UNIFORM TEST OF ALGORITHMIC RANDOMNESS OVER A GENERAL SPACE

PETER GÁCS

ABSTRACT. The algorithmic theory of randomness is well developed when the underlying space is the set of finite or infinite sequences and the underlying probability distribution is the uniform distribution or a computable distribution. These restrictions seem artificial. Some progress has been made to extend the theory to arbitrary Bernoulli distributions (by Martin-Löf), and to arbitrary distributions (by Levin). We recall the main ideas and problems of Levin's theory, and report further progress in the same framework. The issues are the following:
 – Allow non-compact spaces (like the space of continuous functions, underlying the Brownian motion).
 – The uniform test (deficiency of randomness) $\mathbf{d}_P(x)$ (depending both on the outcome x and the measure P) should be defined in a general and natural way.
 – See which of the old results survive: existence of universal tests, conservation of randomness, expression of tests in terms of description complexity, existence of a universal measure, expression of mutual information as "deficiency of independence".
 – The negative of the new randomness test is shown to be a generalization of complexity in continuous spaces; we show that the addition theorem survives.
The paper's main contribution is introducing an appropriate framework for studying these questions and related ones (like statistics for a general family of distributions).

1. INTRODUCTION

1.1. **Problem statement.** The algorithmic theory of randomness is well developed when the underlying space is the set of finite or infinite sequences and the underlying probability distribution is the uniform distribution or a computable distribution. These restrictions seem artificial. Some progress has been made to extend the theory to arbitrary Bernoulli distributions by Martin-Löf in [15], and to arbitrary distributions, by Levin in [11, 12, 13]. The paper [10] by Hertling and Weihrauch also works in general spaces, but it is restricted to computable measures. Similarly, Asarin's thesis [1] defines randomness for sample paths of the Brownian motion: a fixed random process with computable distribution.

The present paper has been inspired mainly by Levin's early paper [12] (and the much more elaborate [13] that uses different definitions): let us summarize part of the content of [12]. The notion of a constructive topological space $\mathbf{X}$ and the space of measures over $\mathbf{X}$ is introduced. Then the paper defines the notion of a uniform test. Each test is a lower semicomputable function $(\mu, x) \mapsto f_\mu(x)$, satisfying $\int f_\mu(x)\mu(dx) \leqslant 1$ for each measure μ. There are also some additional conditions. The main claims are the following.

(a) There is a universal test $\mathbf{t}_\mu(x)$, a test such that for each other test f there is a constant $c > 0$ with $f_\mu(x) \leqslant c \cdot \mathbf{t}_\mu(x)$. The *deficiency of randomness* is defined as $\mathbf{d}_\mu(x) = \log \mathbf{t}_\mu(x)$.

(b) The universal test has some strong properties of "randomness conservation": these say, essentially, that a computable mapping or a computable randomized transition does not decrease randomness.

(c) There is a measure M with the property that for every outcome x we have $\mathbf{t}_M(x) \leqslant 1$. In the present paper, we will call such measures *neutral*.

(d) Semimeasures (semi-additive measures) are introduced and it is shown that there is a lower semicomputable semimeasure that is neutral (so we can assume that the M introduced above is lower semicomputable).

(e) Mutual information $I(x : y)$ is defined with the help of (an appropriate version of) Kolmogorov complexity, between outcomes x and y. It is shown that $I(x : y)$ is essentially equal to $\mathbf{d}_{M \times M}(x, y)$. This interprets mutual information as a kind of "deficiency of independence".

This impressive theory leaves a number of issues unresolved:

1. The space of outcomes is restricted to be a compact topological space, moreover, a particular compact space: the set of sequences over a finite alphabet (or, implicitly in [13], a compactified infinite alphabet). However, a good deal of modern probability theory happens over spaces that are not even locally compact: for example, in case of the Brownian motion, over the space of continuous functions.

2. The definition of a uniform randomness test includes some conditions (different ones in [12] and in [13]) that seem somewhat arbitrary.

3. No simple expression is known for the general universal test in terms of description complexity. Such expressions are nice to have if they are available.

1.2. **Content of the paper.** The present paper intends to carry out as much of Levin's program as seems possible after removing the restrictions. It leaves a number of questions open, but we feel that they are worth to be at least formulated. A fairly large part of the paper is devoted to the necessary conceptual machinery. Eventually, this will also allow to carry further some other initiatives started in the works [15] and [11]: the study of tests that test nonrandomness with respect to a whole class of measures (like the Bernoulli measures).

Constructive analysis has been developed by several authors, converging approximately on the same concepts. We will make use of a simplified version of the theory introduced in [24]. As we have not found a constructive measure theory in the literature fitting our purposes, we will develop this theory here, over (constructive) complete separable metric spaces. This generality is well supported by standard results in measure theoretical probability, and is sufficient for constructivizing a large part of current probability theory.

The appendix recalls some of the needed topology, measure theory and constructive analysis. Constructive measure theory is introduced in Section 2.

Section 3 introduces uniform randomness tests. It proves the existence of universal uniform tests, under a reasonable assumption about the topology ("recognizable Boolean inclusions"). Then it proves conservation of randomness.

Section 4 explores the relation between description (Kolmogorov) complexity and uniform randomness tests. After extending randomness tests over non-normalized measures, its negative logarithm will be seen as a generalized description complexity.

The rest of the section explores the extent to which the old results characterizing a random infinite string by the description complexity of its segments can be extended to the new setting. We will see that the simple formula working for computable measures over infinite sequences does not generalize. However, still rather simple formulas are available

in some cases: namely, the discrete case with general measures, and a space allowing a certain natural cell decomposition, in case of computable measures.

Section 5 proves Levin's theorem about the existence of a neutral measure, for compact spaces. Then it shows that the result does not generalize to non-compact spaces, not even to the discrete space. It also shows that with our definition of tests, the neutral measure cannot be chosen semicomputable, even in the case of the discrete space with one-point compactification.

Section 6 takes up the idea of viewing the negative logarithm of a randomness test as generalized description complexity. Calling this notion *algorithmic entropy*, this section explores its information-theoretical properties. The main result is a (nontrivial) generalization of the addition theorem of prefix complexity (and, of course, classical entropy) to the new setting.

1.3. **Some history.** Attempts to define randomness rigorously have a long but rather sparse history starting with von Mises and continuing with Wald, Church, Ville. Kolmogorov's work in this area inspired Martin-Löf whose paper [15] introduces the notion of randomness used here.

Description complexity has been introduced independently by Solomonoff, Kolmogorov and Chaitin. Prefix complexity has been introduced independently by Levin and Chaitin. See [14] for a discussion of priorities and contributions. The addition theorem (whose generalization is given here) has been proved first for Kolmogorov complexity, with a logarithmic error term, by Kolmogorov and Levin. For the prefix complexity its present form has been proved jointly by Levin and Gács in [6], and independently by Chaitin in [5].

In his PhD thesis, Martin-Löf also characterized randomness of finite sequences via their complexity. For infinite sequences, complete characterizations of their randomness via the complexity of their segments were given by Levin in [11], by Schnorr in [17] and in [5] (attributed). Of these, only Levin's result is formulated for general computable measures: the others apply only to coin-tossing. Each of these works uses a different variant of description complexity. Levin uses monotone complexity and the logarithm of the universal semicomputable measure (see [8] for the difficult proof that these two complexities are different). Schnorr uses "process complexity" (similar to monotone complexity) and prefix complexity. The work [7] by the present author gives characterizations using the original Kolmogorov complexity (for general computable measures).

Uniform tests over the space of infinite sequences, randomness conservation and neutral measures were introduced in Levin's work [12]. The present author could not verify every result in that paper (which contains no proofs); he reproduced most of them with a changed definition in [7]. A universal uniform test with yet another definiton appeared in [13]. In this latter work, "information conservation" is a central tool used to derive several results in logic. In the constellation of Levin's concepts, information conservation becomes a special case of randomness conservation. We have not been able to reproduce this exact relation with our definition here.

The work [9] is based on the observation that Zurek's idea on "physical" entropy and the "cell volume" approach of physicists to the definition of entropy can be unified: Zurek's entropy can be seen as an approximation of the limit arising in a characterization of a randomness test by complexity. The author discovered in this same paper that the negative logarithm of a general randomness test can be seen as a generalization of complexity. He felt encouraged by the discovery of the generalized addition theorem presented here.

The appearence of other papers in the meantime (including [10]) convinced the author that there is no accessible and detailed reference work on algorithmic randomness

for general measures and general spaces, and a paper like the present one, developing the foundations, is needed. (Asarin's thesis [1] does develop the theory of randomness for the Brownian motion. It is a step in our direction in the sense that the space is not compact, but it is all done for a single explicitly given computable measure.)

We do not advocate the uniform randomness test proposed here as necessarily the "definitive" test concept. Perhaps a good argument can be found for some additional conditions, similar to the ones introduced by Levin, providing additional structure (like a semicomputable neutral measure) while preserving naturalness and the attractive properties presented here.

1.4. **Notation for the paper.** (Nothing to do with the formal concept of "notation", introduced later in the section on constructive analysis.) The sets of natural numbers, integers, rational numbers, real numbers and complex numbers will be denoted respectively by $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$. The set of nonnegative real numbers will be denoted by $\mathbb{R}_+$. The set of real numbers with $-\infty, \infty$ added (with the appropriate topology making it compact) will be denoted by $\overline{\mathbb{R}}$. We use $\wedge$ and $\vee$ to denote $\min$ and $\max$, further

$$|x|^+ = x \vee 0, \quad |x|^- = |-x|^+$$

for real numbers x. We partially follow [24], [3] and [10]. In particular, adopting the notation of [24], we denote intervals of the real line as follows, (to avoid the conflict with the notation of a pair (a, b) of objects).

$$[a; b] = \{\, x : a \leqslant x \leqslant b \,\}, \quad (a; b) = \{\, x : a < x < b \,\}, \quad [a; b) = \{\, x : a \leqslant x < b \,\}.$$

If X is a set then X^* is the set of all finite strings made up of elements of X, including the "empty string" Λ. We denote by X^ω the set of all infinite sequences of elements of X. If A is a set then $1_A(x)$ is its indicator function, defined to be 1 if $x \in A$ and to 0 otherwise. For a string x, its length is $|x|$, and

$$x^{\leqslant n} = (x(1), \ldots, x(n)).$$

The relations

$$f \stackrel{+}{<} g, \quad f \stackrel{*}{<} g$$

mean inequality to within an additive constant and multiplicative constant respectively. The first is equivalent to $f \leqslant g + O(1)$, the second to $f = O(g)$. The relation $f \stackrel{*}{=} g$ means $f \stackrel{*}{<} g$ and $f \stackrel{*}{>} g$.

Borrowing from [16], for a function f and a measure μ, we will use the notation

$$\mu f = \int f(x)\mu(dx), \quad \mu^y f(x, y) = \int f(x, y)\mu(dy).$$

2. CONSTRUCTIVE MEASURE THEORY

The basic concepts and results of measure theory are recalled in Section B. For the theory of measures over metric spaces, see Subsection B.6. We introduce a certain fixed, enumerated sequence of Lipschitz functions that will be used frequently. Let $\mathcal{F}_0$ be the set of functions of the form $g_{u,r,1/n}$ where $u \in D$, $r \in \mathbb{Q}$, $n = 1, 2, \ldots$, and

$$g_{u,r,\varepsilon}(x) = |1 - |d(x, u) - r|^+/\varepsilon|^+$$

is a continuous function that is 1 in the ball $B(u, r)$, it is 0 outside $B(u, r + \varepsilon)$, and takes intermediate values in between. Let

$$\mathcal{E} = \{g_1, g_2, \ldots\} \tag{2.1}$$

be the smallest set of functions containing $\mathcal{F}_0$ and the constant 1, and closed under $\vee$, $\wedge$ and rational linear combinations. The following construction will prove useful later.

Proposition 2.1. *All bounded continuous functions can be obtained as the limit of an increasing sequence of functions from the enumerated countable set $\mathcal{E}$ of bounded computable Lipschitz functions introduced in* (2.1).

The proof is routine.

2.1. **Space of measures.** Let $\mathbf{X} = (X, d, D, \alpha)$ be a computable metric space. In Subsection B.6, the space $\mathcal{M}(\mathbf{X})$ of measures over $\mathbf{X}$ is defined, along with a natural enumeration $\nu = \nu_{\mathcal{M}}$ for a subbase $\sigma = \sigma_{\mathcal{M}}$ of the weak topology. This is a constructive topological space $\mathbf{M}$ which can be metrized by introducing, as in B.6.2, the *Prokhorov distance* $p(\mu, \nu)$: the infimum of all those ε for which, for all Borel sets A we have $\mu(A) \leqslant \nu(A^{\varepsilon}) + \varepsilon$, where $A^{\varepsilon} = \{\, x : \exists y \in A \ d(x, y) < \varepsilon \,\}$. Let $D_{\mathbf{M}}$ be the set of those probability measures that are concentrated on finitely many points of D and assign rational values to them. Let $\alpha_{\mathbf{M}}$ be a natural enumeration of $D_{\mathbf{M}}$. Then

$$(\mathcal{M}, p, D_{\mathbf{M}}, \alpha_{\mathbf{M}}) \tag{2.2}$$

is a computable metric space whose constructive topology is equivalent to $\mathbf{M}$. Let $U = B(x, r)$ be one of the balls in $\mathbf{X}$, where $x \in D_{\mathbf{X}}$, $r \in \mathbb{Q}$. The function $\mu \mapsto \mu(U)$ is typically not computable, not even continuous. For example, if $\mathbf{X} = \mathbb{R}$ and U is the open interval $(0; 1)$, the sequence of probability measures $\delta_{1/n}$ (concentrated on $1/n$) converges to δ_0, but $\delta_{1/n}(U) = 1$, and $\delta_0(U) = 0$. The following theorem shows that the situation is better with $\mu \mapsto \mu f$ for computable f:

Proposition 2.2. *Let $\mathbf{X} = (X, d, D, \alpha)$ be a computable metric space, and let $\mathbf{M} = (\mathcal{M}(\mathbf{X}), \sigma, \nu)$ be the effective topological space of probability measures over $\mathbf{X}$. If function $f : \mathbf{X} \to \mathbb{R}$ is bounded and computable then $\mu \mapsto \mu f$ is computable.*

Proof sketch. To prove the theorem for bounded Lipschitz functions, we can invoke the Strassen coupling theorem B.16.

The function f can be obtained as a limit of a computable monotone increasing sequence of computable Lipschitz functions $f_n^{>}$, and also as a limit of a computable monotone decreasing sequence of computable Lipschitz functions $f_n^{<}$. In step n of our computation of μf, we can approximate $\mu f_n^{>}$ from above to within $1/n$, and $\mu f_n^{<}$ from below to within $1/n$. Let these bounds be $a_n^{>}$ and $a_n^{<}$. To approximate μf to within ε, find a stage n with $a_n^{>} - a_n^{<} + 2/n < \varepsilon$. $\qquad\square$

2.2. **Computable measures and random transitions.** A measure μ is called *computable* if it is a computable element of the space of measures. Let $\{g_i\}$ be the set of bounded Lipschitz functions over X introduced in (2.1).

Proposition 2.3. *Measure μ is computable if and only if so is the function $i \mapsto \mu g_i$.*

Proof. The "only if" part follows from Proposition 2.2. For the "if" part, note that in order to trap μ within some Prokhorov neighborhood of size ε, it is sufficient to compute μg_i within a small enough δ, for all $i \leqslant n$ for a large enough n. $\qquad\square$

Example 2.4. Let our probability space be the set $\mathbb{R}$ of real numbers with its standard topology. Let $a < b$ be two computable real numbers. Let μ be the probability distribution with density function $f(x) = \frac{1}{b-a} 1_{[a; b]}(x)$ (the uniform distribution over the interval $[a; b]$). Function $f(x)$ is not computable, since it is not even continuous. However, the measure μ is

computable: indeed, $\mu g_i = \frac{1}{b-a} \int_a^b g_i(x)dx$ is a computable sequence, hence Proposition 2.3 implies that μ is computable. $\diamond$

The following theorem compensates somewhat for the fact mentioned earlier, that the function $\mu \mapsto \mu(U)$ is generally not computable.

Proposition 2.5. *Let μ be a finite computable measure. Then there is a computable map h with the property that for every bounded computable function f with $|f| \leqslant 1$ with the property $\mu(f^{-1}(0)) = 0$, if w is the name of f then $h(w)$ is the name of a program computing the value $\mu\{x : f(x) < 0\}$.*

Proof. Straightforward. $\square$

Remark 2.6. Suppose that there is a computable function that for each i computes a Cauchy sequence $j \mapsto m_i(j)$ with the property that for $i < j_1 < j_2$ we have $|m_i(j_1) - m_i(j_2)| < 2^{-j_1}$, and that for all n, there is a measure ν with the property that for all $i \leqslant n$, $\nu g_i = m_i(n)$. Is there a measure μ with the property that for each i we have $\lim_j m_i(j) = \mu g_i$? Not necessarily, if the space is not compact. For example, let $X = \{1, 2, 3, \dots\}$ with the discrete topology. The sequences $m_i(j) = 0$ for $j > i$ satisfy these conditions, but they converge to the measure 0, not to a probability measure. To guarantee that the sequences $m_i(j)$ indeed define a probability measure, progress must be made, for example, in terms of the narrowing of Prokhorov neighborhoods. $\diamond$

Let now $\mathbf{X}, \mathbf{Y}$ be computable metric spaces. They give rise to measurable spaces with σ-algebras $\mathcal{A}, \mathcal{B}$ respectively. Let $\Lambda = \{\lambda_x : x \in X\}$ be a probability kernel from X to Y (as defined in Subsection B.5). Let $\{g_i\}$ be the set of bounded Lipschitz functions over Y introduced in (2.1). To each g_i, the kernel assigns a (bounded) measurable function

$$f_i(x) = (\Lambda g_i)(x) = \lambda_x^y g_i(y).$$

We will call Λ *computable* if so is the assignment $(i, x) \mapsto f_i(x)$. In this case, of course, each function $f_i(x)$ is continuous. The measure $\Lambda^* \mu$ is determined by the values $\Lambda^* g_i = \mu(\Lambda g_i)$, which are computable from (i, μ) and so the function $\mu \mapsto \Lambda^* \mu$ is computable.

