von Neumann's biased coin revisited

Laurent Bienvenu, Benoit Monin

von Neumann's biased coin revisited

Laurent Bienvenu, Benoît Monin
LIAFA - CNRS & Université de Paris 7
Case 7014, 75205 Paris Cedex 13, France
{laurent.bienvenu,benoit.monin}@liafa.jussieu.fr

***Abstract*—Suppose you want to generate a random sequence of zeros and ones and all you have at your disposal is a coin which you suspect to be biased (but do not know the bias). Can "perfect" randomness be produced with this coin? The answer is positive, thanks to a little trick discovered by von Neumann. In this paper, we investigate a generalization of this question: if we have access to a source of bits produced according to some probability measure μ in some class $\mathcal{C}$ of measures, and suppose we know $\mathcal{C}$ but not μ (in the above example, $\mathcal{C}$ would be the class of all Bernoulli measures), can perfect randomness be produced? We will look at this question from the viewpoint of constructive mathematics and in particular the theory of effective randomness.**

I. Introduction

If one wants to play fair 'head or tail' game with a biased coin, one can use the so-called von Neumann's trick. This trick works as follows. Flip the biased coin twice. If one gets the sequence 'head-tail', declare the result to be 'head', and if you get the sequence 'tail-head', declare the result to be 'tail'. If one gets either 'head-head' or 'tail-tail', start over. Calling p the probability that the biased coin gives a 'head', we see that this process produces 'head' and 'tail' with equal probability $1/2$ (as long as p is neither 0 nor 1 in which case the procedure never produces any output). Of course this procedure can be iterated if one wants to generate a longer (finite or infinite) random sequence of zeroes and ones (where 0='head' and 1='tail'), for example

Input (biased coin): $1111011011100111\ldots$
Output: $0110\ldots$

It is clear that for any given p different from 0 and 1, if the input follows a Bernoulli distribution of parameter p, then the output is uniformly distributed (Bernoulli distribution of parameter $1/2$).

Let us reformulate this result: we are given access to a sequence of zeroes and ones and all we know about this sequence is that it has been generated according to some probability measure μ in some class $\mathcal{C}$ (the class of Bernoulli measures with parameter $p \neq 0, 1$). The class $\mathcal{C}$ is known but not the measure μ. Yet it is possible to design a *single* procedure which, under the sole assumption that the input is μ-random for *some* $\mu \in \mathcal{C}$, produces a random sequence.

The question we are concerned with is the following: what are the classes of measures $\mathcal{C}$ for which such a randomness extraction procedure can be designed? This is still an informal question which can be interpreted in several ways. In this paper, we approach it from a *computability* and *constructive mathematics* viewpoint, interpreting 'extraction procedure' by *computable extraction procedure* and 'randomness' by *Martin-Löf randomness* (whose definition is recalled below).

The two main results we prove are the following:

(1) We show that in case the class $\mathcal{C}$ of measures is effectively compact and effectively orthogonal, then one can extract randomness from any element that is random with respect to any measure in $\mathcal{C}$, in a uniform way. This criterion applies to a wide variety of measures and generalizes von Neumann's theorem.

(2) In order to prove (1), we show that a measure μ belongs to an effectively compact and effectively orthogonal class $\mathcal{C}$ of measures if and only if the measure μ can be computably "guessed" for any of its random reals (in a sense we will make more precise). This answers a question in [1, p.64].

II. Constructive analysis and effective randomness

In von Neumann's trick, both the input and the output belong to the set of infinite binary sequences, also known as Cantor space (and denoted 2^ω). In this paper, due to space restriction, we will stick to this framework, but our results can be extended with a little more effort to more general computable metric spaces. However, we still need to introduce the notions of computable topological space and computable metric spaces, as we will need to consider the space of measures on 2^ω, and the space of open subsets of 2^ω which are respectively a computable metric space and a computable topological space. Let us first briefly present the basic theory of such spaces. The reader who is familiar with the theory can skip this section.

A. Constructive topological spaces and constructive metric spaces

Definition II.1. *A **constructive topological space** a triple $(X, \mathcal{B}, v)$ where*

- *X is a T_0 second countable topological space.*

- *$\mathcal{B}$ is a countable basis for the topology of X (we assume that $\varnothing$ and X belong to $\mathcal{B}$) and $v : \mathbb{N} \to \mathcal{B}$ is a surjection*
- *The intersection operation is effective in the following sense: there exists a computable function*

$$f : \mathbb{N} \times \mathbb{N} \times \mathbb{N} \to \mathbb{N}$$

such that

$$\forall i, j \quad \bigcup_{n \in \mathbb{N}} v(f(i,j,n)) = v(i) \cap v(j)$$

REMARK: The last condition can always be satisfied, by closing the set of basic open sets with all possible finite intersections of them. ∎

In a constructive topological space, the T_0 requirement implies that an element x is entirely determined by the set of basic open set containing it. We denote by $\mathcal{O}(x)$ the set of basic open sets containing x.

Definition II.2. *Let $(X, \mathcal{B}_X, v_X)$ and $(Y, \mathcal{B}_Y, v_Y)$ be two constructive topological spaces. A function $f : X \to Y$ is said to be **(partial) computable** if there is a computably enumerable set A of pairs of integers such that for all i, $f^{-1}(v_Y(i)) \cap Dom\, f = \bigcup_{j:(i,j)\in A} v_X(j) \cap Dom\, f$.* ∎

Let us now move to the setting of constructive metric spaces. These spaces need to be endowed with a richer structure, but they are nicely behaved and easier to study.

Definition II.3. *A **constructive metric space** is given by (X, d, D, γ) where*

- *(X, d) is a separable complete metric space.*
- *D is a countable dense subset of X and $\gamma : \mathbb{N} \to D$ is a surjection.*
- *d is computable i.e. the function:*

$$f : \left\{ \begin{array}{ccc} \mathbb{N} \times \mathbb{N} & \to & \mathbb{R} \\ (i,j) & \mapsto & d(\gamma(i), \gamma(j)) \end{array} \right.$$

is computable. ∎

REMARK: A constructive metric space can always be seen as a constructive topological space by taking the set of open balls of $B(\gamma(i), r)$ for all $i \in \mathbb{N}$ and $r \in \mathbb{Q}^+$ as basic open sets. ∎

As we can see it in the definition, the notion of being computable for a function is an effective version of the continuity. It also induces a generalization of truth-table reducibility. A function $f : 2^\omega \to 2^\omega$ is computable iff there is a Turing functional Φ such that for all oracles x, Φ^x is total and:

$$\forall n \quad \Phi^x(n) = f(x)(n)$$

The next proposition further illustrates this.

Proposition II.1. *Let $(X, \mathcal{B}_X, v_X)$ and $(Y, \mathcal{B}_Y, v_Y)$ be two constructive topological spaces. Let $f : X \to Y$ be a function. The following statements are equivalent:*

1) *There exists a partial computable function $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that: $\forall x \in X \quad g(v_X^{-1}(\mathcal{O}(x)), \mathbb{N}) = v_Y^{-1}(\mathcal{O}(f(x)))$*

2) *There exists a total computable function $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that: $\forall x \in X \quad g(v_X^{-1}(\mathcal{O}(x)), \mathbb{N})) = v_Y^{-1}(\mathcal{O}(f(x)))$*

3) *There is a computably enumerable set A of pairs of integers such that for all i, $f^{-1}(v_Y(i)) = \bigcup_{j:(i,j)\in A} v_X(j)$.* ∎

(see the appendix for the proof).

Definition II.4. *An **effectively open set** (or Σ_1^0 set) of a constructive topological space $(X, \mathcal{B}, v)$ is an open set of the form $\bigcup_{i\in A} v(i)$ where A is a c.e. set of integers. An **effectively closed set** (or Π_1^0 set) is the complement of an effectively open set.* ∎

We can now define the notion of computable point in a constructive topological space.

