathbf{m}(x)$. On the other hand, every lower semicomputable test can be presented in terms of an increase of weights, and the limits of these weights, multiplied by $P(x)$, form a lower semicomputable semimeasure. (Note that the latter transformation is not unique: we can redistribute the weights among a string and its continuations without altering the sum over the infinite sequences.) $\square$

Note that we used that both P (in the second part of the proof) and $1/P$ (in the first part) are lower semicomputable.

In Proposition 2.21, we can replace the sum with the least upper bound. This way, the following theorem connects three quantities, $\mathbf{t}_P$, the supremum and the sum, all of which are equal within a multiplicative constant.

Theorem 2.22. *We have* $\mathbf{t}_P(\omega) \stackrel{*}{=} \sup_{x\sqsubseteq\omega} \frac{\mathbf{m}(x)}{P(x)} \stackrel{*}{=} \sum_{x\sqsubseteq\omega} \frac{\mathbf{m}(x)}{P(x)}$, *or in logarithmic notation*

$$\mathbf{d}_P(\omega) \stackrel{+}{=} \sup_{x\sqsubseteq\omega}(-\log P(x) - \mathrm{Kp}(x)). \tag{3}$$

Proof. The supremum is now smaller, so only the second part of the proof of Proposition 2.21 should be reconsidered.

For a given test t we consider functions t_i (for all $i \in \mathbb{Z}$) defined as follows: $t_i(\omega) = 2^i$ if $t_i(\omega) > 2^i$, and $t_i(\omega) = 0$ otherwise. All t_i are uniformly lower semicomputable, and their sum coincides with t (up to factor 2 in both directions). Also for every ω the sum $\sum_i t_i(\omega)$ exceeds $\sup_i t_i(\omega)$ at most by a factor of 2.

Let us convert every t_i into a sum of weights (as described above). Since t_i has only two values (zero and some other value), we may assume that the vertices of nonzero weight form a prefix-free set (each path meets at most one such vertex).

Summing the weights along every path, we get $\sum_i t_i(\omega)$, or the universal test $\mathbf{t}(\omega)$ (up to a constant factor). And if we replace the sum by the supremum, we get a smaller function, but at most twice smaller (since the summands are different powers of 2). Note that for each t_i the sum is equal to the supremum since only one term is nonzero (for each ω).

More formally, the lower semicomputable function $\lceil \mathbf{d}_P(\omega)\rceil$ can be obtained as the supremum of a sequence of integer-valued basic functions of the form $k_i g_{x_i}(\omega)$, where $g_x(\omega) = 1_{x\Omega}(\omega) = 1$ if $x \sqsubseteq \omega$ and 0 otherwise. We can also require that if $i \neq j$ and $x_i \sqsubseteq x_j$, then $k_i \neq k_j$: indeed, suppose $k_i = k_j$. If $i < j$, then we can delete the jth element, and if $i > j$, then we can replace $2^{k_i} g_{x_i}$ with the sequence of all functions $2^{k_i} g_z$ where z has the same length as x_j but differs from it. We have

$$2\mathbf{t}_P(\omega) \geq 2^{\lceil \mathbf{d}_P(\omega)\rceil} = \sup_i 2^{k_i} g_{x_i}(\omega) = \sup_{x_i\sqsubseteq\omega} 2^{k_i} \geq 2^{-1} \sum_{i\colon x_i\sqsubseteq\omega} 2^{k_i} = 2^{-1} \sum_i 2^{k_i} g_{x_i}(\omega).$$

The last inequality holds since, according to our assumption, all the values k_i belonging to prefixes x_i of the same sequence ω are different, and the sum of different powers of 2 is at most twice larger than its largest element. Integrating with respect to P, we obtain $4 \geq \sum_i 2^{k_i} P(x_i)$; hence $2^{k_i} P(x_i) \stackrel{*}{<} \mathbf{m}(x_i)$ by the maximality of $\mathbf{m}(x)$, so $2^{k_i} \stackrel{*}{<} \frac{\mathbf{m}(x_i)}{P(x_i)}$. We found

$$\mathbf{t}_P(\omega) \stackrel{*}{<} \sup_{i\colon x_i\sqsubseteq\omega} \frac{\mathbf{m}(x_i)}{P(x_i)} \leq \sup_{x\sqsubseteq\omega} \frac{\mathbf{m}(x)}{P(x)}. \quad \square$$

Here is a reformulation:

$$\mathbf{d}_P(\omega) \stackrel{+}{=} \sup_n\bigl(-\log P(\omega(1:n)) - \mathrm{Kp}(\omega(1:n))\bigr).$$

This reformulation can be generalized:

Theorem 2.23. *Let $n_1 < n_2 < \ldots$ be an arbitrary computable sequence of natural numbers. Then*

$$\mathbf{d}_P(\omega) \stackrel{+}{=} \sup_k\bigl(-\log P(\omega(1:n_k)) - \mathrm{Kp}(\omega(1:n_k))\bigr).$$

The constant in $\stackrel{+}{=}$ depends on the sequence n_k.

Proof. Every step of the proof of Theorem 2.22 generalizes to this case straightforwardly. $\square$

This theorem has interesting implications in the case when instead of a sequence ω we consider an infinite two-dimensional array of bits. (From the topological and measure-theoretical viewpoints there is no difference between one-dimensional and two-dimensional bit arrays, so the notion of randomness is naturally defined for two-dimensional arrays.) Then for the randomness deficiency, it is sufficient to compare complexity and probability of squares starting at the origin.

Historical digression. The above formula for the randomness deficiency is a quantitative refinement of the following criterion.

Theorem 2.24 (criterion of randomness in terms of prefix complexity)**.** *A sequence ω is random with respect to a computable measure P if and only if the difference $-\log P(x) - \mathrm{Kp}(x)$ is bounded above for its prefixes.*

(Indeed, the last theorem says that the maximum value of this difference over all prefixes is exactly the average-bounded randomness deficiency.) This characterization of randomness was first announced, without proof, in [4], with the proof attributed to Schnorr. The first proof, for the case of a computable measure, appeared in [9].

The historically first clean characterizations of randomness in terms of complexity were given by Levin and Schnorr independently in [16] and [23]. They have a similar form, but use complexity and a priori probability coming from a different kind of interpreter called "monotonic." (In the cited work, Schnorr uses a slightly different form of complexity, but later he also adopted the version introduced by Levin.)

Definition 2.25 (monotonic complexity)**.** Let us call two strings *compatible* if one is the prefix of the other. An enumerable subset $A \subseteq \{0,1\}^* \times \{0,1\}^*$ is called a *monotonic interpreter* if for any p, p', q and q', if $(p,q) \in A$, $(p',q') \in A$ and p is compatible with p', then q is compatible with q'. For an arbitrary finite or infinite $p \in \{0,1\}^* \cup \Omega$, we define

$$A(p) = \sup\{x \colon \exists p' \sqsubseteq p \ (p',x) \in A\}.$$

The monotonicity property implies that this limit, also in $\{0,1\}^* \cup \Omega$, is well defined.

We define the (monotonic) complexity $\mathrm{Km}_A(x)$ of a string x with respect to A as the length of a shortest string p with $A(p) \sqsupseteq x$. It is known that there is an *optimal* monotonic interpreter, where optimality has the same sense as above, for prefix complexity. We fix an optimal monotonic interpreter V and denote $\mathrm{Km}(x) = \mathrm{Km}_V(x)$.

Remark 2.26 (oracle computation)**.** A monotonic interpreter is a slightly generalized version of what can be accomplished by a Turing machine with a one-way read-only input tape containing a finite or infinite string p. The machine also has a working tape and a one-way output tape. In the process of work, a finite or infinite sequence $T(p)$ appears on this tape. The work may stop, if the machine halts or passes beyond the limit of the input word; it may continue forever otherwise. It is easy to check that the map $p \mapsto T(p)$ is a monotonic interpreter (though not all monotonic interpreters correspond to such machines, resulting in a somewhat narrower class of mappings).

These machines can be viewed as the definition of what we will later call *oracle computation*, namely, a computation that uses p as an oracle.

In our applications, such a machine would have the form $T(p,\omega)$ where the machine works on both infinite strings p and ω as an input, but considers p the oracle and ω the string it is testing for randomness.

The class of mappings is narrower indeed. Let S be an undecidable recursively enumerable set of integers. Set $T(0^n1) = 0$ for all $n \in S$, and $T(0^n10) = 0$ for all n. Now, after reading 0^n1, the machine T has to decide whether to output a 0 before reading the next bit, which is deciding the undecidable set S. It is unknown to us whether this class of mappings also yields a different monotonic complexity.

A monotonic interpreter will also give rise to something like a distribution over the set of finite and infinite strings.

Definition 2.27. Let us feed a sequence of independent random bits into a monotonic interpreter A and consider the output distribution on the finite and infinite sequences. Denote by $M_A(x)$ the probability that the output sequence begins with x. Denote $\mathrm{KM}_A(x) = -\log M_A(x)$.

Recall that Λ denotes the empty string. A function $\mu\colon \{0,1\}^* \to [0,1]$ is called a *continuous semimeasure* over the Cantor space Ω if $\mu(\Lambda) = 1$ and $\mu(x) \geq \mu(x0) + \mu(x1)$ for all $x \in \{0,1\}^*$.

It is easy to check that $M_A(x)$ is a lower semicomputable continuous semimeasure.

Using the same construction as for the optimal interpreter, we can easily prove the following statement (see [30]):

Proposition 2.28. (a) *Every lower semicomputable continuous semimeasure is the output distribution of some monotonic interpreter.*

(b) *Among lower semicomputable continuous semimeasures, there is one that is maximal within a multiplicative constant.*

Definition 2.29 (continuous a priori probability)**.** Let us fix a maximal lower semicomputable continuous semimeasure and denote it $M(x)$. We sometimes call $M(x)$ the *continuous a priori probability* or the *a priori probability on a tree.*

The relation between the continuous a priori probability and monotone complexity is similar in form to the relation between the discrete a priori probability and prefix complexity. However, in this case the function $2^{-\mathrm{Km}(x)}$, still being a lower semicomputable continuous semimeasure, is not maximal (as shown in [10]). In other terms, $\mathrm{KM}(x) = -\log M(x)$ does not exceed $\mathrm{Km}(x)$, but can be smaller (the difference is unbounded).

Now, the characterization by Levin (and a similar one by Schnorr) is the following. Its proof, technically not difficult, can be found in [7, 19, 26].

Proposition 2.30. *Let P be a computable measure over Ω. Then the following properties of an infinite sequence ω are equivalent*:

(i) *ω is random with respect to P*;

(ii) $\limsup_{x\sqsubseteq\omega}[-\log P(x) - \mathrm{Km}(x)] < \infty$;

(iii) $\liminf_{x\sqsubseteq\omega}[-\log P(x) - \mathrm{Km}(x)] < \infty$;

(iv) $\limsup_{x\sqsubseteq\omega}[-\log P(x) - \mathrm{KM}(x)] < \infty$;

(v) $\liminf_{x\sqsubseteq\omega}[-\log P(x) - \mathrm{KM}(x)] < \infty$.

Theorem 2.24 proved above adds to this one more equivalent characterization, namely, that $-\log P(x) - \mathrm{Kp}(x)$ is bounded above. It is different in nature from the one in Proposition 2.30: indeed, the expressions $-\log P(x) - \mathrm{Km}(x)$ and $-\log P(x) - \mathrm{KM}(x)$ are *always bounded from below* by a constant depending only on the measure P (and not on x or ω), while $-\log P(x) - \mathrm{Kp}(x)$ is not.

Moreover, in the latter we cannot replace lim sup with lim inf, as the following example shows. Note that we can add to every string x some bits to achieve $\mathrm{Kp}(y) \geq |y|$ (where $|y|$ is the length of the string y). Indeed, if this was not so, then for the continuations of the string x we would have $\mathbf{m}(y) \geq 2^{-|y|}$, and the sum $\sum_y \mathbf{m}(y)$ would be infinite. Let us build a sequence, adding alternatingly long stretches of zeros to make the complexity substantially less than the length, and then bits that again bring the complexity up to the length (as shown, this is always possible). Such a sequence will not be random with respect to the uniform measure (since the lim sup of the difference is infinite), but has infinitely many prefixes for which the complexity is not less than the length, making the lim inf finite.

The following statement is interesting since no direct proof of it is known: the proof goes through applying Theorem 2.23 and noting that since the permutation of terms of the sequence does not change the coin-tossing distribution, it does not change the notion of randomness. More general theorems of this type, under the name of *randomness conservation*, can be found in [17, 18, 11].

Corollary 2.31. *Consider the uniform distribution* (*coin-tossing*) P *over binary sequences. The maximal difference between* $|x|$ *and* $\mathrm{Kp}(x)$ *for prefixes* x *of a random sequence is invariant* (*up to a constant*) *under any computable permutation of the sequence terms.* (*The constant depends on the permutation, but not on the sequence.*)

Here is another corollary, a reformulation of Proposition 2.21:

Corollary 2.32 (Miller–Yu "ample excess" lemma)**.** *A sequence* ω *is random with respect to a computable measure* P *if and only if*

$$\sum_{x \sqsubseteq \omega} 2^{-\log P(x) - \mathrm{Kp}(x)} < \infty.$$

This corollary also implies the fact already mentioned above:

Corollary 2.33. *Every finite sequence* x *has an extension* y *with* $\mathrm{Kp}(y) > |y|$.

Proof. Take ω random. Then $x\omega$ is random, and therefore by the Miller–Yu lemma $x\omega$ has arbitrarily long prefixes whose complexity is larger than the length. □

2.5. Game interpretation. The formula for the average-bounded deficiency can be interpreted in terms of the following game. Alice and Bob make their moves having no information about the opponent's move. Alice chooses an infinite binary sequence ω, Bob chooses a finite string x. If x turns out to be a prefix of ω, then Alice pays Bob 2^n where n is the length of x. (This version of the game corresponds to the uniform Bernoulli measure; in the general case Alice pays $1/P(x)$.) Recall the game-theoretic notions of *pure strategy*, as a deterministic choice by a player, and *mixed strategy*, as a probability distribution over deterministic choices.

Bob has a trivial strategy (choosing the empty string) that guarantees him 1 whatever Alice does. Also Alice has a mixed strategy (the uniform distribution, or, in the general case, P) that guarantees her the average loss 1 whatever Bob does. Bob can devise a strategy that will benefit him in case (for whatever reason) Alice brings a nonrandom sequence.

A randomized algorithm that has no input and produces a string (or nothing) can be considered a mixed strategy for Bob (if the algorithm does not produce anything, Bob gets no money). For any such algorithm D the expected payment (if Alice produces ω according to the distribution P) does not exceed 1. Therefore, the set of sequences ω where the expected payment (averaged over Bob's random bits) is infinite, is a null set. Observe the following:

(i) For every probabilistic strategy of Bob, his expected gain (as a function of Alice's sequence) is an average-bounded test. (From here it already follows that this expected value will be finite, if Alice's sequence is random in the sense of Martin-Löf.)

(ii) If $m(x)$ is the probability of x as Bob's move with algorithm D, his expected gain against ω is equal to

$$\sum_{x \sqsubseteq \omega} \frac{m(x)}{P(x)}.$$

(iii) Therefore, if we take the algorithm outputting the discrete a priori probability $\mathbf{m}(x)$, then Bob's expected gain will be a universal test (by the proved formula for the universal test).

Using the a priori probability as a mixed strategy enables Bob to punish Alice with an infinite penalty for any non-randomness in her sequence.

One can consider more general strategies for Bob: he can take as a pure strategy not only a string x, but some basic function f on Ω with nonnegative values. Then his gain for the sequence ω brought by Alice is set to $f(\omega)/\int f(\omega)\,dP$. (The denominator makes the expected return equal to 1.) To the move x corresponds the basic function that assigns $2^{|x|}$ to extensions of x and zero elsewhere. This extension does not change anything, since this move is a mixed strategy and we allow Bob to mix his strategies anyway. (After producing f, Bob can make one more randomized step and choose some of the intervals on which f is constant, with an appropriate probability.) In this way we get another formula for the universal test:

$$\mathbf{t}_P(\omega) \stackrel{*}{=} \sum_f \frac{\mathbf{m}(f)f(\omega)}{\int f(\omega)\,dP},$$

where the sum is taken over all basic functions f. This formula might be useful in more general situations (not for the Cantor space) where we do not work with intervals and consider some class of basic functions instead.

On concluding this part, let us point to a similar game-theoretical interpretation of probability theory developed in book [25] by Shafer and Vovk. There, the randomness of an object is not its property but, roughly speaking, a kind of guarantee with which it is being sold.

3. FROM TESTS TO COMPLEXITIES

Formula (3) expresses the randomness deficiency (the logarithm of the universal test) of an infinite sequence in terms of complexities of its finite prefixes. A natural question arises: can we go in the other direction? Is it possible to express the complexity of a finite string x, or some kind of “randomness deficiency” of x, in terms of the deficiencies of x's infinite extensions? Proposition 2.6 and the discussion following it already brought us from infinite sequences to finite ones. This can also be done for the universal test:

Definition 3.1. Fix some computable measure P, and let t be any (average-bounded) test for P. For any finite string x let $\bar{t}(x)$ be the minimal deficiency of all infinite extensions of x:

$$\bar{t}(x) = \inf_{\omega \sqsupseteq x} t(\omega).$$

By Proposition 2.7, $\bar{t}$ is a lower semicomputable function defined on finite strings, and the function t can be reconstructed back from $\bar{t}$; so if $\mathbf{t}_P$ is our fixed universal test, then $\bar{\mathbf{t}}_P$ can be considered as a version of randomness deficiency for finite strings.

The intuitive meaning is clear: a finite sequence z looks nonrandom if *all* infinite sequences that have prefix z look nonrandom.

Question 2. Kolmogorov [14] had a somewhat similar suggestion: for a given sequence z we may consider the minimal deficiency (with respect to the uniform distribution, i.e., a difference between length and complexity) of all its *finite* extensions. Are there any formal connections?

Let us spell out what we found, in more general terms.

Definition 3.2 (extended test for a computable measure)**.** A lower semicomputable monotonic (with respect to the prefix relation) function $T\colon \{0,1\}^* \to [0,\infty]$ is called an *extended test* for a computable measure P if for all N the average over words of length N is bounded by 1:

$$\sum_{x\colon |x|=N} P(x)T(x) \le 1.$$

Monotonicity guarantees that the sum over words of a given length can be replaced by the sum over an arbitrary finite (or even infinite) prefix-free set S:

$$\sum_{x\in S} P(x)T(x) \le 1. \tag{4}$$

(Indeed, extend the words of S to some common greater length.)

Proposition 3.3. *Every extended test generates* (*in the sense of Definition* 2.5) *some average-bounded test on the infinite strings. Conversely, every average-bounded test on the infinite sequences is generated by some extended test.*

Proof. The first part follows immediately from the definition (and the theorem of monotone convergence under the integral sign). In the opposite direction, we can set, for example, $T(x) = \bar{t}(x)$, or refer to Proposition 2.4 if we do not want to rely on compactness. $\square$

The existence of a universal extended test is proved by the usual methods:

Proposition 3.4. *Among the extended tests $T(x)$ for a computable measure $P(x)$, there is a maximal one, up to a multiplicative constant.*

Definition 3.5. Let us fix some dominating extended test and call it the *universal* extended test.

Proposition 3.6. *The universal extended test coincides with $\overline{\mathbf{t}}_P(x)$ to within a bounded factor.*

Proof. Since $\overline{\mathbf{t}}_P$ is an extended test, it is not greater than the universal test (to within a bounded factor). On the other hand, the universal extended test generates a test on the infinite sequences; it just remains to compare it with the maximal one. $\square$

If our space is not compact (say, it is the set of infinite sequences of integers), then $\overline{\mathbf{t}}_P(x)$ is not defined, but there is still a universal extended test, which we will denote by $\mathbf{t}_P(x)$.

Warning: not all extended tests generating $\mathbf{t}_P(\omega)$ are maximal. (For example, one can make the test equal to zero on all short words, transferring its values to its extensions.)

The advantage of the function $\mathbf{t}_P(x)$ is that it is defined on finite strings; condition (4) (for finite sets S) imposed on it is also more elementary than the integral condition, but clearly implies that it generates a test.

The method just shown is not the only way to move to tests on prefixes from tests on infinite sequences:

Definition 3.7. Assume that the computable measure P is positive on all intervals: $P(x) > 0$ for all x. Let $\widehat{\mathbf{t}}_P(x)$ be the conditional expected value of $\mathbf{t}_P(\omega)$ if a random variable $\omega \in \Omega$ has distribution P and the condition is $\omega \sqsupseteq x$. In other terms: let $\widehat{\mathbf{t}}_P(x)$ be the average of $\mathbf{t}_P$ on the interval $x\Omega$, that is, let $\widehat{\mathbf{t}}_P(x) = U(x)/P(x)$ where

$$U(x) = \int\limits_{x\Omega} \mathbf{t}_P(\omega)\, dP(\omega).$$

The function U is a lower semicomputable semimeasure. (It is even a measure, but the measure is not guaranteed to be computable and the measure of the entire space Ω is not necessarily 1. In other words, we get a measure on Ω that has density $\mathbf{t}_P$ with respect to P.) This implies that the function $\widehat{\mathbf{t}}_P(x)$ is a martingale, according to the following definition.