Example 2.7. A computable function $h : X \to Y$ defines an operator Λ_h with $\Lambda_h g = g \circ h$ (as in Example B.12). This is a deterministic computable transition, in which $f_i(x) = (\Lambda_h g_i)(x) = g_i(h(x))$ is, of course, computable from (i, x). We define $h^* \mu = \Lambda_h^* \mu$. $\diamond$

2.3. **Cells.** As pointed out earlier, it is not convenient to define a measure μ constructively starting from $\mu(\Gamma)$ for open cells Γ. The reason is that no matter how we fix Γ, the function $\mu \mapsto \mu(\Gamma)$ is typically not computable. It is better to work with bounded computable functions, since for such a function f, the correspondence $\mu \mapsto \mu f$ is computable.

Under some special conditions, we will still get "sharp" cells. Let f be a bounded computable function over $\mathbf{X}$, let $\alpha_1 < \cdots < \alpha_k$ be rational numbers, and let μ be a computable measure with the property that $\mu f^{-1}(\alpha_j) = 0$ for all j. In this case, we will say that α_j are *regular points* of f with respect to μ. Let $\alpha_0 = -\infty$, $\alpha_{k+1} = \infty$, and for $j = 0, \dots, k$, let Let $U_j = f^{-1}((j, j+1))$. The sequence of disjoint r.e. open sets $(U_0, \dots, U_k)$ will be called the *partition generated by* $f, \alpha_1, \dots, \alpha_k$. (Note that this sequence is not a partition in the sense of $\bigcup_j U_j = \mathbf{X}$, since the boundaries of the sets are left out.) If we have several partitions $(U_{i0}, \dots, U_{i,k})$, generated by different functions f_i ($i = 1, \dots, m$) and different regular cutoff sequences $(\alpha_{ij} : j = 1, \dots, k_i)$, then we can form a new partition generated by all possible intersections

$$V_{j_1,\dots,j_n} = U_{1,j_1} \cap \cdots \cap U_{m,j_m}.$$

A partition of this kind will be called a *regular partition*. The sets $V_{j_1,\ldots,j_n}$ will be called the *cells* of this partition.

Proposition 2.8. *In a regular partition as given above, the values $\mu V_{j_1,\ldots,j_n}$ are computable from the names of the functions f_i and the cutoff points α_{ij}.*

Proof. Straightforward. $\square$

Assume that a computable sequence of functions $b_1(x), b_2(x), \ldots$ over X is given, with the property that for every pair $x_1, x_2 \in X$ with $x_1 \neq x_2$, there is a j with $b_j(x_1) \cdot b_j(x_2) < 0$. Such a sequence will be called a *separating sequence*. Let us give the correspondence between the set $\mathbb{B}^\omega$ of infinite binary sequences and elements of the set

$$X^0 = \{\, x \in X : b_j(x) \neq 0, \ j = 1, 2, \ldots \,\}.$$

For a binary string $s_1 \cdots s_n = s \in \mathbb{B}^*$, let

$$\Gamma_s$$

be the set of elements of X with the property that for $j = 1, \ldots, n$, if $s_j = 0$ then $b_j(\omega) < 0$, otherwise $b_j(\omega) > 0$. This correspondence has the following properties.

(a) $\Gamma_\Lambda = X$.
(b) For each $s \in \mathbb{B}$, the sets Γ_{s0} and Γ_{s1} are disjoint and their union is contained in Γ_s.
(c) For $x \in X^0$, we have $\{x\} = \bigcap_{x \in \Gamma_s} \Gamma_s$.

If s has length n then Γ_s will be called a *canonical n-cell*, or simply canonical cell, or n-cell. From now on, whenever Γ denotes a subset of X, it means a canonical cell. We will also use the notation

$$l(\Gamma_s) = l(s).$$

The three properties above say that if we restrict ourselves to the set X^0 then the canonical cells behave somewhat like binary subintervals: they divide X^0 in half, then each half again in half, etc. Moreover, around each point, these canonical cells become "arbitrarily small", in some sense (though, they may not be a basis of neighborhoods). It is easy to see that if $\Gamma_{s_1}, \Gamma_{s_2}$ are two canonical cells then they either are disjoint or one of them contains the other. If $\Gamma_{s_1} \subset \Gamma_{s_2}$ then s_2 is a prefix of s_1. If, for a moment, we write $\Gamma_s^0 = \Gamma_s \cap X^0$ then we have the disjoint union $\Gamma_s^0 = \Gamma_{s0}^0 \cup \Gamma_{s1}^0$. For an n-element binary string s, for $x \in \Gamma_s$, we will write

$$\mu(s) = \mu(\Gamma_s).$$

Thus, for elements of X^0, we can talk about the n-th bit x_n of the description of x: it is uniquely determined. The 2^n cells (some of them possibly empty) of the form Γ_s for $l(s) = n$ form a partition

$$\mathcal{P}_n$$

of X^0.

Examples 2.9.

1. If **X** is the set of infinite binary sequences with its usual topology, the functions $b_n(x) = x_n - 1/2$ generate the usual cells, and $\mathbf{X}^0 = \mathbf{X}$.
2. If **X** is the interval $[0; 1]$, let $b_n(x) = -\sin(2^n \pi x)$. Then cells are open intervals of the form $(k \cdot 2^{-n}; (k+1) \cdot 2^n)$, the correspondence between infinite binary strings and elements of X^0 is just the usual representation of x as the binary decimal string $0.x_1 x_2 \ldots$.

$\diamondsuit$

When we fix canonical cells, we will generally assume that the partition chosen is also "natural". The bits $x_1, x_2, \dots$ could contain information about the point x in decreasing order of importance from a macroscopic point of view. For example, for a container of gas, the first few bits may describe, to a reasonable degree of precision, the amount of gas in the left half of the container, the next few bits may describe the amounts in each quarter, the next few bits may describe the temperature in each half, the next few bits may describe again the amount of gas in each half, but now to more precision, etc. From now on, whenever Γ denotes a subset of X, it means a canonical cell. From now on, for elements of X^0, we can talk about the n-th bit x_n of the description of x: it is uniquely determined.

The following observation will prove useful.

Proposition 2.10. *Suppose that the space $\mathbf{X}$ is compact and we have a separating sequence $b_i(x)$ as given above. Then the cells Γ_s form a basis of the space $\mathbf{X}$.*

Proof. We need to prove that for every ball $B(x, r)$, there is a cell $x \in \Gamma_s \subset B(x, r)$. Let C be the complement of $B(x, r)$. For each point y of C, there is an i such that $b_i(x) \cdot b_i(y) < 0$. In this case, let $J^0 = \{\, z : b_i(z) < 0 \,\}$, $J^1 = \{\, z : b_i(z) > 0 \,\}$. Let $J(y) = J^p$ such that $y \in J^p$. Then $C \subset \bigcup_y J(y)$, and compactness implies that there is a finite sequence $y_1, \dots, y_k$ with $C \subset \bigcup_{j=1}^k J(y_j)$. Clearly, there is a cell $x \in \Gamma_s \subset B(x, r) \smallsetminus \bigcup_{j=1}^k J(y_j)$. $\qquad\square$

3. Uniform tests

3.1. **Universal uniform test.** Let $\mathbf{X} = (X, d, D, \alpha)$ be a computable metric space, and let $\mathbf{M} = (\mathcal{M}(\mathbf{X}), \sigma, \nu)$ be the constructive topological space of probability measures over $\mathbf{X}$. A *randomness test* is a function $f : \mathbf{M} \times \mathbf{X} \to \overline{\mathbb{R}}$ with the following two properties.

Condition 3.1.

1. The function $(\mu, x) \mapsto f_\mu(x)$ is lower semicomputable. (Then for each μ, the integral $\mu f_\mu = \mu^x f_\mu(x)$ exists.)
2. $\mu f_\mu \leqslant 1$.

$\diamondsuit$

The value $f_\mu(x)$ is intended to quantify the nonrandomness of the outcome x with respect to the probability measure μ. The larger the values the less random is x. Condition 3.1.2 guarantees that the probability of those outcomes whose randomness is $\geqslant m$ is at most $1/m$. The definition of tests is in the spirit of Martin-Löf's tests. The important difference is in the semicomputability condition: instead of restricting the measure μ to be computable, we require the test to be lower semicomputable also in its argument μ.

Just as with Martin-Löf's tests, we want to find a universal test; however, we seem to need a condition on the space $\mathbf{X}$. Let us say that a sequence $i \mapsto U_i$ of sets has *recognizable Boolean inclusions* if the set

$$\{\, (S, T) : S, T \text{ are finite}, \bigcap_{i \in S} U_i \subset \bigcup_{j \in T} U_j \,\}$$

is recursively enumerable. We will say that a computable metric space has recognizable Boolean inclusions if this is true of the enumerated basis consisting of balls of the form $B(x, r)$ where $x \in D$ and $r > 0$ is a rational number.

It is our conviction that the important metric spaces studied in probability theory have recognizable Boolean inclusions, and that proving this in each individual case should not be too difficult. For example, it does not seem difficult to prove this for the space $C[0; 1]$ of

Example C.6, with the set of rational piecewise-linear functions chosen as D. But, we have not carried out any of this work!

Theorem 1. *Suppose that the metric space* $\mathbf{X}$ *has recognizable Boolean inclusions. Then there is a universal test, that is a test* $\mathbf{t}_\mu(x)$ *with the property that for every other test* $f_\mu(x)$ *there is a constant* $c_f > 0$ *with* $c_f f_\mu(x) \leqslant \mathbf{t}_\mu(x)$.

Proof.

1. We will show that there is a mapping that to each name u of a lower semicomputable function $(\mu, x) \mapsto g(\mu, x)$ assigns the name of a lower semicomputable function $g'(\mu, x)$ such that $\mu^x g'(\mu, x) \leqslant 1$, and if g is a test then $g' = g$.

   To prove the statement, let us represent the space $\mathbf{M}$ rather as

$$\mathbf{M} = (\mathcal{M}(\mathbf{X}), p, D, \alpha_{\mathbf{M}}), \tag{3.1}$$

as in (2.2). Since $g(\mu, x)$ is lower semicomputable, there is a computable sequence of basis elements $U_i \subset \mathbf{M}$ and $V_i \subset \mathbf{X}$ and rational bounds r_i such that

$$g(\mu, x) = \sup_i \ r_i 1_{U_i}(\mu) 1_{V_i}(x).$$

Let $h_n(\mu, x) = \max_{i \leqslant n} r_i 1_{U_i}(\mu) 1_{V_i}(x)$. Let us also set $h_0(\mu, x) = 0$. Our goal is to show that the condition $\forall \mu \ \mu^x h_n(\mu, x) \leqslant 1$ is decidable. If this is the case then we will be done. Indeed, we can define $h'_n(\mu, x)$ recursively as follows. Let $h'_0(\mu, x) = 0$. Assume that $h'_n(\mu, x)$ has been defined already. If $\forall \mu \ \mu^x h_{n+1}(\mu, x) \leqslant 1$ then $h'_{n+1}(\mu, x) = h_{n+1}(\mu, x)$; otherwise, it is $h'_n(\mu, x)$. The function $g'(\mu, x) = \sup_n h'_n(\mu, x)$ clearly satisfies our requirements.

   We proceed to prove the decidability of the condition

$$\forall \mu \ \mu^x h_n(\mu, x) \leqslant 1. \tag{3.2}$$

The basis elements V_i can be taken as balls $B(q_i, \delta_i)$ for a computable sequence $q_i \in D$ and computable sequence of rational numbers $\delta_i > 0$. Similarly, the basis element U_i is the set of measures that is a ball $B(\sigma_i, \varepsilon_i)$, in the metric space (3.1). Here, using notation (B.7), σ_i is a measure concentrated on a finite set S_i. According to Proposition B.17, the ball U_i is the set of measures μ satisfying the inequalities

$$\mu(A^{\varepsilon_i}) > \sigma_i(A) - \varepsilon_i$$

for all $A \subset S_i$. For each n, consider the finite set of balls

$$\mathcal{B}_n = \{ B(q_i, \delta_i) : i \leqslant n \} \cup \{ B(s, \varepsilon_i) : i \leqslant n, \ s \in S_i \}.$$

Consider all sets of the form

$$U_{A,B} = \bigcap_{U \in A} U \smallsetminus \bigcup_{U \in B} U$$

for all pairs of sets $A, B \subset \mathcal{B}_n$. These sets are all finite intersections of balls or complements of balls from the finite set $\mathcal{B}_n$ of balls. The space $\mathbf{X}$ has recognizable Boolean inclusions, so it is decidable which of these sets $U_{A,B}$ are nonempty. The condition (3.2) can be formulated as a Boolean formula involving linear inequalities with rational coefficients, for the variables $\mu_{A,B} = \mu(U_{A,B})$, for those A, B with $U_{A,B} \neq \emptyset$. The solvability of such a Boolean condition can always be decided.

2. Let us enumerate all lower semicomputable functions $g_u(\mu, x)$ for all the names u. Without loss of generality, assume these names to be natural numbers, and form the functions $g'_u(\mu, x)$ according to the assertion 1 above. The function $t = \sum_u 2^{-u-1} g'_u$ will be the desired universal test.

$\square$

From now on, when referring to randomness tests, we will always assume that our space $\mathbf{X}$ has recognizable Boolean inclusions and hence has a universal test. We fix a universal test $\mathbf{t}_\mu(x)$, and call the function

$$\mathbf{d}_\mu(x) = \log \mathbf{t}_\mu(x).$$

the *deficiency of randomness* of x with respect to μ. We call an element $x \in X$ *random* with respect to μ if $\mathbf{d}_\mu(x) < \infty$.

Remark 3.2. Tests can be generalized to include an arbitrary parameter y: we can talk about the universal test

$$\mathbf{t}_\mu(x \mid y),$$

where y comes from some constructive topological space $\mathbf{Y}$. This is a maximal (within a multiplicative constant) lower semicomputable function $(x, y, \mu) \mapsto f(x, y, \mu)$ with the property $\mu^x f(x, y, \mu) \leqslant 1$. $\diamondsuit$

3.2. **Conservation of randomness.** For $i = 1, 0$, let $\mathbf{X}_i = (X_i, d_i, D_i, \alpha_i)$ be computable metric spaces, and let $\mathbf{M}_i = (\mathcal{M}(\mathbf{X}_i), \sigma_i, \nu_i)$ be the effective topological space of probability measures over $\mathbf{X}_i$. Let Λ be a computable probability kernel from $\mathbf{X}_1$ to $\mathbf{X}_0$ as defined in Subsection 2.2. In the following theorem, the same notation $\mathbf{d}_\mu(x)$ will refer to the deficiency of randomness with respect to two different spaces, $\mathbf{X}_1$ and $\mathbf{X}_0$, but this should not cause confusion. Let us first spell out the conservation theorem before interpreting it.

Theorem 2. *For a computable probability kernel Λ from $\mathbf{X}_1$ to $\mathbf{X}_0$, we have*

$$\lambda_x^y \mathbf{t}_{\Lambda^*\mu}(y) \overset{*}{<} \mathbf{t}_\mu(x). \tag{3.3}$$

Proof. Let $\mathbf{t}_\nu(x)$ be the universal test over $\mathbf{X}_0$. The left-hand side of (3.3) can be written as

$$u_\mu = \Lambda \mathbf{t}_{\Lambda^*\mu}.$$

According to (B.4), we have $\mu u_\mu = (\Lambda^*\mu)\mathbf{t}_{\Lambda^*\mu}$ which is $\leqslant 1$ since $\mathbf{t}$ is a test. If we show that $(\mu, x) \mapsto u_\mu(x)$ is lower semicomputable then the universality of $\mathbf{t}_\mu$ will imply $u_\mu \overset{*}{<} \mathbf{t}_\mu$.

According to Proposition C.7, as a lower semicomputable function, $\mathbf{t}_\nu(y)$ can be written as $\sup_n g_n(\nu, y)$, where $(g_n(\nu, y))$ is a computable sequence of computable functions. We pointed out in Subsection 2.2 that the function $\mu \mapsto \Lambda^*\mu$ is computable. Therefore the function $(n, \mu, x) \mapsto g_n(\Lambda^*\mu, f(x))$ is also a computable. So, $u_\mu(x)$ is the supremum of a computable sequence of computable functions and as such, lower semicomputable. $\square$

It is easier to interpret the theorem first in the special case when $\Lambda = \Lambda_h$ for a computable function $h : X_1 \to X_0$, as in Example 2.7. Then the theorem simplifies to the following.

Corollary 3.3. *For a computable function $h : X_1 \to X_0$, we have $\mathbf{d}_{h^*\mu}(h(x)) \overset{+}{<} \mathbf{d}_\mu(x)$.*

Informally, this says that if x is random with respect to μ in $\mathbf{X}_1$ then $h(x)$ is essentially at least as random with respect to the output distribution $h^*\mu$ in $\mathbf{X}_0$. Decrease in randomness can only be caused by complexity in the definition of the function h. It is even easier to interpret the theorem when μ is defined over a product space $\mathbf{X}_1 \times \mathbf{X}_2$, and $h(x_1, x_2) = x_1$ is the projection. The theorem then says, informally, that if the pair (x_1, x_2) is random with respect to μ then x_1 is random with respect to the marginal $\mu_1 = h^*\mu$ of μ. This is a very natural requirement: why would the throwing-away of the information about x_2 affect the plausibility of the hypothesis that the outcome x_1 arose from the distribution μ_1?

In the general case of the theorem, concerning random transitions, we cannot bound the randomness of each outcome uniformly. The theorem asserts that the average nonrandomness, as measured by the universal test with respect to the output distribution, does not increase. In logarithmic notation: $\lambda_x^y 2^{\mathbf{d}_{\Lambda^*\mu}(y)} \overset{+}{<} \mathbf{d}_\mu(x)$, or equivalently, $\int 2^{\mathbf{d}_{\Lambda^*\mu}(y)} \lambda_x(dy) \overset{+}{<} \mathbf{d}_\mu(x)$.

Corollary 3.4. *Let Λ be a computable probability kernel from $\mathbf{X}_1$ to $\mathbf{X}_0$. There is a constant c such that for every $x \in \mathbf{X}^1$, and integer $m > 0$ we have*

$$\lambda_x\{\, y : \mathbf{d}_{\Lambda^*\mu}(y) > \mathbf{d}_\mu(x) + m + c \,\} \leqslant 2^{-m}.$$

Thus, in a computable random transition, the probability of an increase of randomness deficiency by m units (plus a constant c) is less than 2^{-m}. The constant c comes from the description complexity of the transition Λ.