Definition II.5. *Let $(X, \mathcal{B}_X, v_X)$ be a constructive topological space, we say that $x \in X$ is **computable** if $\mathcal{O}(x)$ is effectively open. If $(Z, \mathcal{B}_Z, v_Z)$ is some other constructive topological space, and $z \in Z$, we say that x is z-computable (or computable relative to x) if there is a partial computable function f from Z to X such that $f(z) = x$.* ∎

This notion of computability coincides in the Cantor space with the usual notion of computability. We finish this part with an extension of the notion of lower semi-computability.

Definition II.6. *Let $(X, \mathcal{B}, v)$ be a constructive topological space. A function $f : X \to \overline{\mathbb{R}}$ is lower semi-computable iff $\forall r \in \mathbb{Q} \quad f^{-1}((r, +\infty])$ is effectively open uniformly in r.* ∎

Equivalently, a function is lower semi-computable iff it is computable as a function from X to $\mathbb{R}$ endowed with the topology generated by upper sets of the form $(r, +\infty]$.

We present one last notion which will be central in this paper, namely the notion of effective compactness.

Definition II.7. *A subset $\mathcal{K}$ of a constructive topological space $(X, \mathcal{B}, v)$ is **effectively compact** if it is effectively closed and one can enumerate its open covers. This means, given a standard enumeration (W_e) of all computably enumerable subsets of $\mathbb{N}$, that the set*

$$\left\{ e \mid \mathcal{K} \subseteq \bigcup_{i \in W_e} v(i) \right\}$$

is computably enumerable. By the Rice-Shapiro theorem (see [2]), this is equivalent to say that the set of finite sets F (represented as finite objects) of integers such that $\mathcal{K} \subseteq \bigcup_{i\in F} v(i)$ is computably enumerable. ∎

B. The Cantor space

The space of all binary sequences is commonly called *Cantor space* and denoted by 2^ω. We will consider on this space the (canonical) product topology generated by **cylinders**. The cylinder generated by the string s, denoted by $[s]$ is the set of all sequences starting by s, formally: $[s] = \{x \in 2^\omega |\ s \preceq x\}$ where $s \in 2^{<\omega}$. This topology makes 2^ω a compact and 0-dimensional topology Let $s \in 2^{<\omega}$ and let $t \in 2^{<\omega} \cup 2^\omega$. We will write $s \preceq t$ to mean that s is a prefix of t. We will write $|s|$ to denote the length of s and for all $n < |s|$, we denote by $s(n)$ the n-th bit of s (starting from 0 by convention). The empty string (whose length is 0 by convention) is denoted by ϵ. The concatenation of two strings s and t is written $s\hat{}t$.

The following is easy to prove (see [3]).

Proposition II.2. *Let d be the function defined on $2^\omega \times 2^\omega$ by $d(x,y) = 2^{-\min\{i|x(i)\neq y(i)\}}$. Let D be the set of elements of 2^ω whose bits are almost all zeros. Then there exists a bijection $\gamma : \mathbb{N} \to D$ such that $(2^\omega, d, D, \gamma)$ is a computable metric space. Moreover, it is effectively compact.* ∎

C. Probability measures on the Cantor space

We will work with the **Borel σ-algebra** on 2^ω, which is the smallest algebra containing all open sets of 2^ω.

From the Caratheodory theorem of measure theory, it is straightforward that a Borel probability measure on the Cantor space is uniquely determined by the value it takes on cylinders. We will denote by $\mathcal{M}(2^\omega)$ the set of all Borel probability measures on the Cantor space. For our purposes, the topology we need to consider on the space $\mathcal{M}(2^\omega)$ is the **weak topology**, i.e., the smallest topology such that a sequence (μ_n) of measures converges to a measure μ if and only if $\int f(x)\mu_n(dx)$ converges to $\int f(x)\mu(dx)$ for all bounded continuous functions f on 2^ω. A very important result (see for example [3]) is that for any constructive metric space X, the set $\mathcal{M}(X)$ of probability measures over X is itself a constructive metric spaces. This is not an obvious result, and requires quite a bit of work. However, in the Cantor space, measures have a concise repressentation which simplifies things greatly. Indeed, by Carathedory's theorem, a measure μ on 2^ω is uniquely determined by the values taken by μ on cylinders. Therefore, one can identify $\mathcal{M}(2^\omega)$ with the following (effectively) closed subset of $[0,1]^{2^{<\omega}}$:

$$\mathcal{M}(2^\omega) \equiv \left\{ \mu \in [0,1]^{2^{<\omega}} \;\middle|\; \begin{array}{l} \forall s \in 2^{<\omega} \quad \mu(s) = \\ \mu(s\hat{}0) + \mu(s\hat{}1) \\ \wedge \quad \mu(\epsilon) = 1 \end{array} \right\}$$

a natural countable basis for the topology of $\mathcal{M}(2^\omega)$ is the set of cylinders, where this time cylinders are sets of the form

$$[(s_1, I_1); (s_2, I_2); \ldots; (s_n, I_n)] = \{\mu \mid \forall k\, \mu(s_k) \in I_k\}$$

where the s_i's are strings and the I_k are open sub-intervals of $(0,1)$ with rational endpoints (represented as a pair of rationals). This makes $\mathcal{M}(2^\omega)$ a constructive topological space. As we said above, we could even make it a constructive metric space but this is not needed in the rest of the paper. An important point however is that, as an effectively closed subset of an effectively compact space (namely, $[0,1]^{2^{<\omega}}$), the space $\mathcal{M}(2^\omega)$ is effectively compact.

Proposition II.3. *The space $\mathcal{M}(2^\omega)$, as a constructive topological space described above, is effectively compact.* ∎

With this representation, we also get a very simple characterization of computable measures.

Proposition II.4. *A probability measure μ over 2^ω is computable as a point of $\mathcal{M}(2^\omega)$ if and only if the function $\sigma \mapsto \mu([\sigma])$ is a computable function from $2^{<\omega}$ to $[0,1]$ (respectively endowed with the discrete topology and the canonical topology).* ∎

Recall that the goal of this paper is to study randomness extraction, i.e., we want to investigate way to simulate fair, independent, coin tosses. These correspond to the so-called Lebesgue measure on 2^ω.

Definition II.8. *The Lebesgue measure on 2^ω, or uniform measure, denoted by λ is the unique measure on 2^ω such that $\lambda(s) = 2^{-|s|}$, where $|s|$ denote the size of the string s.* ∎

The Lebesgue measure distributes the weight 1 uniformly on 2^ω. On the other extreme, some measures give a positive probability to *single* points of the space 2^ω. Such elements are called **atoms** of the measure.

Definition II.9. *Let $\mu \in \mathcal{M}(2^\omega)$. Let $x \in 2^\omega$. We say that x is an **atom** of 2^ω for the measure μ if $\mu(\{x\}) > 0$.* ∎

D. Algorithmic randomness

Given a probability measure μ on 2^ω, what does it mean for a single point $x \in, 2^\omega$ to be random with respect to μ? This question is at the root of the field of algorithmic randomness. A satisfactory answer was given by Martin-Löf for the uniform measure, which was later extended by Levin [4] and Gács [3]. Intuitively, we want x to be called random if it avoids all the sets of μ measure 0 which can be effectively tested. The next definition follows Gács [3].

Definition II.10. *A **uniform integrable test** is a lower semi-computable function $t : 2^\omega \times \mathcal{M}(2^\omega) \to [0, +\infty]$ such that for any measure μ we have $\int_{2^\omega} t(x,\mu)\mu(dx) \leqslant 1$. We say that x passes the test with the measure μ if $t(x,\mu)$ is finite. We say that x is μ-Martin-Löf random, or simply μ-random, if x passes all uniform integrable tests with the measure μ. We denote by $\mathbf{MLR}_\mu$ the set of μ-random sequences.* ∎

An important result of Levin [4], Gács [3], Hoyrup and Rojás [5] is that there exists a universal uniform test.

