

Layerwise Computability and Image Randomness

Laurent Bienvenu, Mathieu Hoyrup, Alexander Shen

► **To cite this version:**

Laurent Bienvenu, Mathieu Hoyrup, Alexander Shen. Layerwise Computability and Image Randomness. Theory of Computing Systems, Springer Verlag, 2017, 61 (4), pp.1353-1375. 10.1007/s00224-017-9791-8 . hal-01650910

HAL Id: hal-01650910

<https://hal.inria.fr/hal-01650910>

Submitted on 20 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Layerwise computability and image randomness

Laurent Bienvenu* Mathieu Hoyrup† Alexander Shen‡

Abstract

Algorithmic randomness theory starts with a notion of an individual random object. To be reasonable, this notion should have some natural properties; in particular, an object should be random with respect to image distribution if and only if it has a random preimage. This result (for computable distributions and mappings, and Martin-Löf randomness) was known for a long time (folklore); in this paper we prove its natural generalization for layerwise computable mappings, and discuss the related quantitative results.

1 Introduction

Consider some random process, like coin tossing, that generates an infinite sequence of bits (zeros and ones). Assume that we have some theoretical model that describes this process. Such a model is some probability distribution P on the space of possible outcomes; for example, fair coin tossing is modelled by the uniform distribution on the space of infinite bit sequences, also known as Cantor space. Now suppose that we get, in addition to this theoretical model, some experimental data, a sequence ω . The natural question arises: is this data consistent with the model? Is it plausible that a random process with output distribution P produced ω , or the conjecture about probability distribution should be rejected? Sometimes it should — for example, if more than 90% bits in any prefix of ω are zeros, the fair coin assumption does not look acceptable. The same is true if the sequence

*Laboratoire d’Informatique, de Robotique et de Microélectronique de Montpellier, CNRS, Université de Montpellier, laurent.bienvenu@computability.fr. Supported by ANR RaCAF grant.

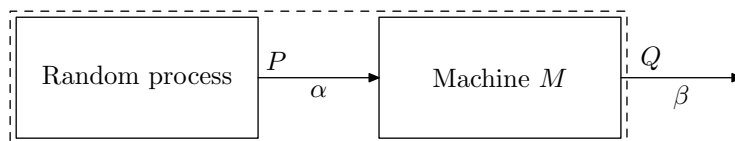
†Laboratoire Lorraine de Recherche en Informatique et ses applications, mathieu.hoyrup@inria.fr

‡Visiting researcher, National Research University Higher School of Economics (HSE), Faculty of Computer Science, Kochnovskiy Proezd 3, Moscow, 125319, Russia; Laboratoire d’Informatique, de Robotique et de Microélectronique de Montpellier, CNRS, Université de Montpellier, alexander.shen@lirmm.fr. Supported by ANR RaCAF grant.

ω turns out to be a binary representation of π . On the other hand, some ω should be considered as plausible outputs (otherwise the conjecture is always rejected). But where is the boundary line?

Classical probability theory does not provide any answer to this question. This is what *algorithmic randomness theory* is about. This theory defines a notion of a random sequence with respect to a probability distribution. More precisely, for a probability distribution P on the Cantor space, it defines a subset of the Cantor space whose elements are called *random sequences with respect to P* , or *P -random sequences*. In fact, there are several definitions; the most basic and popular one is called *Martin-Löf randomness* and was introduced in [6], but there are some others. Having such a notion, we can say that we reject the model (distribution) P if the experimental data, the sequence ω , is not random with respect to P . This interpretation, however, comes with some obligations: the notion of randomness should have some properties to make it reasonable. In this note we consider one of these properties: *image randomness*.

Imagine that we have a machine that consists of two parts: (1) a random process that generates some bit sequence α , and has distribution P , and (2) some algorithmic process M that gets α as input and produces some output sequence β . We may assume that the machine M has an input read-only tape where α is written, unlimited memory (say, infinite work tapes) and write-only one-directional output tape where the bits of β are written sequentially. Such a machine determines a mapping of the Cantor space into the space of finite and infinite binary sequences. When M is applied to a random input having distribution P , the output sequence is a random variable. The distribution of this random variable is what is usually called a *continuous semimeasure*, but we are mainly interested in the case when the output sequence is infinite with P -probability one. In this case the output distribution is some distribution Q on Cantor space (the image of P). Such a scheme is used, for example, when we want to emulate some distribution Q using a fair coin that produces independent uniformly distributed random bits. In this case P is the uniform distribution on the Cantor space.



Now imagine that we observe some sequence β as the output of such a composite random process. When does it look plausible? There are two possible answers. The first one: “when β is random with respect to Q ”, since Q is the output distribution of the composite process. The second one: “when there exists a plausible reconstruction of what

happened inside the black box”, i.e., when *there exists some α that is random with respect to P and is mapped to β by M* . To make the philosophical interpretation consistent, these two properties should be equivalent: for a satisfactory notion of algorithmic randomness the image of the set of algorithmically random with respect to P sequences should coincide with the set of algorithmically random with respect to Q sequences, where Q is the image of P under the transformation.

This property can be split into two parts. One part says that M -image of every P -random sequence is infinite and Q -random. This part can be called *randomness conservation*. The other part claims that every Q -random sequence is an M -image of some P -random sequence: *randomness cannot be created out of nothing*.

It turns out that this property holds for Martin-Löf’s definition of randomness and computable P . Note that Martin-Löf’s definition is applicable only to computable distributions; if P is computable and $M(\alpha)$ is infinite with P -probability 1, then Q is also computable, so Martin-Löf’s definition can be applied to both P and Q . This is a folklore result known for a long time; the third author remembers discussing it with Andrej Muchnik around 1987 (it seems that there was even a short note written by A.S. about that in the proceedings of the First Bernoulli congress in Uzbekistan, then USSR; most probably nobody read it, and we cannot locate this volume now). It was independently discovered by many people, sometimes for some special cases.

Here is one example where this property is useful. Let us prove that for every random sequence ω with respect to the uniform Bernoulli measure $B_{1/2}$ (independent trials, ones have probability $1/2$) there exist a sequence τ that is random with respect to $B_{1/3}$ (Bernoulli distribution where ones have probability $1/3$) that can be obtained from ω by changing some ones to zeros. Indeed, let us imagine a random process that first generates a sequence of random¹ independent real numbers $\vec{\xi} = \xi_1, \xi_2, \dots$ uniformly distributed in $[0, 1]$ and then converts them to bits using threshold $1/2$ (i.e., the i -th bit is 1 if $\xi_i < 1/2$ and 0 otherwise). The output distribution is $B_{1/2}$, so by the ‘no-randomness-from-nothing’ property any $B_{1/2}$ -random sequence ω is can be obtained from some random sequence $\vec{\xi}$ of reals. Now consider the same process, but with a threshold of $1/3$. The measure induced by this process is $B_{1/3}$, so by randomness conservation, we get a $B_{1/3}$ -random sequence, and this sequence is coordinate-wise smaller than the original one.

Recently a notion of *layerwise computable mapping* was introduced [5]; it provides a good level of generality for the image randomness property, and makes the proof nicely balanced. In the following section we reproduce the definition of layerwise computable

¹A sequence of reals can be encoded as a two-dimensional table of bits, if each real is replaced by its binary representation; then this two-dimensional table can be rearranged into a sequence of bits, and the corresponding distribution on the sequences is $B_{1/2}$. So we are still in the same setting with Cantor space and computable transformations, and can apply our property.

mapping and discuss its reformulation in terms of machines with rewritable output. Then (Section 3) we give the proof of image randomness property for layerwise computable mappings. Finally, in Section 4 we consider some extensions and version of the main result (including a quantitative version for randomness deficiencies, and a version for multi-valued mappings).