A randomness conservation result related to Corollary 3.3 was proved in [10]. There, the measure over the space $\mathbf{X}_0$ is not the output measure of the transformation, but is assumed to obey certain inequalities related to the transformation.

4. TESTS AND COMPLEXITY

4.1. **Description complexity.**

4.1.1. *Complexity, semimeasures, algorithmic entropy.* Let $X = \Sigma^*$. For $x \in \Sigma^*$ for some finite alphabet Σ, let $H(x)$ denote the prefix-free description complexity of the finite sequence x as defined, for example, in [14] (where it is denoted by $K(x)$). For completeness, we give its definition here. Let $A : \{0,1\}^* \times \Sigma^* \to \Sigma^*$ be a computable (possibly partial) function with the property that if $A(p_1, y)$ and $A(p_2, y)$ are defined for two different strings p_1, p_2, then p_1 is not the prefix of p_2. Such a function is called a (prefix-free) *interpreter*. We denote

$$H^A(x \mid y) = \min_{A(p,y)=x} |p|.$$

One of the most important theorems of description complexity is the following:

Proposition 4.1 (Invariance Theorem, see for example [14])**.** *There is an optimal interpreter T with the above property: with it, for every interpreter A there is a constant c_A with*

$$H^T(x \mid y) \leqslant H^A(x \mid y) + c_A.$$

We fix an optimal interpreter T and write $H(x \mid y) = H^T(x \mid y)$, calling it the conditional complexity of a string x with respect to string y. We denote $H(x) = H(x \mid \Lambda)$. Let

$$\mathbf{m}(x) = 2^{-H(x)}.$$

The function $\mathbf{m}(x)$ is lower semicomputable with $\sum_x \mathbf{m}(x) \leqslant 1$. Let us call any real function $f(x) \geqslant 0$ over Σ^* with $\sum_x f(x) \leqslant 1$ a *semimeasure*. The following theorem, known as the Coding Theorem, is an important tool.

Proposition 4.2 (Coding Theorem)**.** *For every lower semicomputable semimeasure f there is a constant $c > 0$ with $\mathbf{m}(x) \geqslant c \cdot f(x)$.*

Because of this theorem, we will say that $\mathbf{m}(x)$ is a *universal lower semicomputable semimeasure*. It is possible to turn $\mathbf{m}(x)$ into a measure, by compactifying the discrete space Σ^* into

$$\overline{\Sigma^*} = \Sigma^* \cup \{\infty\}$$

(as in part 1 of Example A.3; this process makes sense also for a constructive discrete space), and setting $\mathbf{m}(\infty) = 1 - \sum_{x \in \Sigma^*} \mathbf{m}(x)$. The extended measure $\mathbf{m}$ is not quite lower semicomputable since the number $\mu(\overline{\Sigma^*} \smallsetminus \{0\})$ is not necessarily lower semicomputable.

Remark 4.3. A measure μ is computable over $\overline{\Sigma^*}$ if and only if the function $x \mapsto \mu(x)$ is computable for $x \in \Sigma^*$. This property does not imply that the number

$$1 - \mu(\infty) = \mu(\Sigma^*) = \sum_{x \in \Sigma^*} \mu(x)$$

is computable. $\diamond$

Let us allow, for a moment, measures μ that are not probability measures: they may not even be finite. Metric and computability can be extended to this case (see [22]), the universal test $\mathbf{t}_\mu(x)$ can also be generalized. The Coding Theorem and other considerations suggest the introduction of the following notation, for an arbitrary measure μ:

$$H_\mu(x) = -\mathbf{d}_\mu(x) = -\log \mathbf{t}_\mu(x). \tag{4.1}$$

Then, with $\#$ defined as the counting measure over the discrete set Σ^* (that is, $\#(S) = |S|$), we have

$$H(x) \stackrel{\pm}{=} H_\#(x).$$

This allows viewing $H_\mu(x)$ as a generalization of description complexity: we will call this quantity the *algorithmic entropy* of x relative to the measure μ. Generalization to conditional complexity is done using Remark 3.2. A reformulation of the definition of tests says that $H_\mu(x)$ is minimal (within an additive constant) among the upper semicomputable functions $(\mu, x) \mapsto f_\mu(x)$ with $\mu^x 2^{-f_\mu(x)} \leqslant 1$. The following identity is immediate from the definitions:

$$H_\mu(x) = H_\mu(x \mid \mu). \tag{4.2}$$

4.1.2. *Computable measures and complexity.* It is known that for computable μ, the test $\mathbf{d}_\mu(x)$ can be expressed in terms of the description complexity of x (we will prove these expressions below). Assume that $\mathbf{X}$ is the (discrete) space of all binary strings. Then we have

$$\mathbf{d}_\mu(x) = -\log \mu(x) - H(x) + O(H(\mu)). \tag{4.3}$$

The meaning of this equation is the following. Due to maximality property of the semimeasure $\mathbf{m}$ following from the Coding Theorem 4.2 above, the expression $-\log \mu(x)$ is an upper bound (within $O(H(\mu))$) of the complexity $H(x)$, and nonrandomness of x is measured by the difference between the complexity and this upper bound. See [26] for a first formulation of this general upper bound relation. As a simple example, consider the uniform distribution μ over the set of binary sequences of length n. Conditioning everything on n, we obtain

$$\mathbf{d}_\mu(x \mid n) \stackrel{\pm}{=} n - H(x \mid n),$$

that is the more the description complexity $H(x \mid n)$ of a binary sequence of length n differs from its upper bound n the less random is x.

Assume that $\mathbf{X}$ is the space of infinite binary sequences. Then equation (4.3) must be replaced with

$$\mathbf{d}_\mu(x) = \sup_n \left(-\log \mu(x^{\leqslant n}) - H(x^{\leqslant n}) \right) + O(H(\mu)). \tag{4.4}$$

For the coin-tossing distribution μ, this characterization has first been first proved by Schnorr, and published in [5].

Remark 4.4. It is possible to obtain similar natural characterizations of randomness, using some other natural definitions of description complexity. A universal semicomputable semimeasure $\mathbf{m}_\Omega$ over the set Ω of infinite sequences was introduced, and a complexity $\mathrm{KM}(x) = -\log \mathbf{m}_\Omega(x)$ defined in [26]. A so-called "monotonic complexity", $\mathrm{Km}(x)$ was introduced, using Turing machines with one-way input and output, in [11], and a closely related quantity called "process complexity" was introduced in [17]. These quantities can also be used in a characterization of randomness similar to (4.3). The nontrivial fact that the complexities KM and Km differ by an unbounded amount was shown in [8]. $\diamond$

For noncomputable measures, we cannot replace $O(H(\mu))$ in these relations with anything finite, as shown in the following example. Therefore however attractive and simple, $\exp(-\log\mu(x) - H(x))$ is not a universal uniform test of randomness.

Proposition 4.5. *There is a measure μ over the discrete space $\mathbf{X}$ of binary strings such that for each n, there is an x with $\mathbf{d}_\mu(x) = n - H(n)$ and $-\log\mu(x) - H(x) \overset{+}{<} 0$.*

Proof. Let us treat the domain of our measure μ as a set of pairs (x, y). Let $x_n = 0^n$, for $n = 1, 2, \dots$. For each n, let y_n be some binary string of length n with the property $H(x_n, y_n) > n$. Let $\mu(x_n, y_n) = 2^{-n}$. Then $-\log\mu(x_n, y_n) - H(x_n, y_n) \leqslant n - n = 0$. On the other hand, let $t_\mu(x, y)$ be the test nonzero only on strings x of the form x_n:

$$t_\mu(x_n, y) = \frac{\mathbf{m}(n)}{\sum_{z \in \mathcal{B}^n} \mu(x_n, z)}.$$

The form of the definition ensures semicomputability and we also have

$$\sum_{x, y} \mu(x, y) t_\mu(x, y) \leqslant \sum_n \mathbf{m}(n) < 1,$$

therefore t_μ is indeed a test. Hence $\mathbf{t}_\mu(x, y) \overset{*}{>} t_\mu(x, y)$. Taking logarithms, $\mathbf{d}_\mu(x_n, y_n) \overset{+}{>} n - H(n)$. $\square$

The same example implies that it is also not an option, even over discrete sets, to replace the definition of uniform tests with the *ad hoc* formula $\exp(-\log\mu(x) - H(x))$:

Proposition 4.6. *The test defined as $f_\mu(x) = \exp(-\log\mu(x) - H(x))$ over discrete spaces $\mathbf{X}$ does not obey the conservation of randomness.*

Proof. Let us use the example of Proposition 4.5. Consider the function $\pi : (x, y) \mapsto x$. The image of the measure μ under the projection is $(\pi\mu)(x) = \sum_y \mu(x, y)$. Thus, $(\pi\mu)(x_n) = \mu(x_n, y_n) = 2^{-n}$. We have seen $\log f_\mu(x_n, y_n) \leqslant 0$. On the other hand,

$$\log f_{\pi\mu}(\pi(x_n, y_n)) = -\log(\pi\mu)(x_n) - H(x_n) \overset{+}{=} n - H(n).$$

Thus, the projection π takes a random pair (x_n, y_n) into an object x_n that is very nonrandom (when randomness is measured using the tests f_μ). $\square$

In the example, we have the abnormal situation that a pair is random but one of its elements is nonrandom. Therefore even if we would not insist on universality, the test $\exp(-\log\mu(x) - H(x))$ is unsatisfactory.

Looking into the reasons of the nonconservation in the example, we will notice that it could only have happened because the test f_μ is too special. The fact that $-\log(\pi\mu)(x_n) - H(x_n)$ is large should show that the pair (x_n, y_n) can be enclosed into the "simple" set $\{x_n\} \times \mathbf{Y}$ of small probability; unfortunately, this observation does not reflect on $-\log\mu(x, y) - H(x, y)$ when the measure μ is non-computable (it does for computable μ).

4.1.3. *Expressing the uniform test in terms of complexity.* It is a natural idea to modify equation (4.3) in such a way that the complexity $H(x)$ is replaced with $H(x \mid \mu)$. However, this expression must be understood properly. The measure μ (especially, when it is not computable) cannot be described by a finite string; on the other hand, it can be described by infinite strings in many different ways. Clearly, irrelevant information in these infinite strings should be ignored. The notion of representation in computable analysis (see Subsection C.1) will solve the problem. An interpreter function should have the property that its output depends only on μ and not on the sequence representing it. Recall the topological space $\mathbf{M}$ of computable measures over our space $\mathbf{X}$. An interpreter $A : \{0,1\}^* \times \mathbf{M} \to \Sigma^*$ is a computable function that is prefix-free in its first argument. The complexity

$$H(x \mid \mu)$$

can now be defined in terms of such interpreters, noting that the Invariance Theorem holds as before. To define this complexity in terms of representations, let $\gamma_{\mathbf{M}}$ be our chosen representation for the space $\mathbf{M}$ (thus, each measure μ is represented via all of its Cauchy sequences in the Prokhorov distance). Then we can say that A is an interpreter if it is $(\mathrm{id}, \gamma_{\mathbf{M}}, \mathrm{id})$-computable, that is a certain computable function $B : \{0,1\}^* \times \Sigma^\omega \to \Sigma^*$ realizes A for every $p \in \{0,1\}^*$, and for every sequence z that is a $\gamma_{\mathbf{M}}$-name of a measure μ, we have $B(p, z) = A(p, \mu)$.

Remark 4.7. The notion of oracle computation and reducibility in the new sense (where the result is required to be independent of which representation of an object is used) may be worth investigating in other settings as well.   ◇

Let us mention the following easy fact:

Proposition 4.8. *If μ is a computable measure then $H(x \mid \mu) \stackrel{+}{=} H(x)$. The constant in $\stackrel{+}{=}$ depends on the description complexity of μ.*

Theorem 3. *If $\mathbf{X}$ is the discrete space Σ^* then we have*

$$\mathbf{d}_\mu(x) \stackrel{+}{=} -\log \mu(x) - H(x \mid \mu). \tag{4.5}$$

Note that in terms of the algorithmic entropy notation introduced in (4.1), this theorem can be expressed as

$$H_\mu(x) \stackrel{+}{=} H(x \mid \mu) + \log \mu(x). \tag{4.6}$$

Proof. In exponential notation, equation (4.5) can be written as $\mathbf{t}_\mu(x) \stackrel{*}{=} \mathbf{m}(x \mid \mu)/\mu(x)$. Let us prove $\stackrel{*}{>}$ first. We will show that the right-hand side of this inequality is a test, and hence $\stackrel{*}{<} \mathbf{t}_\mu(x)$. However, the right-hand side is clearly lower semicomputable in (x, μ) and when we "integrate" it (multiply it by $\mu(x)$ and sum it), its sum is $\leqslant 1$; thus, it is a test.

Let us prove $\stackrel{*}{<}$ now. The expression $\mathbf{t}_\mu(x)\mu(x)$ is clearly lower semicomputable in (x, μ), and its sum is $\leqslant 1$. Hence, it is $\stackrel{+}{<} \mathbf{m}(x \mid \mu)$.   □

Remark 4.9. As mentioned earlier, our theory generalizes to measures that are not probability measures. In this case, equation (4.6) has interesting relations to the quantity called "physical entropy" by Zurek in [25]; it justifies calling $H_\mu(x)$ "fine-grained algorithmic Boltzmann entropy" by this author in [9].   ◇

For non-discrete spaces, unfortunately, we can only provide less intuitive expressions.

Proposition 4.10. *let* $\mathbf{X} = (X, d, D, \alpha)$ *be a complete computable metric space, and let* $\mathcal{E}$ *be the enumerated set of bounded Lipschitz functions introduced in* (2.1)*, but for the space* $\mathbf{M}(\mathbf{X}) \times \mathbf{X}$*. The uniform test of randomness* $\mathbf{t}_\mu(x)$ *can be expressed as*

$$\mathbf{t}_\mu(x) \overset{*}{=} \sum_{f \in \mathcal{E}} f(\mu, x) \frac{\mathbf{m}(f \mid \mu)}{\mu^y f(\mu, y)}. \tag{4.7}$$

Proof. For $\overset{*}{>}$, we will show that the right-hand side of the inequality is a test, and hence $\overset{*}{<} \mathbf{t}_\mu(x)$. For simplicity, we skip the notation for the enumeration of $\mathcal{E}$ and treat each element f as its own name. Each term of the sum is clearly lower semicomputable in (f, x, μ), hence the sum is lower semicomputable in (x, μ). It remains to show that the μ-integral of the sum is $\leqslant 1$. But, the μ-integral of the generic term is $\leqslant \mathbf{m}(f \mid \mu)$, and the sum of these terms is $\leqslant 1$ by the definition of the function $\mathbf{m}(\cdot \mid \cdot)$. Thus, the sum is a test.

For $\overset{*}{<}$, note that $(\mu, x) \mapsto \mathbf{t}_\mu(x)$, as a lower semicomputable function, is the supremum of functions in $\mathcal{E}$. Denoting their differences by $f_i(\mu, x)$, we have $\mathbf{t}_\mu(x) = \sum_i f_i(\mu, x)$. The test property implies $\sum_i \mu^x f_i(\mu, x) \leqslant 1$. Since the function $(\mu, i) \mapsto \mu^x f_i(\mu, x)$ is lower semicomputable, this implies $\mu^x f_i(\mu, x) \overset{*}{<} \mathbf{m}(i \mid \mu)$, and hence

$$f_i(\mu, x) \overset{*}{<} f_i(\mu, x) \frac{\mathbf{m}(i \mid \mu)}{\mu^x f_i(\mu, x)}.$$

It is easy to see that for each $f \in \mathcal{E}$ we have

$$\sum_{i: f_i = f} \mathbf{m}(i \mid \mu) \leqslant \mu(f \mid \mu),$$

which leads to (4.7). $\qquad \square$

Remark 4.11. If we only want the $\overset{*}{>}$ part of the result, then $\mathcal{E}$ can be replaced with any enumerated computable sequence of bounded computable functions. $\qquad \diamondsuit$

4.2. **Infinite sequences.** In this section, we get a nicer characterization of randomness tests in terms of complexity, in special cases. Let $\mathcal{M}_R(X)$ be the set of measures μ with $\mu(X) = R$.

Theorem 4. *Let* $\mathbf{X} = \mathbb{N}^\omega$ *be the set of infinite sequences of natural numbers, with the product topology. For all computable measures* $\mu \in \mathcal{M}_R(X)$*, for the deficiency of randomness* $\mathbf{d}_\mu(x)$*, we have*

$$\mathbf{d}_\mu(x) \overset{+}{=} \sup_n \left(-\log \mu(x^{\leqslant n}) - H(x^{\leqslant n}) \right). \tag{4.8}$$

Here, the constant in $\overset{+}{=}$ *depends on the computable measure* μ*.*

We will be able to prove the $\overset{+}{>}$ part of the statement in a more general space, and without assuming computability. Assume that a separating sequence $b_1, b_2, \ldots$ is given as defined in Subsection 2.3, along with the set X^0. For each $x \in X^0$, the binary sequence $x_1, x_2, \ldots$ has been defined. Let

$$\overline{\mu}(\Gamma_s) = R - \sum \{ \mu(\Gamma_{s'}) : l(s) = l(s'), \ s' \neq s \}.$$

Then $(s, \mu) \mapsto \mu(\Gamma_s)$ is lower semicomputable, and $(s, \mu) \mapsto \overline{\mu}(\Gamma_s)$ is upper semicomputable. And, every time that the functions $b_i(x)$ form a regular partition for μ, we have $\overline{\mu}(\Gamma_s) = \mu(\Gamma_s)$ for all s. Let $\mathcal{M}_R^0(X)$ be the set of those measures μ in $\mathcal{M}_R(X)$ for which $\mu(X \smallsetminus X^0) = 0$.