Definition 3.8. A function $g\colon \{0,1\}^* \to \mathbb{R}$ is called a *martingale* with respect to the probability measure P if

$$P(x)g(x) = P(x0)g(x0) + P(x1)g(x1)$$

for any string x. It is a *supermartingale* if at least the inequality $\ge$ holds here.

Note that, as a martingale, the function $\widehat{\mathbf{t}}_P(x)$ is *not* monotonically increasing with respect to the prefix relation.

Theorem 3.9.

$$\frac{\mathbf{m}(x)}{P(x)} \stackrel{*}{<} \mathbf{t}_P(x) \stackrel{*}{<} \widehat{\mathbf{t}}_P(x) \stackrel{*}{<} \frac{M(x)}{P(x)}, \tag{5}$$

where $\mathbf{m}$ *is the a priori probability on strings as isolated objects* (*whose logarithm is minus the prefix complexity*) *and* M *is the continuous a priori probability as introduced in Definition* 2.29.

Proof. In fact, the first inequality can be made stronger: we can replace $\mathbf{m}(x)/P(x)$ by $\sum_{t\sqsubseteq x}\mathbf{m}(t)/P(t)$. Indeed, this sum is a part of the expression for $\mathbf{t}_P(\omega)$ for every ω that starts with x.

The second inequality uses Proposition 3.6 and relates the minimal and average values of a random variable. The third inequality just compares the lower semicomputable semimeasure $U(x)$ and the maximal semimeasure $M(x)$. $\square$

Note that while $\widehat{\mathbf{t}}_P(x)$ is a martingale, $\frac{M(x)}{P(x)}$ is a supermartingale; it is actually maximal up to a multiplicative constant among the lower semicomputable supermartingales for P.

Remarks 3.10. 1. We may insert

$$\stackrel{*}{<} \max_{t\sqsubseteq x}\frac{\mathbf{m}(t)}{P(t)} \stackrel{*}{<} \sum_{t\sqsubseteq x}\frac{\mathbf{m}(t)}{P(t)} \stackrel{*}{<} \tag{6}$$

between the first and the second terms of (5).

2. Using the logarithmic scale, we get

$$-\log P(x) - \mathrm{Kp}(x) \stackrel{+}{<} \log \mathbf{t}_P(x) \stackrel{+}{<} \log \widehat{\mathbf{t}}_P(x) \stackrel{+}{<} -\log P(x) - \mathrm{KM}(x).$$

3. The measure U depends on P (recall that U is a maximal measure that has density with respect to P), so for different P's, for example with different supports, like the Bernoulli measures with different parameters, we get different measures. But this dependence is bounded by the inequality above: it shows that the possible variations do not exceed the difference between $\mathrm{Kp}(x)$ and $\mathrm{KM}(x)$.

4. The rightmost inequality cannot be replaced by an equality. For example, let P be the uniform (coin-tossing) measure. Then the value of $U(x)$ tends to 0 when x is an increasing prefix of a computable sequence (we integrate over decreasing intervals whose intersection is a singleton that has zero uniform measure). On the other hand, the value $M(x)$ is bounded by a positive constant for all these x.

5. We used compactness (the finiteness of the alphabet $\{0,1\}$) in proving Proposition 2.7. But we could have used Proposition 2.6 and the discussion following it for a starting point, obtaining analogous results for the Baire space of infinite sequences of natural numbers.

All quantities listed in Theorem 3.9 can be used to characterize randomness: a sequence ω is random if the values of the quantity in question are bounded for its prefixes. Indeed, the Levin–Schnorr theorem guarantees that for a random sequence the right-hand side is bounded, and for a nonrandom one the left-hand side is unbounded. The monotonicity of the second term guarantees that all expressions except the first one tend to infinity. As we already mentioned above, one cannot say this about the first quantity.

Question 3. Some quantities used in the theorem ($\mathbf{t}_P(x)$ and two added ones in (6)) are monotonic (with respect to the prefix partial order of x) by definition. We have seen that $\widehat{\mathbf{t}}_P(x)$, as a martingale, is not monotonic. What can be said about $\frac{M(x)}{P(x)}$?

All these quantities are "almost monotonic" since they do not differ much from the monotonic ones.

4. BERNOULLI SEQUENCES

One can try to define randomness not only with respect to some fixed measure but also with respect to some family of measures. Intuitively a sequence is random if we can believe that it is obtained by a random process that respects *one of* these measures. As we show later, this definition can be given for any *effectively compact* class of measures. But to make it more intuitive, we start with a specific example, *Bernoulli measures.*

4.1. Tests for Bernoulli sequences.

The Bernoulli measure B_p arises from independent tossing of a nonsymmetric coin, where the probability of success p is some real number in $[0,1]$ (the same for all trials). Note that we do not require p to be computable.

Definition 4.1 (average-bounded Bernoulli test)**.** A lower semicomputable function t on infinite binary sequences is a *Bernoulli test* if its integral with respect to any B_p does not exceed 1.

Proposition 4.2 (universal Bernoulli test)**.** *There exists a universal* (*maximal up to a constant factor*) *Bernoulli test.*

Proof. A lower semicomputable function is the monotonic limit of basic functions. If the integral of a given basic function with respect to every B_p is less than or equal to 1 for all p, this fact can be established effectively (indeed, the integral is a polynomial in p with rational coefficients). This allows us to eliminate all functions unfit to be tests, and to list all Bernoulli tests. Adding these up with appropriate coefficients, we obtain a universal one. □

Definition 4.3. We fix a universal Bernoulli test and denote it $\mathbf{t}_{\mathcal{B}}(\omega)$. Its logarithm will be called the *Bernoulli deficiency* $\mathbf{d}_{\mathcal{B}}(\omega)$. A sequence is called a *Bernoulli sequence* if its Bernoulli deficiency is finite.

Again, we may modify the definition to within an additive constant, to make it nonnegative and integer.

The informal motivation is the following: ω is a Bernoulli sequence if the claim that it is obtained by independent coin tossing (coin symmetry is not required) looks plausible. And this statement is not plausible if one can formulate some property that is true for ω but defines an "effectively Bernoulli null set" (we did not formally introduce this notion, but could, analogously to effective null sets).

Analogously to the case of computable measures, we can extend the class test to finite sequences:

Definition 4.4 (extended Bernoulli test)**.** A lower semicomputable monotonic function $T\colon \{0,1\}^* \to [0,\infty]$ is called an *extended Bernoulli test* if for all natural numbers N and for all $p \in [0,1]$ the inequality $\sum_{x\colon |x|=N} B_p(x)T(x) \le 1$ holds.

As for computable measures, there is a connection between tests for finite and tests for infinite sequences:

Proposition 4.5. *Every extended Bernoulli test generates a Bernoulli test over Ω. On the other hand, every Bernoulli test over Ω is generated by some extended Bernoulli test.*

There is a dominating universal extended Bernoulli test: it generates a universal Bernoulli test on Ω. As earlier, we will use the same notation $\mathbf{t}_{\mathcal{B}}$ for the maximal tests on the finite and on the infinite sequences. Of course, it generates a universal Bernoulli test.

4.2. Other characterizations of the Bernoulli property.

Just as for the randomness with respect to computable measures, several equivalent definitions exist. One may consider probability-bounded tests (the probability of the event $t(\omega) > N$ on any of the measures B_p must be not greater than $1/N$). Following Martin-Löf's definition for the computable measures, by a test one may mean any computable sequence of effectively open sets U_i with $B_p(U_i) \le 2^{-i}$ for all i and all $p \in [0,1]$. All these variant definitions are equivalent (and this is proved just as for randomness with respect to computable measures).

In fact, the initial definition of Bernoulli sequences given by Martin-Löf in [20] was a bit different. We will show now that in fact it gives an equivalent notion, but this equivalence requires some work.

Notation 4.6. Let $\mathbb{B}(n,k)$ denote the set of binary strings of length n with k ones (and $n-k$ zeros).

Martin-Löf defined a Bernoulli test as a family of sets of words $U_1 \supseteq U_2 \supseteq U_3 \supseteq \dots$; each of these sets is hereditary upward, that is, for every word contains all of its extensions. The following restriction is imposed on these sets: consider arbitrary integer $n \ge 0$ and k from 0 to n; it is required that for all i the share of words in $\mathbb{B}(n,k)$ belonging to U_i is not greater than 2^{-i}.

For convenience of comparison let us replace the sets U_i with an integer-valued lower semicomputable function d for which $U_i = \{x\colon d(i) \ge i\}$. The hereditary property of the sets U_i implies the monotonicity of this function d with respect to the prefix relation. Besides this, it is required that the probability of the event $d \ge i$ within each set $\mathbb{B}(n,k)$ is not greater than 2^{-i}. Clearly, these requirements correspond to probability-bounded extended tests (in the logarithmic scale), only in place of the class B_p on words of length n another set of measures is considered, namely, those concentrated on words of a given length with a given number of ones. The measures in the class B_p take equal values on words of equal lengths with equal number of ones, and are therefore representable by a mixture of uniform measures on $\mathbb{B}(n,k)$ with some coefficients. Replacing B_p with these measures makes the condition stronger.

Let us show that, nonetheless, the set of Bernoulli sequences does not change under such a replacement; moreover, the universal test (as a function on infinite sequences) does not change (as usual, to within a bounded factor). We will show this for the average-bounded variant of tests (changing Martin-Löf's definition accordingly); this does not change the class of Bernoulli sequences. The reasoning is analogous for the probability-bounded tests.

Definition 4.7. A *combinatorial Bernoulli test* is a function $f\colon \{0,1\}^* \to [0,\infty]$ with the following constraints:

(a) it is lower semicomputable;

(b) it is monotonic with respect to the prefix relation;

(c) for all integer n and k with $0 \le k \le n$, the average of the function f on the set $\mathbb{B}(n,k)$ remains below 1:

$$|\mathbb{B}(n,k)|^{-1} \sum_{x \in \mathbb{B}(n,k)} f(x) \le 1.$$

The last condition says that not only is the average of $f(x)$ bounded by 1 over the set $\{0,1\}^n$, as in extended tests for the unbiased coin-tossing measure, but its average is bounded by 1 separately in each of the sets $\mathbb{B}(n,k)$, whose union is $\{0,1\}^n$.

Having such a test for words of bounded length, one can continue it by monotonicity:

Proposition 4.8. *If a combinatorial Bernoulli test $f(x)$ is given on strings x of length less than n, then, extending it to longer strings using monotonicity, we get a function that is still a combinatorial Bernoulli test.*

Proof. We extend f to words of length n, setting $f(x0) = f(x1) = f(x)$ for words x of length $n-1$. The set $\mathbb{B}(n,k)$ consists of two parts: words ending with zero and words ending with one. The first part is in a one-to-one correspondence with $\mathbb{B}(n-1,k)$, and the second part, with $\mathbb{B}(n-1,k-1)$. The function conserves the values in this correspondence; therefore, the average in both parts is not greater than 1. Hence, the average over the whole $\mathbb{B}(n,k)$ is not greater than 1. □

The following is obtained by standard methods:

Proposition 4.9 (universal combinatorial Bernoulli test)**.** *Among combinatorial Bernoulli tests, there is one that is maximal to within a bounded factor.*

Definition 4.10. Let us fix a universal combinatorial Bernoulli test $\mathbf{b}(x)$ and extend it to infinite sequences ω by

$$\mathbf{b}(\omega) = \sup_{x \sqsubseteq \omega} \mathbf{b}(x).$$

We will call the function obtained this way a *universal combinatorial test on* Ω and will also denote it by $\mathbf{b}$.

By monotonicity, the least upper bound in this definition can be replaced with a limit. Let us show that this test coincides (to within a bounded factor) with the Bernoulli tests introduced earlier in Definition 4.3.

Theorem 4.11. $\mathbf{b}(\omega) \overset{*}{=} \mathbf{t}_{\mathcal{B}}(\omega)$.

Proof. We have already seen that $\mathbf{b}(x)$ is an extended Bernoulli test (the bounds on the average on each part $\mathbb{B}(n,k)$ imply the bound on the expected value with respect to the measure B_p, since this measure is constant on each part). Consequently, $\mathbf{b}(\omega) \overset{*}{<} \mathbf{t}_{\mathcal{B}}(\omega)$.

The converse is not true: an extended Bernoulli test may not be a combinatorial test. However, it is possible to construct a combinatorial test that takes the same values (to within a bounded factor) on the infinite sequences, and only this is asserted in the theorem.

Here is the idea. Consider an extended Bernoulli test t on words of length n and transfer it to words of much greater length N (applying the old test to their beginnings of length n). We obtain a certain function t'. We have to show that t' is close to some combinatorial test (that is, only exceeds it by a constant factor). For this, t' must be averaged over the set $\mathbb{B}(N,K)$ for an arbitrary K between 0 and N. In other words, we must average t with respect to the probability distribution on the n-bit prefixes of sequences of length N containing K ones. With $N \gg n$ this distribution will be close to the Bernoulli one with distribution $p = K/N$.

In terms of elementary probability theory, we have an urn with N balls K of which are black, and take out from it n balls. We must compare the probability distribution with the Bernoulli one that would have been obtained at sampling with replacement. Let us show that

> *for* $N = n^2$ *the distribution without replacement does not exceed the one with replacement more than* $O(1)$ *times.*

(The inequality does not hold in the other direction: for $K = 1$ without replacement we cannot obtain a word with two ones, and with replacement we can. But we only need the inequality in the given direction.)

Indeed, in sampling without replacement the probability that a ball of a given color will be drawn is equal to the quotient

$$\frac{\text{The number of remaining balls of this color}}{\text{The number of all remaining balls}}.$$

The number of balls of this color is not more than in the case with replacement; on the other hand, the denominator is at least $N - n$. Therefore, the probability of any combination during sampling without replacement is at most the probability of the same combination with replacement, multiplied by $N/(N-n)$ to the power n. For $N = n^2$ the multiplier $(1 + O(1/n))^n = O(1)$ is obtained.

This way, taking the extended Bernoulli test t and then defining $t'(x)$ on a word x of length N as t on the prefix of x of length $\lfloor \sqrt{N} \rfloor$, the obtained function t' will be a combinatorial test to within a bounded factor. (Note that its monotonicity follows from that of t.) $\square$

4.3. Criterion for Bernoulli sequences. It is natural to compare the notion of Bernoulli sequence (a sequence for which the Bernoulli test is finite) with the notion of a sequence random with respect to the measure B_p. But Martin-Löf's definition of randomness assumes that the measure is computable. Therefore, it cannot be applied directly to B_p if p is noncomputable. However, this definition can be relativized, and if (the binary expansion of) p is given as an oracle (see Remark 2.26), then the measure B_p becomes computable and randomness is well defined. The following theorem supports an intuitive idea of Bernoulli sequence as a sequence that is random with respect to some Bernoulli measure:

Theorem 4.12. *A sequence ω is a Bernoulli sequence if and only if it is random with respect to some measure B_p, with oracle $p \in [0,1]$.*

By "with oracle p," we understand the possibility to obtain from each i the ith bit in the binary expansion of the real number p (which is essentially unique, except in those cases when p is binary-rational, and in these cases both expansions are computable, and the oracle is trivial).

Before proving the theorem (even in a stronger quantitative form), we introduce a new notion of a test depending explicitly on the parameter p of the Bernoulli measure B_p, which will be later extended to arbitrary (not just Bernoulli) measures. The required result will be obtained as the combination of the following claims:

(a) Among the "uniform" randomness tests, there exists a maximal test $\mathbf{t}(\omega, p)$.

(b) The function $\omega \mapsto \inf_p \mathbf{t}(\omega, p)$ coincides (as usual, to within a bounded factor) with the universal Bernoulli test.

(c) For a fixed p, the function $\omega \mapsto \mathbf{t}(\omega, p)$ coincides (to the same precision) with the maximal randomness test for the (p-computable) measure B_p, relativized to p.

These three assertions imply Theorem 4.12 easily: a sequence ω is Bernoulli if the Bernoulli test is finite; the latter is equal to the greatest lower bound of $\mathbf{t}(\omega, p)$; hence its finiteness means $\mathbf{t}(\omega, p) < \infty$ for some p, which is equivalent to the relativized randomness with respect to the measure B_p.

We need some technical preparation. The randomness tests (as functions of two variables) will also be lower semicomputable, but the definition of this concept needs to be extended, since an additional real parameter is involved. (In what follows we will also consider a more general situation in which the second argument is a measure.)

Definition 4.13. In the space $\Omega \times [0,1]$, let us call all sets of the form $x\Omega \times (u, v)$, where $u < v$ are rational numbers, *basic rectangles*. (A technical point: we allow u and v to be outside $[0,1]$, but in this case the rectangle we mean is $x\Omega \times ([0,1] \cap (u, v))$.)

A function $f\colon \Omega \times [0,1] \to [-\infty, \infty]$ is called *lower semicomputable* if there is an algorithm that, given a rational r on its input, enumerates a sequence of basic rectangles whose union is the set of all pairs (ω, p) with $f(\omega, p) > r$.

The notion of *upper semicomputability* is defined analogously, and is equivalent to the lower semicomputability of $-f$. A function with finite real values is called *computable* if it is both upper and lower semicomputable.

This definition, as earlier, requires that the preimage of $(-\infty, r)$ be an effectively open set uniformly in r, only now we consider effectively open sets in $\Omega \times [0,1]$, defined in a natural way.

Since the intersection of effectively open sets is effectively open, the following—more intuitive—formulation is obtained for computability:

Proposition 4.14. *A real function $f\colon \Omega \times [0,1] \to \mathbb{R}$ is computable if and only if for every rational interval (u, v) its preimage is the union of a sequence of basic rectangles that are effectively enumerated, uniformly in u and v.*

The intuitive meaning of this characterization will become clearer after observing that to "give approximations to α with any given precision" is equivalent to "enumerate all intervals containing α." Therefore, for a computable function f we can find approximations to $f(\omega, p)$, if we are given appropriate approximations to ω and p.

We can reformulate the definition of (nonnegative) lower semicomputable function, introducing the notion of basic functions. It is important for us that the basic functions are continuous; therefore, the dependence on the real argument will be piecewise linear, without jumps.

Definition 4.15 (basic functions, Bernoulli case)**.** We define an enumerated list of *basic* functions $\mathcal{E} = \{e_1, e_2, \ldots\}$ over the set $\Omega \times [0,1]$ as follows. For $x \in \{0,1\}^*$, a positive integer k and rational numbers u and v with $u + 2^{-k} < v - 2^{-k}$, define the function $g_{x,u,v,k}(\omega, p)$ as follows. If $x \not\sqsubseteq \omega$, then it is 0. Otherwise, its value does not depend on ω and depends piecewise linearly on p: it is 0 if $p \notin (u, v)$, is 1 if $u + 2^{-k} \le p \le v - 2^{-k}$, and varies linearly in between. Now $\mathcal{E}$ is the smallest set of functions containing all $g_{x,u,v,k}$ and closed under maxima, minima and rational linear combinations.

Lower semicomputable functions admit the following equivalent characterization:

Proposition 4.16. *A function $f\colon \Omega \times [0,1] \to [0,\infty]$ is lower semicomputable if and only if it is the pointwise limit of an increasing computable sequence of basic functions.* (*It follows that basic functions are computable.*)

Proof. This would be completely clear if for basic functions we also allowed the indicator functions of basic rectangles and the maxima of such functions. But we want the basic functions to be continuous (this will be important in what follows). One must note therefore that for $k \to \infty$ the function $g_{x,u,v,k}$ converges to the indicator function of a rectangle. □

The continuity of the basic functions guarantees the following important property:

Proposition 4.17. *Let $f\colon \Omega \times [0,1] \to \mathbb{R}$ be a basic function. The integral $\int f(\omega, p)\, B_p(d\omega)$ is a computable function of the parameter p, uniformly in the code of the basic function f.*

(Computability is understood in the above described sense; we remark that every computable function is continuous. An analogous statement holds for an arbitrary computable function f, not only for basic functions, but we do not need this.)

The following fact, proved in [13], will be used in the present paper a number of times, also in generalizations, but with essentially the same proof.

Proposition 4.18 (trimming)**.** *Let $\varphi\colon \Omega \times [0,1] \to [0,\infty]$ be a lower semicomputable function. There is a lower semicomputable function $\varphi'(\omega, p)$ not exceeding $\varphi(\omega, p)$ with the property that for all p*

(a) $\int \varphi'(\omega, p)\, B_p(d\omega) \le 2$;

(b) *if* $\int \varphi(\omega, p)\, B_p(d\omega) \le 1$, *then* $\varphi'(\omega, p) = \varphi(\omega, p)$ *for all* ω.

Proof. By Proposition 4.16, we can represent $\varphi(\omega, p)$ as a sum of a series of basic functions $\varphi(\omega, p) = \sum_n h_n(\omega, p)$. The integral

$$\int \sum_{i \le n} h_i(\omega, p)\, B_p(d\omega)$$

is computable by Proposition 4.17, as a function of p (uniformly in n); therefore, the set S_n of all p where this integral is less than 2 is effectively open, uniformly in n.

Define now $h'_n(\omega, p)$ as $h_n(\omega, p)$ for all $p \in S_n$ and 0 otherwise. The function $h'_n(\omega, p)$ is lower semicomputable, and the integral $\int \sum_{i \le n} h'_i(\omega, p)\, B_p(d\omega)$ will be less than 2 for all p. Defining $\varphi' = \sum_n h'_n$, we obtain a lower semicomputable function, and the theorem on the integral of monotonic limits implies that $\int \varphi'(\omega, p)\, B_p(d\omega)$ is less than 2 for all p.