2 Layerwise computable mappings

Let P be a computable distribution on the Cantor space $\Omega = \{0, 1\}^{\mathbb{N}}$. Computability means that the function $p(x) = P(x\Omega)$, where $x\Omega$ is the set of all infinite sequences that have prefix x , is computable: given x and rational $\varepsilon > 0$, we can compute some rational ε -approximation to $p(x)$. Let us recall Martin-Löf's definition of randomness. A *Martin-Löf test* is a decreasing sequence

$$U_1 \supset U_2 \supset U_3 \supset \dots$$

of uniformly effectively open² sets such that $P(U_i) \leq 2^{-i}$. A sequence ω is *rejected* by this test if $\omega \in U_n$ for all n . Sequences that are not rejected by any Martin-Löf test are called *Martin-Löf random* with respect to the distribution P . We also call them *Martin-Löf P -random* or just *P -random*.

Consider some mapping F defined on the set of all P -random sequences; the values of F are infinite bit sequences. We say that F is *layerwise computable* if there exists a Martin-Löf test

$$U_1 \supset U_2 \supset U_3 \supset \dots$$

and a machine M that, given i and a sequence $\omega \notin U_i$ (on the input tape), computes $F(\omega)$, i.e., prints $F(\omega)$ bit by bit on the output tape. Note that every random sequence ω is outside U_i for large enough i , so $F(\omega)$ is computed by M for all sufficiently large i . In other terms, for a layerwise computable function F we do not have one algorithm that computes $F(\omega)$ for all ω , but a family of algorithms $M_i(\cdot) = M(i, \cdot)$ that compute F outside the sets U_i of decreasing measure. Note that the behavior of M_i on $\omega \in U_i$ is not specified: $M_i(\omega)$ may be an arbitrary finite or infinite sequence (and may differ from $F(\omega)$ even if $F(\omega)$ is defined).

The original definition of layerwise computable mappings used the notion of *randomness deficiency*. Here is a definition of this notion. Given a Martin-Löf test $U_1 \supset U_2 \supset \dots$, we define the corresponding deficiency function $d_U(\omega) = \sup\{i \mid \omega \in U_i\}$. The function d_U is infinite on the sequences rejected by U , so $d_U(\omega)$ is finite for all random sequences. The

²This means that there is an algorithm that, given n , enumerates a sequence x_0, x_1, x_2, \dots of strings such that U_n is the union of intervals $x_i\Omega$.

definition of layerwise computable mapping can be now reformulated as follows: *there is a machine that computes $F(\omega)$ given ω and some $i > d_U(\omega)$.*

Martin-Löf showed that there exists a *universal* test V for which the deficiency function d_V is maximal (up to an additive constant); this test, therefore, rejects *all* non-random sequences. Let us fix some universal test V . Since d_V is maximal, an upper bound for $d_V(\omega)$ is also (up to some constant) an upper bound for d_U for every test U . So we may say that F is layerwise computable if there is an algorithm that computes $F(\omega)$ given ω and an upper bound for $d_V(\omega)$. This is how the layerwise computable mappings were defined in [5].

To understand better the intuitive meaning of this definition, let us consider layerwise computable mappings with one-bit output (the value is 0 or 1). The definition is essentially the same: a mapping f is layerwise computable if there exists a test U_i and a machine M that, given a sequence ω and i such that $\omega \notin U_i$, computes $f(\omega)$. The difference is that now $f(\omega)$ is just one bit, not a bit sequence.

Let F be a mapping defined on random sequences whose values are infinite sequences, and let F_k are individual bits of this mapping, so

$$F(\omega) = F_0(\omega)F_1(\omega)F_2(\omega)\dots$$

The following simple observation says that we can define layerwise computability in a coordinate-wise way:

Proposition 1. *The mapping F is layerwise computable if and only if all the mappings F_k are uniformly layerwise computable.*

Proof. One direction is obvious: if F is layerwise computable, the same machine can be used for all k to compute F_k . The other direction requires some reasoning: layerwise computability of F_k gives us some test U^k , i.e., a decreasing sequence $U_1^k \supset U_2^k \supset U_3^k \supset \dots$ of uniformly effectively open sets, and some algorithm M^k that, given i , computes $F_k(\omega)$ for all $\omega \notin U_i^k$. We combine these tests U^k into one test U along the diagonals : let

$$U_1 = U_2^0 \cup U_3^1 \cup U_4^2 \cup \dots, \quad U_2 = U_3^0 \cup U_4^1 \cup U_5^2 \cup \dots, \text{ etc.}$$

Then we have

$$P(U_i) \leq P(U_{i+1}^0) + P(U_{i+2}^1) + P(U_{i+3}^2) + \dots \leq 2^{-(i+1)} + 2^{-(i+2)} + 2^{-(i+3)} + \dots \leq 2^{-i},$$

as required. We can compute all the bits of $F(\omega)$ for $\omega \notin U_i$, using upper bound $i + k + 1$ in the algorithm that layerwise-computes $F_k(\omega)$. \square

A layerwise computable mapping f with bit values splits all random P -sequences into two classes (preimage of 1 and preimage of 0), and therefore defines a point in a metric space of all measurable subsets X of Ω . This metric space has distance function $d(X, Y) = P(X \triangle Y)$; the points in this space are equivalence classes (two sets X, Y are equivalent if $d(X, Y) = 0$, i.e., their symmetric difference $X \triangle Y$ is a P -null set). In this space the family of clopen sets (in other terms, the family of sets that are finite unions of intervals) is a countable dense subset on which the distance function is computable in a natural sense (recall that P is a computable distribution). So this space can be considered as a constructive metric space, and computable points in it are defined as points for which we can compute a clopen approximation with any given precision. An equivalent definition of computable points: consider computable sequences of clopen sets C_0, C_1, C_2, \dots such that $d(C_i, C_{i+1}) \leq 2^{-i}$; their limits are computable points.

Proposition 2. *Each layerwise computable mapping f of Cantor space to $\{0, 1\}$ corresponds to a computable point in the metric space of measurable subsets of Ω . This is a one-to-one correspondence: different layerwise computable mappings correspond to different computable points, and each computable point is obtained from some layerwise computable mapping.*

Proof. Let f be a layerwise computable mapping and let $U_1 \supset U_2 \supset \dots$ be a corresponding test. Let M_i be a machine that computes f outside U_i . The domain of M_i (the set of sequences where M_i produces 0 or 1) is an effectively open set; simulating M_i on all inputs, we can enumerate intervals of this domain together with the corresponding values (0 or 1). Denote this domain by V_i ; by definition we know that $V_i \cup U_i$ is the entire Cantor space. Due to compactness, we can wait until finitely many intervals from U_i and V_i are found that cover (together) the entire space. Let C_i be the union of the intervals in V_i that are found up to this moment with value 1. Note that f is not guaranteed to be 1 on all the points in C_i , since some of them may belong to U_i and in this case the value of M_i may differ from the value of f . But we know that the symmetric difference of $f^{-1}(1)$ and C_i is contained in U_i , so it has P -measure at most 2^{-i} . Therefore, the point $f^{-1}(1)$ is computable. Note also that the sets C_i (their indicator functions) converge to f pointwise at every P -random point.

To prove the rest of the proposition, let us construct the inverse mapping. Assume that C is some computable point. Then there exists a computable sequence C_0, C_1, C_2, \dots of clopen sets such that $d(C_i, C) \leq 2^{-i}$. This inequality implies that $d(C_i, C_{i+1}) \leq 2 \cdot 2^{-i}$ (both C_i and C_{i+1} are close to the limit, so they are close to each other). Now the set

$$W_i = (C_i \triangle C_{i+1}) \cup (C_{i+1} \triangle C_{i+2}) \cup (C_{i+2} \triangle C_{i+3}) \cup \dots$$

is an effectively open set of measure at most $2^{i-1} + 2^i + 2^{i+1} + 2^{i+2} + \dots = 4 \cdot 2^{-i}$ that covers all the points ω where the sequence $C_i(\omega)$ does not stabilize. Therefore, these

bad points are all non-random and the limit mapping is well defined on all random points. This limit mapping is layerwise computable, since the sets W_i (more precisely, W_{i+2}) form a test, and C_i computes the limit mapping outside W_i .