Proposition II.5. *There exists a universal uniform integrable test u, that is, a test u such that for any other uniform integrable test t, there exists a positive constant c such that $u \geqslant \frac{t}{c}$.* ∎

As corollary of this result, we get that a sequence x is μ-random if and only if $u(x, \mu) < \infty$. From this point on, we fix a universal integrable test u. The higher $u(x, \mu)$ is, the "less random" the point x is, relative to the measure μ. Therefore, we sometimes refer to the quantity $u(x, \mu)$ as the **randomness deficiency** of x relative to μ.

Some readers may know another definition of Martin-Löf randomness, involving the so-called Martin-Löf and Solovay tests. The two are in fact equivalent as proven recently by Day and Miller [6][1].

Theorem II.1. *For a given measure μ, and $x \in 2^\omega$, the following are equivalent:*

- *x is μ-random*
- *for every μ-computable sequence (O_n) of open sets such that $\mu(O_n) < 2^{-n}$ for all n, $x \notin \bigcap_n O_n$*
- *for every μ-computable sequence (O_n) of open sets such that $\mu(O_n) < 2^{-n}$ for all n, x belongs to only finitely many O_n.* ∎

Here by μ-computable we mean that each O_n can be computably enumerated relatively to μ and uniformly in n. This terminology might seem somewhat non-standard, but we will see in Section IV that there is a natural way to define the concept of computable open set (which will simply coincide with the notion of effectively open set).

E. Formalizing the initial problem in the setting of algorithmic randomness

Now that we have set the framework in which we consider randomness, we can formalize the initial question of the paper.

Question. What are the classes $\mathcal{C} \subseteq \mathcal{M}(2^\omega)$ of measures allowing uniform randomness extraction, i.e., for which there exists a partial computable function $F : 2^\omega \to 2^\omega$ such that

$$\text{For all } x \in 2^\omega, \text{ if } x \text{ is } \mu\text{-random for some } \mu \in \mathcal{C},$$
$$\text{then } F(x) \text{ is } \lambda\text{-random?}$$

While we do not give a full answer, we provide a criterion for a class of measures to have this property which is quite general and applies to many classical classes of measures (including, of course, the class of Bernoulli measures).

III. The Levin-Kautz conversion procedure

So far we have asked, informally and now formally: can we extract randomness from a μ-random sequence x when we have only partial information about the underlying measure μ? We have not yet said what happens if we *do* have full information about μ. In other words, what if we are given an x which is μ-random for some *computable* μ? The next theorem, due to Levin [7] and Kautz [8] gives a full answer to this question: for any computable measure μ, there exists a non-decreasing, in the sense of the lexicographic order, partial computable function $\phi : 2^\omega \to 2^\omega$, defined λ-almost everywhere, such that $\mu = \lambda \circ \phi^{-1}$. As a consequence, ϕ^{-1} transforms every μ-random element into a λ-random element, except for μ-atoms.

Theorem III.1. *Let μ be a computable probability measure on 2^ω. There exists a non-decreasing partial computable function $\phi : 2^\omega \to 2^\omega$ such that ϕ is defined λ-almost everywhere, and for all Borel sets $A \subseteq 2^\omega$, $\mu(A) = \lambda(\phi^{-1}(A))$.* ∎

Corollary III.1. *Let μ be a computable probability measure on 2^ω. There exists a partial computable function $\psi : 2^\omega \to 2^\omega$ such that for every μ-random sequence x which is not an atom of μ, $\psi(x)$ is λ-random.* ∎

This corollary is obtained by taking $\psi = \phi^{-1}$, where ϕ is the partial computable function of the previous theorem. It is not too difficult (although it still requires a bit of work as we will see below) that this function ψ works.

In fact, one can show that the Levin-Kautz conversion procedure is uniform in the measure. Hence one gets a stronger version of Corollary III.1.

Theorem III.2. *There exists a partial computable function $\Gamma : 2^\omega \times \mathcal{M}(2^\omega) \to 2^\omega$ such that if x is μ-random and is not an atom of μ, then $\Gamma(x, \mu)$ is λ-random.* ∎

For completeness, let us present the proof of this theorem.

Proof: For any $\mu \in \mathcal{M}(2^\omega)$, let us build uniformly in μ a partial function $\varphi_\mu : 2^\omega \to 2^\omega$ such that

- φ_μ is μ-computable everywhere on its domain, but on countably many points.
- $\forall s \in 2^{<\omega}$, $\mu([s]) = \lambda(\varphi_\mu^{-1}([s]))$.
- φ_μ is increasing with respect to $<_L$, the lexicographic order on 2^ω.
- The images by φ_μ of points on which φ_μ is not μ-computable are exactly the elements of 2^ω containing finitely many 1.
- If y is not μ-computable then $\varphi_\mu^{-1}(y)$ is a single point and is uniformly computable from y. ∎

For all $n \in \mathbb{N}^*$, let $s_{n,i}$ be the i-th element of $(\{0,1\}^n, <_L)$. We define recursively $a_{n,0} = 0$ and $a_{n,i+1} = a_{n,i} + \mu([s_{n,i}])$. For all $n \in \mathbb{N}$, we have a sequence $0 = a_{n,0} \leqslant a_{n,1} \leqslant \cdots \leqslant a_{n,2^n} = 1$ such that intervals $([a_{n,i}, a_{n,i+1}))_{0 \leqslant i \leqslant 2^n}$ form a partition of $[0,1)$. An interval of the form $[a, a + \mu([s]))$ will be denoted by A_s. Then we define:

$$\psi : \left\{ \begin{array}{ccc} [0,1[& \to & 2^\omega \\ x & \mapsto & \bigcap\limits_{\substack{s \in 2^{<\omega} \\ x \in A_s}} [s] \end{array} \right.$$

[1] the result presented here is a reformulation of the Day-Miller theorem in the context of computable open sets we developed

Let us show that ψ is well defined. We have:

$$\begin{array}{llll} \forall s \forall t \in 2^{<\omega} & A_s \subseteq A_t & \leftrightarrow & [s] \subseteq [t] \\ \forall s \forall t \in 2^{<\omega} & x \in A_s \wedge x \in A_t & \rightarrow & A_t \subseteq A_s \vee A_s \subseteq A_t \end{array}$$

This implies that the intersection is a decreasing intersection of a family of non-empty compact sets. So it is not empty. Since the measure of the space is always 1, we have: $\forall n \quad \exists s \in \{0,1\}^n \ x \in A_s$. Finally there are some s of arbitrarily long size such that $x \in A_s$, which implies that the intersection of all those strings contain only one point. So $\psi(x)$ is well defined.

Let us show that ψ is μ-computable everywhere but on countably many points. Let x be such that $\forall n \in \mathbb{N}^* \quad \forall m \leqslant 2^n \quad x \neq a_{n,m}$. To compute $\psi(x)$ it is enough to approximate more and more all the intervals A_s. If x is in one interval without being at the border, we will know it in a finite amount of time. The set $(A_s)_{s \in 2^{<\omega}}$ is countable so ψ is computable everywhere except on countably many points. Furthermore if $x = a_{n,m}$ and $A_s = [a_{n,m}, a_{n,m+1})$ then $\forall n \ \ x \in A_{s\hat{}0^n}$ and so its image is equal to the string s completed with only zeros.

Let us show that ψ is a morphism (in the measure-theoretic sense). By construction ψ is a morphism from $([0,1], \lambda, \mathcal{B}([0,1]))$ to $(2^\omega, \mu, \mathcal{B}(2^\omega))$. Indeed $\forall n \in N^* \ \forall s \in \{0,1\}^n \quad \psi^{-1}([s]) = A_s$. If $A_s = [a_i, a_{i+1})$ we have $\lambda(A_s) = a_{i+1} - a_i = \mu([s])$ and so $\mu([s]) = \lambda(\psi^{-1}([s]))$. Our measure $\lambda \circ \psi^{-1}$ is equal to the measure μ on all basic open sets of the Cantor space. So by Caratheodory's theorem we can extend the equality to any Borel set.