It remains to note that if for some p the integral $\int \varphi(\omega, p)\, B_p(d\omega)$ does not exceed 1, then this p enters all sets S_n, and the change from h_n to h'_n, as well as the change from φ to φ', does not change it. □

Now we are ready to introduce tests depending explicitly on p:

Definition 4.19. A *uniform test for Bernoulli measures* is a function t of two arguments $\omega \in \Omega$ and $p \in [0,1]$; informally, $t(\omega,p)$ measures the amount of nonrandomness ("regularity") in the sequence ω with respect to the distribution B_p. We require the following:

(a) $t(\omega,p)$ is lower semicomputable jointly as a function of the pair (ω,p);

(b) for every $p \in [0,1]$ the expected value of $t(\omega,p)$ (that is, $\int t(\omega,p)\, B_p(d\omega)$) does not exceed 1.

It remains to prove the three assertions promised earlier:

Lemma 4.20. *There exists a universal uniform test* $\mathbf{t}(\omega,p)$, *that is, a test that multiplicatively dominates all uniform tests for Bernoulli measures.*

Lemma 4.21. *For the universal uniform test* $\mathbf{t}$ *of Lemma* 4.20, *the function* $\mathbf{t}'(\omega) = \inf_p \mathbf{t}(\omega,p)$ *coincides* (*to within a bounded factor in both directions*) *with the universal Bernoulli test of Definition* 4.3.

This lemma implies that ω is a Bernoulli sequence if and only if $\mathbf{t}'(\omega)$ is finite, that is, $\mathbf{t}(\omega,p)$ is finite for some $p \in [0,1]$.

Lemma 4.22. *For a fixed* p *the function* $\mathbf{t}_p(\omega) = \mathbf{t}(\omega,p)$ *coincides* (*to within a bounded factor*) *with the universal randomness test with respect to* B_p *relativized with oracle* p.

Proof of Lemma 4.20. Generate all lower semicomputable functions; using Proposition 4.18, they can then be trimmed to guarantee that all expectations do not exceed, say, 2, and all uniform tests should get through unchanged. Sum up all the trimmed functions with coefficients whose sum is less than 1/2. □

Proof of Lemma 4.21. Let us show that $\mathbf{t}'(\omega)$ is a universal Bernoulli test. The integral of this function with respect to B_p does not exceed 1 since this function does not exceed $\mathbf{t}(\omega,p)$ for that p. The statement that this function is lower semicomputable (as a function of ω) is analogous to Proposition 2.7, and the proof is also analogous, relying on compactness. Both are special cases of the general theorem given in Proposition 7.20.

Therefore, the function $\inf_p \mathbf{t}(\omega,p)$ is a Bernoulli test. The universality (maximality) follows obviously, since any Bernoulli test can be considered a uniform Bernoulli test of two variables that does not depend on the variable p. □

A similar argument shows that for the universal extended uniform Bernoulli test $\mathbf{t}(x,p)$ (defined in a natural way; the first argument x is a binary string) the infimum $\inf_p \mathbf{t}(x,p)$ is a universal extended Bernoulli test.

Proof of Lemma 4.22. Consider first the case when p is a computable real number. Then the function $\mathbf{t}_p\colon \omega \mapsto \mathbf{t}(\omega,p)$ (where $\mathbf{t}$ is a uniform randomness test for Bernoulli measures) is lower semicomputable (we can enumerate all intervals that contain p and combine this with an algorithm for $\mathbf{t}$; in this way we represent $\mathbf{t}_p$ as the least upper bound of the computable sequence of basic functions).

A similar argument works for an arbitrary p and shows that $\mathbf{t}_p$ is lower semicomputable with a p-oracle. Thus, $\mathbf{t}_p$ does not exceed the universal relativized test with respect to B_p.

The reverse implication is a bit more difficult. Assume that t is a lower semicomputable (with oracle p) randomness test with respect to B_p. We need to find a uniform Bernoulli test t' that majorizes it (for a given p). This t' must be lower semicomputable, but now (a subtle but important point) with p as an argument of the function t', not as an oracle. In other words, one has to extend a function defined initially only for a single p to all values of p, while also guaranteeing the bound on the integral.

As a warm-up consider the case of computable p. Then no oracle is needed, and t is lower semicomputable. Adding a dummy variable p, we get a lower semicomputable function of two arguments. But this function may not be a uniform test since its expectation with respect to B_q may be arbitrary if $q \neq p$. However, Proposition 4.18 helps transform it into a t' (which will now really depend on q) with $\int t'(\omega, q) B_q(d\omega) \leq 2$ for all q and $t'(\cdot, p) = t(\cdot, p)$. Halving t' provides a uniform test.

Now consider the case of noncomputable p. In this case p is irrational, so the bits of its binary expansion can be obtained from any sequence of decreasing rational intervals that converge to p. Therefore, an oracle machine that enumerates approximations for t from below (having p as an oracle) can be transformed into a machine that enumerates from below some function $\widetilde{t}(\omega, q)$ that coincides with $t(\omega)$ if $q = p$. The function $\widetilde{t}$ may not be a uniform Bernoulli test (its expectations for $q \neq p$ can be arbitrary); but it again can be trimmed with the help of Proposition 4.18. □

5. ARBITRARY MEASURES OVER BINARY SEQUENCES

In this section, we generalize the theory to arbitrary measures, not only Bernoulli ones, but still stay in the space Ω of binary sequences.

Notation 5.1. The set of all probability measures over the space Ω is denoted by $\mathcal{M}(\Omega)$. (Recall that the measure of the whole space Ω is equal to 1.)

5.1. Uniform randomness tests.

Definition 5.2 (uniform tests)**.** A *uniform* test is a lower semicomputable function $t(\omega, P)$ of two arguments (ω is a sequence and P is a measure on Ω) with

$$\int t(\omega, P)\, P(d\omega) \leq 1$$

for every measure P.

However, we have to define carefully the notion of lower semicomputability in this case. The set $\mathcal{M}(\Omega)$ of all measures is a closed subset of the infinite (countable) product

$$\Xi = [0,1] \times [0,1] \times [0,1] \times \dots$$

(the measure is defined by the values $P(x)$ for all strings x; these values should satisfy equations (2), so we get a closed subset).

Let us introduce basic open sets and computability notions for the set $\Omega \times \mathcal{M}(\Omega)$.

Definition 5.3. An (open) *interval* (basic open set) in the space of measures is given by a finite set of conditions of type $u < P(y) < v$ where y is some binary string and u and v are some rational numbers; the basic open set consists of the measures P that satisfy these conditions. A *basic open set* in $\Omega \times \mathcal{M}(\Omega)$ has the form $x\Omega \times \beta$ (the product of intervals in Ω and $\mathcal{M}(\Omega)$), where β is a basic open set of measures. Now lower and upper semicomputability and computability are defined in terms of these basic open sets just as they were defined for $\Omega \times [0,1]$ in Definition 4.13.

In much of what follows, we will exploit the fact that, due to the finiteness of the alphabet $\{0,1\}$, the space Ω of infinite binary sequences is compact, and also the set of measures $\mathcal{M}(\Omega)$ is compact. Recall that a set C is compact if every cover of C by open sets contains a finite subcover. We need, however, an effective version of compactness:

Definition 5.4 (effective compactness)**.** A compact subset C of $\mathcal{M}(\Omega)$ is called *effectively compact* if the set

$$\left\{ S\colon S \text{ is a finite set of basic open sets and } \bigcup_{E \in S} E \supseteq C \right\}$$

is enumerable.

The set $\mathcal{M}(\Omega)$ itself is easily seen to be compact and effectively compact. It is compact, as said above, as a closed set in the product of compact spaces, and the effectivity follows from the fact that we can check whether some given basic sets cover the whole space (we are dealing with linear equations and inequalities in a finite number of variables, where everything is algorithmically decidable). This also implies the following:

Proposition 5.5. *Every effectively closed subset of $\mathcal{M}(\Omega)$ is effectively compact.*

Proof. Let an effectively closed subset C of $\mathcal{M}(\Omega)$ be the complement of the union of a list $B_1, B_2, \ldots$ of basic open sets. Then a finite set S of basic open sets covers C if and only if, together with a finite set of the B_i, it covers the whole space. And this property is decidable. □

Effective compactness implies effective closedness. This results from the following two properties of our space and our basic open sets:

(a) For every closed set F and every point x outside F there are two disjoint open sets containing F and x.

(b) For every pair of basic open sets, it is uniformly decidable whether they are disjoint.

Let F be an effectively compact set. We call a basic open set B *manifestly disjoint* from F if there is a finite set of basic open sets S disjoint from B that cover F. Due to the effective compactness of F and property (b), the set of all basic open sets manifestly disjoint from F is enumerable. Property (a) implies that it covers the complement of F.

In view of later generalization to cases where the space itself may not be compact, we will refer to some effectively closed sets of $\mathcal{M}(\Omega)$ as effectively compact.

Now we introduce a dense set of computable functions called basic functions on the set $\Omega\times\mathcal{M}(\Omega)$, similarly to Definition 4.15. Their specific form is not too important.

Definition 5.6 (basic functions for binary sequences and arbitrary measures)**.** The set of *basic* functions over the set $\Omega\times\mathcal{M}(\Omega)$ is defined analogously to Definition 4.15, starting from the functions

$$g_{x,y,u,v,k}\colon\ \Omega\times\mathcal{M}(\Omega)\to[0,1]$$

with $x,y\in\{0,1\}^*$ defined as follows. If $x\not\sqsubseteq\omega$, then $g_{x,y,u,v,k}(\omega,P)=0$. Otherwise, its value does not depend on ω and depends piecewise linearly on $P(y)$ in a way that it is 0 if $P(y)\notin(u,v)$ and 1 if $u+2^{-k}\le P(y)\le v-2^{-k}$.

The analogue of Proposition 4.16 holds again: a lower semicomputable function is the monotonic limit of a computable sequence of basic functions (which themselves are computable).

The analogue of Proposition 4.17 also holds: the integral $\int f(\omega,P)\,P(d\omega)$ of a basic function is computable as a function of the measure P, uniformly in the number of the basic function.

Finally, the analogue of Proposition 4.18 holds again:

Theorem 5.7 (trimming)**.** *Let $\varphi(\omega,P)$ be a lower semicomputable function. Then there exists a lower semicomputable function $\varphi'(\omega,P)$ such that for all P*

(a) $\int\varphi'(\omega,P)\,P(d\omega)\le 2$;

(b) *if* $\int\varphi(\omega,P)\,P(d\omega)\le 1$, *then* $\varphi'(\omega,P)=t(\omega,P)$ *for all* ω.

The proof is completely analogous to the proof we gave for Proposition 4.18.

This allows one to construct a universal test as a function of a sequence and an arbitrary measure over Ω:

Theorem 5.8. *There exists a maximal* (*to within a bounded factor*) *uniform randomness test.*

Proof. We use the same approach as before: we trim a lower semicomputable function in such a way that it becomes a test (or almost a test) and remains untouched if it were a test in the first place. □

Definition 5.9. Let us fix a universal uniform randomness test $\mathbf{t}(\omega, P)$.

We call a sequence ω *uniformly random* with respect to a (not necessarily computable) measure P if $\mathbf{t}(\omega, P) < \infty$.

Let us show that for computable measures, the new definition coincides with the old one (Definition 2.13).

Proposition 5.10. *Let P be a computable measure, let $\mathbf{t}_P(\omega)$ be a universal (average-bounded) randomness test for P, and $\mathbf{t}(\omega, P)$ the universal uniform test defined above. Then there are constants $c_1, c_2 > 0$ such that $c_1\mathbf{t}_P(\omega) \le \mathbf{t}(\omega, P) \le c_2\mathbf{t}_P(\omega)$.*

The constants c_1 and c_2 here depend on the choice of the measure P and on the choice of the test $\mathbf{t}_P$ for this measure (this choice was done in an arbitrary way for each computable measure).

This proposition shows that in the case of computable measures, uniform randomness coincides with randomness in the sense of Martin-Löf.

Proof. Let us show $\mathbf{t}(\omega, P) \le c_2\mathbf{t}_P(\omega)$ first. The function $\omega \mapsto \mathbf{t}(\omega, P)$ is lower semicomputable since we can effectively enumerate all intervals in the space of measures that contain P; therefore, it is dominated by $\mathbf{t}_P(\omega)$.

To prove $\mathbf{t}(\omega, P) \ge c_1\mathbf{t}_P(\omega)$, consider the lower semicomputable function

$$t(\omega, Q) = \mathbf{t}_P(\omega).$$

The function $(\omega, Q) \mapsto t(\omega)$ is not guaranteed to be a uniform randomness test, since its integral can be greater than 1 if $Q \ne P$. However, it can be trimmed without changing it at P, and then it still remains an (almost) test. $\square$

We are also interested in tests defined just for one, not necessarily computable, measure P:

Definition 5.11. We will call a function $f\colon \Omega \to [0, \infty]$ *lower semicomputable* relative to a measure P if it is obtained from a lower semicomputable function on the set $\Omega \times \mathcal{M}(\Omega)$ after fixing the second argument at P.

For a measure $P \in \mathcal{M}(\Omega)$, a *$P$-test of randomness* is a function $f\colon \Omega \to [0, \infty]$ lower semicomputable relative to P with the property $\int f(\omega)\,dP \le 1$.

It seems as if a P-test may capture some nonrandomnesses that uniform tests cannot; however, this is not so, since trimming (see Theorem 5.7) generalizes.

Theorem 5.12. *Let P be some measure along with some P-test $t_P(\omega)$. There is a uniform test $t'(\cdot,\cdot)$ with $t_P(\omega) \le 2t'(\omega, P)$. On the other hand, the restriction of any uniform test to the measure P is a P-test.*

The notion of extended text can be generalized to uniform tests:

Definition 5.13 (extended uniform test)**.** A lower semicomputable function $T\colon \{0,1\}^* \times \mathcal{M}(\Omega) \to [0, 1]$ monotonic with respect to the prefix relation is called an *extended uniform test* if for all n and all distributions P we have $\sum_{x\colon |x|=n} T(x, P)P(x) \le 1$.

As earlier, due to monotonicity, we could sum not only over words of a given length, but over an arbitrary prefix-free set.

The following is a corollary to the analogue of Proposition 4.16 (representing a nonnegative lower semicomputable function as a sum of nonnegative basic functions):

Proposition 5.14. *Every uniform test $t(\omega, P)$ can be generated by an extended uniform test in the sense that*

$$t(\omega, P) = \sup_{x \sqsubseteq \omega} T(x, P).$$

Conversely, every extended uniform test T generates a uniform test t.

Among the uniform extended tests, it is also possible to select a maximal one (using an analogous trimming method and summing the results). We fix an extended uniform test and denote it $\mathbf{t}(x, P)$ (where $x \in \{0,1\}^*$ and P is a measure over Ω). It generates a maximal uniform test $\mathbf{t}(\omega, P)$ (to within a bounded factor).

Remark 5.15. Much of the theory worked out at the beginning of this paper for 0–1 sequences also holds for sequences whose elements are arbitrary natural numbers. The extended tests of Definition 5.13 generalize, and the existence of a uniform universal extended test is proved in the same way. But it becomes important to define extended tests directly, not via tests for infinite sequences, since compactness may not hold.

Proposition 5.10 allows us to generalize a result about Bernoulli measures:

Theorem 5.16. *Let P be a measure computable with some oracle A. Assume also that A can be effectively reconstructed as the values of the measure are provided with more and more precision. Then a sequence ω is uniformly random with respect to P if and only if it is random with respect to P with oracle A.*

(Since the oracle A makes P computable, the notion of Martin-Löf randomness is well defined.)

Proof. Assume that $\mathbf{t}(\omega, P) = \infty$ for the universal uniform test $\mathbf{t}$. Note that $\mathbf{t}(\cdot, P)$ is an A-lower semicomputable function and is a P-test, so ω is nonrandom with respect to P with oracle A.

On the other hand, let $t(\omega, A)$ be some A-lower semicomputable P-test with $t(\omega, A) = \infty$. That A can be reconstructed from P means that there is a computable mapping f from measures to binary sequences (oracles) defined at least over P, with $A = f(P)$. But then $(\omega, P) \mapsto t(\omega, f(P))$ is a P-test. The uniformization Theorem 5.12 converts it into a uniform test that is infinite on (ω, P). □

Let us note that not all measures P satisfy the condition of the theorem (it means that the mass problem of "show approximations to the values of P" is equivalent to the decision problem of some set; on the degrees of such mass problems, see [22]). Later, in Theorem 5.36, we show a characterization of uniform randomness for arbitrary measures (in terms of Martin-Löf randomness with oracle).

Another application of the trimming technique: let us show that the notion of uniform randomness test is indeed a generalization of the notion of a uniform Bernoulli test we introduced earlier in Definition 4.19.

Theorem 5.17. *Let $\mathbf{t}(\omega, P)$ be the universal uniform test and let $\mathbf{t}(\omega, p)$ be the universal uniform Bernoulli test defined in Lemma* 4.20. *Then $\mathbf{t}(\omega, B_p) \stackrel{*}{=} \mathbf{t}(\omega, p)$.*

(Here B_p is the Bernoulli measure with parameter p.)

Proof. For the inequality $\stackrel{*}{<}$ note that the function $(\omega, p) \mapsto \mathbf{t}(\omega, B_p)$ is a uniform Bernoulli test, since the mapping $p \mapsto B_p$ is a computable mapping (in a natural sense).

For the other direction, there exists a computable function on measures that maps B_p to p (just take the probability of the one-bit string). Combining this function with $\mathbf{t}(\omega, p)$, we get a lower semicomputable function $f(\omega, P)$ on general measures P with $f(\omega, B_p) = \mathbf{t}(\omega, p)$. The function $f(\omega, p)$ is not a uniform test yet, but again the trimming technique given by Proposition 4.18 yields the desired result. □

5.2. A priori probability with an oracle, and uniform tests. For a computable measure, we had an expression for the universal test via a priori probability in Proposition 2.21. An analogous expression also exists for the universal uniform test:

Theorem 5.18.

$$\mathbf{t}(\omega, P) \stackrel{*}{=} \sum_{x \sqsubseteq \omega} \frac{\mathbf{m}(x|P)}{P(x)}.$$

To be honest, we still owe the reader the definition of the concept of a priori probability with respect to a measure, that is, the quantity $\mathbf{m}(x|P)$. We do this right away, before returning to the proof.

Definition 5.19. A nonnegative function $t(x, P)$ whose arguments are the binary word x and the measure P will be called a *uniform lower semicomputable semimeasure* if it is lower semicomputable and $\sum_x t(x, P) \le 1$ for all measures P over Ω.

Proposition 5.20. *Among the uniform lower semicomputable semimeasures, there is a largest one to within a multiplicative constant.*

This is proved by the same method as the existence of a universal test (and even simpler, since the constraints on the values of the test do not depend on the measure).

Definition 5.21. We will fix one such largest semimeasure, and call it the *a priori probability with respect to* P. We will denote it by $\mathbf{m}(x|P)$.

(The vertical bar in place of a comma emphasizes the similarity to the conditional a priori probability normally considered.)

Proof of Theorem 5.18. We need to check two things. First we need to convince ourselves that the right-hand side of the formula defines a uniform test. Every member of the sum can be considered to be a function of two arguments, equal to 0 outside the cone of extensions of x and equal to $\mathbf{m}(x|P)/P(x)$ inside the cone. For every x, the functions $\mathbf{m}(x|P)$ and $1/P(x)$ are lower semicomputable (uniformly in x), and the sum gives a lower semicomputable function. The integral of this function with respect to any measure P is equal to the sum of the integrals of the members, that is, $\sum_x \mathbf{m}(x|P)$, and therefore does not exceed 1.

There is a special case, when $P(x) = 0$ for some x. In this case the corresponding member of the sum becomes infinite for any ω extending x. But since the measure of this cone is zero, the integral with respect to this measure is by definition zero, and therefore the additive term, if is not equal to $\mathbf{m}(x|P)$, is simply smaller. This way, the right-hand side of the formula is a uniform test, and therefore it does not exceed the universal uniform test: we proved the inequality $\overset{*}{>}$.

The second part of the proof is not so simple: observing the increase of the values of the uniform test, we must distribute this increase among the different members of the sum on the right-hand side, while preserving lower semicomputability. The difficulty is that if, say, the lower semicomputable function was 1 on some effectively open set A, was zero outside it, and then this set was changed to a larger set B, then the difference (the characteristic function of $B \setminus A$) will not in general be lower semicomputable since in the set of measures (just as on a segment) the difference of two intervals will not be open.

This problem is solved by moving to continuous functions. Let us be given an arbitrary uniform test $t(\omega, P)$. Since it is lower semicomputable, it can be represented as the limit of a nondecreasing sequence of basic functions, or—passing to differences—in the form of a sum of a series of nonnegative basic functions: $t(\omega, P) = \sum_i t_i(\omega, P)$.

Being basic, the function t_i of ω depends only on a finite prefix of the sequence ω; denote the length of this prefix by n_i. For every word x of length n_i we get some lower semicomputable function $t_{i,x}(P)$, where $t_i(\omega, P) = t_{i,x}(P)$ if ω begins with x. Now define $m_i(x, P) = t_{i,x}(P)P(x)$ if x has length n_i (for the other lengths, it is zero). The function m_i is lower semicomputable (as the product of two lower semicomputable functions) uniformly in i, and therefore the sum $m(x, P) = \sum_i m_i(x, P)$ will be lower semicomputable.