Now let us show that this correspondence (computable points \rightarrow layerwise computable mappings) is well defined, i.e., we get the same layerwise computable mapping starting from two representations of the same point. If C_0, C_1, \dots and C'_0, C'_1, \dots are two computable sequences of clopen sets that both converge to some computable point C (with prescribed speed of convergence), then $d(C_i, C'_i)$ is also bounded by $2 \cdot 2^{-i}$, and

$$S_i = (C_i \triangle C'_i) \cup (C_{i+1} \triangle C'_{i+1}) \cup (C_{i+2} \triangle C'_{i+2}) \cup \dots$$

is again a test, so for every random sequence ω we have $C_i(\omega) = C'_i(\omega)$ for all sufficiently large i . Therefore, C_i and C'_i give us the same layerwise computable mapping (defined on random sequences).

It remains to note that these two constructions are inverse. Starting with a layerwise computable mapping f , we get a sequence of clopen sets C_i that pointwise converges to f , and also converges to the corresponding point in the metric space (the distance between C_i and the limit is at most 2^{-i}), so it can be used in the second construction, and we get the same mapping on random sequences as its pointwise limit. On the other hand, starting with a computable point, we get a layerwise computable mapping that represents that same point, since the distance between C_i and the pointwise limit of the C_n 's is $O(2^{-i})$. This ends the proof of the proposition. \square

Remark 1. Looking at the construction of a layerwise computable mapping for a given computable sequence C_i , we may note that the test S_i appearing in this construction is in fact a *Schnorr test*; this means that the measure of S_i is uniformly computable given i . Therefore we get convergence for all *Schnorr random* sequences ω , i.e., for all sequences not rejected by any Schnorr test. Combining this observation with Proposition 1, we see that every layerwise computable mapping (with bit or sequence values) originally defined on Martin-Löf random sequences, can be naturally extended to the bigger set of Schnorr random sequences.

Recall that in the original statement about image randomness we considered computable mappings of the Cantor space into a space of finite and infinite sequences whose values are P -almost everywhere infinite. To see that those mappings are indeed a special case of our notion of a layerwise computable mapping, we need the following simple result.

Proposition 3. *Let P be a computable distribution on the Cantor space. Let F be a computable mapping of the Cantor space to the space of finite and infinite sequences. Assume*

that $F(\omega)$ is an infinite sequence for P -almost all ω . Then $F(\omega)$ is infinite for all Martin-Löf random sequences, and F is a layerwise computable mapping on them.

Proof. According to Proposition 1, we may consider each coordinate of the mapping $F(\omega) = F_0(\omega)F_1(\omega)\dots$ separately. We know that $F_i(\omega)$ is defined for P -almost all sequences, so the domain of F_i is an effectively open set of P -measure 1. We enumerate this set until we get a clopen lower bound for it of measure at least $1 - 2^{-i}$. Its complement U_i is a (cl)open set of P -measure at most 2^{-i} , and we get a Martin-Löf test. So every Martin-Löf random sequence belongs to the domain of F_i , and we can compute F_i outside of U_i (in fact, the same machine is used for all i). So all F_i (and F in general) are layerwise computable. \square

We finish this section by giving a natural “machine-language” description of layerwise computable mapping. Consider a machine M that has a read-only input tape, work tapes, and a *rewritable* output tape. Initially the output tape contains all zeros, but the machine can change them to ones back and forth at any time. What matters is only the limit value of a cell if the limit exists; if not, the corresponding output bit is undefined. Without additional restrictions, this model is equivalent to computability relativized to the halting problem (and this is not interesting for us now). We pose an additional restriction: for each i and for each rational $\varepsilon > 0$ one can compute an integer $N(i, \varepsilon)$ such that the probability of the event

machine M with input ω changes i -th output cell after more than $N(i, \varepsilon)$ steps

(taken over P -distribution on all inputs ω) does not exceed ε . This condition implies that the output is well defined (all output values stabilize) for P -almost every input ω .

Proposition 4. *The output sequence is well defined for all Martin-Löf random input sequences. The resulting mapping is layerwise computable; every layerwise computable mapping can be obtained in this way.*

Proof. Again we may use Proposition 1 and consider each coordinate separately. The contents of i -th output cell after t steps is determined by a finite prefix of the input sequence ω , so we get some clopen set. Taking $t = N(i, 2^{-k})$, we get in this way some clopen set C_k and know that the indicator function of C_k coincides with the limit value for all inputs except for a set of measure at most 2^{-k} . Then we use the same argument as in Proposition 2, and conclude that the resulting mapping is layerwise computable.

It remains to show that every layerwise computable mapping can be transformed into a machine of described type. As we have seen, a layerwise computable mapping with bit values is a limit of a sequence of (the indicator functions of) clopen sets C_n such that $P(C_n \triangle C_{n+1}) < 2^{-n}$. The machine M then works in stages; at n -th stage it computes C_n

and then, using C_n as a table, changes (if necessary) the output bit to make it consistent with $C_n(\omega)$, where ω is the input sequence. Then $N(2^{-n})$ is the number of steps sufficient to perform $n + 1$ stages of this procedure (for arbitrary input ω).

This is for one-bit output; we can do the same in parallel for all output bits (interleaving the computations for different output bits) and then define $N(i, 2^{-n})$ in such a way that the machine is guaranteed to terminate $n + 1$ stages for bit i in at most $N(i, 2^{-n})$ steps. \square

This characterization of layerwise computable mappings is not used much in the sequel; we provide it here for two reasons. First, it gives some intuitive understanding of this notion on the “machine language” level. Second, it turns out that this is not only a theoretical characterization of layerwise computable mappings: some natural randomized algorithms really define a layerwise computable mapping in this way. One example where this phenomenon happens is an infinite version of Moser–Tardos algorithm for Lovasz local lemma used by Andrei Rumyantsev to prove a computable version of this lemma (see [8, 9]).

3 Image randomness

In the section we prove the image randomness property for layerwise computable mappings (Theorem 5).

Let P be a computable distribution on the Cantor space Ω , and let $F: \Omega \rightarrow \Omega$ be a layerwise computable mapping defined on P -random sequences and having infinite bit sequences as values. Then an image probability distribution Q on the Cantor space is defined in the usual way: $Q(X) = P(F^{-1}(X))$. Note that since F is defined P -almost everywhere, the distribution Q is well defined.

Theorem 5. (i) *The distribution Q is computable.*

(ii) *If a sequence α is Martin-Löf random with respect to P , then $F(\alpha)$ is Martin-Löf random with respect to Q .*

(iii) *If a sequence β is Martin-Löf random with respect to Q , then there exists some sequence α that is Martin-Löf random with respect to P such that $F(\alpha) = \beta$.*

The first statement (i) is needed to make the notion of Q -randomness well defined. The second statement is the randomness conservation property, and the third statement says that randomness does not “appear from nothing”.

Proof. (i) According to the definition, we need to compute the probability of the event “ $F(\omega)$ starts with u ” for arbitrary bit string u with arbitrary given precision $\varepsilon > 0$. The probability here is taken over P -distribution on input sequence ω . Assume that u consists

of k bits. Using the machine characterization, we can computably find how many steps are needed to finalize each of k first output bits with probability of later change at most ε/k . Then we take the maximal of these numbers:

$$N = \max(N(0, \varepsilon/k), N(1, \varepsilon/k), \dots, N(k-1, \varepsilon/k))$$

Then the probability in question is ε -close to the same probability for the time-bounded computation stopped after N steps, and this probability can be computed by simulating the machine behavior on all possible inputs.