Theorem 5. *Suppose that the space* $\mathbf{X}$ *is compact. Then for all computable measures* $\mu \in \mathcal{M}_R^0(\mathbf{X})$, *for the deficiency of randomness* $\mathbf{d}_\mu(x)$, *the characterization* (4.8) *holds.*

For arbitrary measures and spaces, we can say a little less:

Proposition 4.12. *For all measures* $\mu \in \mathcal{M}_R(X)$, *for the deficiency of randomness* $\mathbf{d}_\mu(x)$, *we have*

$$\mathbf{d}_\mu(x) \stackrel{+}{>} \sup_n \big(-\log \overline{\mu}(x^{\leqslant n}) - H(x^{\leqslant n} \mid \mu) \big). \tag{4.9}$$

Proof. Consider the function

$$f_\mu(x) = \sum_s 1_{\Gamma_s}(x) \frac{\mathbf{m}(s \mid \mu)}{\overline{\mu}(\Gamma_s)} = \sum_n \frac{\mathbf{m}(x^{\leqslant n} \mid \mu)}{\overline{\mu}(x^{\leqslant n})} \geqslant \sup_n \frac{\mathbf{m}(x^{\leqslant n} \mid \mu)}{\overline{\mu}(x^{\leqslant n})}.$$

The function $(\mu, x) \mapsto f_\mu(x)$ is clearly lower semicomputable and satisfies $\mu^x f_\mu(x) \leqslant 1$, and hence

$$\mathbf{d}_\mu(x) \stackrel{+}{>} \log f(x) \stackrel{+}{>} \sup_n \big(-\log \overline{\mu}(x^{\leqslant n}) - H(x^{\leqslant n} \mid \mu) \big).$$

$\square$

Proof of Theorem 4. For binary sequences instead of sequences of natural numbers, the part $\stackrel{+}{>}$ of the inequality follows directly from Proposition 4.12: indeed, look at Examples 2.9. For sequences of natural numbers, the proof is completely analogous.

The proof of $\stackrel{+}{<}$ reproduces the proof of Theorem 5.2 of [7]. The computability of μ implies that $t(x) = \mathbf{t}_\mu(x)$ is lower semicomputable. Let us first replace $t(x)$ with a rougher version:

$$t'(x) = \max\{\, 2^n : 2^n < \mathbf{t}_\mu(x) \,\}.$$

Then $t'(x) \stackrel{*}{=} t(x)$, and it takes only values of the form 2^n. It is also lower semicomputable. Let us abbreviate:

$$1_y(x) = 1_{x\mathbb{N}^\omega}(x), \quad \mu(y) = \mu(y\mathbb{N}^\omega).$$

For every lower semicomputable function f over $\mathbb{N}^\omega$, there are computable sequences $y_i \in \mathbb{N}^*$ and $r_i \in \mathbb{Q}$ with $f(x) = \sup_i r_i 1_{y_i}(x)$, with the additional property that if $i < j$ and $1_{y_i}(x) = 1_{y_j}(x) = 1$ then $r_i < r_j$. Since $t'(x)$ only takes values of the form 2^n, there are computable sequences $y_i \in \mathbb{B}^*$ and $k_i \in \mathbb{N}$ with

$$t'(x) = \sup_i 2^{k_i} 1_{y_i}(x) \stackrel{*}{=} \sum_i 2^{k_i} 1_{y_i}(x),$$

with the property that if $i < j$ and $1_{y_i}(x) = 1_{y_j}(x) = 1$ then $k_i < k_j$. The equality $\stackrel{*}{=}$ follows easily from the fact that for any finite sequence $n_1 < n_2 < \ldots$, $\sum_j 2^{n_j} \leqslant 2\max_j 2^{n_j}$. Since $\mu t' \stackrel{*}{<} 1$, we have $\sum_i 2^{k_i} \mu(y_i) \stackrel{*}{<} 1$. Since the function $i \to 2^{k_i} \mu(y_i)$ is computable, this implies $2^{k_i} \mu(y_i) \stackrel{*}{<} \mathbf{m}(i)$, $2^{k_i} \stackrel{*}{<} \mathbf{m}(i)/\mathbf{m}(y_i)$. Thus,

$$t(x) \stackrel{*}{<} \sup_i 1_{y_i}(x) \frac{\mathbf{m}(i)}{\mu(y_i)}.$$

For $y \in \mathbb{N}^*$ we certainly have $H(y) \stackrel{+}{<} \inf_{y=y_i} H(i)$, which implies $\sup_{y=y_i} \mathbf{m}(i) \leqslant \mathbf{m}(y)$. It follows that

$$t(x) \stackrel{*}{<} \sup_{y \in \mathbb{B}^*} 1_y(x) \frac{\mathbf{m}(y)}{\mu(y)} = \sup_n \frac{\mathbf{m}(x^{\leqslant n})}{\mu(x^{\leqslant n})}.$$

Taking logarithms, we obtain the $\stackrel{+}{<}$ part of the theorem. $\square$

Proof of Theorem 5. The proof of part $\overset{+}{>}$ of the inequality follows directly from Proposition 4.12, just as in the proof of Theorem 4.

The proof of $\overset{+}{<}$ is also similar to the proof of that theorem. The only part that needs to be reproved is the statement that for every lower semicomputable function f over X, there are computable sequences $y_i \in \mathbb{B}^*$ and $q_i \in \mathbb{Q}$ with $f(x) = \sup_i q_i 1_{y_i}(x)$. This follows now, since according to Proposition 2.10, the cells Γ_y form a basis of the space $\mathbf{X}$.   □

5. NEUTRAL MEASURE

Let $\mathbf{t}_\mu(x)$ be our universal uniform randomness test. We call a measure M *neutral* if $\mathbf{t}_M(x) \leqslant 1$ for all x. If M is neutral then no experimental outcome x could refute the theory (hypothesis, model) that M is the underlying measure to our experiments. It can be used as "apriori probability", in a Bayesian approach to statistics. Levin's theorem says the following:

Theorem 6. *If the space $\mathbf{X}$ is compact then there is a neutral measure over $\mathbf{X}$.*

The proof relies on a nontrivial combinatorial fact, Sperner's Lemma, which also underlies the proof of the Brouwer fixpoint theorem. Here is a version of Sperner's Lemma, spelled out in continuous form:

Proposition 5.1 (see for example [20]). *Let $p_1, \dots, p_k$ be points of some finite-dimensional space $\mathbb{R}^n$. Suppose that there are closed sets $F_1, \dots, F_k$ with the property that for every subset $1 \leqslant i_1 < \cdots < i_j \leqslant k$ of the indices, the simplex $S(p_{i_1}, \dots, p_{i_j})$ spanned by $p_{i_1}, \dots, p_{i_j}$ is covered by the union $F_{i_1} \cup \cdots \cup F_{i_j}$. Then the intersection $\bigcap_i F_i$ of all these sets is not empty.*

The following lemma will also be needed.

Lemma 5.2. *For every closed set $A \subset \mathbf{X}$ and measure μ, if $\mu(A) = 1$ then there is a point $x \in A$ with $\mathbf{t}_\mu(x) \leqslant 1$.*

Proof. This follows easily from $\mu t_\mu = \mu^x 1_A(x) t_\mu(x) \leqslant 1$.   □

Proof of Theorem 6. For every point $x \in \mathbf{X}$, let F_x be the set of measures for which $\mathbf{t}_\mu(x) \leqslant 1$. If we show that for every finite set of points $x_1, \dots, x_k$, we have

$$F_{x_1} \cap \cdots \cap F_{x_k} \neq \emptyset, \tag{5.1}$$

then we will be done. Indeed, according to Proposition B.18, the compactness of $\mathbf{X}$ implies the compactness of the space $\mathcal{M}(\mathbf{X})$ of measures. Therefore if every finite subset of the family $\{ F_x : x \in \mathbf{X} \}$ of closed sets has a nonempty intersection, then the whole family has a nonempty intersection: this intersection consists of the neutral measures.

To show (5.1), let $S(x_1, \dots, x_k)$ be the set of probability measures concentrated on $x_1, \dots, x_k$. Lemma 5.2 implies that each such measure belongs to one of the sets F_{x_i}. Hence $S(x_1, \dots, x_k) \subset F_{x_1} \cup \cdots \cup F_{x_k}$, and the same holds for every subset of the indices $\{1, \dots, k\}$. Sperner's Lemma 5.1 implies $F_{x_1} \cap \cdots \cap F_{x_k} \neq \emptyset$.   □

When the space is not compact, there are generally no neutral probability measures, as shown by the following example.

Proposition 5.3. *Over the discrete space $\mathbf{X} = \mathbb{N}$ of natural numbers, there is no neutral measure.*

Proof. It is sufficient to construct a randomness test $t_\mu(x)$ with the property that for every measure μ, we have $\sup_x t_\mu(x) = \infty$. Let

$$t_\mu(x) = \sup\{\, k \in \mathbb{N} : \sum_{y<x} \mu(y) > 1 - 2^{-k} \,\}. \tag{5.2}$$

By its construction, this is a lower semicomputable function with $\sup_x t_\mu(x) = \infty$. It is a test if $\sum_x \mu(x) t_\mu(x) \leqslant 1$. We have

$$\sum_x \mu(x) t_\mu(x) = \sum_{k>0} \sum_{t_\mu(x) \geqslant k} \mu(x) < \sum_{k>0} 2^{-k} \leqslant 1.$$

$\square$

Using a similar construction over the space $\mathbb{N}^\omega$ of infinite sequences of natural numbers, we could show that for every measure μ there is a sequence x with $\mathbf{t}_\mu(x) = \infty$.

Proposition 5.3 is a little misleading, since as a locally compact set, $\mathbb{N}$ can be compactified into $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ (as in Part 1 of Example A.3). Theorem 6 implies that there is a neutral probability measure M over the compactified space $\overline{\mathbb{N}}$. Its restriction to $\mathbb{N}$ is, of course, not a probability measure, since it satisfies only $\sum_{x<\infty} M(x) \leqslant 1$. We called these functions *semimeasures*.

Remark 5.4.

1. It is easy to see that Theorem 3 characterizing randomness in terms of complexity holds also for the space $\overline{\mathbb{N}}$.
2. The topological space of semimeasures over $\mathbb{N}$ is not compact, and there is no neutral one among them. Its topology is not the same as what we get when we restrict the topology of probability measures over $\overline{\mathbb{N}}$ to $\mathbb{N}$. The difference is that over $\mathbb{N}$, for example the set of measures $\{\, \mu : \mu(\mathbb{N}) \geqslant 1/2 \,\}$ is closed, since $\mathbb{N}$ (as the whole space) is a closed set. But over $\overline{\mathbb{N}}$, this set is not closed.

$\diamondsuit$

Neutral measures are not too simple, even over $\overline{\mathbb{N}}$, as the following theorem shows.

Theorem 7. *There is no neutral measure over $\overline{\mathbb{N}}$ that is upper semicomputable over $\mathbb{N}$ or lower semicomputable over $\mathbb{N}$.*

Proof. Let us assume that ν is a measure that is upper semicomputable over $\mathbb{N}$. Then the set

$$\{\, (x,r) : x \in \mathbb{N},\ r \in \mathbb{Q},\ \nu(x) < r \,\}$$

is recursively enumerable: let (x_i, r_i) be a particular enumeration. For each n, let $i(n)$ be the first i with $r_i < 2^{-n}$, and let $y_n = x_{i(n)}$. Then $\nu(y_n) < 2^{-n}$, and at the same time $H(y_n) \overset{+}{<} H(n)$. As mentioned, in Remark 5.4, Theorem 3 characterizing randomness in terms of complexity holds also for the space $\overline{\mathbb{N}}$. Thus,

$$\mathbf{d}_\nu(y_n) \overset{\pm}{=} -\log \nu(y_n) - H(y_n \mid \nu) \overset{+}{>} n - H(n).$$

Suppose now that ν is lower semicomputable over $\mathbb{N}$. The proof for this case is longer. We know that ν is the monotonic limit of a recursive sequence $i \mapsto \nu_i(x)$ of recursive semimeasures with rational values $\nu_i(x)$. For every $k = 0, \dots, 2^n - 2$, let

$$V_{n,k} = \{\, \mu \in \mathcal{M}(\overline{\mathbb{N}}) : k \cdot 2^{-n} < \mu(\{0, \dots, 2^n - 1\}) < (k+2) \cdot 2^{-n} \,\},$$
$$J = \{\, (n,k) : k \cdot 2^{-n} < \nu(\{0, \dots, 2^n - 1\}) \,\}.$$

The set J is recursively enumerable. Let us define the functions $j : J \to \mathbb{N}$ and $x : J \to \{0, \dots, 2^n - 1\}$ as follows: $j(n, k)$ is the smallest i with $\nu_i(\{0, \dots, 2^n - 1\}) > k \cdot 2^{-n}$, and

$$x_{n,k} = \min\{\, y < 2^n : \nu_{j(n,k)}(y) < 2^{-n+1} \,\}.$$

Let us define the function $f_\mu(x, n, k)$ as follows. We set $f_\mu(x, n, k) = 2^{n-2}$ if the following conditions hold:

(a) $\mu \in V_{n,k}$;
(b) $\mu(x) < 2^{-n+2}$;
(c) $(n, k) \in J$ and $x = x_{n,k}$.

Otherwise, $f_\mu(x, n, k) = 0$. Clearly, the function $(\mu, x, n, k) \mapsto f_\mu(x, n, k)$ is lower semicomputable. Condition (b) implies

$$\sum_y \mu(y) f_\mu(y, n, k) \leqslant \mu(x_{n,k}) f_\mu(x_{n,k}, n, k) < 2^{-n+2} \cdot 2^{n-2} = 1. \qquad (5.3)$$

Let us show that $\nu \in V_{n,k}$ implies

$$f_\nu(x_{n,k}, n, k) = 2^{n-2}. \qquad (5.4)$$

Consider $x = x_{n,k}$. Conditions (a) and (c) are satisfied by definition. Let us show that condition (b) is also satisfied. Let $j = j(n, k)$. By definition, we have $\nu_j(x) < 2^{-n+1}$. Since by definition $\nu_j \in V_{n,k}$ and $\nu_j \leqslant \nu \in V_{n,k}$, we have

$$\nu(x) \leqslant \nu_j(x) + 2^{-n+1} < 2^{-n+1} + 2^{-n+1} = 2^{-n+2}.$$

Since all three conditions (a), (b) and (c) are satisfied, we have shown (5.4). Now we define

$$g_\mu(x) = \sum_{n \geqslant 2} \frac{1}{n(n+1)} \sum_k f_\mu(x, n, k).$$

Let us prove that $g_\mu(x)$ is a uniform test. It is lower semicomputable by definition, so we only need to prove $\sum_x \mu(x) f_\mu(x) \leqslant 1$. For this, let $I_{n,\mu} = \{\, k : \mu \in V_{n,k} \,\}$. Clearly by definition, $|I_{n,\mu}| \leqslant 2$. We have, using this last fact and the test property (5.3):

$$\sum_x \mu(x) g_\mu(x) = \sum_{n \geqslant 2} \frac{1}{n(n+1)} \sum_{k \in I_{n,\mu}} \sum_x \mu(x) f_\mu(x, n, k) \leqslant \sum_{n \geqslant 2} \frac{1}{n(n+1)} \cdot 2 \leqslant 1.$$

Thus, $g_\mu(x)$ is a uniform test. If $\nu \in V_{n,k}$ then we have

$$\mathbf{t}_\nu(x_{n,k}) \overset{*}{>} g_\nu(x_{n,k}) \geqslant \frac{1}{n(n+1)} f_\mu(x_{n,k}, n, k) \geqslant \frac{2^{n-2}}{n(n+1)}.$$

Hence ν is not neutral. $\qquad\square$

Remark 5.5. In [12] and [13], Levin imposed extra conditions on tests which allow to find a lower semicomputable neutral semimeasure. A typical (doubtless reasonable) consequence of these conditions would be that if outcome x is random with respect to measures μ and ν then it is also random with respect to $(\mu + \nu)/2$. $\diamond$

Remark 5.6. The universal lower semicomputable semimeasure $\mathbf{m}(x)$ has a certain property similar to neutrality. According to Theorem 3, for every computable measure μ we have $\mathbf{d}_\mu(x) \overset{+}{=} -\log \mu(x) - H(x)$ (where the constant in $\overset{+}{=}$ depends on μ). So, for computable measures, the expression

$$\overline{\mathbf{d}}_\mu(x) = -\log \mu(x) - H(x) \qquad (5.5)$$

can serve as a reasonable deficiency of randomness. (We will also use the test $\bar{\mathbf{t}} = 2^{\bar{\mathbf{d}}}$.) If we substitute $\mathbf{m}$ for μ in $\bar{\mathbf{d}}_\mu(x)$, we get 0. This substitution is not justified, of course. The fact that $\mathbf{m}$ is not a probability measure can be helped, at least over $\mathbb{N}$, using compactification as above, and extending the notion of randomness tests. But the test $\bar{\mathbf{d}}_\mu$ can replace $\mathbf{d}_\mu$ only for computable μ, while $\mathbf{m}$ is not computable. Anyway, this is the sense in which all outcomes might be considered random with respect to $\mathbf{m}$, and the heuristic sense in which $\mathbf{m}$ may still be considered "neutral". $\diamondsuit$

Remark 5.7. Solomonoff proposed the use of a universal lower semicomputable semimeasure (actually, a closely related structure) for inductive inference in [18]. He proved in [19] that sequences emitted by any computable probability distribution can be predicted well by his scheme. It may be interesting to see whether the same prediction scheme has stronger properties when used with the truly neutral measure M of the present paper. $\diamondsuit$

6. RELATIVE ENTROPY

Some properties of description complexity make it a good expression of the idea of individual information content.

6.1. **Entropy.** The entropy of a discrete probability distribution μ is defined as

$$\mathcal{H}(\mu) = -\sum_x \mu(x) \log \mu(x).$$

To generalize entropy to continuous distributions the *relative entropy* is defined as follows. Let μ, ν be two measures, where μ is taken (typically, but not always), to be a probability measure, and ν another measure, that can also be a probability measure but is most frequently not. We define the *relative entropy* $\mathcal{H}_\nu(\mu)$ as follows. If μ is not absolutely continuous with respect to ν then $\mathcal{H}_\nu(\mu) = -\infty$. Otherwise, writing

$$\frac{d\mu}{d\nu} = \frac{\mu(dx)}{\nu(dx)} =: f(x)$$

for the (Radon-Nikodym) derivative (density) of μ with respect to ν, we define

$$\mathcal{H}_\nu(\mu) = -\int \log \frac{d\mu}{d\nu} d\mu = -\mu^x \log \frac{\mu(dx)}{\nu(dx)} = -\nu^x f(x) \log f(x).$$

Thus, $\mathcal{H}(\mu) = \mathcal{H}_\#(\mu)$ is a special case.