Let us show that m is a semimeasure, that is, $\sum_x m(x, P) \le 1$ for all P. Indeed, in $\sum_i m_i(x, P)$ the nonzero terms correspond to words of length n_i, and this sum is equal to $\sum_x t_{i,x}(P)P(x)$, that is, exactly to the integral $\int t_i(\omega, P)\, P(d\omega)$, and the sum of these integrals does not exceed 1 by our condition.

Moreover, if for all prefixes x of the sequence ω the measure $P(x)$ is not equal to zero, then

$$\sum_{x\sqsubseteq\omega}\frac{m_i(x,P)}{P(x)}=\frac{t_{i,x_i}(P)P(x_i)}{P(x_i)}=t_i(\omega,P)$$

(here x_i is the prefix of length n_i of ω); hence after summing over i

$$\sum_{x\sqsubseteq\omega}\frac{m(x,P)}{P(x)}=t(\omega,P),$$

and it just remains to apply the maximality of the a priori probability to obtain the $\overset{*}{<}$-inequality in the case when all prefixes of ω have nonzero P-measure. On the other hand, if one of these has zero P-measure, then the right-hand side is infinite, and so here the inequality is also satisfied. □

Question 4. For the universal randomness test with respect to a computable measure, in this formula one could replace the sum with a maximum. Is this possible for uniform tests? (The reasoning applied there encounters difficulties in the uniform case.) Can one define the a priori probability on the tree in a reasonable way, and prove a uniform variant of the Levin–Schnorr theorem?

We return to the notion of a priori probability with oracle (and its relation to prefix complexity) in Subsection 7.4.

5.3. Effectively compact classes of measures. We have considered Bernoulli tests, that is, lower semicomputable functions $t(\omega)$ that are tests with respect to all Bernoulli measures. In this definition, in place of Bernoulli measures, an arbitrary effectively compact class can be taken:

Definition 5.22. Let $\mathcal{C}$ be an effectively compact class of measures over Ω. We say that a lower semicomputable function t on Ω is a *$\mathcal{C}$-test* if $\int t(\omega)\,dP\le 1$ for every $P\in\mathcal{C}$.

Theorem 5.23. *Let $\mathcal{C}$ be an effectively compact class of measures.*

(a) *There exists a universal $\mathcal{C}$-test $\mathbf{t}_{\mathcal{C}}(\cdot)$.*

(b) $\mathbf{t}_{\mathcal{C}}(\omega)=\inf_{P\in\mathcal{C}}\mathbf{t}(\omega,P)$.

Proof. Both of these statements are proved analogously to Lemmas 4.20 and 4.21. □

Remark 5.24. Since $\mathcal{C}$ is compact and the function $\mathbf{t}(\omega,P)$ is lower semicomputable, the inf-operation can be replaced by the min-operation.

Question 5. Can we give criteria for randomness with respect to natural closed classes of measures (for example, in terms of complexity)? How can we describe Bernoulli sequences in terms of complexities of their initial segments? It is known that the main term of the randomness deficiency is

$$\log\binom{n}{k}-\mathrm{Kp}(x|n,k)$$

for the beginning of x of length n with k ones. The lecture notes [7] contain a characterization of Bernoulli sequences, but it is rather messy.

What about Markov measures? Shift-invariant measures?

5.4. Sparse sequences. There are several situations closely related to some intuitive understanding of randomness, but not fitting directly into the framework of the question of randomness of a given outcome ω relative to a given model (measure P). Our example here is a natural notion of sparsity, introduced in [3], but another example, online tests, will be considered in Section 9.

It is natural to call a sequence ω "p-sparse" if its 1's come from some p-random sequence ω', but we allow some of its 0's to also come from the 1's of ω'. For example, the 1's of ω' may be a sequence of miracles, and ω is the sequence of those miracles that have been reported. The tacit hypothesis is, of course, that all reported miracles actually happened.

Definition 5.25 (sparse sequences)**.** Let us introduce a coordinate-wise order between infinite binary sequences (or binary sequences of the same length): we say that $\omega \le \omega'$ if this is true coordinate-wise, that is, $\omega(i) \le \omega'(i)$ for all i: in other words, ω' is obtained from ω by replacing some 0's with 1's.

Let B_p be a Bernoulli measure with some computable p. We say that a binary sequence ω is *p-sparse* if $\omega \le \omega'$ for some B_p-random sequence ω'. (In terms of sets, p-sparse sets are subsets of p-random sets.)

We will show that in the definition of sparsity, the existential quantifier can be eliminated, giving a criterion in terms of monotonic tests.

Definition 5.26. A real function f on Ω is called *monotonic* if $\omega' \ge \omega$ implies $f(\omega') \ge f(\omega)$.

A monotonic lower semicomputable function $t\colon \Omega \to [0,\infty]$ is a *p-sparsity test* if $\int t(\omega)\,dB_p \le 1$. A p-sparsity test is *universal* if it multiplicatively dominates all other sparsity tests for p.

The monotonicity of tests guarantees, informally speaking, that only the presence of some 1's is counted as regularity, not their absence. (Note that earlier we spoke of an entirely different kind of monotonicity, while defining extended tests: there we compared the values of a function on a finite word and its extension.)

Proposition 5.27. *Consider the universal test* $\mathbf{t}(\omega, P)$. *The expression*

$$r_p(\omega) = \min_{\omega' \ge \omega} \mathbf{t}(\omega', B_p)$$

defines a universal p-sparsity test.

Proof. Each p-sparsity test is by definition a test with respect to the measure B_p. Using its monotonicity and comparing it with the universal test, we find that no sparsity test exceeds r_p (to within a bounded factor).

In the other direction it must be shown that the minimum in the expression for r_p is achieved, and that this function is a p-sparsity test. The lower semicomputability is proved using the fact that the property $\omega \le \omega'$ gives an effectively closed set of the effectively compact space $\Omega \times \Omega$ (cf. Proposition 7.20 below). The monotonicity and the integral inequality follow immediately from the definition. $\square$

This implies the following characterization in terms of tests:

Theorem 5.28. *A sequence ω is p-sparse* (*is obtained from a p-random one by replacing some* 1*'s with zeros*) *if and only if the universal sparsity test $r_p(\omega)$ is finite.*

Sparsity is equivalent to randomness with respect to a certain class of measures. To define this class, we introduce the notion of coupling of measures.

Definition 5.29. For measures P and Q we write $P \preceq Q$ and say that P can be *coupled below* Q if there exists a probability distribution R on pairs of sequences (ω, ω') such that

(a) the first projection (marginal distribution) of R is P and the second one is Q;

(b) the measure R is entirely concentrated on pairs (ω, ω') with $\omega \le \omega'$ (the probability of this event with respect to the measure R is 1).

The following characterization of coupling is well known; it has many proofs, but all seem to go back to [28, Theorem 11, p. 436]. A proof can be found in [3].

Proposition 5.30. *The property $P \preceq Q$ is equivalent to the following*: *for all monotonic basic functions f the following inequality holds*:

$$\int f(\omega)\,dP \le \int f(\omega)\,dQ.$$

In this characterization, we could have said "all monotonic integrable functions" as well (on the other hand, we could also consider only indicator functions of upward closed sets).

Notation 5.31. Let $\mathcal{S}_p$ be the set of measures that can be coupled below B_p.

Proposition 5.32. *The set $\mathcal{S}_p$ of measures is effectively closed (and thus effectively compact).*

Proof. For each function f in Proposition 5.30, the condition defines an effectively closed set, and their intersection will also be effectively closed. □

Theorem 5.33. *The universal p-test $r_p(\omega)$ is a universal class test for the class $\mathcal{S}_p$.*

Thus, a sequence is p-sparse if and only if it is random with respect to some measure that can be coupled below B_p.

The following lemma will be key to the proof.

Lemma 5.34 (monotonization)**.** *Let $t\colon \Omega \to \mathbb{R}$ be a basic function with $\int t(\omega)\,dQ \le 1$ for all $Q \in \mathcal{S}_p$. Define the monotonic function $\widehat{t}(\omega) = \max_{\omega' \le \omega} t(\omega')$ (the maximum is achieved since $t(\omega)$ depends only on finitely many positions of ω). Then $\int \widehat{t}(\omega)\,dB_p \le 1$.*

Proof. Let the function t depend only on the first n coordinates. For each $x \in \{0,1\}^n$ fix $x' \le x$ for which $t(x')$ reaches the maximum (among all such x'). Besides the distribution B_p, consider a distribution Q in which the Bernoulli measure of x is transferred to x' (the measures of several x may be transferred to the same x' and then be added). We described the behavior of Q on the first n bits; the following bits are chosen independently, and the probability of 1 in each position is equal to p. Note also that for the expected values of the functions t and $\widehat{t}$ only the first n bits count.

By the construction, $Q \preceq B_p$ (essentially, we described a measure on pairs); therefore, $\int t(\omega)\,dQ \le 1$. But this integral is equal to $\int \widehat{t}(\omega)\,dB_p$. □

Proof of Theorem 5.33. Every p-sparsity test t is a class test for $\mathcal{S}_p$. Indeed, its integral with respect to a measure in the class $\mathcal{S}_p$ does not exceed its integral with respect to the measure B_p, by the monotonicity of the test and the possibility of coupling.

On the other hand, let us show that for every test t and for the class $\mathcal{S}_p$, there is a p-sparsity test that is not smaller. Indeed, the test t is the limit of an increasing sequence t_n of basic functions. Applying to them the monotonization Lemma 5.34, we obtain a sequence of basic functions $\widehat{t_n}$ that are everywhere greater than or equal to t_n and have integrals bounded by 1 with respect to the measure B_p. Their limit is the needed p-sparsity test. □

5.5. Different kinds of randomness. There are several ways to define randomness with respect to an arbitrary (not necessarily computable) measure. We have already defined uniform randomness. Here are some other ways.

Oracles. We can use the Martin-Löf definition (or its average-bounded version) with oracles. We would call a sequence ω random with respect to P if there exists an oracle A that makes P computable such that ω is Martin-Löf random with respect to P with oracle A. (We say "there exists an oracle that makes P computable" but not "for all oracles that make P computable"; indeed, some powerful oracle can always make ω computable and therefore nonrandom, unless ω is an atom of P.) As Adam Day and Joseph Miller have shown [6], this definition turns out to be equivalent to uniform randomness. The proof of this equivalence needs some preparation.

First let us look into why it is not possible to take as an oracle the measure itself (as was done for the Bernoulli measures, where we chose a binary expansion of the number p as an oracle). Well, the choice of such a representation is not unique ($0.01111\ldots = 0.10000\ldots$). When all we have is a single number p, this is not important, as the nonuniqueness arises only for rational p, and in this case both representations are computable. But for measures this is not so: a measure is represented by a countable number of reals (say, the probabilities of individual words, or the conditional probabilities), and the arbitrariness in the choice of a representation is not reduced to a finite number of variants.

Definition 5.35. Fix some representation of measures by infinite binary sequences, that is, a computable (and therefore continuous) mapping $\pi \mapsto R_\pi$ of Ω onto the space of measures. For example, we may split the binary sequence π into countably many parts and use these parts as binary representations of the probability that the sequence continues with 1 after a certain prefix.

Define an *r-test* (representation test, test of randomness relative to a given representation of the measure) as a lower semicomputable function $t(\omega,\pi)$ with $\int t(\omega,\pi)\,R_\pi(d\omega) \le 1$ for all π.

This notion of r-test depends on the representation method chosen; there are no intuitive reasons to choose one specific representation and declare it to be "natural," but any representation is good for the argument below and we assume some representation fixed. The following statements can be proved just as similar statements before:

(a) Every lower semicomputable function $t(\omega,\pi)$ can be trimmed and thus made not greater than twice an r-test (without changing it for those π where it already was an r-test).

(b) There exists a universal (maximal to within a bounded factor) r-test $\mathbf{t}(\omega,\pi)$.

For a fixed π, the function $\mathbf{t}(\cdot,\pi)$ is universal among the π-computable average-bounded tests with respect to the measure R_π. Indeed, it is such a test; on the other hand, any such test can be lower semicomputed by the oracle machine. This machine is applicable to any oracle (though may not give a test), giving a lower semicomputable function $t'(\omega,\pi)$ that is equal to the starting test for the given π. It remains to apply property (a).

As a consequence of this simple reasoning, we find that the quantity $\mathbf{t}(\omega,\pi)$ is finite if and only if the sequence ω is random relative to the oracle π, with respect to the measure R_π.

Theorem 5.36 (Day–Miller)**.** *A sequence ω is uniformly random with respect to a measure P if and only if there is an oracle computing P that makes ω random* (*in the original Martin-Löf sense*)*. More precisely,*

$$\mathbf{t}(\omega,P) \stackrel{*}{=} \inf_{R_\pi=P} \mathbf{t}(\omega,\pi).$$

Proof. Let us prove the equality shown in the theorem. Note that if t is a uniform test, then $t(\omega,R_\pi)$ as a function of ω and π is an r-test and is therefore dominated by the universal r-test.

The other direction is somewhat more difficult. We have to show that the function on the right-hand side is lower semicomputable as a function of the sequence ω and the measure P. (The integral condition is obtained easily afterwards, as the measure P has at least one representation π.) This can be proved using the effective compactness of the set of pairs (P,π) with $P=R_\pi$. In the general form (for constructive metric spaces) this statement forms the content of Lemma 7.21.

It remains to explain the connection between the given equality and randomness relative to an oracle. If $\mathbf{t}(\omega,P)$ is finite, then by the proved equality a π exists with $R_\pi=P$ and finite $\mathbf{t}(\omega,\pi)$. As we have seen, this in turn means that ω is random with respect to the measure P, with an oracle π that makes P computable. Conversely, if $\mathbf{t}(\omega,P)$ is infinite and some oracle A makes P computable, then the function $\mathbf{t}(\cdot,P)$ becomes A-lower semicomputable and its integral with respect to the measure P does not exceed 1; hence the sequence ω will not be random relative to the oracle A and with respect to the measure P. $\square$

Blind (*oracle-free*) *tests.* We can define the notion of an effectively null set as before, even if the measure is not computable. The maximal effectively null set may not exist. For example, if a measure P may be concentrated on some noncomputable sequence π, then all intervals not containing π will be effective null sets, and their union (the complement of the singleton $\{\pi\}$) will not be, since otherwise π would be computable.

However, we can still define a random sequence as a sequence that does not belong to *any* effectively null set. Kjos-Hanssen suggested the name "Hippocratic randomness" for this definition (referring to a certain legend about the doctor Hippocrates), but we prefer the more neutral name "blind randomness."

Definition 5.37 (blind tests)**.** A lower semicomputable function $t(\omega)$ with $\int t(\omega)\,dP(\omega) \le 1$ will be called a *blind, or oracle-free, test* for the measure P. A sequence ω is *blindly random* with respect to the measure P if $t(\omega) < \infty$ for all such blind tests.

As we have seen, there may not exist a maximal blind test.

This oracle-free notion of randomness can be characterized in the terms introduced earlier:

Theorem 5.38. *A sequence ω is blindly random with respect to a measure P if and only if ω is random with respect to any effectively compact class of measures that contains P.*

Proof. Assume first that ω is not random with respect to some effectively compact class of measures that contains P. Then the universal test with respect to this class is a blind test that shows that ω is not blindly random with respect to P.

Now assume that there exists some blind test t for the measure P with $t(\omega) = \infty$. Then just consider the class $\mathcal{C}$ of measures Q with $\int t(\omega)\,dQ \le 1$. This class is effectively closed (and thus effectively compact). Indeed, t is the supremum of a computable sequence of basic functions t_n. The class of measures Q with $\int t_n(\omega)\,dQ > 1$ is effectively open, uniformly in n, and $\mathcal{C}$ is the complement of the union of these sets. $\square$

It is easy to see from the definition (or from the preceding theorem) that uniform randomness implies blind randomness. The converse statement is not true:

Theorem 5.39. *There exist a sequence ω and a measure P such that ω is blindly random with respect to P but not uniformly random.*

Proof. Indeed, blind randomness does not change if we change the measure slightly (up to an $O(1)$-factor). On the other hand, the changed measure may have much more oracle power that makes a sequence nonrandom. For example, we may start with the uniform Bernoulli random measure $B_{1/2}$ (coin tosses with probabilities $1/2, 1/2, 1/2, \dots$) and fix some random sequence $\omega = \omega(1)\omega(2)\dots$. Then consider a (slightly) different measure B' with probabilities $1/2 + \omega(1)\varepsilon_1, 1/2 + \omega(2)\varepsilon_2, \dots$ where $\varepsilon_1, \varepsilon_2, \dots$ are so small and converge to zero so fast that they do not change the measure more than by an $O(1)$-factor while being all positive. Then B' encodes ω, which makes it easy to construct a uniform test t with $t(\omega, B') = \infty$. $\square$

However, there are some special cases (including Bernoulli measures) where uniform and blind randomness are equivalent. In order to formulate the sufficient conditions for such a coincidence, let us start with some definitions.

Definition 5.40 (effective orthogonality)**.** For a probability measure P, let $\mathrm{Randoms}(P)$ denote the set of sequences uniformly random with respect to P. A class of measures is called *effectively orthogonal* if $\mathrm{Randoms}(P) \cap \mathrm{Randoms}(Q) = \varnothing$ for any two different measures in it.

Theorem 5.41. *Let $\mathcal{C}$ be an effectively compact and effectively orthogonal class of measures. Then for every measure P in $\mathcal{C}$ the uniform randomness with respect to P is equivalent to the blind randomness with respect to P.*

The statement looks strange: we claim something about randomness with respect to a measure P, but the condition of the claim is that P can be included into a class of measures with some properties. (It would be natural to have a more direct requirement for P instead.) The theorem implies that the measures of Theorem 5.39 do not belong to any such class.

Proof. We have already noted that in one direction the statement is obviously true. Let us prove the converse. Assume that a sequence ω is blindly random with respect to a measure P. By

Theorem 5.38, it is random with respect to the class $\mathcal{C}$. So, ω is uniformly random with respect to some measure P' from the class $\mathcal{C}$. It remains to show $P = P'$.

Imagine that this is not the case. Then we can construct an effectively compact class of measures $\mathcal{C}'$ that contains P but not P'. Indeed, since P and P' are different, they assign different measures to some finite string, and this fact can be used, in the form of a closed condition separating P from P', to construct $\mathcal{C}'$. Consider now the effectively compact class $\mathcal{C} \cap \mathcal{C}'$. It contains P, and therefore ω will be random with respect to this class. Hence the class contains some measure P'' with respect to which ω is uniformly random. But $P' \neq P''$ (one measure is in $\mathcal{C}'$, while the other one is not), so we get a contradiction with the assumption on the effective orthogonality of the class $\mathcal{C}$. □

Remark 5.42. The proved theorem is applicable in particular to the class of Bernoulli measures. It is tempting to think that there is a simpler proof, at least for this case: if ω is random with respect to p, we can compute p from ω as the limit of the relative frequency, and no additional oracle is needed. This is not so: though p is *determined* by ω, it does not even depend continuously on ω. Indeed, no initial segment of the sequence guarantees that its limiting frequency is in some given interval. However, we can apply an analogous reasoning to sequences ω with the randomness deficiency bounded by some constant. (See [12], which introduces the notion of *layerwise computability.*) In particular, it can be shown that if ω is random with respect to the measure B_p, then p is computable with oracle ω.

6. NEUTRAL MEASURE

The following theorem, first published in [17] and then again in [11], points to a curious property of uniform randomness which distinguishes it from randomness using an oracle.

Definition 6.1. A measure is called *neutral* if every sequence is uniformly random with respect to it.

Theorem 6.2. *There exists a neutral measure; moreover, there is a measure N such that $\mathbf{t}(\omega, N) \le 1$ for all sequences ω.*

Before proving this theorem we note that all computable sequences are atoms of any neutral measure (have positive probability). Indeed, a test can be constructed that is looking for long segments of the computable sequence that have small measure (one can achieve this having the sequence and the measure as arguments), and assigns them large values of deficiency.

It follows that a neutral measure cannot be computable. Indeed, for a computable measure there exists a computable sequence that is not an atom (adding bits sequentially, we choose the next bit in such a way that its conditional probability is at most $2/3$). Such a sequence cannot be random with respect to N. For the same reason a neutral measure cannot be equivalent to an oracle (for a neutral measure N one cannot find an oracle A that makes it computable and at the same time can be uniformly reconstructed from every approximation of N). Indeed, in this case uniform randomness (as we have shown) is equivalent to randomness with respect to N with oracle A, and the same argument works.

A neutral measure cannot be lower or upper semicomputable either, but this statement does not seem interesting, since here a semicomputable measure over Ω is also computable. Some more meaningful (and less trivial) versions of this fact are proved in [11].

Proof of Theorem 6.2. Consider the universal test $\mathbf{t}(\omega, P)$. We claim that there exists a measure N with $\mathbf{t}(\omega, N) \le 1$ for every ω. In other terms, for every ω we have a condition on N saying that $\mathbf{t}(\omega, N) \le 1$ and we need to prove that these conditions (there is a continuum of them) have nonempty intersection. Each of these conditions is a closed set in a compact space (recall that $\mathbf{t}$ is lower semicontinuous), so it is enough to show that finite intersections are nonempty.