(ii) Assuming that $\beta = F(\alpha)$ is not Q -random for some P -random α , we need to get a contradiction. Consider a Martin-Löf test $U_1 \supset U_2 \supset \dots$ provided by the layerwise computability of F , and a Martin-Löf test $V_1 \supset V_2 \supset \dots$ that rejects β . Using them, we construct a test that rejects α . By definition, there exists a computable mapping F_i that coincides with F outside U_i . Let us consider $F_i^{-1}(V_i)$. It requires some care since V_i is an open set in the Cantor space of infinite sequences, while the values of F_i may be finite. Still we may represent V_i as a union of intervals $x_{i,0}\Omega \cup x_{i,1}\Omega \cup x_{i,2}\Omega \cup \dots$, and consider all ω such that $F_i(\omega)$ has one of the strings $x_{i,0}, x_{i,1}, x_{i,2}, \dots$ as a prefix. In this way we get an effectively open set that is denoted $F_i^{-1}(V_i)$ in the sequel. This notation is formally incorrect since the set depends not only on V_i , but also on its representation as the union of intervals, but this will not create a problem.

The set $F_i^{-1}(V_i)$ consists of two parts: sequences $\omega \in U_i$, and sequences $\omega \notin U_i$. The first part is small since U_i is small; the second part is small since F_i coincides with F outside U_i , so the second part is contained in $F^{-1}(V_i)$. So the measures of both parts are bounded by 2^{-i} , and we get at most $2 \cdot 2^{-i}$ in total. Now we add U_i to $F_i^{-1}(V_i)$ and get an effectively open set of measure $O(2^{-i})$ that covers α both for the case $\alpha \in U_i$ and for the case $\alpha \notin U_i$. Since this can be done for arbitrary i , the sequence α is not random.

(iii) Let β be a Q -random sequence. We want to show that β belongs to the image $F(R)$ where R is the set of all P -random sequences. Martin-Löf's definition of randomness guarantees that the complement of R can be represented as the intersection of a decreasing sequence of effectively open sets from the universal Martin-Löf test, so R can be presented as the union of an increasing sequence of their complements, i.e., a sequence $R_1 \subset R_2 \subset R_3 \subset \dots$, of effectively closed sets where P -measure of R_i is at least $1 - 2^{-i}$. The mapping F is defined everywhere on R , so it is defined everywhere on R_i . Moreover, since we started from the *universal* Martin-Löf test, the definition of layerwise computability guarantees that on R_i the mapping F may be computed by a machine with write-only output tape, so F is continuous on R_i . The set R_i is compact (being a closed set in a compact space), so its image $F(R_i)$ is compact and therefore closed. The Q -measure of $F(R_i)$ is at least $1 - 2^{-i}$, since the preimage $F^{-1}(F(R_i))$ of $F(R_i)$ contains R_i .

It remains to show that the sets $F(R_i)$ are uniformly effectively closed (i.e., their complements are uniformly effectively open). Indeed, in this case their intersection contains only non- \mathcal{Q} -random sequences, so the sequence β belongs to some $F(R_i)$ and is therefore an image of a (random) sequence from R_i .

For that we may use the effective version of an argument proving that an image of a compact set is compact. We enumerate intervals V that are disjoint with $F(R_i)$ in such a way that their union is the complement of $F(R_i)$. For that, we look for a tuple (U_1, \dots, U_k) of intervals and an interval V such that (a) V is disjoint with all U_i ; (b) $F^{-1}(U_1), F^{-1}(U_2), \dots, F^{-1}(U_k)$ cover R_i . The conditions (a) and (b) guarantee that V is disjoint with $F(R_i)$: the image of every element in R_i belongs to one of the U_i and therefore does not belong to V . Note that we can replace F by F_i in (b) and get an equivalent condition, since $F = F_i$ on R_i , so a cover remains a cover. Using this observation, we can enumerate all the tuples U_1, \dots, U_k, V (since the complement of R_i and $F_i^{-1}(U_1), \dots, F_i^{-1}(U_k)$ are effectively open and Ω is compact, we can look for cases when some intervals inside the preimages together with some intervals inside the complement of R_i cover the entire Ω). When such a tuple U_1, \dots, U_k, V is found, we include the interval V in the enumeration. It does not intersect $F(R_i)$, as we have seen; the standard compactness argument shows that every point outside $F(R_i)$ is covered by one of those intervals V , so $F(R_i)$ indeed has an effectively open complement. This construction is uniform in i . \square

4 Remarks and enhancements

4.1 Image randomness and van Lambalgen theorem

Every sequence $\alpha = \alpha_0 \alpha_1 \alpha_2 \alpha_3 \dots$ can be split into two sequences $\alpha^e = \alpha_0 \alpha_2 \dots$ and $\alpha^o = \alpha_1 \alpha_3 \dots$ by separating its even and odd terms. The mapping $\alpha \mapsto (\alpha^e, \alpha^o)$ provides an isomorphism between the Cantor space Ω and its square $\Omega \times \Omega$, where Ω is considered as probability space equipped with uniform Bernoulli distribution. Applying Theorem 5 to the function $\alpha \mapsto \alpha^e$ (the composition of our isomorphism and the projection function), we see that

- if α is random, then α^e is random;
- every random sequence β is equal to α^e for some random α .

The second statement means that for every random sequence α^e we can also find some α^o in such a way that together they make some random sequence α . One may ask which sequences α^o can be used to achieve this. The answer is given by the well known *van Lambalgen theorem* that says that α is random if and only if α^e is random and α^o is

random *with respect to the oracle* α^e . The oracle randomness is defined in the usual way: the sets of small measure that form the test should be effectively open with the oracle (the enumerating algorithm may consult the oracle).

In general, we have two computable distributions P_1 and P_2 on two copies of Ω and consider randomness with respect to their product $P_1 \times P_2$. The van Lambalgen theorem says that a pair (α_1, α_2) is $(P_1 \times P_2)$ -random if and only if α_1 is P_1 -random and α_2 is P_2 -random with respect to oracle α_1 . It implies that for every P_1 -random sequence α_1 there exists some α_2 that makes the pair (α_1, α_2) random with respect to $P_1 \times P_2$. This (much weaker) statement is also a direct corollary of Theorem 5. The advantage of using Theorem 5 is that it can be applied to every computable distribution P on $\Omega \times \Omega$, not necessarily the product distribution: if a pair (α_1, α_2) is P -random, then α_1 is random with respect to the marginal distribution P_1 ; if α_1 is P_1 -random, then (α_1, α_2) is P -random *for some* α_2 .

One can ask for which α_2 this happens, i.e., how one can generalize van Lambalgen's theorem to non-product distributions. There are some results in this direction [12, 11], but we do not go into details here; let us only note that this happens for almost all α_2 with respect to the conditional distribution defined according to Takahashi [11].

4.2 Quantitative version

As many other results about algorithmic randomness, Theorem 5 has a quantitative version. Recall that *randomness deficiency* of a sequence ω may be defined as the maximal number i such that $\omega \in U_i$, where $U_1 \supset U_2 \supset U_3 \supset \dots$ is the universal Martin-Löf test (see [2, 10]; this version of randomness deficiency is called *probability-bounded* deficiency there). Using this notion, we can state the promised quantitative version.

Let P be a computable distribution on Ω , let F be an Ω -valued layerwise computable mapping defined on P -random sequences, and let Q be the F -image of P .

Theorem 6.

$$d_Q(\beta) = \inf\{d_P(\alpha) \mid F(\alpha) = \beta\} + O(1).$$

Here d_P and d_Q stand for the randomness deficiency functions for distributions P and Q ; the constant in $O(1)$ depends on the choice of deficiency functions and on F . Note that this result directly implies Theorem 5 (recall that the infimum of the empty set is $+\infty$).