Example 6.1. Let $f(x)$ be a probability density function for the distribution μ over the real line, and let λ be the Lebesgue measure there. Then

$$\mathcal{H}_\lambda(\mu) = -\int f(x) \log f(x) dx.$$

$\diamondsuit$

In information theory and statistics, when both μ and ν are probability measures, then $-\mathcal{H}_\nu(\mu)$ is also denoted $D(\mu \parallel \nu)$, and called (after Kullback) the information divergence of the two measures. It is frequently used in the role of a distance between μ and ν. It is not symmetric, but can be shown to obey the triangle inequality, and to be nonnegative. Let us prove the latter property: in our terms, it says that relative entropy is nonpositive when both μ and ν are probability measures.

Proposition 6.2. *Over a space* $\mathbf{X}$, *we have*

$$\mathcal{H}_\nu(\mu) \leqslant -\mu(X) \log \frac{\mu(X)}{\nu(X)}. \tag{6.1}$$

In particular, if $\mu(X) \geqslant \nu(X)$ *then* $\mathcal{H}_\nu(\mu) \leqslant 0$.

Proof. The inequality $-a \ln a \leqslant -a \ln b + b - a$ expresses the concavity of the logarithm function. Substituting $a = f(x)$ and $b = \mu(X)/\nu(X)$ and integrating by ν:

$$(\ln 2)\mathcal{H}_\nu(\mu) = -\nu^x f(x) \ln f(x) \leqslant -\mu(X) \ln \frac{\mu(X)}{\nu(X)} + \frac{\mu(X)}{\nu(X)} \nu(X) - \mu(X) = -\mu(X) \ln \frac{\mu(X)}{\nu(X)},$$

giving (6.1). $\qquad\square$

The following theorem generalizes an earlier known theorem stating that over a discrete space, for a computable measure, entropy is within an additive constant the same as "average complexity": $\mathcal{H}(\mu) \stackrel{+}{=} \mu^x H(x)$.

Theorem 8. *Let* μ *be a probability measure. Then we have*

$$\mathcal{H}_\nu(\mu) \leqslant \mu^x H_\nu(x \mid \mu). \tag{6.2}$$

If X *is a discrete space then the following estimate also holds:*

$$\mathcal{H}_\nu(\mu) \stackrel{+}{>} \mu^x H_\nu(x \mid \mu). \tag{6.3}$$

Proof. Let δ be the measure with density $\mathbf{t}_\nu(x \mid \mu)$ with respect to ν: $\mathbf{t}_\nu(x \mid \mu) = \frac{\delta(dx)}{\nu(dx)}$. Then $\delta(X) \leqslant 1$. It is easy to see from the maximality property of $\mathbf{t}_\nu(x \mid \mu)$ that $\mathbf{t}_\nu(x \mid \mu) > 0$, therefore according to Proposition B.9, we have $\frac{\nu(dx)}{\delta(dx)} = \left(\frac{\delta(dx)}{\nu(dx)} \right)^{-1}$. Using Proposition B.9 and 6.2:

$$\mathcal{H}_\nu(\mu) = -\mu^x \log \frac{\mu(dx)}{\nu(dx)},$$

$$-\mu^x H_\nu(x \mid \mu) = \mu^x \log \frac{\delta(dx)}{\nu(dx)} = -\mu^x \log \frac{\nu(dx)}{\delta(dx)},$$

$$\mathcal{H}_\nu(\mu) - \mu^x H_\nu(x \mid \mu) = -\mu^x \log \frac{\mu(dx)}{\delta(dx)} \leqslant -\mu(X) \log \frac{\mu(X)}{\delta(X)} \leqslant 0.$$

This proves (6.2).

Over a discrete space $\mathbf{X}$, the function $(x, \mu, \nu) \mapsto \frac{\mu(dx)}{\nu(dx)} = \frac{\mu(x)}{\nu(x)}$ is computable, therefore by the maximality property of $H_\nu(x \mid \mu)$ we have $\frac{\mu(dx)}{\nu(dx)} \stackrel{*}{<} \mathbf{t}_\nu(x \mid \mu)$, hence $\mathcal{H}_\nu(\mu) = -\mu^x \log \frac{\mu(dx)}{\nu(dx)} \stackrel{+}{>} \mu^x H_\nu(x \mid \mu)$. $\qquad\square$

6.2. **Addition theorem.** The most important information-theoretical property of description complexity is the following theorem (see for example [14]):

Proposition 6.3 (Addition Theorem). *We have* $H(x, y) \stackrel{+}{=} H(x) + H(y \mid x, H(x))$.

Mutual information is defined as $I(x : y) = H(x) + H(y) - H(x, y)$. By the Addition theorem, we have $I(x : y) \stackrel{+}{=} H(y) - H(y \mid x, H(x)) \stackrel{+}{=} H(x) - H(x \mid y, H(y))$. The two latter expressions show that in some sense, $I(x : y)$ is the information held in x about y as well as the information held in y about x. (The terms $H(x)$, $H(y)$ in the conditions are logarithmic-sized corrections to this idea.) Using (5.5), it is interesting to view mutual information

$I(x:y)$ as a deficiency of randomness of the pair (x,y) in terms of the expression $\overline{\mathbf{d}}_\mu$, with respect to $\mathbf{m} \times \mathbf{m}$:

$$I(x:y) = H(x) + H(y) - H(x,y) = \overline{\mathbf{d}}_{\mathbf{m} \times \mathbf{m}}(x,y).$$

Taking $\mathbf{m}$ as a kind of "neutral" probability, even if it is not quite such, allows us to view $I(x:y)$ as a "deficiency of independence". Is it also true that $I(x:y) \overset{+}{=} \mathbf{d}_{\mathbf{m} \times \mathbf{m}}(x)$? This would allow us to deduce, as Levin did, "information conservation" laws from randomness conservation laws.[1]

Expression $\mathbf{d}_{\mathbf{m} \times \mathbf{m}}(x)$ must be understood again in the sense of compactification, as in Section 5. There seem to be two reasonable ways to compactify the space $\mathbb{N} \times \mathbb{N}$: we either compactify it directly, by adding a symbol ∞, or we form the product $\overline{\mathbb{N}} \times \overline{\mathbb{N}}$. With either of them, preserving Theorem 3, we would have to check whether $H(x,y \mid \mathbf{m} \times \mathbf{m}) \overset{+}{=} H(x,y)$. But, knowing the function $\mathbf{m}(x) \times \mathbf{m}(y)$ we know the function $x \mapsto \mathbf{m}(x) \overset{*}{=} \mathbf{m}(x) \times \mathbf{m}(0)$, hence also the function $(x,y) \mapsto \mathbf{m}(x,y) = \mathbf{m}(\langle x,y \rangle)$, where $\langle x,y \rangle$ is any fixed computable pairing function. Using this knowledge, it is possible to develop an argument similar to the proof of Theorem 7, showing that $H(x,y \mid \mathbf{m} \times \mathbf{m}) \overset{+}{=} H(x,y)$ does not hold.

Question 1. *Is there a neutral measure M with the property $I(x:y) = \mathbf{d}_{M \times M}(x,y)$? Is this true maybe for all neutral measures M? If not, how far apart are the expressions $\mathbf{d}_{M \times M}(x,y)$ and $I(x:y)$ from each other?*

The Addition Theorem (Proposition 6.3) can be generalized to the algorithmic entropy $H_\mu(x)$ introduced in (4.1) (a somewhat similar generalization appeared in [23]). The generalization, defining $H_{\mu,\nu} = H_{\mu \times \nu}$, is

$$H_{\mu,\nu}(x,y) \overset{+}{=} H_\mu(x \mid \nu) + H_\nu(y \mid x, \ H_\mu(x \mid \nu), \ \mu). \tag{6.4}$$

Before proving the general addition theorem, we establish a few useful facts.

Proposition 6.4. *We have*

$$H_\mu(x \mid \nu) \overset{+}{<} -\log \nu^y 2^{-H_{\mu,\nu}(x,y)}.$$

Proof. The function $f(x,\mu,\nu)$ that is the right-hand side, is upper semicomputable by definition, and obeys $\mu^x 2^{-f(x,\mu,\nu)} \leqslant 1$. Therefore the inequality follows from the minimum property of $H_\mu(x)$. $\square$

Let us generalize the minimum property of $H_\mu(x)$.

Proposition 6.5. *Let $(x,y,\nu) \mapsto f_\nu(x,y)$ be a nonnegative lower semicomputable function with $F_\nu(x) = \log \nu^y f_\nu(x,y)$. Then for all x with $F_\nu(x) > -\infty$ we have*

$$H_\nu(y \mid x, \lfloor F_\nu(x) \rfloor) \overset{+}{<} -\log f_\nu(x,y) + F_\nu(x).$$

Proof. Let us construct a lower semicomputable function $(x,y,m,\nu) \mapsto g_\nu(x,y,m)$ for integers m with the property that $\nu^y g_\nu(x,y,m) \leqslant 2^{-m}$, and for all x with $F_\nu(x) \leqslant -m$ we have $g_\nu(x,y,m) = f_\nu(x,y)$. Such a g can be constructed by watching the approximation of f grow and cutting it off as soon as it would give $F_\nu(x) > -m$. Now $(x,y,m,\nu) \mapsto 2^m g_\nu(x,y,m)$ is a uniform conditional test of y and hence it is $\overset{*}{<} 2^{-H_\nu(y \mid x,m)}$. To finish the proof, substitute $-\lfloor F_\nu(x) \rfloor$ for m and rearrange. $\square$

---

[1]We cannot use the test $\overline{\mathfrak{t}}_\mu$ for this, since—as it can be shown easily–it does not obey randomness conservation.

Let $z \in \mathbb{N}$, then the inequality

$$H_\mu(x) \overset{+}{<} H(z) + H_\mu(x \mid z) \tag{6.5}$$

will be a simple consequence of the general addition theorem. The following lemma, needed in the proof of the theorem, generalizes this inequality somewhat:

Lemma 6.6. *For a computable function $(y, z) \mapsto f(y, z)$ over $\mathbb{N}$, we have*

$$H_\mu(x \mid y) \overset{+}{<} H(z) + H_\mu(x \mid f(y, z)).$$

Proof. The function

$$(x, y, \mu) \mapsto g_\mu(x, y) = \sum_z 2^{-H_\mu(x \mid f(y,z)) - H(z)}$$

is lower semicomputable, and $\mu^x g_\mu(x, y) \leqslant \sum_z 2^{-H(z)} \leqslant 1$. Hence $g_\mu(x, y) \overset{*}{<} 2^{-H_\mu(x \mid y)}$. The left-hand side is a sum, hence the inequality holds for each element of the sum: just what we had to prove. $\qquad\square$

As mentioned above, the theory generalizes to measures that are not probability measures. Taking $f_\mu(x, y) = 1$ in Proposition 6.5 gives the inequality

$$H_\mu(x \mid \lfloor \log \mu(X) \rfloor) \overset{+}{<} \log \mu(X),$$

with a physical meaning when μ is the phase space measure. Using (6.5), this implies

$$H_\mu(x) \overset{+}{<} \log \mu(X) + H(\lfloor \log \mu(X) \rfloor). \tag{6.6}$$

The following simple monotonicity property will be needed:

Lemma 6.7. *For $i < j$ we have*

$$i + H_\mu(x \mid i) \overset{+}{<} j + H_\mu(x \mid j).$$

Proof. From Lemma 6.6, with $f(i, n) = i + n$ we have

$$H_\mu(x \mid i) - H_\mu(x \mid j) \overset{+}{<} H(j - i) \overset{+}{<} j - i.$$

$\qquad\square$

Theorem 9 (General addition). *The following inequality holds:*

$$H_{\mu,\nu}(x, y) \overset{+}{=} H_\mu(x \mid \nu) + H_\nu(y \mid x, \ H_\mu(x \mid \nu), \ \mu).$$

Proof. To prove the inequality $\overset{+}{<}$, let us define

$$G_{\mu,\nu}(x, y, m) = \min_{i \geqslant m} \ i + H_\nu(y \mid x, i, \mu).$$

Function $G_{\mu,\nu}(x, y, m)$ is upper semicomputable and decreasing in m. Therefore

$$G_{\mu,\nu}(x, y) = G_{\mu,\nu}(x, y, H_\mu(x \mid \nu))$$

is also upper semicomputable since it is obtained by substituting an upper semicomputable function for m in $G_{\mu,\nu}(x, y, m)$. Lemma 6.7 implies

$$G_{\mu,\nu}(x, y, m) \overset{+}{=} m + H_\nu(y \mid x, m, \mu),$$
$$G_{\mu,\nu}(x, y) \overset{+}{=} H_\mu(x \mid \nu) + H_\nu(y \mid x, H_\mu(x \mid \nu), \mu).$$

Now, we have

$$\nu^y 2^{-m-H_\nu(y|x,m,\mu)} \leqslant 2^{-m},$$

$$\nu^y 2^{-G_{\mu,\nu}(x,y)} \overset{*}{<} 2^{-H_\mu(x|\mu)}.$$

Therefore $\mu^x \nu^y 2^{-G} \overset{*}{<} 1$, implying $H_{\mu,\nu}(x,y) \overset{+}{<} G_{\mu,\nu}(x,y)$ by the minimality property of $H_{\mu,\nu}(x,y)$. This proves the $\overset{+}{<}$ half of our theorem.

To prove the inequality $\overset{+}{>}$, let

$$f_\nu(x,y,\mu) = 2^{-H_{\mu,\nu}(x,y)},$$

$$F_\nu(x,\mu) = \log \nu^y f_\nu(x,y,\mu).$$

According to Proposition 6.5,

$$H_\nu(y \mid x, \lfloor F \rfloor, \mu) \overset{+}{<} -\log f_\nu(x,y,\mu) + F_\nu(x,\mu),$$

$$H_{\mu,\nu}(x,y) \overset{+}{>} -F + H_\nu(y \mid x, \lceil -F \rceil, \mu).$$

Proposition 6.4 implies $-F_\nu(x,\mu) \overset{+}{>} H_\mu(x \mid \nu)$. The monotonity lemma 6.7 implies from here the $\overset{+}{>}$ half of the theorem. $\qquad\square$

6.3. **Some special cases of the addition theorem; information.** The function $H_\mu(\cdot)$ behaves quite differently for different kinds of measures μ. Recall the following property of complexity:

$$H(f(x) \mid y) \overset{+}{<} H(x \mid g(y)) \overset{+}{<} H(x). \tag{6.7}$$

for any computable functions f, g This implies

$$H(y) \overset{+}{<} H(x,y).$$

In contrast, if μ is a probability measure then

$$H_\nu(y) \overset{+}{>} H_{\mu,\nu}(x,y).$$

This comes from the fact that $2^{-H_\nu(y)}$ is a test for $\mu \times \nu$.

Let us explore some of the consequences and meanings of the additivity property. As noted in (4.2), the subscript μ can always be added to the condition: $H_\mu(x) \overset{+}{=} H_\mu(x \mid \mu)$. Similarly, we have

$$H_{\mu,\nu}(x,y) := H_{\mu \times \nu}(x,y) \overset{+}{=} H_{\mu \times \nu}(x,y \mid \mu \times \nu) \overset{+}{=} H_{\mu \times \nu}(x,y \mid \mu,\nu) =: H_{\mu,\nu}(x,y \mid \mu,\nu),$$

where only before-last inequality requires new (easy) consideration.

Let us assume that $X = Y = \Sigma^*$, the discrete space of all strings. With general μ, ν such that $\mu(x), \nu(x) \neq 0$ for all x, using (4.6), the addition theorem specializes to the ordinary addition theorem, conditioned on μ, ν:

$$H(x,y \mid \mu,\nu) \overset{+}{=} H(x \mid \mu,\nu) + H(y \mid x, \ H(x \mid \mu,\nu), \ \mu,\nu).$$

In particular, whenever μ, ν are computable, this is just the regular addition theorem.

Just as above, we defined mutual information as $I(x:y) = H(x) + H(y) - H(x,y)$, the new addition theorem suggests a more general definition

$$I_{\mu,\nu}(x:y) = H_\mu(x \mid \nu) + H_\nu(y \mid \mu) - H_{\mu,\nu}(x,y).$$

In the discrete case $X = Y = \Sigma^*$ with everywhere positive $\mu(x), \nu(x)$, this simplifies to

$$I_{\mu,\nu}(x:y) = H(x \mid \mu,\nu) + H(y \mid \mu,\nu) - H(x,y|\mu,\nu),$$

which is $\stackrel{\pm}{=} I(x : y)$ in case of computable μ, ν. How different can it be for non-computable μ, ν?

In the general case, even for computable μ, ν, it seems worth finding out how much this expression depends on the choice of μ, ν. Can one arrive at a general, natural definition of mutual information along this path?

7. CONCLUSION

When uniform randomness tests are defined in as general a form as they were here, the theory of information conservation does not fit nicely into the theory of randomness conservation as it did with [12] and [13]. Still, it is worth laying the theory onto broad foundations that, we hope, can serve as a basis for further development.

APPENDIX A. TOPOLOGICAL SPACES

Given two sets X, Y, a *partial function* f from X to Y, defined on a subset of Y, will be denoted as

$$f :\subseteq X \to Y.$$

A.1. **Topology.** A *topology* on a set X is defined by a class τ of its subsets called *open sets*. It is required that the empty set and X are open, and that arbitrary union and finite intersection of open sets is open. The pair (X, τ) is called a *topological space*. A topology τ' on X is called *larger*, or *finer* than τ if $\tau' \supseteq \tau$. A set is called *closed* if its complement is open. A set B is called the *neighborhood* of a set A if B contains an open set that contains A. We denote by $\overline{A}, A^\mathrm{o}$ the closure (the intersection of all closed sets containing A) and the interior of A (the union of all open sets in A) respectively. Let

$$\partial A = \overline{A} \smallsetminus A^\mathrm{o}$$

denote the boundary of set A. A *base* is a subset β of τ such that every open set is the union of some elements of β. A *neighborhood* of a point is a base element containing it. A *base of neighborhoods of a point* x is a set N of neighborhoods of x with the property that each neighborhood of x contains an element of N. A *subbase* is a subset σ of τ such that every open set is the union of finite intersections from σ.

Examples A.1.