It is easy to see that ψ is increasing with respect to the lexicographic order. Let us show that if y is not μ-computable then $\psi^{-1}(y)$ is a single point and is uniformly computable from y. First we know that if y is an atom for the measure μ then y is μ-computable. Indeed, if $\mu(y) > x > 0$ then there exists an open neighborhood of y such that y is the only point of the neighborhood with measure greater than x. Then it is enough to search for all basic open set included in this neighborhood such that their measure is greater than x. Thus if y is not μ-computable it cannot be an atom of the measure μ. But since ψ is an increasing function for lexicographic order, it cannot have more than one pre-image on y, otherwise y would be an atom of the measure. To compute $\psi^{-1}(y)$ it is enough to compute ψ on all elements of $2^{<\omega}$. Only one path can give y and for all other paths, we know in a finite amount of time if it does not give y.

Now let $f : 2^\omega \to [0,1]$ be the usual identification of a binary sequence s with the real in $[0,1]$ having binary expansion s. We define $\varphi_\mu : 2^\omega \to 2^\omega$ by $\varphi_\mu(x) = \psi(f(x))$. f is total computable so $\varphi(x)$ is μ-computable for every point such that ψ can compute its image by f. Each point of $[0,1)$ has at most two pre-images by f, so φ_μ is computable everywhere but on countably many points. Since f is an increasing function with respect to lexicographic order then so is φ_μ. We easily see that φ_μ^{-1} is still a well defined and computable function on all non μ-computable points. To conclude, f and ψ are morphisms, then so is φ_μ.

Now, to get the functional Γ as in the theorem, set $\Gamma(x, \mu) = \varphi_\mu^{-1}(x)$. Suppose now that x is μ-random and is not an atom of μ. Then, $y = \varphi_\mu^{-1}(x)$ exists (as we have shown). Let us show that it is λ-random. Suppose not and take a test as in theorem II.1 such that $y \in \bigcap_n O_n$. We have that $\varphi_\mu(y) \in \bigcap \varphi_\mu(O_n^c)^c$. Since O_n^c is effectively closed, it is effectively compact. The image of an effectively compact by a μ-computable function is a μ-effectively compact. In the Cantor space, effectively compact sets are effectively closed sets and so $\varphi_\mu(O_n^c)^c$ is a μ-effectively open set[2] Besides we have that $V_i^c \subseteq \varphi_\mu^{-1}(\varphi_\mu(V_i^c))$ and so $\varphi_\mu^{-1}(\varphi_\mu(V_i^c))^c \subseteq V_i$. Then:

$$\begin{array}{rrcl} & \lambda(\varphi_\mu^{-1}(\varphi_\mu(V_i^c))^c) & \leqslant & \lambda(V_i) \\ \Rightarrow & \mu(\varphi_\mu(V_i^c)^c) & \leqslant & \lambda(V_i) \\ \Rightarrow & \mu(\varphi_\mu(V_i^c)^c) & \leqslant & 2^{-i} \end{array}$$

Which implies, using theorem II.1 that $\varphi_\mu(y) = x$ is not μ-random, which is a contradiction.

IV. Learnability of probability measures

In view of the Levin-Kautz theorem, one can make the following (informal) conjecture regarding the original problem: suppose we had a class of measures $\mathcal{C}$ such that from any sequence x which is μ-random for some $\mu \in \mathcal{C}$, one could computably "guess" (in a uniform way) which measure of that class x is random relative to. Then, applying the Levin-Kautz procedure, one could computably transform x into a λ-random sequence and therefore the class $\mathcal{C}$ would allow uniform randomness extraction. The Bernoulli measures are an example of such measures: from a sequence x which is B_p-random for some p, one can compute p: indeed the law of large numbers tells us that the frequency of zeroes and ones tends to p, and the law of iterated logarithm gives a bound on the speed of convergence (it is well known that Martin-Löf random sequences satisfy both laws). The computation of p is not completely uniform however. Consider the following sequence:

$$x = 0011010101110000100\ldots$$

which we know is B_p random for some p. Can we say anything about p after having read the first twenty bits? No: there are two possible scenarios (a) either p is close to $1/2$, or (b) p is not close to $1/2$ but the first twenty bits are very atypical. Reading more bits will not help; one cannot, computably and after reading finitely many bits of x, distinguish between the two cases, therefore p cannot be computed uniformly.

[2]Technically speaking the function ϕ_μ is not total, it is defined up to a computable sets of points, which essentially allows us to treat the function ϕ_μ as total.

Suppose on the other hand that we had an upper bound on the randomness deficiency of $u(x, B_p)$. Then after reading sufficiently many bits, scenario (b) can be ruled out. This idea is precisely the base for the theory of *layerwise computable* functions introduced by Hoyrup and Rojás [9] (similar ideas can be found in Eddalat [10], [11]). A function is said to be μ-layerwise computable if it is defined on all μ-random sequences, and it is uniformly computable up to an "advice" on an upper bound for the randomness deficiency of x.

Definition IV.1. *Let μ be a probability measure on 2^ω. Let Y be a computable metric space. A function $f : 2^\omega \to Y$ is said to be **layerwise computable** if*

- *it is defined on all μ-random sequences*
- *there exists a partial computable function $F : 2^\omega \times \mathbb{N} \to Y$ such that, for all $x \in 2^\omega$ and $c \in \mathbb{N}$, if $u(x, \mu) < c$, then $F(x, c) = f(x)$* ∎

Definition IV.2. *A measure $\mu \in \mathcal{M}(2^\omega)$ is (layerwise) learnable if there exists a total computable function $F : 2^\omega \times \mathbb{N} \to \mathcal{M}(2^\omega)$ such that:*

$$\forall x \in 2^\omega \ \forall c \in \mathbb{N} \ \ u(x, \mu) \leqslant c \to F(x, c) = \mu$$

*A class of measure $\mathcal{C} \subseteq \mathcal{M}(2^\omega)$ is **(layerwise) learnable** if there exists a computable function $F : 2^\omega \times \mathbb{N} \to \mathcal{M}(2^\omega)$ such that:*

$$\forall \mu \in C \ \ \forall x \in 2^\omega \ \forall c \in \mathbb{N} \ \ u(x, \mu) \leqslant c \to F(x, c) = \mu$$ ∎

Theorem IV.1. *Let $\mathcal{C}$ be an effectively compact class of effectively orthogonal measures. Then $\mathcal{C}$ is learnable.* ∎

Proof: Suppose that $x \in \mathbf{MLR}_\mu$ for $\mu \in \mathcal{C}$ and suppose that $u(x, \mu) \leqslant c$. Since u is lower semi-computable, the set

$$\{(z, \nu) \in 2^\omega \times \mathcal{M}(2^\omega) |\ u(z, \nu) \leqslant c\}$$

is an effectively closed set, and thus the set:

$$A_c = \mathcal{C} \cap \{\nu \mid u(x, \nu) \leqslant c\}$$

is x-effectively compact, uniformly in x, as $\mathcal{C}$ is effectively compact. To see this, notice that the set $\{\nu \mid u(x, \nu) \leqslant c\}$ is effectively compact relative to x. This follows from the fact (relativized to x) that the image of an effectively compact set under a computable function (here the projection over x) is an effectively compact set, and this is is uniform, i.e. from a code of an effectively compact set one can compute a code for its image. Of course the set $\mathcal{C}$ is effectively compact, therefore also effectively compact relative to x, therefore the intersection A_c is effectively compact relative to x, and a code for A_c can be found uniformly in x and c.

Moreover, since $\mathcal{C}$ is effectively orthogonal, A_c contains only one point, which is μ. It remains to use the fact that if an effectively compact set A contains only one element, this element is computable uniformly in the code for A (see Lemma VIII.1 in the appendix). ∎

In fact, Theorem IV.1 is a *characterization* of learnability, which is both rather surprising and much more difficult to prove.