Proof. For the proof we should just look closely at the argument we used. First, we have shown by contraposition that an image $F(\alpha)$ for P -random α is Q -random. Now we need to show that if the deficiency of $F(\alpha)$ is high, then the deficiency of α is high. Indeed, if $\beta = F(\alpha)$ is covered by V_i (in the notation used to prove Theorem 5), then α is covered by

the set $W_i = U_i \cup F_i^{-1}(V_i)$ whose measure is $O(2^{-i})$. The sets W_i also form a Martin-Löf test (up to a $O(1)$ -change in numbering), so

$$d_Q(F(\alpha)) \leq d_P(\alpha) + O(1)$$

(the quantitative version of randomness conservation).

To prove the reverse inequality, note that the complements of the sets $F(R_i)$ form a Q -test. So if $d_Q(\beta)$ is small and i is large, the sequence β belongs to $F(R_i)$ and not to its complement. More formally, β belongs to $F(R_i)$ for i that is only $O(1)$ -larger than $d_Q(\beta)$, and all elements in $F(R_i)$ have preimages in R_i whose deficiency is bounded by i . \square

4.3 Quantitative version for expectation-bounded deficiency

There is another version of randomness deficiency introduced by Levin. This version is called in [2, 10] the *expectation-bounded* deficiency, while the original Martin-Löf version (described above) is called *probability-bounded*. In this version $d_P(\cdot)$ is defined as the logarithm of the maximal (up to a $O(1)$ -factor) non-negative lower semicomputable function with finite integral with respect to a given distribution (see [2, 10] for details). These two versions of randomness deficiency differ more than by a constant (the probability-bounded one is bigger, and the difference is unbounded), but the formula of Theorem 6 is true also for the expectation-bounded deficiency. Let us prove this (assuming that the reader is familiar with the basic techniques from [2]).

Let us show first that the randomness conservation property $d_Q(F(\alpha)) \leq d_P(\alpha) + O(1)$ for this version of the randomness deficiency. Note that if α is not random, $F(\alpha)$ is undefined and $d_P(\alpha)$ is infinite, so the claim makes sense only if α is P -random, and $F(\alpha)$ is defined for every P -random α . Let $t_Q(\cdot)$ be the maximal lower semicomputable integrable function (the universal randomness test) for distribution Q , so $d_Q(\cdot) = \log t_Q(\cdot)$. We assume that $\int t_Q(\beta) dQ(\beta) \leq 1$. We may try to consider a function $t(\alpha) = t_Q(F(\alpha))$ and show that it is a P -test and therefore is bounded by the universal P -test (so we get the inequality we want to prove). However, this plan does not work “as is”, because the function $t(\alpha)$ is well defined only for random α and there is no reason to expect that t can be extended to a lower semicomputable function on the entire Ω (though the other requirement for tests, bounded integral, is true for t).

To overcome this difficulty, we recall that F is a layerwise computable mapping. We may assume that F coincides with computable F_i outside some effectively open set U_i whose P -measure is at most 2^{-i} . Now we may consider a function

$$t_i(\alpha) = t_Q(F_i(\alpha))$$

defined on random α and, moreover, on all α where $F_i(\alpha)$ is infinite. It is easy to see that t_i can be naturally extended to a lower semicomputable function defined on the entire Ω . We denote this extension also by t_i . Now the problem is that there is no reason to expect that $\int_{\Omega} t_i(\alpha) dP(\alpha)$ is finite; the only thing we are sure about is that $\int_{\Omega \setminus U_i} t_i(\alpha) dP(\alpha) \leq 1$. This happens because $t_i = t$ outside U_i , and for t the similar inequality is true (recall that Q is the image of P).

So we take the next step and consider the function t'_i defined as t_{2^i} artificially cut at 2^i (all bigger values are replaced by 2^i). It is still lower semicomputable, but in this way we guarantee that the integral of t'_i inside U_{2^i} does not exceed $2^{-2^i} \times 2^i = 2^{-i}$. Let $t'(\alpha) = \sup_i t'_i(\alpha)$. The function t' is lower semicomputable since all t'_i are uniformly lower semicomputable. To get the bound for the integral, let us consider $t^N(\alpha) = \sup_{i \leq N} t'_i(\alpha)$. When can $t^N(\alpha)$ exceed $t(\alpha)$ for a random α ? If this happens because of some $t'_i(\alpha)$, then α belongs to U_{2^i} , and the excess is bounded by a function that equals 2^i in U_{2^i} and equals 0 outside U_{2^i} . The integral of this function is at most 2^{-i} , and the total excess is bounded by the sum of these integrals, so it is bounded by 1. We conclude that $\int t'(\alpha) dP(\alpha) \leq 2$ (since $\int t(\alpha) dP(\alpha) \leq 1$ and the excess is bounded by 1), so $\log t'(\alpha) \leq d_P(\alpha) + O(1)$. It remains to note that $t(\alpha) = t_Q(F(\alpha)) \leq t'(\alpha)$ for random α , since $\alpha \notin U_{2^i}$ for large i , and for those i we have $t'_i(\alpha) = \min(t(\alpha), 2^i)$, so the supremum over i is at least $t(\alpha)$.

Now the reverse inequality. Consider the universal test $t_P(\cdot)$ for distribution P , i.e., a maximal non-negative P -integrable lower semicomputable function; its logarithm is $d_P(\cdot)$. Now consider the function

$$u(\beta) = \inf\{t_P(\alpha) \mid F(\alpha) = \beta\},$$

where the infimum is taken over random α (otherwise $F(\alpha)$ is not defined). If α is a P -distributed random variable, then $\beta = F(\alpha)$ is a Q -distributed random variable, and $u(F(\alpha)) \leq t_P(\alpha)$, so

$$\int u(\beta) dQ(\beta) = \int u(F(\alpha)) dP(\alpha) \leq \int t_P(\alpha) dP(\alpha) \leq 1$$

(assuming that u is measurable, otherwise the integral may be not well defined).

It remains to prove that u is lower semicomputable. Indeed, in this case the integral we consider is well defined, the function u is a Q -test, so $\log u(\beta) \leq d_Q(\beta)$, and that is exactly what we need. To prove the lower semicomputability of u let us recall the definition. We need to prove that the set $\{\beta \mid u(\beta) > r\}$ is uniformly effectively open for every rational r . Note that for this set all the values of u greater than r are equivalent, so it remains unchanged if we let t_P be infinite on U_i for some very large i (where $t_P(\alpha)$ is large anyway). Here U_i is the effectively open set of small measure provided by the definition of layerwise computability. After that we may change F on U_i making it a total computable

mapping, i.e., a mapping that outputs infinite sequences on all inputs (as soon as the input of F is covered by U_i , we extend the output arbitrarily, e.g., by zeros). Then we may apply the standard reasoning for the case of a total computable mapping; here it is.

We have $u(\beta) = \inf\{t_P(\alpha) \mid F(\alpha) = \beta\} > r$ if and only if there exists some rational $s > r$ such that $t_P(\alpha) > s$ for all α such that $F(\alpha) = \beta$. We assume now that F is total (as discussed above), and consider all α , not only random ones, since $t_P(\alpha)$ is infinite for all non-random α . The condition “ $t_P(\alpha) > s$ for all α such that $F(\alpha) = \beta$ ” means that β does not belong to $F(\{\alpha \mid t_P(\alpha) \leq s\})$. The latter set is an image of an effectively closed set under a total computable mapping, so (as we have seen in the proof of Theorem 5) it is an effectively closed set, and its complement is effectively open. The construction is uniform, so we get what we wanted.

4.4 An application to random closed sets

Now we want to present an example where the generalization to layerwise computable mappings is important (i.e., the result for computable mappings is not enough). Let us consider two definitions of randomness for trees and show that they are equivalent.