1. Let X be a set, and let β be the set of all points of X. The topology with base β is the *discrete topology* of the set X. In this topology, every subset of X is open (and closed).
2. Let X be the real line $\mathbb{R}$, and let $\beta_\mathbb{R}$ be the set of all open intervals $(a; b)$. The topology $\tau_\mathbb{R}$ obtained from this base is the usual topology of the real line. When we refer to $\mathbb{R}$ as a topological space without qualification, this is the topology we will always have in mind.
3. Let $X = \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, and let $\beta_{\overline{\mathbb{R}}}$ consist of all open intervals $(a; b)$ and in addition of all intervals of the forms $[-\infty; a)$ and $(a; \infty]$. It is clear how the space $\overline{\mathbb{R}}_+$ is defined.
4. Let X be the real line $\mathbb{R}$. Let $\beta_\mathbb{R}^>$ be the set of all open intervals $(-\infty; b)$. The topology with base $\beta_\mathbb{R}^>$ is also a topology of the real line, different from the usual one. Similarly, let $\beta_\mathbb{R}^<$ be the set of all open intervals $(b; \infty)$.
5. On the space Σ^ω, let $\tau_C = \{ A\Sigma^\omega : A \subseteq \Sigma^* \}$ be called the topology of the *Cantor space* (over Σ).

$\diamondsuit$

A set is called a G_δ set if it is a countable intersection of open sets, and it is an F_σ set if it is a countable union of closed sets.

For two topologies τ_1, τ_2 over the same set X, we define the topology $\tau_1 \vee \tau_2 = \tau_1 \cap \tau_2$, and $\tau_1 \wedge \tau_2$ as the smallest topology containing $\tau_1 \cup \tau_2$. In the example topologies of the real numbers above, we have $\tau_\mathbb{R} = \tau_\mathbb{R}^< \wedge \tau_\mathbb{R}^>$.

We will always require the topology to have at least the T_0 property: every point is determined by the class of open sets containing it. This is the weakest one of a number of other possible separation properties: both topologies of the real line in the example above have it. A stronger such property would be the T_2 property: a space is called a *Hausdorff* space, or T_2 space, if for every pair of different points x, y there is a pair of disjoint open sets A, B with $x \in A$, $y \in B$. The real line with topology $\tau_\mathbb{R}^>$ in Example A.1.4 above is not a Hausdorff space. A space is Hausdorff if and only if every open set is the union of closed neighborhoods.

Given two topological spaces (X_i, τ_i) $(i = 1, 2)$, a function $f :\subseteq X_1 \to X_2$ is called *continuous* if for every open set $G \subset X_2$ its inverse image $f^{-1}(G)$ is also open. If the topologies τ_1, τ_2 are not clear from the context then we will call the function (τ_1, τ_2)-continuous. Clearly, the property remains the same if we require it only for all elements G of a subbase of X_2. If there are two continuous functions between X and Y that are inverses of each other then the two spaces are called *homeomorphic*. We say that f is continuous *at point* x if for every neighborhood V of $f(x)$ there is a neighborhood U of x with $f(U) \subseteq V$. Clearly, f is continuous if and only if it is continuous in each point.

A *subspace* of a topological space (X, τ) is defined by a subset $Y \subseteq X$, and the topology $\tau_Y = \{ G \cap Y : G \in \tau \}$, called the *induced* topology on Y. This is the smallest topology on Y making the identity mapping $x \mapsto x$ continuous. A partial function $f :\subseteq X \to Z$ with $\mathrm{dom}(f) = Y$ is continuous iff $f : Y \to Z$ is continuous.

For two topological spaces (X_i, τ_i) $(i = 1, 2)$, we define the *product topology* on their product $X \times Y$: this is the topology defined by the subbase consisting of all sets $G_1 \times X_2$ and all sets $X_1 \times G_2$ with $G_i \in \tau_i$. The product topology is the smallest topology making the projection functions $(x, y) \mapsto x$, $(x, y) \mapsto y$ continuous. Given topological spaces X, Y, Z we call a two-argument function $f : X \times Y \mapsto Z$ continuous if it is continuous as a function from $X \times Y$ to Z. The product topology is defined similarly for over the product $\prod_{i \in I} X_i$ of an arbitrary number of spaces, indexed by some index set I. We say that a function is $(\tau_1, \dots, \tau_n, \mu)$-continuous if it is $(\tau_1 \times \cdots \times \tau_n, \mu)$-continuous.

Examples A.2.

1. The space $\mathbb{R} \times \mathbb{R}$ with the product topology has the usual topology of the Euclidean plane.
2. Let X be a set with the discrete topology, and X^ω the set of infinite sequences with elements from X, with the product topology. A base of this topology is provided by all sets of the form uX^ω where $u \in X^*$. The elements of this base are closed as well as open. When $X = \{0, 1\}$ then this topology is the usual topology of infinite binary sequences.

$$\diamond$$

A real function $f : X_1 \to \mathbb{R}$ is called continuous if it is $(\tau_1, \tau_\mathbb{R})$-continuous. It is called *lower semicontinuous* if it is $(\tau_1, \tau_\mathbb{R}^<)$-continuous. The definition of upper semicontinuity is similar. Clearly, f is continuous if and only if it is both lower and upper semicontinuous. The requirement of lower semicontinuity of f is that for each $r \in \mathbb{R}$, the set $\{ x : f(x) > r \}$ is open. This can be seen to be equivalent to the requirement that the single set $\{ (x, r) : f(x) > r \}$ is open. It is easy to see that the supremum of any set of lower semicontinuous functions is lower semicontinuous.

Let (X, τ) be a topological space, and B a subset of X. An *open cover* of B is a family of open sets whose union contains B. A subset K of X is said to be *compact* if every open cover of K has a finite subcover. Compact sets have many important properties: for example, a continuous function over a compact set is bounded.

Example A.3.

1. Every finite discrete space is compact. An infinite discrete space $\mathbf{X} = (X, \tau)$ is not compact, but it can be turned into a compact space $\overline{\mathbf{X}}$ by adding a new element called ∞: let $\overline{X} = X \cup \{\infty\}$, and $\overline{\tau} = \tau \cup \{\overline{X} \smallsetminus A : A \subset X \text{ closed }\}$. More generally, this simple operation can be performed with every space that is *locally compact*, that each of its points has a compact neighborhood.
2. In a finite-dimensional Euclidean space, every bounded closed set is compact.
3. It is known that if $(\mathbf{X}_i)_{i \in I}$ is a family of compact spaces then their direct product is also compact.

$\diamond$

A subset K of X is said to be *sequentially compact* if every sequence in K has a convergent subsequence with limit in K. The space is *locally compact* if every point has a compact neighborhood.

A.2. **Metric spaces.** In our examples for metric spaces, and later in our treatment of the space of probability measures, we refer to [2]. A *metric space* is given by a set X and a distance function $d : X \times X \to \mathbb{R}_+$ satisfying the *triangle inequality* $d(x, z) \leqslant d(x, y) + d(y, z)$ and also property that $d(x, y) = 0$ implies $x = y$. For $r \in \mathbb{R}_+$, the sets

$$B(x, r) = \{\, y : d(x, y) < r \,\}, \quad \overline{B}(x, r) = \{\, y : d(x, y) \leqslant r \,\}$$

are called the open and closed *balls* with radius r and center x. A metric space is also a topological space, with the base that is the set of all open balls. Over this space, the distance function $d(x, y)$ is obviously continuous. Each metric space is a Hausdorff space; moreover, it has the following stronger property. For every pair of different points x, y there is a continuous function $f : X \to \mathbb{R}$ with $f(x) \neq f(y)$. (To see this, take $f(z) = d(x, z)$.) This is called the T_3 property. A metric space is *bounded* when $d(x, y)$ has an upper bound on X. A topological space is called *metrizable* if its topology can be derived from some metric space.

Notation. For an arbitrary set A and point x let

$$d(x, A) = \inf_{y \in A} d(x, y),$$
$$A^\varepsilon = \{\, x : d(x, A) < \varepsilon \,\}. \tag{A.1}$$

$\diamond$

Examples A.4.

1. The real line with the distance $d(x, y) = |x - y|$ is a metric space. The topology of this space is the usual topology $\tau_{\mathbb{R}}$ of the real line.
2. The space $\mathbb{R} \times \mathbb{R}$ with the Euclidean distance gives the same topology as the product topology of $\mathbb{R} \times \mathbb{R}$.
3. An arbitrary set X with the distance $d(x, y) = 1$ for all pairs x, y of different elements, is a metric space that induces the discrete topology on X.

4. Let X be a bounded metric space, and let $Y = X^\omega$ be the set of infinite sequences $x = (x_1, x_2, \dots)$ with distance function $d^\omega(x, y) = \sum_i 2^{-i} d(x_i, y_i)$. The topology of this space is the same as the product topology defined in Example A.2.2.

5. Let X be a metric space, and let $Y = X^\omega$ be the set of infinite bounded sequences $x = (x_1, x_2, \dots)$ with distance function $d(x, y) = \sup_i d(x_i, y_i)$.

6. Let X be a metric space, and let $C(X)$ be the set of bounded continuous functions over X with distance function $d'(f, g) = \sup_x d(f(x), g(x))$. A special case is $C[0; 1]$ where the interval $[0; 1]$ of real numbers has the usual metric.

7. Let l_2 be the set of infinite sequences $x = (x_1, x_2, \dots)$ of real numbers with the property that $\sum_i x_i^2 < \infty$. The metric is given by the distance $d(x, y) = (\sum_i |x_i - y_i|^2)^{1/2}$.

$\diamond$

A topological space has the *first countability property* if each point has a countable base of neighborhoods. Every metric space has the first countability property since we can restrict ourselves to balls with rational radius. Given a topological space (X, τ) and a sequence $x = (x_1, x_2, \dots)$ of elements of X, we say that x *converges* to a point y if for every neighborhood G of y there is a k such that for all $m > k$ we have $x_m \in G$. We will write $y = \lim_{n \to \infty} x_n$. It is easy to show that if spaces (X_i, τ_i) $(i = 1, 2)$ have the first countability property then a function $f : X \to Y$ is continuous if and only if for every convergent sequence (x_n) we have $f(\lim_n x_n) = \lim_n f(x_n)$. A topological space has the *second countability property* if the whole space has a countable base. For example, the space $\mathbb{R}$ has the second countability property for all three topologies $\tau_\mathbb{R}, \tau_\mathbb{R}^<, \tau_\mathbb{R}^>$. Indeed, we also get a base if instead of taking all intervals, we only take intervals with rational endpoints. On the other hand, the metric space of Example A.4.5 does not have the second countability property. In a topological space (X, τ), a set B of points is called *dense* at a point x if it intersects every neighborhood of x. It is called *everywhere dense*, or *dense*, if it is dense at every point. A metric space is called *separable* if it has a countable everywhere dense subset. This property holds if and only if the space as a topological space has the second countability property.

Example A.5. In Example A.4.6, for $X = [0; 1]$, we can choose as our everywhere dense set the set of all polynomials with rational coefficients, or alternatively, the set of all piecewise linear functions whose graph has finitely many nodes at rational points. $\diamond$

Let X be a metric space, and let $C(X)$ be the set of bounded continuous functions over X with distance function $d'(f, g) = \sup_x d(f(x), g(x))$. A special case is $C[0; 1]$ where the interval $[0; 1]$ of real numbers has the usual metric.

Let (X, d) be a metric space, and $a = (a_1, a_1, \dots)$ an infinite sequence. A metric space is called *complete* if every Cauchy sequence in it has a limit. It is well-known that every metric space can be embedded (as an everywhere dense subspace) into a complete space.

It is easy to see that in a metric space, every closed set is a G_δ set (and every open set is an F_σ set).

Example A.6. Consider the set $D[0; 1]$ of functions over $[0; 1]$ that are right continuous and have left limits everywhere. The book [2] introduces two different metrics for them: the Skorohod metric d and the d_0 metric. In both metrics, two functions are close if a slight monotonic continuous deformation of the coordinate makes them uniformly close. But in the d_0 metric, the slope of the deformation must be close to 1. It is shown that the two metrics give rise to the same topology; however, the space with metric d is not complete, and the space with metric d_0 is. $\diamond$

Let (X, d) be a metric space. It can be shown that a subset K of X is compact if and only if it is sequentially compact. Also, K is compact if and only if it is closed and for every ε, there is a finite set of ε-balls (balls of radius ε) covering it.

We will develop the theory of randomness over separable complete metric spaces. This is a wide class of spaces encompassing most spaces of practical interest. The theory would be simpler if we restricted it to compact or locally compact spaces; however, some important spaces like $C[0; 1]$ (the set of continuouos functions over the interval $[0; 1]$, with the maximum difference as their distance) are not locally compact.

Given a function $f : X \to Y$ between metric spaces and $\beta > 0$, let $\mathrm{Lip}_\beta(X, Y)$ denote the set of functions (called the Lipschitz(β) functions, or simply Lipschitz functions) satisfying

$$d_Y(f(x), f(y)) \leqslant \beta d_X(x, y). \tag{A.2}$$

All these functions are uniformly continuous. Let $\mathrm{Lip}(X) = \mathrm{Lip}(X, \mathbb{R}) = \bigcup_\beta \mathrm{Lip}_\beta$ be the set of real Lipschitz functions over X.

APPENDIX B. MEASURES

For a survey of measure theory, see for example [16].

B.1. **Set algebras.** A (Boolean set-) *algebra* is a set of subsets of some set X closed under intersection and complement (and then, of course, under union). It is a *σ-algebra* if it is also closed under countable intersection (and then, of course, under countable union). A *semialgebra* is a set $\mathcal{L}$ of subsets of some set X closed under intersection, with the property that the complement of every element of $\mathcal{L}$ is the disjoint union of a finite number of elements of $\mathcal{L}$. If $\mathcal{L}$ is a semialgebra then the set of finite unions of elements of $\mathcal{L}$ is an algebra.

Examples B.1.

1. The set $\mathcal{L}_1$ of left-closed intervals of the line (including intervals of the form $(-\infty; a)$) is a semialgebra.
2. The set $\mathcal{L}_2$ of all intervals of the line (which can be open, closed, left-closed or right-closed), is a semialgebra.
3. In the set $\{0, 1\}^\omega$ of infinite 0-1-sequences, the set $\mathcal{L}_3$ of all subsets of the form $u\{0, 1\}^\omega$ with $u \in \{0, 1\}^*$, is a semialgebra.
4. The σ-algebra $\mathcal{B}$ generated by $\mathcal{L}_1$, is the same as the one generated by $\mathcal{L}_2$, and is also the same as the one generated by the set of all open sets: it is called the family of *Borel sets* of the line. The Borel sets of the extended real line $\overline{\mathbb{R}}$ are defined similarly.
5. Given σ-algebras $\mathcal{A}, \mathcal{B}$ in sets X, Y, the product σ-algebra $\mathcal{A} \times \mathcal{B}$ in the space $X \times Y$ is the one generated by all elements $A \times Y$ and $X \times B$ for $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

$\Diamond$

B.2. **Measures.** A *measurable space* is a pair $(X, \mathcal{S})$ where $\mathcal{S}$ is a σ-algebra of sets of X. A *measure* on a measurable space $(X, \mathcal{S})$ is a function $\mu : B \to \overline{\mathbb{R}}_+$ that is *σ-additive*: this means that for every countable family $A_1, A_2, \ldots$ of disjoint elements of $\mathcal{S}$ we have $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$. A measure μ is *σ-finite* if the whole space is the union of a countable set of subsets whose measure is finite. It is *finite* if $\mu(X) < \infty$. It is a *probability measure* if $\mu(X) = 1$.

It is important to understand how a measure can be defined in practice. Algebras are generally simpler to grasp constructively than σ-algebras; semialgebras are yet simpler. Suppose that μ is defined over a semialgebra $\mathcal{L}$ and is additive. Then it can always be uniquely extended to an additive function over the algebra generated by $\mathcal{L}$. The following is an important theorem of measure theory.

Proposition B.2. *Suppose that a nonnegative set function defined over a semialgebra $\mathcal{L}$ is σ-additive. Then it can be extended uniquely to the σ-algebra generated by $\mathcal{L}$.*

Examples B.3.

1. Let x be point and let $\mu(A) = 1$ if $x \in A$ and 0 otherwise. In this case, we say that μ is *concentrated* on the point x.
2. Consider the the line $\mathbb{R}$, and the algebra $\mathcal{L}_1$ defined in Example B.1.1. Let $f : \mathbb{R} \to \mathbb{R}$ be a monotonic real function. We define a set function over $\mathcal{L}_1$ as follows. Let $[a_i; b_i)$, $(i = 1, \ldots, n)$ be a set of disjoint left-closed intervals. Then $\mu(\bigcup_i [a_i; b_i)) = \sum_i f(b_i) - f(a_i)$. It is easy to see that μ is additive. It is σ-additive if and only if f is left-continuous.
3. Let $B = \{0, 1\}$, and consider the set B^ω of infinite 0-1-sequences, and the semialgebra $\mathcal{L}_3$ of Example B.1.3. Let $\mu : B^* \to \mathbb{R}^+$ be a function. Let us write $\mu(uB^\omega) = \mu(u)$ for all $u \in B^*$. Then it can be shown that the following conditions are equivalent: μ is σ-additive over $\mathcal{L}_3$; it is additive over $\mathcal{L}_3$; the equation $\mu(u) = \mu(u0) + \mu(u1)$ holds for all $u \in B^*$.
4. The nonnegative linear combination of any finite number of measures is also a measure. In this way, it is easy to construct arbitrary measures concentrated on a finite number of points.
5. Given two measure spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ it is possible to define the product measure $\mu \times \nu$ over the measureable space $(X \times Y, \mathcal{A} \times \mathcal{B})$. The definition is required to satisfy $\mu \times \nu(A \times B) = \mu(A) \times \nu(B)$, and is determined uniquely by this condition. If ν is a probability measure then, of course, $\mu(A) = \mu \times \nu(A \times Y)$.

$$\diamondsuit$$

Remark B.4. Example B.3.3 shows a particularly attractive way to define measures. Keep splitting the values $\mu(u)$ in an arbitrary way into $\mu(u0)$ and $\mu(u1)$, and the resulting values on the semialgebra define a measure. Example B.3.2 is less attractive, since in the process of defining μ on all intervals and only keeping track of finite additivity, we may end up with a monotonic function that is not left continuous, and thus with a measure that is not σ-additive. In the subsection on probability measures over a metric space, we will find that even on the real line, there is a way to define measures in a step-by-step manner, and only checking for consistency along the way. $\diamondsuit$

A *probability space* is a triple $(X, \mathcal{S}, P)$ where $(X, \mathcal{S})$ is a measurable space and P is a probability measure over it.