So let $\omega_1, \dots, \omega_k$ be k sequences. We want to prove that there exists a measure N such that $\mathbf{t}(\omega_i, N) \le 1$ for every i. This measure will be a convex combination of measures concentrated on $\omega_1, \dots, \omega_k$. So we need to prove that k closed subsets (corresponding to k inequalities) of a k-vertex simplex have a common point. It is a direct consequence of the classical topology result formulated in the following lemma (which is used in the standard proof of Brouwer's fixed-point theorem).

Lemma 6.3 (Knaster–Mazurkievicz–Kuratowski). *Let a simplex with vertices* $1, \dots, k$ *be covered by closed sets* $A_1, \dots, A_k$ *in such a way that the vertex* i *belongs to* A_i *(for every* i*), the edge* i*–*j *is covered by* $A_i \cup A_j$*, and so on (formally, the face* $(i_1, \dots, i_s)$ *of the simplex is a subset of* $A_{i_1} \cup \ldots \cup A_{i_s}$*; in particular, the union* $A_1 \cup \ldots \cup A_k$ *is the entire simplex). Then the intersection* $A_1 \cap \ldots \cap A_k$ *is not empty.*

For completeness, let us reproduce the standard proof of this lemma.

Proof. Consider a simplicial subdivision of the simplex into smaller simplices of the same dimension and mark every vertex of this subdivision by a number i from 1 to k with the condition that this vertex is in A_i. Moreover, we require that the vertex i of the original simplex be marked by number i, the vertices on the segment i–j be marked either by number i or by number j, and so on. Sperner's combinatorial lemma implies that there is a simplex in the subdivision whose vertices are labeled with all k colors. Decreasing to null the size of the maximal simplex of the subdivision and selecting a limit point in the sequence of the obtained many-colored simplices, we find a point in the intersection of all A_i. □

To show that the lemma gives us what we want, consider any point of some face. For example, let X be a measure that is a mixture of, say, ω_1, ω_5 and ω_7. We need to show that X belongs to $A_1 \cup A_5 \cup A_7$; in our terms, this means that one of the numbers $\mathbf{t}(\omega_1, X)$, $\mathbf{t}(\omega_5, X)$ and $\mathbf{t}(\omega_7, X)$ does not exceed 1. This is easy since we know $\int t(\omega, X)\, dX(\omega) \le 1$ (by the definition of the test), and this integral is a convex combination of the above three numbers with some coefficients (the weights of ω_1, ω_5 and ω_7 in X). □

7. RANDOMNESS IN A METRIC SPACE

Most of the theory presented above for infinite binary sequences generalizes to infinite sequences of natural numbers. Much of it generalizes even further, to an arbitrary metric space. In what follows below we not only generalize; some of the results are new also for the binary sequence case.

7.1. Constructive metric spaces. We rely on the definition of a constructive metric space and the space of measures on it, as defined in [11] and [13] (the lecture notes [7] are also recommended).

Definition 7.1. A *constructive metric space* is a tuple $\mathbf{X} = (X, d, D, \alpha)$ where (X, d) is a complete separable metric space, with a countable dense subset D and an enumeration α (a total function mapping natural numbers onto D). It is assumed that the real function $d(\alpha(v), \alpha(w))$ is computable. Open balls with center in D and rational radius are called *ideal balls*, or *basic open sets*, or *basic balls*. The (countable) set of basic balls will also be called the *canonical basis* in the topology of the metric space.

An infinite sequence $s_1, s_2, \ldots$ with $s_i \in D$ is called a *strong Cauchy* sequence if for all $m < n$ we have $d(s_m, s_n) \le 2^{-m}$. Since the space is complete, such a sequence always has a (unique) limit, which we will say is *represented* by the sequence.

We will generally use the notational convention of this definition: if there is a constructive metric space with an underlying set X, then we will use $\mathbf{X}$ (boldface) to denote the whole structure (X, d, D, α). But frequently, we just use X when the structure is automatically understood.

Examples 7.2. 1. A set $X = \{s_1, s_2, \dots\}$ can be turned into a constructive *discrete* metric space by making the distance between any two different points equal to 1. The set D consists of all points $\alpha(i) = s_i$.

2. The set $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ can be turned into a constructive metric space by measuring the distance between any two different points with the distance function $d(x, y) = \left|\frac{1}{x} - \frac{1}{y}\right|$, where of course, $\frac{1}{\infty} = 0$. The set D consists of all points of $\mathbb{N}$. This metric space is called the *one-point compactification*, in a topological sense, of the *discrete metric space* $\mathbb{N}$ of Example 1.

3. The real line $\mathbb{R}$ with the distance $d(x, y) = |x - y|$ is a constructive metric space, and so is $\mathbb{R}_+ = [0, \infty)$. We can add the element ∞ to get $\overline{\mathbb{R}}_+ = [0, \infty]$. This is not a metric space now, but is still equipped with a natural constructive topology (see Remark 7.4 below). It could be equipped with a new metric in a way that would not change this constructive topology.

4. If $\mathbf{X}$ and $\mathbf{Y}$ are constructive metric spaces, then we can define a constructive metric space $\mathbf{Z} = \mathbf{X} \times \mathbf{Y}$ with one of its natural metrics, for example, the sum of distances in both coordinates. In the case when $\mathbf{X} = \mathbf{Y} = \mathbb{R}$, this is called the L_1 metric. Let $D_{\mathbf{Z}}$ be the product $D_{\mathbf{X}} \times D_{\mathbf{Y}}$.

5. Let X be a finite or countable (enumerated) alphabet, with a fixed numbering, and let $X^{\mathbb{N}}$ be the set of infinite sequences $x = (x(1), x(2), \dots)$ with distance function $d^{\mathbb{N}}(x, y) = 2^{-n}$ where n is the first i with $x(i) \neq y(i)$. This space generalizes the binary Cantor space of Definition 2.1 to the case mentioned in Remark 5.15. The balls in it are cylinder sets: for a given finite sequence z, we take all continuations of z.

Remark 7.3. Each point x of a constructive metric space $\mathbf{X}$ can be viewed as an "approximation mass problem": its solutions are total functions that for any given rational $\varepsilon > 0$ produce an ε-approximation to x by a point of the canonical dense set D. This is a mass problem in the sense of [21]. One can also note that this mass problem is Medvedev equivalent to the enumeration problem: enumerate all basic balls that contain x.

Remark 7.4. A constructive metric space is a special case of a more general concept, a constructive topological space, which is often useful.

A *constructive topological space* $\mathbf{X} = (X, \tau, \nu)$ is a topological space over a set X with a basis τ effectively enumerated (not necessarily without repetitions) as a list $\tau = \{\nu(1), \nu(2), \dots\}$.

For every nonempty subset Z of the space X, we can equip Z with a constructive topology: we intersect all basic sets with Z, without changing their numbering. On the other hand, not every subset of a constructive metric space naturally has the structure of a constructive metric space (the everywhere dense set D is not inherited).

But instead of introducing constructive topological spaces formally, we prefer not to burden the present paper with more abstractions, and will speak about some concepts like effective open sets and continuous functions, as defined on an arbitrary subset Z of the constructive metric space X.

Definition 7.5. An open subset of a constructive metric space is *lower semicomputable open* (or r.e. open, or c.e. open), or *effectively open* if it is the union of an enumerable set of elements of the canonical basis. It is *upper semicomputable closed*, or *effectively closed* if its complement is effectively open. Given any set $A \subseteq X$, a set U is *effectively open on* A if there is an effectively open set V such that $U \cap A = V \cap A$.

Note that in the last definition U is not necessarily a part of A, but only its intersection with A matters.

Computable functions can be defined in terms of effectively open sets.

Definition 7.6 (computable function)**.** Let X and Y be constructive metric spaces and $f \colon X \to Y$ a function. Then f is *continuous* if for each element U of the canonical basis of Y the set $f^{-1}(U)$ is an open set. It is *computable* if $f^{-1}(U)$ is also an effectively open set, uniformly in U. A partial function $f\colon X \to Y$ defined at least on a set A is *computable on A* if for each element U of the canonical basis of Y the set $f^{-1}(U)$ is effectively open on A, uniformly in U.

An element $x \in X$ is called *computable* if the function $f\colon \{0\} \to X$ with $f(0) = x$ is computable.

When $f(x)$ is defined only at a single point x_0, we say that the element $y_0 = f(x_0)$ is x_0-*computable*. When $f\colon X \times Z \to Y$, defined on $X \times \{z_0\}$, is computable, we say that the function $g\colon X \to Y$ defined by $g(x) = f(x, z_0)$ is z_0-*computable*, or computable *from* z_0.

There are several alternative characterizations of a computable element.

Proposition 7.7. *The following statements are equivalent for an element x of a constructive metric space $\mathbf{X} = (X, d, D, \alpha)$:*

(i) *x is computable*;

(ii) *the set of basic balls containing x is enumerable*;

(iii) *there is a computable sequence $z_1, z_2, \dots$ of elements of D with $d(x, z_n) \le 2^{-n}$.*

The following proposition connects computability with a more intuitive concept based on representation by strong Cauchy sequences.

Proposition 7.8. *Let $\mathbf{X}$ and $\mathbf{Y}$ be constructive metric spaces and $f\colon X \to Y$ a function. Then f is computable if and only if there is a computable transformation that turns each strong Cauchy sequence $s_1, s_2, \dots$ with $s_i \in D_{\mathbf{X}}$ converging to a point $x \in X$ into a strong Cauchy sequence $t_1, t_2, \dots$ with $t_i \in D_{\mathbf{Y}}$ converging to $f(x)$.*

If f is a partial function with domain Z, then f is computable if and only if there is a computable transformation that turns each strong Cauchy sequence $s_1, s_2, \dots$ with $s_i \in D_{\mathbf{X}}$ converging to some point $x \in Z$ into a strong Cauchy sequence $t_1, t_2, \dots$ with $t_i \in D_{\mathbf{Y}}$ converging to $f(x)$.

We omit the proof of this statement, which is not difficult.

Remark 7.9. Though x_0-computability means computability from a strong Cauchy sequence $s_1, s_2, \dots$ converging to x_0, it should not be considered the same as computability using a machine that treats this sequence as an "oracle." In case of x_0-computability, the resulting output must be independent of the strong Cauchy sequence $s_1, s_2, \dots$ representing x_0.

The following definition of lower semicomputability is also a straightforward generalization of the special case in Definition 2.2.

Definition 7.10 (lower semicomputability)**.** Let $\mathbf{X} = (X, d, D, \alpha)$ be a constructive metric space. A function $f\colon X \to [-\infty, \infty]$ is *lower semicontinuous* if the sets $\{x\colon f(x) > r\}$ are open, for every rational number r (this implies that they are open for all r, not only rational).

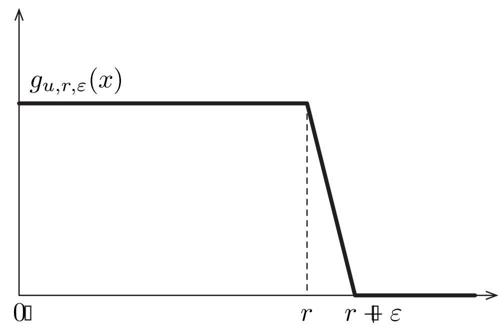
It is *lower semicomputable* if these sets are effectively open, uniformly in the rational number r. It is *upper semicomputable* if $-f$ is lower semicomputable.

A partial function $f\colon X \to Y$ defined at least on a set A is *lower semicomputable on A* if the sets $\{x\colon f(x) > r\}$ are effectively open in A, uniformly for every rational number r.

It is easy to check that a real function over a constructive metric space is computable if and only if it is lower and upper semicomputable. As before, one can define semicomputability equivalently with the help of basic functions.

Let us introduce an everywhere dense set of simple functions.

Definition 7.11 (hat functions, basic functions)**.** We define an enumerated list of *basic* functions $\mathcal{E} = \{e_1, e_2, \dots\}$ in the constructive metric space $\mathbf{X} = (X, d, D, \alpha)$ as follows. For each point $u \in D$ and positive rational numbers r and ε let us define the *hat function* $g_{u,r,\varepsilon}$: its value at x is determined by the distance from x to u and is equal to 1 if this distance is at most r, equal to zero if the distance is not less than $r + \varepsilon$, and varies linearly as the distance runs through the segment $[r, r + \varepsilon]$ (see the figure).



A hat function.

Let $\mathcal{E}$ be the smallest set of functions containing all hat functions that is closed under min, max and rational linear combinations.

Proposition 7.12. *A function $f\colon X \to [0,\infty]$ defined on a constructive metric space is lower semicomputable if and only if it is the limit of a computable increasing sequence of basic functions.*

Note that the above characterization also holds for lower semicontinuous functions, if we omit the requirement that the sequence is computable.

Definition 7.13. We can introduce the notion of lower semicomputability *from* z_0, or z_0*-lower semicomputability*, similarly to the z_0-computability of Definition 7.6, as the lower semicomputability of a function defined on the set $X \times \{z_0\}$.

Sometimes two metrics on a space are equivalent from the point of view of computability questions. Let us formalize this notion.

Definition 7.14 (uniform continuity, equivalence)**.** Let X and Y be two metric spaces and $f\colon X \to Y$ a function. We say that f is *uniformly continuous* if for each $\varepsilon > 0$ there is a $\delta > 0$ such that $d_X(x_1, x_2) \le \delta$ implies $d_Y(f(x_1), f(x_2)) \le \varepsilon$.

If $\mathbf{X}$ and $\mathbf{Y}$ are constructive metric spaces and a function f is computable, we will call it *effectively uniformly continuous* if δ can be computed from ε effectively.

Two metrics d_1 and d_2 over the same space are (*effectively*) *equivalent* if the identity map is (effectively) uniformly continuous in both directions.

For example, the Euclidean metric and the L_1 metric introduced in Example 7.2.4 are equivalent in the space $\mathbb{R}^2$.

Effective compactness was introduced in Definition 5.4; this generalizes immediately to arbitrary metric spaces. A weaker notion, local compactness, also has an effective version.

Definition 7.15 (effective compactness and local compactness)**.** A compact subset C of a constructive metric space $\mathbf{X} = (X, d, D, \alpha)$ is called *effectively compact* if the set

$$\left\{S\colon S \text{ is a finite set of basic open sets and } \bigcup_{E \in S} E \supseteq C\right\}$$

is enumerable.

A subset C of a metric space is called *locally compact* if it is covered by the union of a set of balls B such that $\overline{B} \cap C$ is compact. Here $\overline{B}$ is the closure of B. It is *effectively locally compact* if it is covered by the union of an enumerated sequence of basic balls B_k such that $\overline{B}_k \cap C$ is effectively compact, uniformly in k.

Examples 7.16. 1. The countable discrete space of Example 7.2.1 is effectively compact if it is finite, and effectively locally compact otherwise.

2. The segment $[0,1]$ is effectively compact. The line $\mathbb{R}$ is effectively locally compact.

3. If the alphabet X is finite, then the space $X^{\mathbb{N}}$ of infinite sequences is effectively compact. Otherwise it is not even locally compact.

4. Let $\alpha \in [0,1]$ be a lower semicomputable real number that is not computable. (It is known that there are such numbers, for example, $\sum_{x\in\mathbb{N}} 2^{-\mathrm{Kp}(x)}$.) The lower semicomputability of α allows one to enumerate the rationals less than α and allows the segment $[0,\alpha]$ to inherit the constructive metric (and topology) from the real line. This space is compact, but not effectively compact.

The following is a useful characterization of effective compactness.

Proposition 7.17. (a) *A compact subset C of a constructive metric space $\mathbf{X} = (X, d, D, \alpha)$ is effectively compact if and only if from each* (*rational*) ε *one can compute a finite set of ε-balls covering C.*

(b) *For an effectively compact subset C of a constructive metric space, in every enumerable set of basic open sets covering C one can effectively find a finite covering.*

Proof. Assume that for all ε we can show a finite covering S_ε of the set C by balls of radius ε. Along with such a covering, we can enumerate all coverings with *manifestly* large balls (this means that for any ball $B(x,\varepsilon)$ in the covering S_ε there is a ball $B(y,\sigma)$ in the new covering with $\sigma > \varepsilon + d(x,y)$). The compactness of C guarantees that while S_ε runs through all ε-coverings of C, this way all coverings of C will be enumerated. (Indeed, if there is some covering S' not falling into the enumeration, then for all ε there is a ball of the covering S_ε not manifestly contained in any ball of S'. Applying compactness and taking a limit point of the centers of these non-contained balls, we obtain a contradiction.)

The remaining statements are proved quite easily. □

The following statement generalizes Proposition 5.5, with the same proof.

Proposition 7.18. *Every effectively closed subset E of an effectively compact set C is also effectively compact.*

According to our definitions, this means that the intersection of an effectively compact subset of X with an effectively closed subset of X is effectively compact.

As earlier, the converse also holds: every effectively compact subset of a constructive metric space is effectively closed. Indeed, we can consider all possible coverings of this set by basic balls, as well as outside balls that are manifestly (by the relation between the distances from their centers and their radii) disjoint from the balls of the covering. The union of all these outside balls provides the complement of our effectively compact set.

It is known that a continuous function maps compact sets into compact ones. This statement also has a constructive counterpart, also provable by a standard argument:

Proposition 7.19. *Let C be an effectively compact subset of a constructive metric space X, and f a computable* (*on C*) *function from C into another constructive metric space Y. Then $f(C)$ is effectively compact.*

The statement that a lower semicontinuous function on a compact set reaches its minimum also has a computable analog (we provide a parametrized variant):

Proposition 7.20 (parametrized minimum)**.** *Let $\mathbf{Y}$ and $\mathbf{Z}$ be constructive metric spaces*; *let $f\colon Y \times Z \to [0,\infty]$ be a lower semicomputable function and C an effectively closed subset of $Y \times Z$. If it is also effectively compact, then the function*

$$g(y) = \inf_{z\colon (y,z)\in C} f(y,z)$$

is lower semicomputable (*and the* inf *can be replaced with* min *due to compactness*; *the minimal element of the empty set is $+\infty$ by definition*).

Instead of effective compactness of C, it is sufficient to require that its projection $C_Y = \{y\colon \exists z\,(y,z) \in C\}$ is effectively closed and covered by a computable sequence of basic balls B_k such that $(\overline{B}_k \times Z) \cap C$ is effectively compact, uniformly in k.

The weaker condition formulated at the end holds, for example, if Y is effectively locally compact and Z is effectively compact.

Proof. To begin with, we reproduce the classical proof of lower semicontinuity. One needs to check that the set $\{y\colon r < g(y)\}$ is open for all r. This set can be represented in the form of a union, since the condition $r < g(y)$ is equivalent to the condition

$$(\exists r' > r)\,\forall z\,\bigl[(y,z) \in C \Rightarrow f(y,z) > r'\bigr],$$

and it is sufficient to check the openness of the set

$$U = \bigl\{y\colon \forall z\,\bigl[(y,z) \in C \Rightarrow f(y,z) > r'\bigr]\bigr\}.$$

Now, $U = (Y \setminus C_Y) \cup \bigcup_k (B_k \cap U)$. Since $Y \setminus C_Y$ is assumed to be open, it is sufficient to show that each $B_k \cap U$ is open. Let $F_k = \overline{B}_k \times Z$; then, by the assumption, $F_k \cap C$ is compact. The condition $f(y,z) > r'$ by the assumption defines a certain open set V of pairs; hence $F_k \cap C \setminus V$ is closed and, as a subset of a compact set, compact. It follows that its projection

$$\bigl\{y \in \overline{B}_k\colon \exists z\,(y,z) \in F_k \cap C \setminus V\bigr\},$$

as a continuous image of a compact set, is also compact, and so closed. Its complement in B_k, which is $B_k \cap U$, is then open.

Now this argument must be translated into an effective language. First of all, note that it is sufficient to consider rational r and r'. Then the set V is effectively open, and the set $F_k \cap C \setminus V$ is effectively closed and, as a subset of an effectively compact set, also effectively compact. Its projection, as a computable image of an effectively compact set, is also effectively compact and so effectively closed. The complement of the projection is then effectively open. □

The following lemma is an application:

Lemma 7.21. *Let X, Z and Z' be metric spaces, where X is locally compact and Z is compact. Let $f\colon Z \to Z'$ be continuous and surjective, and $t\colon X \times Z \to [0,\infty]$ a lower semicontinuous function. Then the function $t_f\colon X \times Z' \to [0,\infty]$ defined by the formula*

$$t_f(x,z') = \inf_{z\colon f(z)=z'} t(x,z)$$

is lower semicontinuous.

If X, Z and Z' are constructive metric spaces, X is effectively locally compact, Z is effectively compact, f is computable, and t is lower semicomputable, then t_f is lower semicomputable.

Proof. We will prove just the effective version. We will apply Proposition 7.20 with $Y = X \times Z'$, and $C = X \times \{(f(z),z)\colon z \in Z\}$. Then

$$t_f(x,z') = \inf_{(x,z',z)\in C} t(x,z).$$

The set Y is effectively locally compact, as the product of an effectively locally compact set and an effectively compact set. The projection of the set C onto Y is the whole set Y, and hence it is closed. Hence the proposition is applicable, according to the remark following it. □

7.2. Measures over a constructive metric space. On a metric space, the *Borel sets* are the smallest σ-algebra containing the open sets. We can define measures on Borel sets. These measures have the following *regularity* property:

Proposition 7.22 (regularity)**.** *Let P be a measure over a complete separable metric space. Then every measurable set A can be approximated from above by open sets*:

$$P(A) = \inf_{G \supseteq A} P(G),$$

where G is open.