Consider the following two random processes:

1. Take the full binary tree. Cut independently each edge with probability $1/3$. Then consider the connected component of the root in the remaining graph. It is a binary tree that may be finite or infinite (as we will see, the first case happens with probability $1/4$). Delete all vertices that have finitely many descendants. If the connected component of the root was finite, this gives an empty tree, otherwise we get an infinite tree without leaves.
2. Consider three possible elements



Select one of them randomly (all three have probability $1/3$) and place it at the root of the tree. Then at every vertex of level 1 (there can be one or two, depending on the root choice) select one of three elements independently (using fresh randomness), and so on.

Both random processes define a probability distribution on the set of infinite subtrees of binary tree that have no leaves; in the first case we should consider the conditional

distribution under the condition “process generates a non-empty tree”. As we will see, this multiplies all probabilities by $4/3$, since the probability to get an infinite tree is $3/4$.

It turns out that these two distributions on trees without leaves are the same. Moreover, an algorithmic version of the same statement is also true. Imagine that we use bits from a Martin-Löf random sequence when the random process asks for fresh randomness. It turns out that we get the same class of trees (called Martin-Löf random trees) in both cases. This was proven in [4] by some ad hoc construction; in this section we explain why this follows from Theorem 5.

Let us first compare the distributions. First of all let us explain how the probabilities $1/4$ and $3/4$ (for finite and infinite trees) are obtained. Let p_n be a probability to have a tree of height at least n when cutting each edge of a full binary tree independently with probability $1/3$. The inductive definition of the process gives the recurrence:

$$p_{n+1} = 2q(1-q)p_n + q^2(2p_n - p_n^2) = -\frac{4}{9}p_n^2 + \frac{4}{3}p_n$$

where $q = 2/3$ is the probability that a given edge exists (is not cut). The first term corresponds to two disjoint cases when one of the root’s children exists and the other one is cut, the second term is the probability that both children exist and at least one of them has subtree of height at least n (these events have probability p_n and are independent). We start with $p_0 = 1$ and iterate the function $x \mapsto -\frac{4}{9}x^2 + \frac{4}{3}x$; it is easy to see that these iterations converge to the fixed point $3/4$ exponentially fast. So the event “tree is infinite after edges are randomly cut with probability $1/3$ ”, being the intersection of a decreasing sequence of events with probabilities p_n , has probability $3/4$.

Knowing this, we may compute the probabilities of three possibilities shown above. The first case (where the root has two sons in the final tree) has probability $(2/3)^2(3/4)^2 = 1/4$, the second and the third cases have probability

$$(2/3)(3/4)(1 - (2/3)(3/4)) = 1/4$$

each. So all three cases have the same probability $1/4$ and conditional probability $1/3$, and the same is true at every vertex. So two distributions described above are the same.

Now we switch to algorithmic randomness. First of all, we need to fix the representation of trees by bit sequences. A tree without leaves can be considered as a set of vertices of a full binary tree (that contains with every vertex x its father and one of its sons). Let $\{v_0, v_1, v_2, \dots\}$ be the vertices of the full binary tree (=binary strings). Then each tree without leaves can be represented as a sequence of bits: i th bit is 1 if v_i belongs to the tree.

We start with the second random process (in the list above). It uses uniform random choice between three possibilities. So we consider ternary sequences (Martin-Löf randomness can be defined for any finite alphabet in the same way, and all the results can be

easily extended to this case), and the uniform distribution on $\{0, 1, 2\}^\infty$ where each term is uniformly distributed (each of three possible values has probability $1/3$) and the terms are independent. Then for each sequence $\omega \in \{0, 1, 2\}^\infty$ we construct a tree without leaves as described (using terms of the sequence to make choice between three extensions). In this way we get a total computable mapping $\{0, 1, 2\}^\infty \rightarrow \Omega$. Then we consider the image of the uniform distribution on $\{0, 1, 2\}^\infty$ and get the image distribution on sequences representing trees (and therefore on trees). Random sequences with respect to this image distribution are exactly the images of Martin-Löf random sequences in $\{0, 1, 2\}^\infty$ due to Theorem 5; corresponding trees are called (Martin-Löf) random trees.

In fact, this description can be understood in two ways. We may use the terms of a ternary sequence to make choices in all vertices in advance (and then some choices do not matter since the vertex does not appear in the tree). Or we may use the terms of a ternary sequence only for the choices that really matter (starting from the root and then adding layers one by one); in the latter case the same term of the sequence may be used in different vertices depending on the preceding terms. These two possibilities give different mappings $\{0, 1, 2\}^\infty \rightarrow \Omega$. However, standard probability theory arguments show that the image distributions are the same. As we know, this implies that the image of the set of random ternary sequences is the same.

The first random process (generating trees by cutting edges) is more complicated, and layerwise computable mappings appear. We start with a binary sequence (generated according to $B_{1/3}$). Then we cut the edges according to the terms of this binary sequence, and get a tree without leaves or the empty set of vertices (that corresponds to a zero sequence when we encode trees by binary sequences). In this way we get a mapping $\Omega \rightarrow \Omega$.

This mapping is layerwise computable. To prove this, consider a rewriting machine that constructs a tree using random input bits to cut the edges of the full binary tree. This machine starts from the root and goes upwards, but when a vertex x with a finite subtree above x is discovered, this subtree (including x itself) is deleted. So there is always a possibility that some part of the tree, or the entire tree, will be deleted in the future. However, the probability of the event “the root survived up to height n , but later was deleted” is bounded by $p_n - \lim p_n$ (in fact, is equal to it) and computably converges to zero. Similar statement is true for every vertex. So the mapping is layerwise computable.

What is the image distribution (of $B_{1/3}$)? As we have seen, it has an atom (zero sequence, or empty tree) that has probability $1/4$, and the rest is the previous distribution on trees multiplied by $3/4$. It is easy to see that random sequences are the same (plus a zero sequence) as in the first case. It remains to apply Theorem 5.

4.5 Many-valued mappings

We considered a random process that consists of two parts: a source of random bits (with some distribution P) that generates some bit sequence α , and some machine M that transforms these bits (the sequence α) into an output sequence β . Now we want to consider a more general situation when the transformation is non-deterministic. We consider M as a black box, and the only thing we know is some relation between its input α and output β . Formally, we have some closed subset F of $\Omega \times \Omega$, and we know that $\langle \alpha, \beta \rangle$ always belongs to F .

The question remains the same: which sequences are plausible as outputs of such a device? For which β there exists some P -random sequence α such that $\langle \alpha, \beta \rangle \in F$?

Let us give two motivating examples for this scheme. First, consider a bit sequence that is Martin-Löf random with respect to B_p (independent bits, each bit equals 1 with probability p). Then replace some ones by zeros. Which sequences can be obtained by this two-stage process? These sequences (corresponding sets) were called *p-sparse* in [3]. In our language, we consider B_p as P (source distribution), and coordinate-wise ordering as F :

$$F(\alpha_0\alpha_1\dots, \beta_0\beta_1\dots) \Leftrightarrow \forall i (\beta_i \leq \alpha_i).$$

Another example: let α be a random tree (in the sense described above) and let $F(\alpha, \beta)$ mean that β is a path in α . Then we get the question: which sequences may appear as paths in random trees? This question was studied in [1, 4, 7]; finally these sequences were characterized in terms of continuous a priori probability.