Theorem IV.2. *If $\mathcal{D} \subseteq \mathcal{M}(2^\omega)$ be a class of layerwise learnable measures, then $\mathcal{D}$ is contained in an effectively compact class of effectively orthogonal measures.* ∎

The remainder of this section will be dedicated to the proof of Theorem IV.2. Before we present the core of the argument, we will need some preliminary discussion. To prove Theorem IV.2, we want to be able to say that the function which takes as input a measure μ and an open set O and returns $\mu(O)$ is lower semi-computable in the input (μ, O). However, it is not clear how to make sense of this. To do so, we need to consider the *space of open sets of* 2^ω as a constructive topological space. This is achieved by taking as basis the sets are the set of open sets containing one given cylinders. Formally we denote by $\mathcal{T}(2^\omega)$ the set of open sets of 2^ω and we denote:

$$[\![s]\!] = \{O \in \mathcal{T}(2^\omega) |\ [s] \subseteq O\}$$

It is easy to see that $\{[\![s]\!] |\ s \in 2^{<\omega}\}$ is a subbasis for a T_0, second countable topology. The basis is given by finite intersections of elements of the subbasis. Then we have the following proposition:

Proposition IV.1. *Let $O \subseteq 2^\omega$ be an open set. Then O is effectively open if and only if it is a computable point of $\mathcal{T}(2^\omega)$.* ∎

Proof: Let $O = \bigcup_{i \in \mathbb{N}} [s_i]$ be an effectively enumerable open set with its enumeration $\{[s_i]\}_{i \in \mathbb{N}}$. One can compute an enumeration $\{[t_i]\}_{i \in \mathbb{N}}$ of all basic open sets of 2^ω included in one of the $[s_i]$. Then all intersections of finite subsequences of $\{[\![t_i]\!]\}_{i \in \mathbb{N}}$ are an effective enumeration of basic open sets of $\mathcal{T}(2^\omega)$ containing O. The converse is similar. ∎

We can use the two previous constructive topological spaces in order to state this useful proposition:

Proposition IV.2. *The evaluation map $ev : \mathcal{M}(2^\omega) \times \mathcal{T}(2^\omega) \to \mathbb{R}$ defined by $ev(\mu, O) = \mu(O)$ is lower semi-computable.* ∎

Proof: Let $r \in [0, 1]$. Let $\mathcal{O}(\mathcal{M}(2^\omega))$ denote the set of basic open sets of $\mathcal{M}(2^\omega)$ and $\mathcal{O}(\mathcal{T}(2^\omega))$ denote the set of basic open sets of $\mathcal{T}(2^\omega)$. Let $A_r =$

$$\left\{ \begin{array}{l|l} \begin{pmatrix} [(s_1, I_1); \ldots; (s_n, I_n)] \\ [[\![s_1]\!]; \ldots; [\![s_n]\!]] \end{pmatrix} \in \begin{array}{l} \mathcal{O}(\mathcal{M}(2^\omega)) \times \\ \mathcal{O}(\mathcal{T}(2^\omega)) \end{array} & \\ \\ n \in \mathbb{N}, \ \ (\sum_{k \leqslant n} \min(I_k)) > r & \end{array} \right\}$$

be an effective enumeration of basic open sets of $\mathcal{M}(2^\omega) \times \mathcal{T}(2^\omega)$ such that any measure in the basic open set has a value strictly greater than r on the corresponding basic open set of 2^ω. The enumeration A_r is effectively uniform in r. Let us show that $A_r = ev^{-1}((r;1])$. We trivially have $A_r \subseteq ev^{-1}((r;1])$. We now show that $ev^{-1}((r;1]) \subseteq A_r$. Suppose that for $\mu \in \mathcal{M}(2^\omega)$ and $O \in \mathcal{T}(2^\omega)$ we have $\mu(O) > r$. Let us show that $(\mu, O) \in A_r$. Since $\mu(O) > r$ then there exists $n \in \mathbb{N}$ and $\{s_i\}_{i \leqslant n}$ a finite list of basic open sets included in O such that $\sum_i \mu([s_i]) > r$. Besides one can find n rational numbers $\{q_i\}_{i \leqslant n}$ such that $q_i < \mu([s_i])$ and $\sum_i q_i > r$. Then (μ, O) belongs to the open sets $[(s_0,(q_0,1)); (s_1,(q_1,1)); \dots; (s_n,(q_n,1))]$ which is in A_r. ∎

We can now prove Theorem IV.1.

Proof: Let $F : 2^\omega \times \mathbb{N} \to \mathcal{M}(2^\omega)$ be a partial computable function withnessing the learnability of the class $\mathcal{D}$, i.e., a function such that $\forall \mu \in \mathcal{D} \;\; \forall x \in 2^\omega \;\; \forall c \in \mathbb{N} \;\; t(x,\mu) \leqslant c \to F(x,c) = \mu$. We define:

$$\mathcal{C} = \left\{ \mu \in \mathcal{M}(2^\omega);\ \forall c \in \mathbb{N} \;\; \mu(\{x;\ F(x,c) = \mu\}) > 1 - \frac{1}{c} \right\}$$

Let us show that $\mathcal{D} \subseteq \mathcal{C}$. We know that for any μ we have:

$$\int t(x,\mu)\mu(dx) \leqslant 1$$

So using the Chebychev inequality we can deduce that:

$$\begin{array}{llll} \forall \mu \in \mathcal{M}(2^\omega) \;\; \forall c \in \mathbb{N} & \mu(\{x;\ t(x,\mu) \geqslant c\}) & \leqslant & \frac{1}{c} \\ \forall \mu \in \mathcal{M}(2^\omega) \;\; \forall c \in \mathbb{N} & \mu(\{x;\ t(x,\mu) < c\}) & > & 1 - \frac{1}{c} \end{array}$$

From the hypothesis we know that:

$$\forall \mu \in \mathcal{D} \;\; \forall c \in \mathbb{N} \;\; \{x;\ t(x,\mu) < c\} \subseteq \{x;\ F(x,c) = \mu\}$$

Then we have:

$$\forall \mu \in \mathcal{D} \;\; \forall c \in \mathbb{N} \;\; \mu\big(\{x;\ F(x,c) = \mu\}\big) > 1 - \frac{1}{c}$$

So $\mathcal{D} \subseteq \mathcal{C}$.

Let us now show that $\mathcal{C}$ is an effectively closed set. For this, we show that $\overline{\mathcal{C}}$, the complement of $\mathcal{C}$, is an effectively open set. We have:

$$\begin{array}{rcl} \mathcal{C} & = & \big\{\mu \in \mathcal{M}(2^\omega);\ \forall c \;\; \mu(\{x;\ F(x,c) = \mu\}) > 1 - \frac{1}{c}\big\} \\ & = & \big\{\mu \in \mathcal{M}(2^\omega);\ \forall c \;\; \mu(\{x;\ F(x,c) \neq \mu\}) \leqslant \frac{1}{c}\big\} \end{array}$$

and thus

$$\overline{\mathcal{C}} = \big\{\mu \in \mathcal{M}(2^\omega);\ \exists c \;\; \mu(\{x;\ F(x,c) \neq \mu\}) > \tfrac{1}{c}\big\}$$

We need to show that $\{x;\ F(x,c) \neq \mu\}$ is effectively open relative to μ. Suppose $F(y,c) \neq \mu$. Since F is computable, having an enumeration of open sets containing μ, we will know in a finite time, with a finite prefix t of y if $F(y,c) \neq \mu$. So $y \in [t] \subseteq \{x;\ F(x,c) \neq \mu\}$ which makes the required set μ-effectively open. We can now define:

$$\gamma : \left\{ \begin{array}{rcl} \mathcal{M}(2^\omega) \times \mathbb{N} & \to & \mathcal{T}(2^\omega) \\ (\mu, c) & \mapsto & \{x;\ F(x,c) \neq \mu\} \end{array} \right.$$