Let $(X_i, \mathcal{S}_i)$ $(i = 1, 2)$ be measurable spaces, and let $f : X \to Y$ be a mapping. Then f is *measurable* if and only if for each element B of $\mathcal{S}_2$, its inverse image $f^{-1}(B)$ is in $\mathcal{S}_1$. If μ_1 is a measure over $(X_1, \mathcal{S}_1)$ then μ_2 defined by $\mu_2(A) = \mu_1(f^{-1}(A))$ is a measure over X_2 called the measure *induced* by f.

B.3. Integral. A measurable function $f : X \to \mathbb{R}$ is called a *step function* if its range is finite. The set of step functions is closed with respect to linear combinations and also with respect to the operations $\wedge, \vee$. Such a set of functions is called a *Riesz space*.

Given a step function which takes values x_i on sets A_i, and a finite measure μ, we define

$$\mu(f) = \mu f = \int f \, d\mu = \int f(x)\mu(dx) = \sum_i x_i \mu(A_i).$$

This is a linear positive functional on the set of step functions. Moreover, it can be shown that it is continuous on monotonic sequences: if $f_i \searrow 0$ then $\mu f_i \searrow 0$. The converse can

also be shown: Let μ be a linear positive functional on step functions that is continuous on monotonic sequences. Then the set function $\mu(A) = \mu(1_A)$ is a finite measure.

Proposition B.5. *Let $\mathcal{E}$ be any Riesz space of functions with the property that $1 \in \mathcal{E}$. Let μ be a positive linear functional on $\mathcal{E}$ continuous on monotonic sequences, with $\mu 1 = 1$. The functional μ can be extended to the set $\mathcal{E}_+$ of monotonic limits of nonnegative elements of $\mathcal{E}$, by continuity. In case when $\mathcal{E}$ is the set of all step functions, the set $\mathcal{E}_+$ is the set of all nonnegative measurable functions.*

Let us fix a finite measure μ over a measurable space $(X, \mathcal{S})$. A measurable function f is called *integrable* with respect to μ if $\mu|f|^+ < \infty$ and $\mu|f|^- < \infty$. In this case, we define $\mu f = \mu|f|^+ - \mu|f|^-$. The set of integrable functions is a Riesz space, and the positive linear functional μ on it is continuous with respect to monotonic sequences. The continuity over monotonic sequences also implies the following *bounded convergence theorem*.

Proposition B.6. *Suppose that functions f_n are integrable and $|f_n| < g$ for some integrable function g. Then $f = \lim_n f_n$ is integrable and $\mu f = \lim_n \mu f_n$.*

Two measurables functions f, g are called *equivalent* with respect to μ if $\mu(f - g) = 0$.
For two-dimensional integration, the following theorem holds.

Proposition B.7. *Suppose that function $f(\cdot, \cdot)$ is integrable over the space $(X \times Y, \mathcal{A} \times \mathcal{B}, \mu \times \nu)$. Then for μ-almost all x, the function $f(x, \cdot)$ is integrable over $(Y, \mathcal{B}, \nu)$, and the function $x \mapsto \nu^y f(x, y)$ is integrable over $(X, \mathcal{A}, \mu)$ with $(\mu \times \nu) f = \mu^x \mu^y f$.*

B.4. Density. Let μ, ν be two measures over the same measurable space. We say that ν is *absolutely continuous* with respect to μ, or that μ *dominates* ν, if for each set A, $\mu(A) = 0$ implies $\nu(A) = 0$. It can be proved that this condition is equivalent to the condition that there is a positive real number c with $\nu \leqslant c\mu$. Every nonnegative integrable function f defines a new measure ν via the formula $\nu(A) = \mu(f \cdot 1_A)$. This measure ν is absolutely continuous with respect to μ. The Radon-Nikodym theorem says that the converse is also true.

Proposition B.8 (Radon-Nikodym theorem)**.** *If ν is dominated by μ then there is a nonnegative integrable function f such that $\nu(A) = \mu(f \cdot 1_A)$ for all measurable sets A. The function f is defined uniquely to within equivalence with respect to μ.*

The function f of the Radom-Nikodym Theorem above is called the *density* of ν with respect to μ. We will denote it by

$$f(x) = \frac{\mu(dx)}{\nu(dx)} = \frac{d\mu}{d\nu}.$$

The following theorem is also standard.

Proposition B.9.
(a) Let μ, ν, η be measures such that η is absolutely continuous with respect to μ and μ is absolutely continuous with respect to ν. Then the "chain rule" holds:

$$\frac{d\eta}{d\nu} = \frac{d\eta}{d\mu} \frac{d\mu}{d\nu}. \tag{B.1}$$

(b) If $\frac{\nu(dx)}{\mu(dx)} > 0$ for all x then μ is also absolutely continuous with respect to ν and $\frac{\mu(dx)}{\nu(dx)} = \left(\frac{\nu(dx)}{\mu(dx)} \right)^{-1}$.

Let μ, ν be two measures, then both are dominated by some measure η (for example by $\eta = \mu + \nu$). Let their densities with respect to η be f and g. Then we define the *total variation distance* of the two measures as

$$D(\mu, \nu) = \eta(|f - g|).$$

It is independent of the dominating measure η.

Example B.10. Suppose that the space X can be partitioned into disjoint sets A, B such that $\nu(A) = \mu(B) = 0$. Then $D(\mu, \nu) = \mu(A) + \nu(B) = \mu(X) + \nu(X)$.   ◇

B.5. **Random transitions.** Let $(X, \mathcal{A})$, $(Y, \mathcal{B})$ be two measureable spaces (defined in Subsection B.2). We follow the definition given in [16]. Suppose that a family of probability measures $\Lambda = \{ \lambda_x : x \in X \}$ on $\mathcal{B}$ is given. We call it a *probability kernel*, (or Markov kernel, or conditional distribution) if the map $x \mapsto \lambda_x B$ is measurable for each $B \in \mathcal{B}$. When X, Y are finite sets then λ is a Markov transition matrix. The following theorem shows that λ assigns a joint distribution over the space $(X \times Y, \mathcal{A} \times \mathcal{B})$ to each input distribution μ.

Proposition B.11. *For each nonnegative $\mathcal{A} \times \mathcal{B}$-measureable function f over $X \times Y$,*

1. *the function $y \to f(x, y)$ is $\mathcal{B}$-measurable for each fixed x;*
2. *$x \to \lambda_x^y f(x, y)$ is $\mathcal{A}$-measurable;*
3. *the integral $f \to \mu^x \lambda_x^y f(x, y)$ defines a measure on $\mathcal{A} \times \mathcal{B}$.*

According to this proposition, given a probability kernel Λ, to each measure μ over $\mathcal{A}$ corresponds a measure over $\mathcal{A} \times \mathcal{B}$. We will denote its marginal over $\mathcal{B}$ as

$$\Lambda^* \mu. \tag{B.2}$$

For every measurable function $g(y)$ over Y, we can define the measurable function $f(x) = \lambda_x g = \lambda_x^y g(y)$: we write

$$f = \Lambda g. \tag{B.3}$$

The operator Λ is linear, and monotone with $\Lambda 1 = 1$. By these definitions, we have

$$\mu(\Lambda g) = (\Lambda^* \mu) g. \tag{B.4}$$

Example B.12. Let $h : X \to Y$ be a measureable function, and let λ_x be the measure $\delta_{h(x)}$ concentrated on the point $h(x)$. This operator, denoted Λ_h is, in fact, a deterministic transition, and we have $\Lambda_h g = g \circ h$. In this case, we will simplify the notation as follows:

$$h^* \mu = \Lambda_h^*.$$

◇

B.6. **Probability measures over a metric space.** We follow the exposition of [2]. Whenever we deal with probability measures on a metric space, we will assume that our metric space is complete and separable (Polish). Let $\mathbf{X} = (X, d)$ be a complete separable metric space. It gives rise to a measurable space, where the measurable sets are the Borel sets of $\mathbf{X}$. It can be shown that, if A is a Borel set and μ is a finite measure then there are sets $F \subseteq A \subseteq G$ where F is an F_σ set, G is a G_δ set, and $\mu(F) = \mu(G)$. Let $\mathcal{B}$ be a base of open sets closed under intersections. Then it can be shown that μ is determined by its values on elements of $\mathcal{B}$. The following proposition follows then essentially from Proposition B.2.

Proposition B.13. *Let $\mathcal{B}^*$ be the set algebra generated by the above base $\mathcal{B}$, and let μ be any σ-additive set function on $\mathcal{B}^*$ with $\mu(X) = 1$. Then μ can be extended uniquely to a probability measure.*

We say that a set A is a *continuity set* of measure μ if $\mu(\partial A) = 0$: the boundary of A has measure 0.

B.6.1. *Weak topology.* Let

$$\mathcal{M}(\mathbf{X})$$

be the set of probability measures on the metric space $\mathbf{X}$. Let

$$\delta_x$$

be a probability measure concentrated on point x. Let x_n be a sequence of points converging to point x but with $x_n \neq x$. We would like to say that δ_{x_n} converges to δ_x. But the total variation distance $D(\delta_{x_n}, \delta_x)$ is 2 for all n. This suggests that the total variation distance is not generally the best way to compare probability measures over a metric space. We say that a sequence of probability measures μ_n over a metric space (X, d) *weakly converges* to measure μ if for all bounded continuous real functions f over X we have $\mu_n f \to \mu f$. This *topology of weak convergence* $(\mathcal{M}, \tau_w)$ can be defined using a number of different subbases. The one used in the original definition is the subbase consisting of all sets of the form

$$A_{f,c} = \{\, \mu : \mu f < c \,\}$$

for bounded continuous functions f and real numbers c. We also get a subbase (see for example [16]) if we restrict ourselves to the set $\mathrm{Lip}(X)$ of Lipschitz functions defined in (A.2). Another possible subbase giving rise to the same topology consists of all sets of the form

$$B_{G,c} = \{\, \mu : \mu(G) > c \,\} \tag{B.5}$$

for open sets G and real numbers c. Let us find some countable subbases. Since the space $\mathbf{X}$ is separable, there is a sequence $U_1, U_2, \ldots$ of open sets that forms a base. We can restrict the subbase of the space of measures to those sets $B_{G,c}$ where G is the union of a finite number of base elements U_i and c is rational. Thus, the space $(\mathcal{M}, \tau_w)$ itself has the second countability property.

It is more convenient to define a countable subbase using bounded continuous functions f, since $\mu \mapsto \mu f$ is continuous on such functions, while $\mu \mapsto \mu U$ is typically not continuous when U is an open set. Let $\mathcal{F}_0$ be the set of functions introduced before (2.1). Let

$$\mathcal{F}_1$$

be the set of functions f with the property that f is the minimum of a finite number of elements of $\mathcal{F}_0$. Note that each element f of $\mathcal{F}_1$ is bounded between 0 and 1, and from its definition, we can compute a bound β such that $f \in \mathrm{Lip}_\beta$.

Proposition B.14. *The following conditions are equivalent:*

1. *μ_n weakly converges to μ.*
2. *$\mu_n f \to \mu f$ for all $f \in \mathcal{F}_1$.*
3. *For every Borel set A, that is a continuity set of μ, we have $\mu_n(A) \to \mu(A)$.*
4. *For every closed set F, $\liminf_n \mu_n(F) \geqslant \mu(F)$.*
5. *For every open set G, $\limsup_n \mu_n(G) \leqslant \mu(G)$.*

As a subbase

$$\sigma_{\mathcal{M}} \tag{B.6}$$

for $\mathcal{M}(x)$, we choose the sets $\{\, \mu : \mu f < r \,\}$ and $\{\, \mu : \mu f > r \,\}$ for all $f \in \mathcal{F}_1$ and $r \in \mathbb{Q}$. Let $\mathcal{E}$ be the set of functions introduced in (2.1). It is a Riesz space as defined in Subsection B.3. A reasoning combining Propositions B.2 and B.5 gives the following.

Proposition B.15. *Suppose that a positive linear functional μ with $\mu 1 = 1$ is defined on $\mathcal{E}$ that is continuous with respect to monotone convergence. Then μ can be extended uniquely to a probability measure over $\mathbf{X}$ with $\mu f = \int f(x)\mu(dx)$ for all $f \in \mathcal{E}$.*

B.6.2. *Prokhorov distance.* The definition of measures in the style of Proposition B.15 is not sufficiently constructive. Consider a gradual definition of the measure μ, extending it to more and more elements of $\mathcal{E}$, while keeping the positivity and linearity property. It can happen that the function μ we end up with in the limit, is not continuous with respect to monotone convergence. Let us therefore metrize the space of measures: then an arbitrary measure can be defined as the limit of a Cauchy sequence of simple meaures.

One metric that generates the topology of weak convergence is the *Prokhorov distance* $p(\mu, \nu)$: the infimum of all those ε for which, for all Borel sets A we have (using the notation (A.1))

$$\mu(A) \leqslant \nu(A^\varepsilon) + \varepsilon.$$

It can be shown that this is a distance and it generates the weak topology. The following result helps visualize this distance:

Proposition B.16 (Coupling Theorem, see [21]). *Let μ, ν be two probability measures over a complete separable metric space $\mathbf{X}$ with $p(\mu, \nu) \leqslant \varepsilon$. Then there is a probability measure P on the space $\mathbf{X} \times \mathbf{X}$ with marginals μ and ν such that for a pair of random variables (ξ, η) having joint distribution P we have*

$$P\{\, d(\xi, \eta) > \varepsilon \,\} \leqslant \varepsilon.$$

Since this topology has the second countability property, the metric space defined by the distance $p(\cdot, \cdot)$ is separable. This can also be seen directly. Let S be a countable everywhere dense set of points in X. Consider the set of $\mathcal{M}_0(X)$ of those probability measures that are concentrated on finitely many points of S and assign rational values to them. It can be shown that $\mathcal{M}_0(X)$ is everywhere dense in the metric space $(\mathcal{M}(X), p)$; so, this space is separable. It can also be shown that $(\mathcal{M}(X), p)$ is complete. Thus, a measure can be given as the limit of a sequence of elements $\mu_1, \mu_2, \ldots$ of $\mathcal{M}_0(X)$, where $p(\mu_i, \mu_{i+1}) < 2^{-i}$.

The definition of the Prokhorov distance quantifies over all Borel sets. However, in an important simple case, it can be handled efficiently.

Proposition B.17. *Assume that measure ν is concentrated on a finite set of points $S \subset X$. Then the condition $p(\nu, \mu) < \varepsilon$ is equivalent to the finite set of conditions*

$$\mu(A^\varepsilon) > \nu(A) - \varepsilon \tag{B.7}$$

for all $A \subset S$.

B.6.3. *Relative compactness.* A set Π of measures in $(\mathcal{M}(X), p)$ is called *relatively compact* if every sequence of elements of Π contains a convergent subsequence. Relative compactness is an important property for proving convergence of measures. It has a useful characterization. A set of Π of measures is called *tight* if for every ε there is a compact set K such that $\mu(K) > 1 - \varepsilon$ for all μ in Π. Prokhorov's theorem states (under our assumptions of the separability and completeness of (X, d)) that a set of measures is relatively compact if and only if it is tight and if and only if its closure is compact in $(\mathcal{M}(X), p)$. In particular, the following fact is known.

Proposition B.18. *The space $(\mathcal{M}(\mathbf{X}), p)$ of measures is compact if and only if the space (X, d) is compact.*

So, if (X, d) is not compact then the set of measures is not compact. But still, each measure μ is "almost" concentrated on a compact set. Indeed, the one-element set $\{\mu\}$ is compact and therefore by Prokhorov's theorem tight. Tightness says that for each ε a mass of size $1 - \varepsilon$ of μ is concentrated on some compact set.

APPENDIX C. COMPUTABLE ANALYSIS

If for some finite or infinite sequences x, y, z, w, we have $z = wxy$ then we write $w \sqsubseteq z$ (w is a *prefix* of z) and $x \triangleleft z$. For integers, we will use the toupling functions

$$\langle i, j \rangle = \frac{1}{2}(i+1)(i+j+1) + j, \quad \langle n_1, \ldots, n_{k+1} \rangle = \langle \langle n_1, \ldots, n_k \rangle, n_{k+1} \rangle.$$

Inverses: $\pi_i^k(n)$.

Unless said otherwise, the alphabet Σ is always assumed to contain the symbols 0 and 1. After [24], let us define the *wrapping function* $\iota : \Sigma^* \to \Sigma^*$ by

$$\iota(a_1 a_2 \cdots a_n) = 110a_1 0a_2 0 \cdots a_n 011. \tag{C.1}$$

Note that

$$|\iota(x)| = (2|x| + 5) \vee 6. \tag{C.2}$$

For strings $x, x_i \in \Sigma^*$, $p, p_i \in \Sigma^\omega$, $k \geqslant 1$, appropriate tupling functions $\langle x_1, \ldots, x_k \rangle$, $\langle x, p \rangle$, $\langle p, x \rangle$, etc. can be defined with the help of $\langle \cdot, \cdot \rangle$ and $\iota(\cdot)$.

C.1. **Notation and representation.** The concepts of notation and representation, as defined in [24], allow us to transfer computability properties from some standard spaces to many others. Given a countable set C, a *notation* of C is a surjective partial mapping $\delta :\subseteq \mathbb{N} \to C$. Given some finite alphabet $\Sigma \supseteq \{0, 1\}$ and an arbitrary set S, a *representation* of S is a surjective mapping $\chi :\subseteq \Sigma^\omega \to S$. A *naming system* is a notation or a representation. Here are some standard naming systems:

1. id, the identity over Σ^* or Σ^ω.
2. $\nu_{\mathbb{N}}, \nu_{\mathbb{Z}}, \nu_{\mathbb{Q}}$ for the set of natural numbers, integers and rational numbers.
3. $\mathrm{Cf} : \Sigma^\omega \to 2^{\mathbb{N}}$, the *characteristic function representation* of sets of natural numbers, is defined by $\mathrm{Cf}(p) = \{\, i : p(i) = 1 \,\}$.
4. $\mathrm{En} : \Sigma^\omega \to 2^{\mathbb{N}}$, the *enumeration representation* of sets of natural numbers, is defined by $\mathrm{En}(p) = \{\, w \in \Sigma^* : 110^{n+1}11 \triangleleft p \,\}$.
5. For $\Delta \subseteq \Sigma$, $\mathrm{En}_\Delta : \Sigma^\omega \to 2^{\Delta^*}$, the *enumeration representation* of subsets of Δ^*, is defined by $\mathrm{En}_\Delta(p) = \{\, w \in \Sigma^* : \iota(w) \triangleleft p \,\}$.