Let us return to the general question mentioned above, for arbitrary measure P and closed set $F \subset \Omega \times \Omega$. Some answer can be provided in terms of *class randomness* (see [2] for the definition and the basic properties; we assume that the reader is familiar with this notion). Consider some effectively closed set \mathcal{L} in the effective metric space of all distributions on the Cantor space. An *expectation-bounded class test with respect to \mathcal{L}* is a lower semicomputable function $t(\cdot)$ with non-negative values defined on the Cantor space such that the integral of t with respect to any distribution in \mathcal{L} does not exceed 1. For every effectively closed class \mathcal{L} there exists a maximal (up to a constant factor) test $t_{\mathcal{L}}$. The sequences ω such that $t_{\mathcal{L}}(\omega)$ is finite are called *random with respect to class \mathcal{L}* , and $d_{\mathcal{L}}(\omega) = \log t_{\mathcal{L}}(\omega)$ is called the *expectation-bounded randomness deficiency of ω with respect to \mathcal{L}* . If \mathcal{L} is a singleton, its only element is a computable distribution and \mathcal{L} -randomness is equivalent to Martin-Löf randomness. In the general case, a sequence ω is \mathcal{L} -random if and only if it is uniformly random with respect to some distribution $Q \in \mathcal{L}$ (this distribution may not be computable, so a notion of randomness with respect to non-computable distributions is needed, see [2]).

The notion of class randomness can be equivalently defined in terms of Martin-Löf

class tests: a decreasing sequence $U_1 \supset U_2 \supset \dots$ is a \mathcal{L} -test if $Q(U_i) \leq 2^{-i}$ for every $Q \in \mathcal{L}$. For each test a deficiency function is introduced whose value on ω is $\sup\{i \mid \omega \in U_i\}$; there exist universal tests for which the deficiency function is maximal, and this maximal function is called the *probability-bounded randomness deficiency with respect to \mathcal{L}* . (We can also use probability-bounded tests defined as lower semicomputable functions t such that for every c the event $t(\omega) > c$ has probability at most $1/c$ with respect to every distribution in \mathcal{L} ; this gives an equivalent definition of probability-bounded deficiency.)

Now we can formulate the promised answer about sequences β such that $F(\alpha, \beta)$ for some P -random α , assuming that the relation F satisfies some conditions. Namely, we assume that

- (i) F is an effectively closed subset of $\Omega \times \Omega$.
- (ii) The projection of F on the first coordinate is Ω : for every α there exists some β such that $F(\alpha, \beta)$.
- (iii) The F -preimage of every interval $[y]$ in the Cantor space, i.e., the set of all α such that $F(\alpha, \beta)$ for some β starting with y , is a clopen set (a finite union of intervals) that can be computed given y .

It is easy to check that all these conditions are satisfied in the two examples given above.

Recall that we consider some computable distribution P on Ω (the first coordinate of the product $\Omega \times \Omega$). Consider the class \mathcal{L} of all distributions Q on the second copy of Ω that are *F-couplable with P* . This means, by definition, that there exists some distribution S on $\Omega \times \Omega$ concentrated inside F (the complement of F has S -measure 0) such that the first projection (marginal distribution) of S is P and the second projection is Q .

Proposition 7. *This class \mathcal{L} is a non-empty effectively closed class of distributions on Ω .*

Proof. The condition “the first projection of S equals P ” is effectively closed (defines an effectively closed class of distributions S). The same can be said about the condition “ S is concentrated inside F ” (the complement of F is a union of a computable sequence of intervals in $\Omega \times \Omega$, and we require the S -measure of each of these intervals to be 0, thus getting a sequence of uniformly effectively closed conditions). Let us check that these two conditions have non-empty intersection. By compactness, it is enough to consider some clopen $F' \supset F$ instead of F : if for each F' the intersection is non-empty, then it is non-empty for F . The projection of F' contains the projection of F and therefore covers the entire Ω ; since F' is clopen (a finite union of intervals), we can represent Ω as a disjoint union $\Omega = U_1 + \dots + U_n$ of intervals, and find some intervals V_1, \dots, V_n such that $U_i \times V_i \subset F'$ for every i . Then on each $U_i \times V_i$ we consider a distribution that is a product

of P restricted to U_i , and some computable distribution on V_i , and combine these products to get a distribution in the desired class (projection is P and support is in F').

It remains to note that this intersection is an effectively closed (and therefore compact) class of distributions on $\Omega \times \Omega$; the projection to the second coordinate is a computable continuous mapping, so its image is again an effectively closed non-empty class of distributions. \square

This proposition allows us to consider the randomness deficiency function $d_{\mathcal{L}}$ for the class \mathcal{L} , and compare it with the randomness deficiency function d_P for distribution P .

Theorem 8. *Let P be a computable probability distribution; let F be a subset of $\Omega \times \Omega$ satisfying the conditions (i)–(iii), and let \mathcal{L} be the class of distributions Q which are F -couplable with P . The following equality holds with $O(1)$ -precision (both for expectation- and probability-bounded versions of deficiency function):*

$$d_{\mathcal{L}}(\beta) = \inf_{\alpha} \{d_P(\alpha) \mid F(\alpha, \beta)\},$$

In particular, a sequence β is random with respect to the class \mathcal{L} if and only if $F(\alpha, \beta)$ is true for some P -random sequence α .

Proof. Let t_P be the maximal expectation-bounded randomness test with respect to P , i.e., maximal lower semicomputable function such that $\int t_P(\omega) dP(\omega) \leq 1$. Then we may consider the function

$$t(\beta) = \inf\{t_P(\alpha) \mid F(\alpha, \beta)\}.$$

It is easy to check (see [2]) that $t(\cdot)$ is lower semicomputable. Let us check that

$$\int t(\beta) dQ(\beta) \leq 1 \text{ for every distribution } Q \in \mathcal{L}.$$

According to the definition of \mathcal{L} , for every $Q \in \mathcal{L}$ there exists some distribution S on $\Omega \times \Omega$ that is zero outside F and has projections P and Q (for the first and second coordinates). Then the integral in question is equal to $\int_{\Omega \times \Omega} t(\beta) dS(\alpha, \beta)$. Since S vanishes outside F and $t(\beta) \leq t_P(\alpha)$ when $(\alpha, \beta) \in F$, the latter integral does not exceed $\int_{\Omega \times \Omega} t_P(\alpha) dS(\alpha, \beta) = \int t_P(\alpha) dP(\alpha) \leq 1$. So the function $t(\beta)$ is a test with respect to \mathcal{L} and does not exceed $t_{\mathcal{L}}(\beta)$. One inequality of the theorem (\geq) is proven for expectation-bounded tests.

To prove the same inequality for probability-bounded tests, it is enough to show that if $t_P(\cdot)$ is a probability-bounded test with respect to P , then $t(\cdot)$ is a probability-bounded test with respect to \mathcal{L} . Indeed, for a given threshold c consider the set $U = \{\beta \mid t(\beta) > c\}$. We need to prove that $Q(U) \leq 1/c$ for every $Q \in \mathcal{L}$. Let S be the corresponding measure

on the product space. Then $Q(U) = S(\Omega \times U) = S((\Omega \times U) \cap F)$. The definition of $t(\cdot)$ guarantees that $\Omega \times U \cap F \subset V \times \Omega$, where $V = \{\alpha \mid t_P(\alpha) > c\}$ (if the infimum exceeds c , then all the elements exceed c). And $S(V \times \Omega) = P(V) \leq 1/c$ since t_P is a probability-bounded test.

So one direction is proven for both types of tests. The other direction is more difficult: here we have to use the assumptions about F . Again we consider expectation-bounded tests first. We need to show that

$$t_{\mathcal{L}}(\beta) \leq \inf_{\alpha} \{t_P(\alpha) \mid F(\alpha, \beta)\}$$

(for maximal tests, up to a constant factor). This means that

$$t_{\mathcal{L}}(\beta) \leq t_P(\alpha) \quad \text{if} \quad F(\alpha, \beta).$$

This can be rewritten also as

$$\sup_{\beta} \{t_{\mathcal{L}}(\beta) \mid F(\alpha, \beta)\} \leq t_P(\alpha).$$

It would be guaranteed if the left-hand side, denoted by $t(\alpha)$ in the sequel, were an expectation-bounded P -test, i.e., were lower semicomputable and had a bounded integral.