Since $\gamma(\mu,c)$ is μ-effectively open, by Proposition IV.1, $\gamma(\mu,c)$ is μ-computable and so γ is computable. Recall the definition of the evaluation map:

$$ev : \left\{ \begin{array}{rcl} \mathcal{M}(2^\omega) \times \mathcal{T}(2^\omega) & \to & \mathbb{R} \\ (\mu, B) & \mapsto & \mu(B) \end{array} \right.$$

By Proposition IV.2, ev is lower semi-computable. For a fixed $c \in \mathbb{N}$, we now define:

$$\delta_c : \left\{ \begin{array}{rcl} \mathcal{M}(2^\omega) & \to & \mathbb{R} \\ \mu & \mapsto & ev(\mu, \gamma(\mu,c)) \end{array} \right.$$

As a composition of lower semi-computable functions, δ_c is lower semi-computable. Therefore:

$$\begin{array}{rcl} \overline{\mathcal{C}} & = & \big\{\mu \in \mathcal{M}(2^\omega);\ \exists c \in \mathbb{N} \;\; \mu(\{x;\ F(x,c) \neq \mu\}) > \frac{1}{c}\big\} \\ & = & \bigcup_{c \in \mathbb{N}} \big\{\mu \in \mathcal{M}(2^\omega);\ \delta_c(\mu) > \frac{1}{c}\big\} \\ & = & \bigcup_{c \in \mathbb{N}} \delta_c^{-1}((\frac{1}{c}, 1]) \end{array}$$

We deduce that this is a countable union of effectively open sets. It is not hard to see that the open sets are uniformly effective in c.

Let us show that the class is effectively orthogonal. Suppose that $x \in \mathbf{MLR}_\mu$ and consider the sets $V_{\mu,c} = \{x;\ F(x,2^c) \neq \mu\}$. As shown above, all $V_{\mu,c}$ are μ-effectively open sets, uniformly in c. Furthermore we have that $\mu(V_{\mu,c}) \leqslant 2^{-c}$. From Theorem II.1, since x is μ-random, it can only belong to finitely many sets $V_{\mu,c}$. So for any other measure $\nu \in C$ we have that $x \in V_{\nu,c}$ for infinitely many c and so $x \notin \mathbf{MLR}_\nu$. Then the class is effectively orthogonal. ∎

V. On the possibility of randomness extraction

We can now put everything together to get, as promised, a generalization of von Neumann's trick to more general classes of measures.

Theorem V.1. *Given a class $\mathcal{C}$ of measures on 2^ω, if $\mathcal{C}$ is Π^0_1 and has the effective orthogonality property, then uniform randomness extraction is possible on $\mathcal{C}$-random sequences. That is, there exists a partial computable function $E : 2^\omega \to 2^\omega$ such that for every sequence x, if x is μ-random for some $\mu \in \mathcal{C}$ and x is not an atom of μ, then $E(x)$ is λ-random.* ∎

Proof: On an input $x \in 2^\omega$, the function E does the following:

(i) It assumes that x is random with respect to some $\mu \in \mathcal{C}$, and starts by making a "guess" on the value of the randomness deficiency of x with respect to μ. Say it starts with $c = 1$.

(ii) It computes a code for the Π^0_1 subset of $\mathcal{C}$

$$A_c = \mathcal{C} \cap \{\nu \mid u(x,\nu) \leqslant c\}$$

(iii) The set A_c is effectively compact relative to x and, under the assumption that x is random; is either (a) empty or (b) contains only one element. Thus E runs the procedure described in $\Gamma(x,\mu)$ described in the Levin-Kautz conversion theorem (Theorem III.2) and computes an element μ such that, if we are in case (b), $\mathcal{A}_c = \{\mu\}$

(iv) It then runs the Levin-Kautz conversion procedure on the pair (x,μ) and produces a binary sequence (i.e. starts outputting bits one by one).

(v) Meanwhile, E keeps enumerating the complement of A_c. If A_c is in fact empty, then by effective compactness, this will be recognized at some finite stage. In this case, interrupt the running procedure (iv), increases c by 1 and goes back to step (ii).

Let us show that this algorithm is correct. Assume x is μ-random for some $\mu \in \mathcal{C}$ with randomness deficiency at most d, and is not an atom of μ. For all $c < d$, the class A_c will be empty hence for each such c the production of bits performed by step (iv) will be interrupted by (v). Hence the variable c in the algorithm will eventually reach the value d, and by that time, only a finite string σ has been produced in the output. When the variable c reaches the value d, by construction, we have $A_c = \{\mu\}$, thus E correctly computes the measure μ at step (iii) and therefore step (iv) correctly produces a λ-random sequence z. Therefore the full output of the algorithm is the sequence σz. Using the classical fact that λ-randomness is invariant under finite changes of bits (adding finitely many bits, deleting finitely many bits, or replacing finitely many bits), this shows that σz is random as z is by construction. ∎

One can even extend the previous theorem to *computable unions* of Π^0_1 classes with the orthogonality property.

Theorem V.2. *The conclusion of Theorem V.1 still holds under the weaker assumption that $\mathcal{C}$ is a computable union $\bigcup_m \mathcal{C}_m$ of Π^0_1 classes all of whom have the effective orthogonality property. In particular Theorem V.1 holds if $\mathcal{C}$ is a Σ^0_2 class with the orthogonality property.* ∎

Proof: Let $(\mathcal{C}_m)$ be an enumeration of the Π^0_1 classes whose union is $\mathcal{C}$. Without loss of generality, assume that every element $\mathcal{C}_m$ is repeated infinitely often in this enumeration. Now, modify the algorithm of the previous proof as follows: in step (ii), the "search space" A_c is taken to be equal to $\mathcal{C}_c \cap \{\nu \mid u(x,\nu) \leqslant c\}$. This is enough because if $u(x,\mu) < d$ for some $\mu \in \mathcal{C}$, then $\mu \in C_m$ for some m which (because of the repetition assumption), can be taken as large as wanted, in particular bigger than d. Then, $x \in A_m$, and hence the algorithm will eventually run step (iv) forever, without being interrupted by step (v). And of course, as all the $\mathcal{C}_m$ have the effective orthogonality property, the variable c eventually reaches a value such that A_c is exactly a singleton. The rest of the proof is the same. ∎

In terms of the arithmetical complexity of the class $\mathcal{C}$, Theorem V.2 is optimal, i.e., it does not hold under the assumption that $\mathcal{C}$ is Π^0_2 instead of Σ^0_2. Indeed, Levin proved that there exists on 2^ω a *neutral measures*, i.e., a measures ν such that *all* sequences $x \in 2^\omega$ are random with respect to ν. No neutral measure is computable, but there are neutral measures that are computable using **0'**, the halting set, as an oracle. It is well-known that if z is a **0'**-computable element of a computable metric space, then the singleton $\{z\}$ is a Π^0_2 subset of that space. Therefore, take the class $\mathcal{C} = \{\nu\}$ where ν is a **0'**-computable neutral measure. Then $\mathcal{C}$ is Π^0_2 and (obviously!) has the effective orthogonality property. Now take a 1-generic sequence[3] $x \in 2^\omega$ which is not an atom of ν (there are uncountable many 1-generic and only countably many atoms). Thus x is ν-random (by definition of a neutral measure), is not an atom of ν, and there is no partial computable function E such that $E(x)$ is λ-random, as no 1-generic sequence computes a λ-random real (see [12, Exercise 4.1.6]).