It is possible to introduce a metric over measures:

Definition 7.23 (Prokhorov distance)**.** For a set A and a point x let us define the distance from x to A as $d(x,A) = \inf_{y \in A} d(x,y)$. The ε-neighborhood of a set A is defined as $A^\varepsilon = \{x \colon d(x,A) < \varepsilon\}$.

The *Prokhorov distance* $\rho(P,Q)$ between two measures is the infimum of all ε such that for all Borel sets A we have $P(A) \le Q(A^\varepsilon) + \varepsilon$ and $Q(A) \le P(A^\varepsilon) + \varepsilon$.

It is known that $\rho(P,Q)$ is indeed a metric, and it turns the set of probability measures over a metric space X into a metric space. There is a number of other metrics for measures that are equivalent in the sense of Definition 7.14.

Definition 7.24 (space of measures)**.** For a constructive metric space $\mathbf{X}$, let $\mathbf{M} = \mathcal{M}(\mathbf{X})$ be the metric space of probability measures over $\mathbf{X}$, with the metric $\rho(P,Q)$. The dense set $D_{\mathbf{M}}$ is the set of those probability measures that are concentrated on finitely many points of $D_{\mathbf{X}}$ and assign rational values to them. Let $\alpha_{\mathbf{M}}$ be a natural enumeration of $D_{\mathbf{M}}$; this turns $\mathbf{M}$ into a constructive metric space.

A probability measure is called *computable* when it is a computable element of the space $\mathbf{M}$.

Computability of measures is a particularly simple property for the Cantor space of binary sequences in Definition 2.8 (which is easily shown to be equivalent to the definition given here); it is just as simple for the Baire space of sequences over a countable alphabet.

The analogue of Proposition 4.17 holds again: the integral $\int f(\omega, P)\, P(d\omega)$ of a basic function is computable as a function of the measure P, uniformly in the code of the basic function. Here is a closely related result:

Proposition 7.25. *If f is a bounded effectively uniformly continuous function, then its integral with respect to the measure P is an effectively uniformly continuous function of P.*

Proof. It can be assumed without loss of generality that f is nonnegative (add a constant). Let measures P and P' be close. Then $P'(A) \le P(A_\varepsilon) + \varepsilon$, where A_ε denotes the ε-neighborhood of A. Then

$$\int f\, dP' \le \int f_\varepsilon\, dP + \varepsilon,$$

where $f_\varepsilon(x)$ is the supremum of f on the ε-neighborhood of x. (The integral of a nonnegative function g is defined by the measures of the sets $G_t = \{x \colon g(x) \ge t\}$; by Fubini's theorem on the change of the order of integration, this measure must be integrated in t as a function of t. Now, if $f(x) \ge t$, then $f_\varepsilon(x) \ge t$ in the ε-neighborhood of x.) It remains to apply the effective uniform continuity of f to find out the precision with which the measure must be given in order to obtain a given precision in the integral. $\square$

On the other hand, the measure $P(B)$ of a basic ball B is not necessarily computable, only lower semicomputable. It is shown in [13] that this property also characterizes the computability of measures: P is computable if and only if $P(B)$ is lower semicomputable, uniformly in the basic ball B.

It is known that if a complete separable metric space is compact, then so is the set of measures with the described metric. The following constructive version is proved by standard means:

Proposition 7.26. *If a constructive metric space* $\mathbf{X}$ *is effectively compact, then its space of probability measures* $\mathcal{M}(\mathbf{X})$ *is also effectively compact.*

For the binary Cantor space, this was proved in Proposition 5.5. There, the topology of the space of measures was simply derived from the topology of the space $[0,1] \times [0,1] \times \ldots$. It can be seen that the Prokhorov metric leads to the same topology.

Example 7.27. Another interesting simple metric space is the infinite discrete space, say, on the set of natural numbers $\mathbb{N}$. This is not a compact space, and the set of measures, namely, the set of all functions $P(x) \geq 0$ with $\sum_{x\in\mathbb{N}} P(x) = 1$, is not compact either.

On the other hand, the set of semimeasures (see Definition 2.19) is compact. Indeed, recall that the space $[0,1] \times [0,1] \times \ldots$ of functions $P\colon \mathbb{N} \to [0,1]$ is compact. Hence for each n the subset F_n of this set consisting of functions P obeying the restriction $P(0) + P(1) + \ldots + P(n) \leq 1$ is also compact, as the product of a compact finite-dimensional set $\{(P(0), P(1), \ldots, P(n)) \in [0,1]^{n+1}\colon P(0) + \ldots + P(n) \leq 1\}$ and the compact infinite product set $\{(P(n+1), P(n+2), \ldots)\colon 0 \leq P(x) \leq 1 \text{ for } x > n\}$. The intersection of all sets F_n is then also compact, and is equal to the set of semimeasures.

Equivalently, we can consider the one-point compactification $\overline{\mathbb{N}}$ of $\mathbb{N}$ given in Example 7.2.2. Measures P on this space can be identified with semimeasures over $\mathbb{N}$: we simply set $P(\infty) = 1 - \sum_{n<\infty} P(n)$.

7.3. Randomness in a metric space. In the Cantor space Ω of infinite binary sequences we defined

- randomness with respect to computable measures (in the sense of Martin-Löf); see Definition 2.9;
- uniform randomness with respect to arbitrary measures (when the test is a function of the sequence and the measure), Definition 5.2;
- randomness with respect to an effectively compact class of measures, Definition 5.22;
- blind (oracle-free) randomness in Definition 5.37;

All these notions carry over with minor changes to an arbitrary constructive metric space. In the present section (Subsections 7.3 and 7.4) we discuss these generalizations and their properties, and then in Section 8 we consider in more detail randomness with respect to an orthogonal class of measures.

For computable measures, a test is defined as a lower semicomputable function on a constructive metric space with the integral bounded by 1. Among such tests, there is a maximal one to within a multiplicative constant. As earlier, this is proved with the help of trimming: we list all lower semicomputable functions, forcing them into tests or almost tests, and then add them up with coefficients from a converging series.

This is done as before, by considering lower semicomputable functions as monotonic limits of basic ones. It is used that the integral of a basic function with respect to a measure is computable as a function of the measure: see Propositions 7.25 and the discussion preceding it.

The uniform tests introduced in Definition 5.2 generalize immediately to the case of constructive metric spaces. Such a test is a lower semicomputable function of two arguments $t(x, P)$, where x is a point of our metric space and P is a measure over this space. The integral condition has the same form as earlier: $\int t(x,P)\, P(dx) \leq 1$.

As earlier, there exists a universal test, and this can be proved by the technique of trimming:

Theorem 7.28 (trimming in metric spaces)**.** *Let* $u(x, P)$ *be a lower semicomputable function whose first argument is a point of a constructive metric space and the second one is measure over* *this space. Then there exists a uniform tests $t(x,P)$ satisfying $u(x,Q) \le 2t(x,Q)$ for all Q such that the function $u_Q\colon x \mapsto u(x,Q)$ is a test in the measure Q, that is, $\int u(x,Q)\,Q(dx) \le 1$.*

The proof repeats the reasoning of the proof of Theorem 5.7, while using the fact that for a basic function $b(x,P)$ on the product space the integral $\int b(x,P)\,P(dx)$ is a computable (continuous) function of P (which is proved analogously to our above argument on the computability of the integral).

We will denote the universal uniform test again by $\mathbf{t}(x,P)$. Strictly speaking, it also depends on the constructive metric space on which it is defined, but in general it is evident which space is being considered; therefore, it is not shown in the notation.

Definition 5.11 and Proposition 5.12 extend without difficulty.

Definition 7.29 (tests for arbitrary measures)**.** Let $\mathbf{X} = (X,d,D,\alpha)$ be a constructive metric space. For a measure $P \in \mathcal{M}(\mathbf{X})$, a *P-test of randomness* is a function $f\colon X \to [0,\infty]$ lower semicomputable from P with the property $\int f(x)\,dP \le 1$.

It seems as if a P-test may capture some nonrandomnesses that uniform tests cannot; however, this is not so, since trimming (see Theorem 5.7) generalizes:

Theorem 7.30 (uniformization)**.** *Let P be some measure over a constructive metric space X, along with some P-test $t_P(x)$. There is a uniform test $t'(\cdot,\cdot)$ with $t_P(x) \le 2t'(x,P)$.*

Proof. By definition, a lower semicomputable (with respect to P) function can be represented as $c \mapsto t(x,P)$ where t is a lower semicomputable function of two arguments. It remains to use the preceding theorem. □

Theorem 5.36 generalizes to the case of constructive metric spaces. Let us mention one of the facts that generalize to uniform tests.

Proposition 7.31 (uniform Kurtz tests)**.** *Let X be a constructive metric space and S an effectively open subset of the space $X \times \mathcal{M}(X)$. If the set $S_P = \{x\colon (x,P) \in S\}$ has P-measure 1 for some measure $P \in \mathcal{M}(X)$, then the set S_P contains all uniformly P-random points.*

Proof. The indicator function $1_S(x,P)$ of the set S, which is equal to unity on S and to zero outside, is lower semicomputable. According to Proposition 7.12, it can be written as the limit of a computable increasing sequence of basic functions $0 \le g_n(x,P) \le 1$. The sequence $G_n\colon P \mapsto \int g_n(x,P)\,dP$ is an increasing sequence of functions computable uniformly in n. The monotone convergence theorem implies $G_n(P) \to 1$ for all $P \in \mathcal{M}(X)$. Let us define for each measure P the numbers $n_k(P)$ as the minimal values of n for which $G_n(P) > 1 - 2^{-k}$. These numbers are upper semicomputable as functions of P (in a natural sense; for measures P with $P(S_P) < 1$, some of these $n_k(P)$ are infinite). Correspondingly, the functions $1 - g_{n_k(P)}(x,P)$, as functions of x and P (define such a function to be zero for infinite $n_k(P)$, independently of x) are lower semicomputable, uniformly in k. Then $t(x,P) = \sum_{k>0}(1 - g_{n_k(P)}(x,P))$ is a uniform test, since at a given P its kth summand is zero for infinite $n_k(P)$ and has integral not greater than 2^{-k} for finite $n_k(P)$.

The conditions of the theorem tell about a measure P with $P(S_P) = 1$. Then all numbers $n_k(P)$ are finite. Consider an x outside S_P: then $g_{n_k(P)}(x,P) = 0$ by definition. Therefore, all terms of the test sum are equal to unity; thus x is not a P-random point. Consequently, S_P includes all uniformly P-random points. □

7.4. A priori probability, with an oracle. In Subsection 5.2 we defined the a priori probability with a condition whose role was played by a measure over the Cantor space Ω. Now, having introduced the notion of a constructive metric space, we can note that this definition extends naturally to an arbitrary such space $\mathbf{X}$: we consider nonnegative lower semicomputable functions $m\colon \mathbb{N} \times X \to [0,\infty]$ for which $\sum_i m(i,x) \le 1$, for all $x \in X$.

Among such functions, there is a maximal one to within a multiplicative constant. This is proved by the method of trimming: the lower semicomputable function $m(i,x)$ can be obtained as a sum of a series of basic functions each of which differs form zero only for one i; these basic functions must be multiplied by correcting coefficients that depend on the sum over all i. (At each stage, this sum has only finitely many members.)

We will call the maximal function of this kind the *a priori probability with condition* x, and denote it $\mathbf{m}(i|x)$. We consider the first argument a natural number, but this is not essential: it is possible to consider words (or any other discrete constructive objects). As a special case we obtain the definition of a priori probability conditioned on a measure (Subsection 5.2), as well as the standard notions of a priori probability with an oracle (which corresponds to $\mathbf{X} = \Omega$, the Cantor space of infinite sequences) and the conditional a priori probability (corresponding to $\mathbf{X} = \mathbb{N}$).

In analogy with the Day–Miller theorem, the a priori probability with a condition is expressible (in an arbitrary *effectively compact* constructive metric space $\mathbf{X}$) in terms of the a priori probability with an oracle.

Proposition 7.32. *Let $F\colon \Omega \to X$ be a computable map whose image is the whole space X. Then*

$$\mathbf{m}(i|x) \stackrel{*}{=} \min_{\pi\colon F(\pi)=x} \mathbf{m}(i|\pi).$$

Proof. We reason as in the proof of Theorem 5.36. The function $(i, \pi) \mapsto \mathbf{m}(i|F(\pi))$ is lower semicomputable on $\mathbb{N} \times \Omega$, which implies the $\stackrel{*}{<}$-inequality.

In order to obtain the reverse inequality, we use Lemma 7.21 and note that the function on the right-hand side is well defined (the minimum is achieved) and is lower semicomputable. □

Note that $\mathbf{m}(i|\pi)$, the a priori probability with an oracle on the right-hand side of Proposition 7.32, is expressible by prefix complexity with an oracle. For the case of conditions in metric spaces it is not clear how to define the prefix complexity with such a condition (one can speak of functions whose graph is enumerable with respect to x, but it is not clear how to build a universal one). One can define formally $\mathrm{Kp}(i|x)$ as $\max_{\pi\colon F(\pi)=x} \mathrm{Kp}(i|\pi)$, and then $\mathrm{Kp}(i|x) \stackrel{+}{=} -\log \mathbf{m}(i|x)$, but it is questionable whether this can be considered a satisfactory definition of prefix complexity (say, the usual arguments using the self-delimiting property of programs are not applicable with such a definition). It is more honest to simply speak of the logarithm of the a priori probability. Many results still remain true: for example, the formula $\mathrm{Kp}(i, j|x) \stackrel{+}{<} \mathrm{Kp}(i|x) + \mathrm{Kp}(j|x)$ can be proved without introducing self-delimiting programs, by just reasoning about probabilities.

Remark 7.33. Analogously, it is possible to supply points in constructive metric spaces as conditions in some of our other definitions. For example, we can consider uniform tests over the Cantor space Ω of infinite binary sequences, with condition in an arbitrary constructive metric space $\mathbf{X}$: these will be lower semicomputable functions $t(\omega, P, x)$ with $\int t(\omega, P, x)\, P(d\omega) \le 1$ for all P and x. It is also possible to fix a computable measure P, say the uniform one, and define tests with respect to this measure with conditions in $\mathbf{X}$.

8. CLASSES OF ORTHOGONAL MEASURES

The definition of a class test for an effectively compact class of measures, as well as Theorem 5.23 about the expression of a class test, generalizes, with the same proof.

The set of Bernoulli measures has an important property shared by many classes considered in practice; namely, a random sequence determines the measure to which it belongs. A consequence of this was spelled out in Theorem 5.41. This section explores the topic in a more general setting.

There are some examples naturally generalizing the Bernoulli case: finite or infinite ergodic Markov chains and ergodic stationary processes. Below, we will dwell a little more on the latter, since it brings up a rich complex of new questions.

We will consider orthogonal classes in the general setting of metric spaces: from now on, our measurable space is the one obtained from a constructive metric space $\mathbf{X} = (X, d, D, \alpha)$. The following classical concept is analogous to effective orthogonality, introduced in Definition 5.40.

Definition 8.1 (orthogonal measures)**.** Let P and Q be two measures over a measurable space $(X, \mathcal{A})$, that is, a space X with a σ-algebra $\mathcal{A}$ of measurable sets on it. We say that they are *orthogonal* if the space can be partitioned into disjoint measurable sets U and V with the property $P(V) = Q(U) = 0$.

Let $\mathcal{C}$ be a class of measures. We say that $\mathcal{C}$ is *orthogonal* if there is a measurable function $\varphi\colon X \to \mathcal{C}$ such that $P(\varphi^{-1}(P)) = 1$ for every $P \in \mathcal{C}$.

Note that the space $\mathcal{M}(X)$, as a metric space, also allows the definition of Borel sets, and it is in this sense that we can talk about f being measurable.

Examples 8.2. 1. In an orthogonal class, any two (different) measures P and Q are orthogonal. Indeed, the sets $\{P\}$ and $\{Q\}$ are Borel (since closed); hence their preimages are measurable (and obviously disjoint). The converse statement is false: A class $\mathcal{C}$ of mutually orthogonal probability measures is not necessarily orthogonal, even if the class is effectively compact. For example, let λ be the uniform distribution over the interval $[0, 1]$, and let for each $x \in [0, 1]$ the probability measure δ_x be concentrated at x. Then the class $\{\lambda\} \cup \{\delta_x \colon x \in [0, 1]\}$ is effectively compact, and its elements are mutually orthogonal. But the whole class is not orthogonal: the orthogonality condition requires $\varphi(x) = \delta_x$, but then $\varphi^{-1}(\lambda)$ will be empty.

2. Let P and Q be two probability measures. Of course, if Randoms(P) and Randoms(Q) are disjoint, then P and Q are orthogonal. The converse is not always true: for example it fails if λ and δ_x are as above, where x is random with respect to λ.

The following definition introduces the important example of stationary ergodic processes.

Definition 8.3. The Cantor space Ω of infinite binary sequences is equipped with an operation $T\colon \omega(1)\omega(2)\omega(3)\ldots \mapsto \omega(2)\omega(3)\omega(4)\ldots$ called the *shift*. A probability distribution P over Ω is *stationary* if for every Borel subset A of Ω we have $P(A) = P(T^{-1}(A))$. It is easy to see that this property is equivalent to requiring

$$P(x) = P(0x) + P(1x)$$

for every binary string x.

A Borel set $A \subseteq \Omega$ is called *invariant* with respect to the shift operation if $T(A) \subseteq A$. For example, the set of all sequences in which the relative frequency converges to $1/2$ is an invariant set. A stationary distribution is called *ergodic* if every invariant Borel set has measure 0 or 1.

Here is a new example of a stationary process (all Bernoulli measures and stationary Markov chains are also examples).

Example 8.4. Let $Z_1, Z_2, \ldots$ be a sequence of independent identically distributed random variables taking values 0 and 1 with probabilities 0.9 and 0.1, respectively. Let $X_0, X_1, X_2, \ldots$ be defined as follows: X_0 takes values 0, 1 and 2 with equal probabilities, and independently of all Z_i; further $X_n = X_0 + \sum_{i=1}^{n} Z_i \bmod 3$. Finally, let $Y_n = 0$ if $X_n = 0$ and $Y_n = 1$ otherwise. The process $Y_0, Y_1, \ldots$ is clearly stationary, and can also be proved to be ergodic. As a function of the Markov chain $X_0, X_1, \ldots$, it is also called a *hidden Markov chain*.

The following statement is a consequence of Birkhoff's pointwise ergodic theorem. For each binary string x let $g_x(\omega) = 1_{x\Omega}(\omega)$ be the indicator function of the set $x\Omega$: it is 1 if and only if x is a prefix of ω.

Proposition 8.5. *Let P be a stationary process over the Cantor space Ω.*

(a) *With probability* 1, *the average*

$$A_{x,n}(\omega) = \frac{1}{n}\bigl(g_x(\omega) + g_x(T\omega) + \ldots + g_x(T^{n-1}\omega)\bigr) \tag{7}$$

converges.

(b) *If the process is ergodic, then the sequence converges to $P(x)$.*

(For nonergodic processes, the limit may depend on ω.)

Birkhoff's theorem is more general, pertaining to more general spaces and measure-preserving transformations T, arbitrary integrable functions in place of g_x, and convergence to the expected value in the ergodic case. But the proposition captures its essence (and can also be used in the derivation of the more general versions).

Part (b) of Proposition 8.5 implies that the class $\mathcal{C}$ of ergodic measures is an orthogonal class. Indeed, let us call a sequence ω "stable" if for all strings x the averages $A_{x,n}(\omega)$ of (7) converge. It is easy to see that in this case the numbers $P(x)$ determine some probability measure Q_ω. Now, let $\varphi\colon \Omega \to \mathcal{C}$ be a function that assigns to each stable sequence ω the measure Q_ω provided Q_ω is ergodic. If the sequence is not stable or Q_ω is not ergodic, then let $\varphi(\omega)$ be some arbitrary fixed ergodic measure. It can be shown that φ is a measurable function: here, we use the fact that the set of stable sequences is a Borel set. By part (b) of Proposition 8.5, the relation $P(\varphi^{-1}(P)) = 1$ holds for all ergodic measures.

Note that the class of all ergodic measures is not closed, but we did not rely on the closedness of this class in the definition.

Example 8.2.2 shows that two measures can be orthogonal and still have common random sequences. But for computable measures, as we will show right away, this is not possible.

In Definition 5.40 we called a class of measures P effectively orthogonal if all sets of random sequences Randoms(P) for measures P in the class are disjoint from each other.

Theorem 8.6. *Two computable probability measures on a constructive metric space are orthogonal if and only they are effectively orthogonal.*

Speaking of the effective orthogonality of two measures, we mean that they have no common (uniform) random sequences. In the effective case, pairwise orthogonality within the class and the orthogonality of the whole class are equivalent by definition.