Imagine for a while that the supremum is achieved for some measurable function

$$\beta(\alpha) = \operatorname{argmax}\{t_{\mathcal{L}}(\beta) \mid F(\alpha, \beta)\},$$

so $t(\alpha) = t_{\mathcal{L}}(\beta(\alpha))$. By construction, the pair $(\alpha, \beta(\alpha))$ belongs to F for every α , so the random pair $(\alpha, \beta(\alpha))$ defines a distribution on $\Omega \times \Omega$ whose support is in F and whose first projection is P . Then its second projection Q is in \mathcal{L} , so the integral $\int t_{\mathcal{L}}(\beta) dQ(\beta)$ does not exceed 1. On the other hand,

$$\int t_{\mathcal{L}}(\beta) dQ(\beta) = \int t_{\mathcal{L}}(\beta(\alpha)) dP(\alpha) = \int t(\alpha) dP(\alpha),$$

so we get the required bound for the function t .

Of course, this argument is not valid: the maximum may be never achieved, so the function $\beta(\alpha)$ is not defined; another problem is that the function t may not be lower semicomputable (it can be guaranteed for the infimum, but not for the supremum). So we need to replace this argument by a valid one.

We start in the same way and consider the maximal lower semicomputable \mathcal{L} -test $t_{\mathcal{L}}$. Let $t_{\mathcal{L}}^1 \leq t_{\mathcal{L}}^2 \leq \dots$ be a computable sequence of basic functions (functions with finitely

many values having clopen preimages) that converges to $t_{\mathcal{L}}$. Now we may consider the functions

$$t^i(\alpha) = \max\{t_{\mathcal{L}}^i(\beta) \mid F(\alpha, \beta)\}$$

they are well defined since $t_{\mathcal{L}}^i$ has only finitely many values. Moreover, the conditions on F guarantee that $t^i(\cdot)$ is a basic function that can be computed given i . The sequence t^i increases, so we may consider its supremum, some lower semicomputable function t . We will show (see below) that $\int t^i(\alpha) dP(\alpha) \leq 1$ for every i . Then the same is true for t , so $t(\alpha) \leq ct_P(\alpha)$ for some c and for all α (due to the maximality of t_P). Then, for $(\alpha, \beta) \in F$ we have

$$t_{\mathcal{L}}^i(\beta) \leq t^i(\alpha) \leq t(\alpha) \leq ct_P(\alpha),$$

and it remains to take infimum over all α and then supremum over i to get the reverse inequality we need.

So it remains to prove that $\int t^i(\alpha) dP(\alpha) \leq 1$ for every i . This argument follows the informal explanation above, but instead of the function $\beta(\alpha)$ we use Proposition 7 to construct a distribution in \mathcal{L} .

Fix some i . Recall that the function $t^i(\alpha)$ is defined as $\max\{t_{\mathcal{L}}^i(\beta) \mid F(\alpha, \beta)\}$. Let $c_1 > c_2 > \dots > c_n$ be all the values of $t_{\mathcal{L}}^i$ and $U_k = \{\beta \mid t_{\mathcal{L}}^i(\beta) \geq c_k\}$. Let $V_k = F^{-1}(U_k) = \{\alpha \mid t^i(\alpha) \geq c_k\}$. The sets U_k and V_k increase as k increases; all these sets are clopen. The function $t^i(\alpha)$ equals c_1, \dots, c_n when α belongs to the sets $V_1, V_2 \setminus V_1, \dots, V_n \setminus V_{n-1}$ (respectively). Consider now the sets $V_1 \times U_1, (V_2 \setminus V_1) \times U_2, \dots, (V_n \setminus V_{n-1}) \times U_n$. On these sets $t^i(\alpha) \leq t_{\mathcal{L}}^i(\beta)$. Consider the intersections of these sets with F ; by construction their first projections coincide with $V_1, V_2 \setminus V_1, \dots$. Take the union of these intersections and apply Proposition 7 to it; we get a distribution whose first projection is P ; its support is a subset of F where $t^i(\alpha) \leq t_{\mathcal{L}}^i(\beta)$. The second projection of this distribution is in \mathcal{L} and it can be used to prove the inequality as described above.

Similar argument can be used for probability-bounded tests: to show that t' is probability-bounded, it is enough to show that all t^i are; the set $V = \{\alpha \mid t^i(\alpha) > c\}$ is the F -preimage of the set $U = \{\beta \mid t_{\mathcal{L}}^i(\beta) > c\}$; we apply Proposition 7 to the set $(V \times U) \cup ((\Omega \setminus V) \times \Omega)$ and note that the inequality $\chi_V(\alpha) \leq \chi_U(\beta)$ is true for every pair $\langle \alpha, \beta \rangle$ in this set. \square

Acknowledgments

We would like to thank Bruno Bauwens, Wolfgang Merkle, Jason Rute and other participants of the Heidelberg Focus Semester, Mikhail Andreev, Alexey Milovanov, Gleb Novikov, Andrey Rumyantsev, Nikolay Vereshchagin and other researchers from Kolmogorov seminar in Moscow, Bruno Durand, Andrei Romashchenko and other colleagues

in the ESCAPE group in Montpellier, as well as ESCAPE visitors, especially Peter Gács, and all the people in the HSE and Poncelet laboratory in Moscow for their hospitality.

References

- [1] Barmpalias, George; Broadhead, Paul; Cenzer, Douglas; Dashti, Seyyed; and Weber, Rebecca, Algorithmic randomness and closed sets, *Journal of Logic and Computation*, **17**(6), 2007, 1041–1062.
- [2] Bienvenu, Laurent; Gács, Péter; Hoyrup, Mathieu; Rojas, Cristobal; and Shen, Alexander, Algorithmic tests and randomness with respect to a class of measures, *Proceedings of the Steklov Institute of Mathematics*, **274**, 2011, 34–89. arXiv:1103.1529
- [3] Bienvenu, Laurent; Romashchenko, Andrei and Shen, Alexander, Sparse sets, *Journées Automates Cellulaires 2008 (Uzes)*, 18–28. hal.archives-ouvertes.fr/hal-00274010/en/
- [4] Diamondstone, David and Kjos-Hanssen, Bjørn, Martin-Löf randomness and Galton–Watson processes, *Annals of Pure and Applied Logic*, **163**(5), 2012, 519–529. arXiv:1408.2875
- [5] Hoyrup, Mathieu and Rojas, Cristóbal, An Application of Martin-Löf Randomness to Effective Probability Theory, *Mathematical Theory and Computational Practice, 5th Conference on Computability in Europe, CiE 2009, Heidelberg, Germany, July 19–24, 2009. Proceedings* (Lecture Notes in Computer Science, 5635), p. 260–269.
- [6] Martin-Löf, Per, The definition of random sequences, *Information and Control* (now *Information and Computation*), **9**, 1966, 602–619.
- [7] Miller, Joseph S. and Rute, Jason, Energy Randomness, arXiv:1509.00524
- [8] Rumyantsev, Andrei, *Infinite computable version of Lovasz Local Lemma*, preprint, 2010, arXiv:1012.0557.
- [9] Rumyantsev, Andrei and Shen, Alexander, Probabilistic Constructions of Computable Objects and a Computable Version of Lovász Local Lemma, *Fundamenta Informaticae*, **32**(1), 2014, 1–14.

- [10] Vereshchagin, Nikolai; Uspensky, Vladimir and Shen, Alexander, *Kolmogorov complexity and algorithmic randomness*. Moscow, MCCME, 2013. (In Russian; draft English translation is available at www.lirmm.shen/~ashen/kolmbook-eng.pdf.)
- [11] Takahashi, Hayato, Algorithmic randomness and monotone complexity on product space, *Information and Computation*, **209**(2), 2011, 183–197.
- [12] Vovk, Vladimir and Vyugin, Vladimir, On the empirical validity of the Bayesian method, *Journal of the Royal Statistical Society. Series B (Methodological)*, **55** (1), 1993, 253–266.