One can define names for all computable functions between spaces that are Cartesian products of terms of the kind Σ^* and Σ^ω. Then, the notion of computability can be transferred to other spaces as follows. Let $\delta_i : Y_i \to X_i$, $i = 1, 0$ be naming systems of the spaces X_i. Let $f :\subseteq X_1 \to X_0$, $g :\subseteq Y_1 \to Y_0$. We say that function g *realizes* function f if

$$f(\delta_1(y)) = \delta_0(g(y)) \tag{C.3}$$

holds for all y for which the left-hand side is defined. Realization of multi-argument functions is defined similarly. We say that a function $f : X_1 \times X_2 \to X_0$ is $(\delta_1, \delta_2, \delta_0)$-*computable* if there is a computable function $g :\subseteq Y_1 \times Y_2 \to Y_0$ realizing it. In this case, a name for f is naturally derived from a name of g.[2]

For representations ξ, η, we write $\xi \leqslant \eta$ if there is a computable function $f :\subseteq \Sigma^\omega \to \Sigma^\omega$ with $\xi(x) = \eta(f(x))$. In words, we say that ξ is *reducible* to η, or that f reduces (translates) ξ to η. There is a similar definition of reduction for notations. We write $\xi \equiv \eta$ if $\xi \leqslant \eta$ and $\eta \leqslant \xi$.

C.2. **Constructive topological space.**

---

[2]Any function g realizing f via (C.3) automatically has a certain *extensivity* property: if $\delta_1(y) = \delta_1(y')$ then $g(y) = g(y')$.

C.2.1. *Definitions.* Section A gives a review of topological concepts. A *constructive topological space* $\mathbf{X} = (X, \sigma, \nu)$ is a topological space over a set X with a subbase σ effectively given as a list $\sigma = \{\nu(1), \nu(2), \dots\}$, and having the T_0 property (thus, every point is determined uniquely by the subset of elements of σ containing it). By definition, a constructive topological space satisfies the second countability axiom.[3] We obtain a base

$$\sigma^\cap$$

of the space $\mathbf{X}$ by taking all possible finite intersections of elements of σ. It is easy to produce an effective enumeration for $\sigma^\cap$ from ν. We will denote this enumeration by $\nu^\cap$.

The *product operation* is defined over constructive topological spaces in the natural way.

Examples C.1.

1. A discrete topological space, where the underlying set is finite or countably infinite, with a fixed enumeration.
2. The real line, choosing the base to be the open intervals with rational endpoints with their natural enumeration. Product spaces can be formed to give the Euclidean plane a constructive topology.
3. The real line $\mathbb{R}$, with the subbase $\sigma_\mathbb{R}^>$ defined as the set of all open intervals $(-\infty; b)$ with rational endpoints b. The subbase $\sigma_\mathbb{R}^<$, defined similarly, leads to another topology. These two topologies differ from each other and from the usual one on the real line, and they are not Hausdorff spaces.
4. Let X be a set with a constructive discrete topology, and X^ω the set of infinite sequences with elements from X, with the product topology: a natural enumerated basis is also easy to define.

$$\diamondsuit$$

Due to the T_0 property, every point in our space is determined uniquely by the set of open sets containing it. Thus, there is a representation $\gamma_\mathbf{X}$ of $\mathbf{X}$ defined as follows. We say that $\gamma_\mathbf{X}(p) = x$ if $\text{En}_\Sigma(p) = \{\, w : x \in \nu(w) \,\}$. If $\gamma_\mathbf{X}(p) = x$ then we say that the infinite sequence p is a *complete name* of x: it encodes all names of all subbase elements containing x. From now on, we will call $\gamma_\mathbf{X}$ the *complete standard representation of the space* $\mathbf{X}$.[4]

C.2.2. *Constructive open sets, computable functions.* In a constructive topological space $\mathbf{X} = (X, \sigma, \nu)$, a set $G \subseteq X$ is called *r.e. open* in set B if there is a r.e. set E with $G = \bigcup_{w \in E} \nu^\cap(w) \cap B$. It is r.e. open if it is r.e. open in X. In the special kind of spaces in which randomness has been developed until now, constructive open sets have a nice characterization:

Proposition C.2. *Assume that the space* $\mathbf{X} = (X, \sigma, \nu)$ *has the form* $Y_1 \times \cdots \times Y_n$ *where each* Y_i *is either* Σ^* *or* Σ^ω. *Then a set* G *is r.e. open iff it is open and the set* $\{\, (w_1, \dots, w_n) : \bigcap_i \nu(w_i) \subset G \,\}$ *is recursively enumerable.*

Proof. The proof is not difficult, but it relies on the discrete nature of the space Σ^* and on the fact that the space Σ^ω is compact and its base consists of sets that are open and closed at the same time. □

---

[3]A constructive topological space is an effective topological space as defined in [24], but, for simplicity we require the notation ν to be a total function.

[4] The book [24] denotes $\gamma_\mathbf{X}$ as $\delta'_\mathbf{X}$ instead. We use $\gamma_\mathbf{X}$ only, dispensing with the notion of a "computable" topological space.

It is easy to see that if two sets are r.e. open then so is their union. The above remark implies that a space having the form $Y_1 \times \cdots \times Y_n$ where each Y_i is either Σ^* or Σ^ω, also the intersection of two recursively open sets is recursively open. We will see that this statement holds, more generally, in all computable metric spaces.

Let $\mathbf{X}_i = (X_i, \sigma_i, \nu_i)$ be constructive topological spaces, and let $f :\subseteq X_1 \to X_0$ be a function. As we know, f is continuous iff the inverse image $f^{-1}(G)$ of each open set G is open. Computability is an effective version of continuity: it requires that the inverse image of subbase elements is uniformly constructively open. More precisely, $f :\subseteq X_1 \to X_0$ is *computable* if the set

$$\bigcup_{V \in \sigma_0^\cap} f^{-1}(V) \times \{V\}$$

is a r.e. open subset of $X_1 \times \sigma_0^\cap$. Here the base $\sigma_0^\cap$ of $\mathbf{X}_0$ is treated as a discrete constructive topological space, with its natural enumeration. This definition depends on the enumerations ν_1, ν_0. The following theorem (taken from [24]) shows that this computability coincides with the one obtained by transfer via the representations $\gamma_{\mathbf{X}_i}$.

Proposition C.3. *For $i = 0, 1$, let $\mathbf{X}_i = (X_i, \sigma_i, \nu_i)$ be constructive topological spaces. Then a function $f :\subseteq X_1 \to X_0$ is computable iff it is $(\gamma_{\mathbf{X}_1}, \gamma_{\mathbf{X}_0})$-computable for the representations $\gamma_{\mathbf{X}_i}$ defined above.*

As a name of a computable function, we can use the name of the enumeration algorithm derived from the definition of computability, or the name derivable using this representation theorem.

Remark C.4. As in Proposition C.2, it would be nice to have the following statement, at least for total functions: "Function $f : X_1 \to X_0$ is computable iff the set

$$\{ (v, w) : \nu_1^\cap(w) \subset f^{-1}[\nu_0(v)] \}$$

is recursively enumerable." But such a characterization seems to require compactness and possibly more. $\diamond$

Let us call two spaces X_1 and X_0 *effectively homeomorphic* if there are computable maps between them that are inverses of each other. In the special case when $X_0 = X_1$, we say that the enumerations of subbases ν_0, ν_1 are *equivalent* if the identity mapping is a effective homeomorphism. This means that there are recursively enumerable sets F, G such that

$$\nu_1(v) = \bigcup_{(v,w) \in F} \nu_0^\cap(w) \text{ for all } v, \quad \nu_0(w) = \bigcup_{(w,v) \in G} \nu_1^\cap(v) \text{ for all } w.$$

Lower semicomputability is a constructive version of lower semicontinuity. Let $\mathbf{X} = (X, \sigma, \nu)$ be a constructive topological space. A function $f :\subseteq X \to \overline{\mathbb{R}}_+$ is called *lower semicomputable* if the set $\{ (x, r) : f(x) > r \}$ is r.e. open. Let $\mathbf{Y} = (\overline{\mathbb{R}}_+, \sigma_{\mathbb{R}}^<, \nu_{\mathbb{R}}^<)$ be the effective topological space introduced in Example C.1.2, in which $\nu_{\mathbb{R}}^>$ is an enumeration of all open intervals of the form $(r; \infty]$ with rational r. It can be seen that f is lower semicomputable iff it is $(\nu, \nu_{\mathbb{R}}^>)$-computable.

C.2.3. *Computable elements and sequences.* Let $\mathbf{U} = (\{0\}, \sigma_0, \nu_0)$ be the one-element space turned into a trivial constructive topological space, and let $\mathbf{X} = (X, \sigma, \nu)$ be another constructive topological space. We say that an element $x \in X$ is *computable* if the function $0 \mapsto x$ is computable. It is easy to see that this is equivalent to the requirement that the set $\{ u : x \in \nu(u) \}$ is recursively enumerable. Let $\mathbf{X}_j = (X_j, \sigma_j, \nu_j)$, for $i = 0, 1$ be constructive

topological spaces. A sequence f_i, $i = 1, 2, \ldots$ of functions with $f_i : X_1 \to X_0$ is a *computable sequence of computable functions* if $(i, x) \mapsto f_i(x)$ is a computable function. Using the s-m-n theorem of recursion theory, it is easy to see that this statement is equivalent to the statement that there is a recursive function computing from each i a name for the computable function f_i. The proof of the following statement is not difficult.

Proposition C.5. *Let $\mathbf{X}_i = (X_i, \sigma_i, \nu_i)$ for $i = 1, 2, 0$ be constructive topological spaces, and let $f : X_1 \times X_2 \to X_0$, and assume that $x_1 \in X_1$ is a computable element.*

1. *If f is computable and then $x_2 \mapsto f(x_1, x_2)$ is also computable.*
2. *If $\mathbf{X}_0 = \overline{\mathbb{R}}$, and f is lower semicomputable then $x_2 \mapsto f(x_1, x_2)$ is also lower semicomputable.*

C.3. **Computable metric space.** Following [4], we define a computable metric space as a tuple $\mathbf{X} = (X, d, D, \alpha)$ where (X, d) is a metric space, with a countable dense subset D and an enumeration α of D. It is assumed that the real function $d(\alpha(v), \alpha(w))$ is computable. As x runs through elements of D and r through positive rational numbers, we obtain the enumeration of a countable basis $\{\, B(x, r) : x \in D, r \in \mathbb{Q} \,\}$ (of balls or radius r and center x) of $\mathbf{X}$, giving rise to a constructive topological space $\tilde{\mathbf{X}}$. Let us call a sequence $x_1, x_2, \ldots$ a *Cauchy* sequence if for all $i < j$ we have $d(x_i, x_j) \leqslant 2^{-i}$. To connect to the type-2 theory of computability developed above, the *Cauchy-representation* $\delta_{\mathbf{X}}$ of the space can be defined in a natural way. It can be shown that as a representation of $\tilde{\mathbf{X}}$, it is equivalent to $\gamma_{\tilde{\mathbf{X}}}$: $\delta_{\mathbf{X}} \equiv \gamma_{\tilde{\mathbf{X}}}$.

Example C.6. Example A.5 is a computable metric space, with either of the two (equivalent) choices for an enumerated dense set. $\diamondsuit$

Similarly to the definition of a computable sequence of computable functions in C.2.3, we can define the notion of a computable sequence of bounded computable functions, or the computable sequence f_i of computable Lipschitz functions: the bound and the Lipschitz constant of f_i are required to be computable from i. The following statement shows, in an effective form, that a function is lower semicomputable if and only if it is the supremum of a computable sequence of computable functions.

Proposition C.7. *Let $\mathbf{X}$ be a computable metric space. There is a computable mapping that to each name of a nonnegative lower semicomputable function f assigns a name of a computable sequence of computable bounded Lipschitz functions f_i whose supremum is f.*

Proof sketch. Show that f is the supremum of a computable sequence of computable functions $c_i 1_{B(u_i, r_i)}$ where $u_i \in D$ and $c_i, r_i > 0$ are rational. Clearly, each indicator function $1_{B(u_i, r_i)}$ is the supremum of a computable sequence of computable functions $g_{i,j}$. We have $f = \sup_n f_n$ where $f_n = \max_{i \leqslant n} c_i g_{i,n}$. It is easy to see that the bounds on the functions f_n are computable from n and that they all are in Lip_{β_n} for a β_n that is computable from n. $\square$

The following is also worth noting.

Proposition C.8. *In a computable metric space, the intersection of two r.e. open sets is r.e. open.*

Proof. Let $\beta = \{\, B(x, r) : x \in D, r \in \mathbb{Q} \,\}$ be a basis of our space. For a pair (x, r) with $x \in D$, $r \in \mathbb{Q}$, let

$$\Gamma(x, r) = \{\, (y, s) : y \in D, \ s \in \mathbb{Q}, \ d(x, y) + s < r \,\}.$$

If U is a r.e. open set, then there is a r.e. set $S_U \subset D \times \mathbb{Q}$ with $U = \bigcup_{(x,r) \in S_U} B(x,r)$. Let $S'_U = \bigcup \{ \Gamma(x,r) : (x,r) \in S_U \}$, then we have $U = \bigcup_{(x,r) \in S'_U} B(x,r)$. Now, it is easy to see

$$U \cap V = \bigcup_{(x,r) \in S'_U \cap S'_V} B(x,r).$$

$\square$

References

[1] Yevgeniy A. Asarin, *Individual random signals: a complexity approach*, Ph.D. thesis, Moscow State University, Moscow, Russia, 1988, In Russian. 1.1, 1.3

[2] Patrick Billingsley, *Convergence of probability measures*, Wiley, 1968, First edition. Second edition 1999. A.2, A.6, B.6

[3] Vasco Brattka, *Computability over topological structures*, Computability and Models (S. Barry Cooper and Sergey S. Goncharov, eds.), Kluwer Academic Publishers, New York, 2003, pp. 93–136. 1.4

[4] Vasco Brattka and Gero Presser, *Computability on subsets of metric spaces*, Theoretical Computer Science **305** (2003), 43–76. C.3

[5] Gregory J. Chaitin, *A theory of program size formally identical to information theory*, J. Assoc. Comput. Mach. **22** (1975), 329–340. 1.3, 4.1.2

[6] Peter Gács, *On the symmetry of algorithmic information*, Soviet Math. Dokl. **15** (1974), 1477–1780. 1.3

[7] _____, *Exact expressions for some randomness tests*, Z. Math. Log. Grdl. M. **26** (1980), 385–394, Short version: Springer Lecture Notes in Computer Science 67 (1979) 124-131. 1.3, 4.2

[8] _____, *On the relation between descriptional complexity and algorithmic probability*, Theoretical Computer Science **22** (1983), 71–93, Short version: Proc. 22nd IEEE FOCS (1981) 296-303. 1.3, 4.4

[9] _____, *The Boltzmann entropy and randomness tests*, Proceedings of the Workshop on Physics and Computation, IEEE Computer Society Press, 1994, Extended abstract., pp. 209–216. 1.3, 4.9

[10] Peter Hertling and Klaus Weihrauch, *Randomness spaces*, Proc. of ICALP'98, Lecture Notes in Computer Science, vol. 1443, Springer, 1998, pp. 796–807. 1.1, 1.3, 1.4, 3.2

[11] Leonid A. Levin, *On the notion of a random sequence*, Soviet Math. Dokl. **14** (1973), no. 5, 1413–1416. 1.1, 1.2, 1.3, 4.4

[12] _____, *Uniform tests of randomness*, Soviet Math. Dokl. **17** (1976), no. 2, 337–340. 1.1, 2, 1.3, 5.5, 7

[13] _____, *Randomness conservation inequalities: Information and independence in mathematical theories*, Information and Control **61** (1984), no. 1, 15–37. 1.1, 1, 2, 1.3, 5.5, 7

[14] Ming Li and Paul M. B. Vitányi, *Introduction to Kolmogorov complexity and its applications (second edition)*, Springer Verlag, New York, 1997. 1.3, 4.1.1, 4.1, 6.2

[15] Per Martin-Löf, *The definition of random sequences*, Information and Control **9** (1966), 602–619. 1.1, 1.2, 1.3

[16] David Pollard, *A user's guide to measure-theoretic probability*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, U.K., 2001. 1.4, B, B.5, B.6.1

[17] Claus Peter Schnorr, *Process complexity and effective random tests*, J. Comput. Syst. Sci **7** (1973), 376. 1.3, 4.4

[18] Raymond J. Solomonoff, *A formal theory of inductive inference i*, Information and Control **7** (1964), 1–22. 5.7

[19] _____, *Complexity-based induction systems: Comparisons and convergence theorems*, IEEE Transactions on Information Theory **IT-24** (1978), no. 4, 422–432. 5.7

[20] Edwin H. Spanier, *Algebraic topology*, Mc Graw-Hill, New York, 1971. 5.1

[21] Volker Strassen, *The existence of probability measures with given marginals*, Annals of Mathematical Statistics **36** (1965), 423–439. B.16

[22] Flemming Topsoe, *Topology and measure*, Lecture Notes in Mathematics, vol. 133, Springer Verlag, Berlin, 1970. 4.1.1

[23] Volodimir G. Vovk and V. V. Vyugin, *On the empirical validity of the Bayesian method*, Journal of the Royal Statistical Society B **55** (1993), no. 1, 253–266. 6.2

[24] Klaus Weihrauch, *Computable analysis*, Springer, 2000. 1.2, 1.4, C, C.1, 3, 4, C.2.2

[25] Woitiech H. Zurek, *Algorithmic randomness and physical entropy*, Physical Review A **40** (1989), no. 8, 4731–4751. 4.9

[26] Alexander K. Zvonkin and Leonid A. Levin, *The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms*, Russian Math. Surveys **25** (1970), no. 6, 83–124. 4.1.2, 4.4

BOSTON UNIVERSITY
E-mail address: gacs@bu.edu