Proof. We only need to prove one direction. Assume that P and Q are orthogonal, that is, there is a measurable set A with $P(A) = 1$ and $Q(A) = 0$. By Proposition 7.22, these measures are regular, so there is a sequence $G_n \supseteq A$ of open sets with $Q(G_n) < 2^{-n}$. Then for every n there is also a finite union $H_n \subseteq G_n$ of basic balls with $P(H_n) > 1 - 2^{-n}$ and $Q(H_n) < 2^{-n}$; moreover, there is a computable sequence H_n with this property. Let $U_m = \bigcup_{n>m} H_n$. By Proposition 7.31, $\bigcap_m U_m$ contains all random points of P. On the other hand, the sets U_m form a Martin-Löf test for the measure Q, so the intersection contains no random points of Q. □

We have shown above that ergodic measures form an orthogonal class. Careful analysis shows that this is also true effectively.

Theorem 8.7. *The set of ergodic measures over the Cantor set Ω forms an effectively orthogonal class.*

Proof. Paper [29] (more precisely, an analysis of it that will create uniform tests) shows that

(a) sequences uniformly random with respect to some stationary measure are stable (in the sense that the above indicated limit of averages exists for them);

(b) uniformly random sequences with respect to an ergodic measure are "typical" in the sense that these averages converge to $P(x)$.

To show (a), the paper introduces the function $\sigma(\omega, \alpha, \beta)$ for rationals $0 < \alpha < \beta$, which is the maximum number of times that $A_{x,n}(\omega)$ crosses the interval (α, β) from below α to above β. This function is lower semicomputable, uniformly in the rationals α and β. Then it shows

$$(1+\alpha^{-1})(\beta-\alpha)\int \sigma(\omega,\alpha,\beta)\,dP \le 1,$$

that is, that $(1+\alpha^{-1})(\beta-\alpha)\sigma(\omega,\alpha,\beta)$ is an average-bounded test, implying that for Martin-Löf random sequences (and even for blind-random sequences), the average $A_{x,n}(\omega)$ crosses (α, β) from below α to above β only a finite number of times. Now one can combine all these tests, for all strings x and all rational $0 < \alpha < \beta$, into a single test (not depending on P).

To express (b), in view of part (a), it is sufficient, for each x, to prove

$$\liminf_n A_{x,n}(\omega) \le P(x) \le \limsup_n A_{x,n}(\omega)$$

for random ω. Take for example the statement for the lim inf. It is sufficient to show for any k and m that $\inf_{n\ge m} A_{x,n}(\omega) \le P(x) + 2^{-k}$ for a random ω. The set

$$S_{x,k,m} = \bigl\{(\omega, P)\colon \exists n \ge m\ A_{x,i}(\omega) < P(x) + 2^{-k}\bigr\}$$

is effectively open, and the Birkhoff theorem implies $P(S_{x,k,m}(P)) = 1$ for all ergodic measures P, for the set $S_{x,k,m}(P) = \{x\colon (x,P) \in S_{x,k,m}\}$. Proposition 7.31 implies that then for each P the set $S_{x,k,m}(P)$ contains all P-random points.

Another approach is a proof that just shows (b) for computable ergodic measures (in a relativizable way), without an explicit test, as done in [2]. Then a reference to Theorem 5.36 allows us to conclude the same about uniformly random sequences. □

It is convenient to treat orthogonality of a class in terms of separator functions. For this, note that by a measurable real function we mean a Borel-measurable real function, that is, a function with the property that the inverse images of Borel sets are Borel sets.

Definition 8.8 (separator function)**.** Let $\mathcal{C}$ be a class of measures over the metric space X. A measurable function $s\colon X \times \mathcal{M}(X) \to [0,\infty]$ is called a *separator function* for the class $\mathcal{C}$ if for all measures P we have $\int s(x,P)\,dP \le 1$ and further, for $P, Q \in \mathcal{C}$, $P \ne Q$ implies that only one of the values $s(x,P)$ and $s(x,Q)$ is finite.

In case we have a constructive metric space $\mathbf{X}$, a separator function $s(x,P)$ is called a *separator test* if it is lower semicomputable in (x,P).

We could have required the integral to be bounded only for measures in the class, since trimming allows the extension of the boundedness property to all measures, just as in the remark after Definition 7.29.

The following observation connects orthogonality with separator functions and also shows that in the case of effective orthogonality, each measure can be effectively reconstructed from any of its random elements.

Theorem 8.9. *Let $\mathcal{C}$ be a class of measures.*

(a) *If the class $\mathcal{C}$ is Borel and orthogonal, then there is a separator function for it.*

(b) *The class $\mathcal{C}$ is effectively orthogonal if and only if there is a separator test for it.*

The converse of part (a) might not hold: this needs further investigation.

Proof. Let us prove (a). If $\varphi(x)$ is a measurable function assigning measure $P \in \mathcal{C}$ to each element $x \in X$ as required in the definition of orthogonality, then by a general theorem of topological measure theory (see [15]), its graph is measurable. This allows the following definition: for $P \notin \mathcal{C}$ set $s(x,P) = 1$, while for $P \in \mathcal{C}$ set $s(x,P) = 1$ if $\varphi(x) = P$ and $s(x,P) = \infty$ otherwise.

Now let us prove (b). If $\mathcal{C}$ is effectively orthogonal, then the uniform test $\mathbf{t}(x, P)$ is a separator test for the class $\mathcal{C}$. Suppose now that there is a separator test s for the class $\mathcal{C}$, and let $P, Q \in \mathcal{C}$, $P \neq Q$, and $x \in \mathrm{Randoms}(P)$. Since s is a randomness test, $s(x, P) < \infty$, which implies $s(x, Q) = \infty$; hence $x \notin \mathrm{Randoms}(Q)$. $\square$

The following result is less expected: it shows that if the class of measures is effectively compact, then the existence of a lower semicontinuous separator function implies the existence of a lower semicomputable one (that is, a separator test).

Theorem 8.10. *If for an effectively compact class of measures there is a lower semicontinuous separator function* $s(x, P)$, *then this class is effectively orthogonal.*

Proof. Let $\mathcal{C}$ be an effectively compact class of measures on a constructive metric space. We need to show that under the conditions of the theorem, for any two distinct measures P_1 and P_2 in $\mathcal{C}$, the sets of random sequences are disjoint:

$$\mathrm{Randoms}(P_1) \cap \mathrm{Randoms}(P_2) = \varnothing.$$

In the constructive metric space $\mathbf{M}$ of measures, take two disjoint closed basic balls B_1 and B_2 that contain the measures P_1 and P_2. The classes $\mathcal{C}_i = \mathcal{C} \cap B_i$, $i = 1, 2$, of measures are disjoint effectively compact classes of measures containing P_1 and P_2. Consider the functions

$$t_i(x) = \inf_{P \in \mathcal{C}_i} s(x, P).$$

For all x at least one of the values $t_1(x)$ and $t_2(x)$ is infinite. By (a version of) Proposition 7.20, the functions $t_i(x)$ are lower semicontinuous and hence are $\mathcal{C}_1$- and $\mathcal{C}_2$-tests, respectively.

Now we follow some of the reasoning of the proof of Proposition 7.31. For an integer $k > 1$, consider the open set $S_k = \{x \colon t_1(x) > 2^k\}$. Since t_1 is a $\mathcal{C}_1$-test, $P(S_k) < 2^{-k}$ for all $P \in \mathcal{C}_1$. On the other hand, since for all x one of the two values $t_1(x)$ and $t_2(x)$ is infinite, $P(S_k) = 1$ for all $P \in \mathcal{C}_2$. The indicator function $1_{S_k}(x)$ of the set S_k is lower semicontinuous; therefore, it can be written as the limit of an increasing sequence (now not necessarily computable!) of basic functions $g_{k,n}(x)$. We conclude, as in the proof of Proposition 7.31, that for each P there is an $n = n_k(P)$ with $\int g_{k,n}(x)\, dP > 1 - 2^{-k}$ for all $P \in \mathcal{C}_2$. The effective compactness of $\mathcal{C}_2$ implies then that there is an n independent of P with the same property. In summary, for each $k > 0$ a basic function h_k is found with

$$\int h_k\, dP < 2^{-k} \quad \text{for all } P \in \mathcal{C}_1, \qquad \int h_k\, dP > 1 - 2^{-k} \quad \text{for all } P \in \mathcal{C}_2.$$

Such a basic function h_k can be found effectively from k, by complete enumeration. Now we can construct a lower semicomputable function

$$t_1'(x) = \sum_k h_k(x).$$

It is a test for the class $\mathcal{C}_1$, while $t_2'(x) = \sum_k (1 - h_k(x))$ is a test for all $P \in \mathcal{C}_2$ for the same reasons. These tests must be finite for elements random for P_1 and P_2, and this cannot happen simultaneously for both tests. $\square$

The meaning of separator tests introduced above can be clarified as follows. Due to the effective orthogonality of $\mathcal{C}$, the universal uniform test $\mathbf{t}(\omega, P)$ allows one to separate the sequences into random ones according to different measures of the class $\mathcal{C}$: looking at a sequence ω, random with respect to some measure of this class (= random with respect to the class), we are looking for a $P \in \mathcal{C}$ for which $\mathbf{t}(\omega, P)$ is finite. This measure is unique in the class $\mathcal{C}$ (by the definition of effective orthogonality).

This separation property, however, can also be satisfied by a nonuniversal test, and we called such tests separator tests. The nonuniversal test is less demanding about the idea of randomness, giving it, so to say, a "first approximation": it might accept a sequence as random that will be rejected by a more serious test. (The converse is impossible, since the universal test is maximal.) What matters is only that this preliminary crude triage separates the measures of the class $\mathcal{C}$, that is, that no sequence should appear "random," even "in the first approximation," with respect to two measures at the same time.

For brevity, just for the purposes of the present paper, we will call "typicality" this "randomness in the first approximation":

Definition 8.11. Given a separator test $s(x,P)$, we call an element x *typical* for $P \in \mathcal{C}$ (with respect to the test s) if $s(x,P) < \infty$.

A typical element determines uniquely the measure P for which it is typical.

For an example, consider the class $\mathcal{B}$ of Bernoulli measures. For a test "in the first approximation," we may recall von Mises, who called the stability of relative frequencies the first property of a random sequence ("Kollektiv" in his words). The stability of relative frequencies (strong law of large numbers in today's terminology) means $S_n(\omega)/n \to p$. Here $S_n(\omega)$ is the number of ones in the initial segment of length n of the sequence ω, and p is the parameter of the Bernoulli measure B_p.

There are several close requirements in this spirit:

(1) $S_n(\omega)/n \to p$ with a certain convergence speed.

(2) $S_n(\omega)/n \to p$.

(3) In the case when $\mathcal{C}$ is the class of all ergodic stationary measures over the Cantor space Ω, convert the proof of Theorem 8.7 into a test, implying $A_{x,n}(\omega) \to P(x)$ for all x.

Among these requirements, the one that seems most natural to a mathematician, namely (2), is not expressible in a semicomputable way. Requirement (1) has many possible formulations, depending on the convergence speed; we will show an example below.

Requirement (3) is significantly more complicated to understand, but is still much simpler than a universal test. It *does not* imply a computable convergence speed directly; indeed, as V'yugin showed in [29], a computable convergence speed does not exist for the case of computable nonergodic measures. But later works, starting with [1], have shown that the convergence for ergodic measures has a speed computable from P.

Here is an example of a test expressing requirement (1). (For simplicity, we obtain the convergence of relative frequencies not on all segments, but only on lengths that are powers of two. With more care, one could obtain similar bounds on all initial segments.) By Chebyshev's inequality

$$B_p\bigl(\{x \in \mathbb{B}^n\colon |S_n(x) - np| > \lambda n^{1/2}(p(1-p))^{1/2}\}\bigr) \le \lambda^{-2}.$$

Here $S_n(x)$ is the number of ones among the first n bits. Since $p(1-p) \le 1/4$, this implies

$$B_p\bigl(\{x \in \mathbb{B}^n\colon |S_n(x) - np| > \lambda n^{1/2}/2\}\bigr) \le \lambda^{-2}.$$

Setting $\lambda = n^{0.1}$ and ignoring the factor $1/2$ gives

$$B_p\bigl(\{x \in \mathbb{B}^n\colon |S_n(x) - np| > n^{0.6}\}\bigr) \le n^{-0.2}.$$

For $n = 2^k$,

$$B_p\bigl(\{x \in \mathbb{B}^{2^k}\colon |S_{2^k}(x) - 2^k p| > 2^{0.6k}\}\bigr) \le 2^{-0.2k}.$$

Now, for a sequence ω in $\mathbb{B}^{\mathbb{N}}$ and for $p \in [0,1]$ let

$$g(\omega, B_p) = \sup\{k\colon |S_{2^k}(\omega) - 2^k p| > 2^{0.6k}\}.$$

Then

$$\int g(\omega, B_p)\, B_p(d\omega) \le \sum_k k \cdot 2^{-0.2k} = c < \infty.$$

Dividing by c, we obtain a test. This is a separator test, since $g(\omega, B_p) < \infty$ implies that $2^{-k} S_{2^k}(\omega)$ converges to p, and this cannot happen for two different p.

Theorem 5.41 generalizes, with essentially the same proof (using basic balls instead of initial sequences): it says that in an effectively compact effectively orthogonal class of measures, blind randomness is the same as uniform Martin-Löf randomness. This raises the question as to whether every ergodic measure belongs to some effectively compact class. The answer is negative:

Theorem 8.12. *Consider stationary measures over* Ω (*with the shift transformation*). *Among these, there are some ergodic measures that do not belong to any effectively compact class of ergodic measures.*

Before proving the theorem, let us prove some preparatory statements.

Proposition 8.13. *Both the ergodic measures and the nonergodic measures are dense in the set of stationary measures* $\mathcal{M}(\Omega)$ *over* Ω.

Proof. First we will show how to approximate an arbitrary stationary measure P by ergodic measures. Without loss of generality assume that all probabilities $P(x)$ for finite strings x are positive. (If not, then we can mix in a little of the uniform measure.) For a fixed n, consider the values $P(x)$ on strings x of length at most n. There is a process that reproduces these probabilities and that is isomorphic to an ergodic Markov process on $\{0,1\}^{n-1}$. In this process, for an arbitrary $x \in \{0,1\}^{n-2}$ and any $b, b' \in \{0,1\}$ the transition probability from bx to xb' is $P(bxb')/P(bx)$. Since both transition probabilities are positive, this Markov process is ergodic.

Now we show how to approximate an arbitrary ergodic measure by nonergodic measures. Let P be ergodic. Let us fix some $n > 0$ and $\varepsilon > 0$. By the pointwise ergodic theorem, there is a sequence in which the limiting frequencies of all words converge to the measure (almost all sequences—with respect to this measure—are such). Taking a long piece of this sequence and repeating it leads to a periodic sequence in which the frequencies of words of length not exceeding n differ from the measure P by at most ε (for any given n and $\varepsilon > 0$). (The repetition forms new words on the boundaries, but for a large length this effect is negligible.) Consider now the measure concentrated on the shifts of this sequence, assigning the same weight to each of them (their number is equal to the minimum period). This measure is not ergodic, but is close to P. □

Proposition 8.14. *The set of ergodic measures is a* G_δ *set in the metric space of all measures over* Ω.

Proof. We can restrict attention to the (closed) set of stationary measures. Let P be a stationary probability measure over Ω. Consider the function $A_{x,n}$ over Ω, defining $A_{x,n}(\omega)$ to be equal to the average number of occurrences of the word x in the first n possible positions of ω. By the ergodic theorem, the sequence of functions $A_{x,1}, A_{x,2}, \ldots$ converges in the L_1 sense. Moreover, the stationary measure P is ergodic if and only if the limit of this convergence is the constant function with value $P(x)$.

Since the limit exists for all stationary measures, it is sufficient to check that the constant $P(x)$ is a cluster point. For any x, N and each rational ε the set $S_{x,N,\varepsilon}$ of those P for which there is an $n \ge N$ with

$$\int |A_{x,n}(\omega) - P(x)|\, P(d\omega) < \varepsilon$$

is open, and the set of ergodic stationary measures is the intersection of these sets for all x, N and ε. □

Proof of Theorem 8.12. The union of all effectively compact classes of ergodic measures is F_σ. Suppose that it is equal to the set of all ergodic measures. Then the set of nonergodic measures is a G_δ set, which is also dense by Proposition 8.13.

As shown in Propositions 8.13 and 8.14, the set of ergodic measures is a dense G_δ set. However, by the Baire category theorem, two dense G_δ sets cannot have an empty intersection. This contradiction proves the theorem. □

The following question still remains open:

Question 6. Is there an ergodic measure over Ω for which uniform and blind randomness are different?

Returning to arbitrary effectively compact effectively orthogonal classes, we can connect the universal tests with class tests of Theorem 5.23 and separator tests.

Theorem 8.15. *Let $\mathcal{C}$ be an effectively compact effectively orthogonal class of measures, let $\mathbf{t}(x,P)$ be the universal uniform test, and let $\mathbf{t}_{\mathcal{C}}(x)$ be a universal class test for $\mathcal{C}$. Assume that $s(x,P)$ is a separator test for $\mathcal{C}$. Then we have the representation*

$$\mathbf{t}(x,P) \stackrel{*}{=} \max(\mathbf{t}_{\mathcal{C}}(x), s(x,P))$$

for all $P \in \mathcal{C}$ and $x \in X$.

Proof. Let us note first that $\mathbf{t}_{\mathcal{C}}(x)$ and $s(x,P)$ do not exceed the universal uniform test $\mathbf{t}(x,P)$. Indeed, $s(x,P)$ is a uniform test by definition. Also by definition, the universal class test $\mathbf{t}_{\mathcal{C}}(x)$ is a uniform test.

On the other hand, let us show that if $\mathbf{t}_{\mathcal{C}}(x)$ and $s(x,P)$ are finite, then $\mathbf{t}(x,P)$ does not exceed the greater one of them (to within a multiplicative constant). The finiteness of the first test guarantees that $\min_{Q\in\mathcal{C}} \mathbf{t}(x,Q)$ is finite: this minimum is equal to $\mathbf{t}_{\mathcal{C}}(x)$ to within a constant factor. If this minimum was achieved on some measure $Q \neq P$, then both values $s(x,Q)$ and $s(x,P)$ would be finite, contradicting to the definition of a separator. (Note that we proved a statement slightly stronger than promised: in place of "greater of the two," one can write "the first of the two, if the second one is finite.") □

The above theorem splits the randomness test into two parts (points at two possible causes of non-randomness). First, we must convince ourselves that x is random with respect to the class $\mathcal{C}$. For example, in the case of a measure B_p in the class $\mathcal{B}$ of Bernoulli measures, we must first be convinced that $\mathbf{t}_{\mathcal{B}}(\omega)$ is finite. This encompasses all the irregularity criteria. If the independence of the sequence is taken for granted, we may assume that the class test is satisfied.

After this, we know that our sequence is Bernoulli, and some kind of simple test of the type of the law of large numbers is sufficient to find out with respect to which Bernoulli measure is it random: B_p or some other one. This second part, typicality testing, is analogous to parameter testing in statistics.

Separation is the only requirement of the separator test: its numerical value is irrelevant. For example, in the Bernoulli test case, no matter how crude the convergence criterion expressed by the separator test $s(x,P)$ is, the maximum is always (essentially) the same universal test.

9. ARE UNIFORM TESTS TOO STRONG?

9.1. Monotonicity and/or quasi-convexity. Uniform tests may seem too strong, in case P is a noncomputable measure. In particular, randomness with respect to computable measures (in the sense of Martin-Löf or in the uniform sense; they are the same for computable measures) has certain intuitively desirable properties that uniform randomness lacks. One of these is monotonicity: roughly, if Q is greater than P and if x is random with respect to P, then it should also be random with respect to Q.

Proposition 9.1. *For computable measures P and Q and for all rational $\lambda > 0$, if $\lambda P(A) \le Q(A)$ for all A, then*

$$\mathbf{m}(\lambda) \cdot \lambda\, \mathbf{t}(x, Q) \overset{*}{<} \mathbf{t}(x, P).$$

Here $\mathbf{m}(\lambda)$ is the discrete a priori probability of the rational λ. To make the constant in $\overset{*}{<}$ independent of P and Q, one also needs to multiply the left-hand side by $\mathbf{m}(P, Q)$.

Proof. We have

$$1 \ge \int \mathbf{t}(x, Q)\, dQ \ge \int \lambda\, \mathbf{t}(x, Q)\, dP;$$

hence $\lambda\, \mathbf{t}(x, Q)$ is a P-test. Using the trimming method of Theorem 7.28 in finding universal tests, one can show that the sum

$$\sum_{\lambda :\, \lambda \int \mathbf{t}(x,Q)\, dP < 2} \mathbf{m}(\lambda) \cdot \lambda\, \mathbf{t}(x, Q)$$

is a P-test, and hence does not exceed $\mathbf{t}(x, P)$ up to a multiplicative constant. Therefore, this is true of each member of the sum, which is just what the theorem claims. It is easy to see that the multiplicative constants depend here on P and Q only via inserting a factor $\mathbf{m}(P, Q)$. □

The intuitive motivation for monotonicity is as follows: if there are two devices with internal randomness generators, having output distributions P and Q, and $\lambda P \le Q$, then it can be imagined that the second device simulates the first one with probability λ and does its own thing otherwise. Then every outcome intuitively plausible as the outcome of the first device must also be deemed a plausible outcome of the second one, since this could have simulated the first one by chance. (The numerical value of the randomness deficiency may be, of course, somewhat larger, since we must believe in addition that the λ-probability event occurred.)