VI. An example: Markov chains

We now give a concrete example of Σ^0_2 effectively orthogonal class of measure on which we can extract randomness: the set of Markov measures. A Markov measure on 2^ω is a measure such that the relative probability of each bit of the sequence to be 0 only depends on the values of the m-th previous bits for a fixed constant m. Formally, μ is a Markov chain if there exists an integer m such that for all strings σ_1, σ_2, for all strings τ of length m

$$\mu(\sigma_1\hat{}\tau\hat{}0)\mu(\sigma_2\hat{}\tau) = \mu(\sigma_2\hat{}\tau\hat{}0)\mu(\sigma_1\hat{}\tau)$$

It is clear from this definition that the class of Markov measures on 2^ω is Σ^0_2. Unfortunately, it does not have the effective orthogonality property, so we cannot directly apply the results of the previous section. However, we can overcome this problem by finding a computable union $\mathcal{C} = \bigcup_m \mathcal{C}_m$ of effectively closed classes with the effective orthogonality property such that any sequence x which is random with respect to some Markov measure μ, is ν-random for some $\nu \in \mathcal{C}$.

This is done using some classical results on the theory of Markov chains. Let us begin by a few definitions.

Definition VI.1. *A matrix $P \in \mathbb{M}_m(\mathbb{R})$ is called a **stochastic matrix** if its entries are non-negative and the sum of entry of each row is 1 i.e. $\forall 0 < i,j \leqslant m \;\; P_{i,j} \geqslant 0$ and $\forall 0 < i \leqslant m$ we have $\sum_{j=1}^m P_{i,j} = 1$. Analogously, a **stochastic vector** is a vector taking its values in $[0,1]$ and such that the sum of its element is 1.* ∎

[3] see for example [12] for a definition of such reals

A stochastic matrix, together with a stochastic vector, both of size 2^m, uniquely define a Markov chain of memory m over the Cantor space. The case i-th coordinate of the stochastic vector corresponds to the probability that our process starts with the i-th element of 2^m, while the entry at the i-th line and j-th column of the matrix gives us the probability that our process produces the j-th element of 2^m, knowing that it just produced the i-th element of 2^m. If P is a stochastic matrix of size 2^m we will write $P(s,t)$ to denote $P_{i,j}$ where s is the i-th element of 2^m and t is the j-th element of 2^m (say in the lexicographic order).

As we will see, this class is not effectively orthogonal by itself, but we can find a Σ_2^0 subclass of effectively orthogonal Markov measures such that any element random with respect to some measure in the whole class is also random with respect to a measure in the subclass. In order to introduce Markov measures we first need a few notions. In this section, $\mathbb{M}_m(X)$ will denote the space of square matrix of size m taking their values in X. $\mathbb{V}_m(X)$ will denote the space of vector of size m. If $P \in \mathbb{M}_m(X)$ we write $P_{i,j}$ to denote the value of P on the i-th line and the j-th column. If $V \in \mathbb{V}_m(X)$ we write V_i to denote the value of V on the i-th line. By convention we start the indexation by 1. In our case we will only be interested in finite Markov processes and so in finite stochastic matrices.

Definition VI.2. *Let $N \in \mathbb{N}$. Let $P \in \mathbb{M}_{2^N}(\mathbb{R})$ be a stochastic matrix. Let π be a stochastic vector of size 2^N. Then $\mu \in \mathcal{M}(2^\omega)$ is the **Markov measure** with **memory** N associated to P and π if $\forall m \quad \forall x \in \{s \in 2^{<\omega};\ \ |s| = mN\}$ we have:*

$$\mu(x) = \pi(x_1) P(x_2, x_3) \dots P(x_{m-1}, x_m)$$

where $(x_i)_{i \leqslant m}$ denotes $x(iN+1) \dots x((i+1)N)$. ∎

Proposition VI.1. *Let $N \in \mathbb{N}^*$. The set $Mark_N$ of Markov measures with memory N is effectively compact.* ∎

Proof: Let $X = \bigcup_{m \in \mathbb{N}} 2^{mN}$ be the set of all finite strings having a multiple of N as length. We have that $\mu \in Mark_N$ iff

$$\begin{array}{l} \forall x_1, x_2 \in X \quad \forall s \in 2^N \quad \forall t \in 2^N \\ (\mu(x_1\hat{}s) \neq 0 \wedge \mu(x_2\hat{}s) \neq 0) \rightarrow \frac{\mu(x_1\hat{}s\hat{}t)}{\mu(x_1\hat{}s)} = \frac{\mu(x_2\hat{}s\hat{}t)}{\mu(x_2\hat{}s)} \\ \leftrightarrow \\ \forall x_1, x_2 \in X \quad \forall s \in 2^N \quad \forall t \in 2^N \\ \mu(x_1\hat{}s\hat{}t)\mu(x_2\hat{}s) = \mu(x_2\hat{}s\hat{}t)\mu(x_1\hat{}s) \end{array}$$

which is a closed condition. As the set of measure is effectively compact, then $Mark_N$ is also effectively compact. ∎

Now in order to extract an effectively orthogonal subclass of $Mark_N$, we introduce the notion of a state in a stochastic matrix.

Definition VI.3. *Let $P \in \mathbb{M}(\mathbb{R})$ be a stochastic matrix. A state j is said to be **accessible** from a state i if $\exists n \ \ (P^n)_{i,j} > 0$. We write $\boldsymbol{j \leftarrow i}$ if j is accessible from i. We write $\boldsymbol{j \leftrightarrow i}$ for the equivalence relation $(j \leftarrow i) \wedge (i \leftarrow j)$. An equivalence class H for the equivalence relation $\leftrightarrow$ is called an accessibility class is said to be **final** if any other equivalence class is inaccessible from it.* ∎

Two different issues can make Markov measures to share their random. The first issue is due to final accessibility classes. Only the states in such classes will matter for the randomness of a sequence, since with probability 1, the process will eventually fall into one of them. But if a Markov measure has n several final accessibility classes, its random sequences will be exactly the union of all randoms for n simpler Markov measures, each of them having one of the final accessibility class.

So we can restrict ourselves to the Markov measures having one final accessibility class.

The second issue is to deal with states which are not in a final accessibility class. This is done by replacing all these states by only one initial state, which will force an arbitrary large number of bit to have a fixed value. This way for any possible random sequence starting by a string depending on a non final accessibility class, it will be random for the Markov chain having the right corresponding initial and final states.

Proposition VI.2. *Let $N \in \mathbb{N}^*$. Let $C \subseteq 2^N$. Let $s \in (2^N \backslash C) \cup \{\epsilon\}$. The class $Mark_{N,C,p}$ of Markov measures such that:*

- *The first N bits are fixed to the string p, i.e. $\mu(p) = 1$*
- *The set C is an final accessibility class such that there is a probability of $\frac{1}{|C|}$ to go from s to each state in C*

is a Σ_2^0 class of measures. ∎

Proof: Let $X = \bigcup_{m \in \mathbb{N}} 2^{mN}$ be the set of all finite strings having a multiple of N as length. By definition, the class $Mark_{N,C,p}$ can be defined by three conditions:

$$\begin{array}{ll} 1: & \forall r \in C \quad \forall s \in C \quad \exists t_1, \dots, t_N \in C \\ & \mu(p\hat{}r\hat{}s) \neq 0 \vee \dots \vee \mu(p\hat{}r\hat{}t_1 \dots t_n\hat{}s) \neq 0 \end{array}$$

$$\begin{array}{ll} 2: & \forall r \in C \quad \forall s \notin C \quad \forall t_1, \dots, t_N \in 2^N \\ & \mu(p\hat{}r\hat{}s) = 0 \wedge \dots \wedge \mu(p\hat{}r\hat{}t_1 \dots t_n\hat{}s) = 0 \end{array}$$

$$3: \quad \mu(p) = 1 \wedge \forall t \in C \quad \mu(p\hat{}t) = \frac{1}{|C|}$$

(conditions 1 and 2 express the fact that C is a final accessibility class).