Uniform randomness violates, alas, this property: if the measure Q is larger but computationally more complex, then the randomness tests with respect to Q can exploit this additional information to make nonrandom some outcomes that were random with respect to P (see the proof of Theorem 5.39). This is just the source of the difference between uniform and blind (oracle-free) randomness, for which the analogous monotonicity property is obviously satisfied.

Another situation for which we have some intuition on randomness is the mixture (convex combination) of measures. Imagine two devices with output measures P and Q, and an outer box which triggers one of them with some probabilities λ and $1 - \lambda$. As a whole, we obtain a system whose outcome is distributed according to the measure $\lambda P + (1 - \lambda)Q$. About which outcomes can we assert that they are obtained randomly as a result of this experiment? Clearly both the outcomes random with respect to P and those random with respect to Q must be accepted (with the understanding that if the coefficient is small, then some additional but finite suspicion is added). And there should not be any other outcomes. A quantitative elaboration of this result (which in one direction follows from monotonicity) is given below.

Proposition 9.2. *Let P and Q be two computable measures.*

(a) *For a rational $0 < \lambda < 1$,*

$$\mathbf{m}(\lambda) \cdot \mathbf{t}(x, \lambda P + (1 - \lambda)Q) \overset{*}{<} \max(\mathbf{t}(x, P), \mathbf{t}(x, Q)).$$

(b) *For arbitrary $0 < \lambda < 1$,*

$$\mathbf{t}(x, \lambda P + (1 - \lambda)Q) \overset{*}{>} \min(\mathbf{t}(x, P), \mathbf{t}(x, Q)).$$

The constants in $\overset{*}{<}$ depend on the length of the shortest programs defining P and Q (their complexities), but not on λ (or other aspects of P and Q).

Statement (a) could be called the *quasi-convexity* of randomness tests (to within a multiplicative constant). For a test with an exact quasi-convexity property (without any multiplicative constants) there is a lower semicomputable semimeasure that is neutral (after extending tests to semimeasures, see [17, 9]).

Statement (b) implies that no other random outcomes exist for the mixture of P and Q. This could be called the *quasi-concavity* of randomness tests (to within a multiplicative constant).

Proof. Part (a) follows from Proposition 9.1. Indeed, if $\lambda \geq 1/2$, then Proposition 9.1 implies $\mathbf{m}(\lambda) \cdot \mathbf{t}(x, \lambda P + (1-\lambda)Q) \stackrel{*}{<} \mathbf{t}(x, P)$ (absorbing $1/2$ into the $\stackrel{*}{<}$). If $\lambda < 1/2$, then it implies $\mathbf{m}(1-\lambda) \cdot \mathbf{t}(x, \lambda P + (1-\lambda)Q) \stackrel{*}{<} \mathbf{t}(x, Q)$ similarly, and we just recall $\mathbf{m}(\lambda) \stackrel{*}{=} \mathbf{m}(1-\lambda)$.

Part (b) follows from the fact that the right-hand side is a test with respect to an arbitrary mixture of the measures P and Q, and trimming can convert it into a uniform test. $\square$

It is easy to see that all these statements exploit the computability of the measures and the mixing coefficients in an essential way. The corresponding counterexamples are easy to build once it is recognized that the mixture of measures can be stronger from an oracle-computational point of view than any of them, as well as can be weaker. For example, let us divide the segment $[0, 1]$ into two halves and consider the measures P and Q that are uniformly distributed over these halves. Their mixture with coefficients λ and $1 - \lambda$ will make the number λ obviously nonrandom (since it can be computed from this measure), though with respect to one of the measures it can very well be random. Taking instead of P and Q their mixtures, say, with coefficients $1/3$ and $2/3$ and then reversed, one can make λ random with respect to both measures.

In this example the mixture contains more information than each of the original measures. It can also be the other way: bend the interval $[0, 1]$ with the uniform measure into a circle and cut it into two half-circles by the points p and $p + 1/2$. Then the uniform measures on these half-circles make p computable with respect to them and thus nonrandom, while the average of these measures is the uniform measure on the circle, with respect to which p can very well be random.

Let us note that for blind (oracle-free) randomness, we can guarantee without any restrictions that the set of points random with respect to the mixture of P and Q is the union of points random with respect to P and Q. (In one direction this follows from monotonicity, which we already mentioned. In the other one: if an outcome is not random with respect to P and not random with respect to Q, then there are two tests proving this, and their minimum will be a lower semicomputable test proving its non-randomness with respect to the mixture.)

These are strong motives for modifying the concept of randomness test in order to reproduce these properties, while conserving other desirable properties (say, the existence of a universal test and with it the notion of deficiency of randomness). Some of such modifications can be seen in [17, 9, 18].

9.2. Locality. Imagine that some sequence ω is uniformly random with respect to a measure P and starts with 0. Change the values of the measure on sequences that start with 1. It is not guaranteed that ω remains uniformly random, since now the measure may become stronger as an oracle (allowing one to compute more). But this looks strange since the changes in the measure are in the part of the universe that does not touch ω.

For blind (oracle-free) randomness, specifically this example is impossible (one can force the test to zero on sequences beginning with unity), but in principle the concept of test depends not only on the measure along the sequence (not only on the probabilities of occurrences of nulls and ones after its start).

For randomness with respect to computable measures, the situation is again better.

Proposition 9.3 (prequentiality)**.** *Let P and Q be two computable measures on the space Ω of binary sequences. If P and Q coincide on all initial segments of some sequence ω, then this sequence is simultaneously random or nonrandom with respect to P and Q.*

Proof. This follows immediately from the randomness criterion in terms of the complexity of the initial segments (Levin–Schnorr theorem) in any of its variants (Theorem 2.24, Proposition 2.30, or Corollary 2.32). □

On the other hand, it is easy to modify one of the counterexamples in Subsection 9.1 to violate prequentiality as well.

In the case of computable measures on an arbitrary constructive metric space a similar statement holds, though with a stronger requirement: we assume that two computable measures are equal on all sets contained in some neighborhood of the outcome ω. (In this case it is possible to multiply the test by a basic function without changing it in ω, while making it zero outside the neighborhood of coincidence).

Here is yet another way to obtain a clearly prequential definition of randomness, in which the randomness deficiency is a function of the sequence itself and the measures of its initial segments. For a given sequence ω and a given sequence $\{q(i)\}$ of real numbers with

$$1 = q(0) \geq q(1) \geq q(2) \geq \ldots \geq 0,$$

let

$$\mathbf{t}'(\omega, q) = \inf \mathbf{t}(\omega, P),$$

where the infimum is taken over all measures P with $P(\omega(1:n)) = q(n)$. The corresponding sets are effectively compact, so that this minimum will be a lower semicomputable function of ω and the sequence q. If for the sequence ω and the measures $q(i)$ of its initial segments the value $\mathbf{t}'(\omega, q)$ is finite, then the sequence ω can be called *prequentially random*.

In other words, a sequence ω is prequentially random with respect to a measure P if there is a (in general different) measure Q with respect to which ω is random and which coincides with P on all initial segments of ω.

The requirement of prequentiality has been invoked in connection with a theory that extends probability theory and statistics to models of forecasting (see, for example, [5, 27]). An example situation is the following. Let $\omega(n) = 1$ mean that there is rain on day n and $\omega(n) = 0$ mean the absence of rain. Suppose that a forecasting office makes daily forecasts $p(1), p(2), \ldots$ of the probability of rain. It is not necessarily proposing a coherent probability model of global weather (a global probability distribution). It just provides forecasts for the conditional probabilities along the path corresponding to the weather that actually takes place.

Is it possible to estimate the quality of the forecast? It seems that in some situations, yes: if, for example, all forecasts are close to zero (say, are less than 10%) and the majority of days (say, more than 90%) are rainy. (It is said that the forecast is poorly *calibrated*.) Naturally, there are other possible inconsistencies, not related to the frequencies: the general question is whether the given sequence can be accepted as randomly obtained with the predicted probabilities. (Such a question also arises in the situation of estimating the quality of a random number generator each of whose output values is claimed to occur with whatever distribution the customer requires at that time of the process, for that particular bit.)

An additional circumstance to consider at the estimation of the quality of forecasts is that the forecaster can use a variety of information accessible to them at the moment of prediction (say, the evening of the preceding day), and not only the members of the sequence ω. The presence of such information must also be taken into account at the estimation of the quality of the forecast.

Paper [27] proposes several different approaches to this question, which turn out to be equivalent. One involves a generalization of the notion of martingale (see Definition 3.8). It would be interesting to establish a connection with uniform randomness tests in the spirit of the above defined prequential deficiency. (Admittedly, in place of probabilities of initial segments, one must deal here with conditional probabilities, which is not quite the same, if these are not separated from zero.)

10. QUESTIONS FOR FUTURE DISCUSSION

We have already noted some questions that (in our view) would be interesting to study. In this section we collected a few more such questions.

1. Consider the following method for generating a sequence $\xi \in \Omega$ using an arbitrary distribution P on Ω in which the probabilities of all words are positive. Take a random sequence ρ of independent reals $\rho(1), \rho(2), \ldots$ uniformly distributed over $[0,1]$. At stage n, after outputting $\xi(1:n-1)$, set $\xi(n) = 1$ if

$$\rho(n) < P(\xi(1:n-1)1)/P(\xi(1:n-1)).$$

If we consider this as a random process, the output distribution will be exactly P. What sequences can be obtained at the output from a Martin-Löf random sequence of real numbers at the input? (It can be verified that for computable measures P one gets exactly the sequences that are Martin-Löf random with respect to P.)

2. Recall the formula for the deficiency for computable measures:

$$\mathbf{t}(\omega, P) \stackrel{*}{=} \sum_{x \sqsubseteq \omega} \frac{\mathbf{m}(x)}{P(x)}. \tag{8}$$

Both sides make sense for noncomputable P, but this formula is no more true. Indeed, the right-hand side does not change significantly if a measure P is replaced by some other one that is close to P but is much more powerful as an oracle; and the left-hand side can become infinite while it was finite for P.

Denote the right-hand side by $t'(\omega, P)$. Does it make sense to take the finiteness of $t'(\omega, P)$ as a definition of randomness with respect to a noncomputable measure? It will be at least monotonic (an increase of the measure will only increase randomness). With respect to mixtures of measures, we can say that it is quasi-convex; moreover, it is proved in [8] that $1/t'(\omega, P)$ is a concave function of P.

Another possibility is to define the randomness deficiency for an infinite sequence ω as $\log \sup_{x \sqsubseteq \omega} [M(x)/P(x)]$ (and consider the corresponding definition of randomness). For computable measures we obtain a definition equivalent to Martin-Löf's standard one. Paper [11] shows that the uniform tests defined by this expression (using either $\mathbf{m}(x)$ or $M(x)$) do not obey randomness conservation, while the universal uniform test does. Work [8] shows that, on the other hand, an expression related to the right-hand side of (8), with the summation running over all positive basic functions instead of only the functions $1_{x\Omega}(\omega)$, obeys randomness conservation.

3. Can we define a reasonable class of tests with the property in Proposition 9.1 holding for all measures P (or some stronger version of it) so that there exists a universal class? For example, one may require

$$P \le c \cdot Q \;\Rightarrow\; t(\omega, P) \ge t(\omega, Q)/c$$

(motivation: this is true for the right-hand side of formula (8)). Could one also require the quasi-convexity, as in Proposition 9.2? Papers [17] and [9], as well as [18], provide some such examples.

How about the quasi-concavity of Proposition 9.2? A uniform test with this property seems less likely, since our counterexample seems more robust.

4. Relativization in recursion theory means that we take some set A and artificially declare it "decidable" by adding some oracle that tells us whether $x \in A$ for any given x. Almost all the theorems of classical recursion theory can be relativized. It is more delicate to declare some set E "enumerable." This means that we have some enumeration oracle that enumerates the set E. The problem is, of course, that there are many enumerations. Still we can give the definition of an E-enumerable set. Let W be a set of pairs of the form (x, S) where x is an integer and S

is a finite set of integers; assume W to be enumerable in the classical sense. Then consider the set $\Gamma(E, W)$ of all x such that $(x, S) \in W$ for some $S \subset E$. The sets $\Gamma(E, W)$ (for fixed E and all enumerable W) are called *enumerable with respect to the enumeration oracle* E. (The relation $(x, S) \in W$ means that we add x to the E-enumeration as soon as we see all elements of S in E.) A standard (decision) oracle for a set A can be considered a special case of an enumeration oracle (say, for the set $\{2n\colon n \in A\} \cup \{2n+1\colon n \notin A\}$).

For some purposes, an enumeration oracle is as meaningful as a decision oracle: for example, we can speak about a lower semicomputable function with respect to enumeration oracle E, since it can be defined in terms of enumerable sets. But what can be proved for this kind of relativized notions?

For example, is there (for an arbitrary E) a maximal lower E-semicomputable semimeasure? Can one define prefix complexity with oracle E, and will it coincide with the logarithm of the maximal semimeasure lower semicomputable relative to E (if the latter exists)? What if we assume, in addition, that E is the set of all basic balls in a constructive metric space that contain a given point?

(For comparison: we could define an E-computable function as a function whose graph is E-enumerable. Then some familiar properties will hold; say, the composition of E-computable functions is again E-computable. On the other hand, we cannot guarantee that every nonempty E-enumerable set is the range of a total E-computable function: for some E this is not so.)

5. We may try to extend the definition of randomness in a different direction: to lower semicomputable semimeasures (that is, output distributions of probabilistic machines that generate an output sequence bit by bit). Levin's motivation for his definition was his goal to define the independence of a pair (ξ, η) of infinite sequences as randomness with respect to the semimeasure $M \times M$. Correspondingly, the deficiency of randomness of the pair (ξ, η) with respect to $M \times M$ could be called the quantity of mutual information between the sequences ξ and η. This is motivated by the fact that the algorithmic mutual information

$$\mathrm{Kp}(x) + \mathrm{Kp}(y) - \mathrm{Kp}(x, y) = -\log(\mathbf{m}(x) \times \mathbf{m}(y)) - \mathrm{Kp}(x, y)$$

between finite objects x and y indeed looks like deficiency of randomness with respect to $\mathbf{m} \times \mathbf{m}$.

One possibility is to say that ω is random if $M(z)/Q(z)$ is bounded for the prefixes z of ω, where M is the a priori probability on the tree and Q is the semimeasure in question. Another possibility is to use random sequences for unbiased coin tossing and consider the output sequences in all these cases. It is not clear whether these two definitions coincide or if the second notion is well-defined (that is, for two different machines with the same output distribution the image of the set of random sequences is the same). For *computable measures* it is indeed the case.

6 (Steven Simpson). Can we use uniform tests (modified in a proper way) for defining, say, 2-randomness? (The standard definition uses non-semicontinuous tests, but maybe it can be reformulated.)

ACKNOWLEDGMENTS

The authors are grateful to their colleagues with whom they have discussed the questions considered in the paper: first to L. Levin, who introduced (or inspired) many of the concepts considered in the paper, to A. Bufetov and A. Klimenko, and also to V. V'yugin and other participants of the Kolmogorov seminar (at the Faculty of Mechanics and Mathematics, Moscow State University).

The paper was written with the financial support of the grant NAFIT ANR-08-EMER-008-01 and of the Russian Foundation for Basic Research (project no. 09-01-00709-a).

REFERENCES

1. J. Avigad, P. Gerhardy, and H. Towsner, "Local Stability of Ergodic Averages," Trans. Am. Math. Soc. **362** (1), 261–288 (2010).
2. L. Bienvenu, A. Day, M. Hoyrup, I. Mezhirov, and A. Shen, "A Constructive Version of Birkhoff's Ergodic Theorem for Martin-Löf Random Points," arXiv: 1007.5249 [math.DS].
3. L. Bienvenu, A. Romashchenko, and A. Shen, "Sparse Sets," in *Journées Automates Cellulaires: Proc. First Symp. on Cellular Automata, Uzès, France, 2008*, pp. 18–28, http://hal.archives-ouvertes.fr/docs/00/27/40/10/PDF/18-28.pdf
4. G. J. Chaitin, "A Theory of Program Size Formally Identical to Information Theory," J. ACM **22**, 329–340 (1975).
5. A. Chernov, A. Shen, N. Vereshchagin, and V. Vovk, "On-line Probability, Complexity and Randomness," in *Algorithmic Learning Theory: Proc. 19th Int. Conf., Budapest, 2008*, Ed. by Y. Freund et al. (Springer, Berlin, 2008), Lect. Notes Comput. Sci. **5254**, pp. 138–153.
6. A. R. Day and J. S. Miller, "Randomness for Non-computable Measures," Trans. Am. Math. Soc. (in press); http://www.math.wisc.edu/~jmiller/Papers/neutral.pdf
7. P. Gács, "Lecture Notes on Descriptional Complexity and Randomness," Tech. Rep. (Boston Univ., Comput. Sci. Dept., Boston, MA, 1988), http://www.cs.bu.edu/~gacs/papers/ait-notes.pdf
8. P. Gács, "Komplexität und Zufälligkeit," PhD Thesis (J.W. Goethe Univ., Frankfurt, 1978).
9. P. Gács, "Exact Expressions for Some Randomness Tests," Z. Math. Logik Grundlagen Math. **26**, 385–394 (1980). Short version in *Theoretical Computer Science: Proc. 4th GI Conf.* (Springer, Berlin, 1979), Lect. Notes Comput. Sci. **67**, pp. 124–131.
10. P. Gács, "On the Relation between Descriptional Complexity and Algorithmic Probability," Theor. Comput. Sci. **22**, 71–93 (1983). Short version in *Proc. 22nd Annu. Symp. on Foundations of Computer Science, Nashville, 1981* (IEEE Comput. Soc., New York, 1981), pp. 296–303.
11. P. Gács, "Uniform Test of Algorithmic Randomness over a General Space," Theor. Comput. Sci. **341** (1–3), 91–137 (2005).
12. M. Hoyrup and C. Rojas, "An Application of Martin-Löf Randomness to Effective Probability Theory," in *Mathematical Theory and Computational Practice: Proc. Conf. CiE 2009* (Springer, Berlin, 2009), Lect. Notes Comput. Sci. **5635**, pp. 260–269.
13. M. Hoyrup and C. Rojas, "Computability of Probability Measures and Martin-Löf Randomness over Metric Spaces," Inf. Comput. **207** (7), 830–847 (2009).
14. A. N. Kolmogorov, "Logical Basis for Information Theory and Probability Theory," Probl. Peredachi Inf. **5** (3), 3–7 (1969) [Probl. Inf. Transm. **5** (3), 1–4 (1969)]; IEEE Trans. Inf. Theory **14** (5), 662–664 (1968).
15. K. Kuratowski, *Topology* (Academic Press, New York, 1966, 1968), Vols. 1, 2.
16. L. A. Levin, "On the Notion of a Random Sequence," Dokl. Akad. Nauk SSSR **212** (3), 548–550 (1973) [Sov. Math., Dokl. **14** (5), 1413–1416 (1973)].
17. L. A. Levin, "Uniform Tests of Randomness," Dokl. Akad. Nauk SSSR **227** (1), 33–35 (1976) [Sov. Math., Dokl. **17** (2), 337–340 (1976)].
18. L. A. Levin, "Randomness Conservation Inequalities: Information and Independence in Mathematical Theories," Inf. Control **61** (1), 15–37 (1984).
19. M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, 3rd ed. (Springer, New York, 2008).
20. P. Martin-Löf, "The Definition of Random Sequences," Inf. Control **9**, 602–619 (1966).
21. Yu. T. Medvedev, "Degrees of Difficulty of Mass Problems," Dokl. Akad. Nauk SSSR **104** (4), 501–504 (1955).
22. J. S. Miller, "Degrees of Unsolvability of Continuous Functions," J. Symb. Log. **69** (2), 555–584 (2004).
23. C. P. Schnorr, "Process Complexity and Effective Random Tests," J. Comput. Syst. Sci. **7** (4), 376–388 (1973). Conf. version in *Proc. 4th Annu. ACM Symp. on Theory of Computing, Denver, 1972* (ACM, New York, 1972), pp. 168–176.
24. G. Shafer, A. Shen, N. Vereshchagin, and V. Vovk, "Test Martingales, Bayes Factors, and p-Values," arXiv: 0912.4269v2 [math.ST].
25. G. Shafer and V. Vovk, *Probability and Finance: It's Only a Game!* (Wiley, Chichester, 2001).
26. A. Shen, "Algorithmic Information Theory and Kolmogorov Complexity," Tech. Rep. 2000-034 (Uppsala Univ., Uppsala, 2000), http://www.it.uu.se/research/publications/reports/2000-034/

27. V. Vovk and A. Shen, "Prequential Randomness and Probability," Theor. Comput. Sci. **411** (29–30), 2632–2646 (2010).
28. V. Strassen, "The Existence of Probability Measures with Given Marginals," Ann. Math. Stat. **36**, 423–439 (1965).
29. V. V. V'yugin, "Ergodic Theorems for Individual Random Sequences," Theor. Comput. Sci. **207** (2), 343–361 (1998).
30. A. K. Zvonkin and L. A. Levin, "The Complexity of Finite Objects and the Development of the Concepts of Information and Randomness by Means of the Theory of Algorithms," Usp. Mat. Nauk **25** (6), 85–127 (1970) [Russ. Math. Surv. **25** (6), 83–124 (1970)].

Translated by the authors