It is clear that conditions 2 and 3 are effectively closed. As to condition 1, it can be rewritten as:

$$\begin{array}{ll} 1': & \exists q \in]0,1] \cap \mathbb{Q} \quad \forall r \in C \quad \forall s \in C \quad \exists t_1, \dots, t_N \in C \\ & \mu(p\hat{}r\hat{}s) \geqslant q \vee \dots \vee \mu(p\hat{}r\hat{}t_1 \dots t_n\hat{}s) \geqslant q \end{array}$$

and therefore is a Σ_2^0 condition. ∎

We now prove the desired proposition:

Proposition VI.3. *The class*

$$Mark = \bigcup \{Mark_{N,C,s} |\ N \in \mathbb{N}^*\ C \subseteq 2^{\mathbb{N}}\ s \in (2^N \backslash C) \cup \{\epsilon\}\}$$

is is an effectively orthogonal Σ_2^0 class. Furthermore any sequence random for an extended Markov measure will be random for a measure in $Mark$. ∎

Proof: We can easily see that the class is Σ_2^0. Now Let $N \in \mathbb{N}^*$ and let

$$Mark_N = \bigcup_{\substack{C \in 2^N \\ s \in (2^N \backslash C) \cup \{\epsilon\}}} Mark_{N,C,s}$$

Note that $Mark = \bigcup_{N \in \mathbb{N}} Mark_N$ is an increasing sequence, indeed we have $Mark_N \subseteq Mark_{N+1}$. So it is enough to show that for any $N \in \mathbb{N}^*$, $Mark_N$ is orthogonal. If two different measures $\mu, \nu \in Mark_N$ do not have the same N bits fixed, they are orthogonal since they have disjoint support. If they have the same first N bits fixed then some conditional probability to go from one state i to another state j will differ between μ and ν, and this will be reflected by their random elements. That is, the relative frequency of occurrence of the patter t after the pattern s, with t the j-th string and s the i-th string, will have some fixed value $f_\mu = P^\mu_{i,j}$ on all μ-random elements, and $f_\nu = P^\nu_{i,j}$ on ν-random elements. Since these two values differ, μ and ν will have no common random element.

Finally, the fact that any sequence random for a Markov measure will be random for a measure in $Mark$ comes from our construction and was discussed previously: the states which do not belong to any final equivalent class is handled by fixing the corresponding N first bits, and the fact that one Markov measure can have several final accessibility class is handled by the fact that their random will be the union of those of the corresponding measures in $Mark$, each of them having one of the final accessibility class. ∎

Corollary VI.1. *Uniform randomness extraction is possible on the class of Markov measures, i.e., there exists a partial computable function $E : 2^\omega \to 2^\omega$ such that $E(x)$ is λ-random whenever x is μ-random for some Markov measure μ.* ∎

VII. CONCLUSION

We have obtained in this paper a strong generalization of von Neumann's trick for Bernoulli measures by showing that any computable union of Π_1^0 classes of measures with the effective orthogonality property allows a uniform randomness extraction. We believe this is interesting for several reasons. First, while it does not fully characterize the classes of measures allowing randomness extraction, the criterion given is quite general and applies to a wide variety of classes: Bernoulli measures, Markov chains, etc, and the list could be made much longer if we had worked on a compact computable metric space (all the results presented in this paper extend to this setting): Poisson processes, brownian motion, etc. Second, this is an example of a theorem of constructive mathematics which does not seem to have an analogue (at least not an obvious one) in classical mathematics. Indeed, one could classically interpret von Neumann's trick as follows: there exists a function F such that $\mu \circ F^{-1} = \lambda$ for all Bernoulli measures (except the trivial ones). We on the other hand present a uniform procedure to extract randomness from all Markov chains, while it is clear that there is no function F such that $\mu \circ F^{-1} = \lambda$ for all Markov measures μ. The fact that our extraction procedure is only partial computable (which it has to be because of potential atoms in the measures) makes even less likely the existence of a natural analogue to our results in classical mathematics. We also introduced and gave a complete characterization of the notion of learnability for a class of probability measures which we believe is interesting in its own right, and should have further applications in the field of algorithmic randomness.

ACKNOWLEDGMENT

The authors wish to thank Mathieu Hoyrup, Jason Rute and Alexander Shen for helpful comments and suggestions.

REFERENCES

[1] L. Bienvenu, P. Gács, M. Hoyrup, C. Rojas, and A. Shen, "Algorithmic tests and randomness with respect to a class of measures," *Proceedings of the Steklov Institute of Mathematics*, vol. 274, pp. 41–102, 2011.

[2] P. Odifreddi, *Classical Recursion Theory*. North-Holland Publishing Co., 1989, vol. 1.

[3] P. Gács, "Uniform test of algorithmic randomness over a general space," *Theoretical Computer Science*, vol. 341, pp. 91–137, 2005.

[4] L. Levin, "Some theorems on the algorithmic approach to probability theory and information theory," in *Dissertation in mathematics, Moscow*, 1971.

[5] M. Hoyrup and C. Rojas, "Computability of probability measures and martin-löf randomness over metric spaces," *Information and Computation*, vol. 207(7), pp. 830–847, 2009.

[6] A. R. Day and J. S. Miller, "Randomness for non-computable measures," to appear in Transactions of the American Mathematical Society.

[7] L. A. Levin and A. K. Zvonkin, "The complexity of finite objects and the basing of the concepts of information and randomness on the theory of algorithms," *Uspehi Mat. Nauk*, vol. 25, no. 6(156), pp. 85–127, 1970.

[8] S. M. Kautz, "Degrees of random sets," Ph.D. dissertation, Cornell University, 1991.

[9] M. Hoyrup and C. Rojas, "An application of randomness to effective probability theory," *Lecture Notes in Computer Science*, vol. 5555, pp. 549–561, 2009.

[10] A. Edalat, "A computable approach to measure and integration theory," in *Annual ACM/IEEE Symposium on Logic In Computer Science (LICS)*. IEEE Computer Society, 2007, pp. 463–472.

[11] ——, "A computable approach to measure and integration theory," *Information and Computation*, vol. 207, no. 5, pp. 642–659, 2009.

[12] A. Nies, *Computability and Randomness*. Oxford University Press, 2009.

VIII. Appendix

Proof of Propostion II.1

Proof: $1 \to 2$: Assuming that $f : X \to Y$ is computable with some partial computable function $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. One can build a total function g' extending g such that $g'(v^{-1}(\mathcal{O}(x), \mathbb{N}))) = v^{-1}(\mathcal{O}(f(x)))$. To compute $g'(O, n)$, we compute for all $m \leqslant n$ the first n steps of $g(O, m)$. If at least one terminates and have never been assigned to a $g'(O, m)$ for some $m < n$, then we assign it to $g'(O, n)$. Otherwise we assign Y to $g'(O, n)$.
$2 \to 1$: trivial

$1 \to 3$: Assuming that f is computable, let $O \in \mathcal{O}(Y)$. If $f(x) \in O$ then there is an open set $V \in \mathcal{O}(x)$ and an integer $n \in \mathbb{N}$ such that $g(V, n) = O$. Note that $g(V, n) = O \to f(V) \subseteq O$. So by enumerating all $V \in \mathcal{O}(X)$ and outputting V each time $\exists n \;\; g(V, n) = O$, we have an enumeration of $f^{-1}(O)$.

$3 \to 1$: Assuming that $f^{-1}(O)$ is an enumerable open set, one can enumerate $\bigcup_{O \in \mathcal{O}(Y)} f^{-1}(O) \times O$. For each basic open set U enumerated in $f^{-1}(O)$, define $g(U, n) = O$, where n is the smallest integer such that $g(U, n)$ is undefined. $\operatorname{Im} g$ will contain all basic open sets containing a point in $\operatorname{Im} f$. So for any open set U containing some $f(x)$, we will have an open set O containing x and a $n \in \mathbb{N}$ such that $g(O, n) = U$ which implies $g(v^{-1}(\mathcal{O}(x), \mathbb{N}))) = v^{-1}(\mathcal{O}(f(x)))$. ∎

With truth-table reducibility, with an approximation more and more precise of any Cantor space element, one can uniformly compute an approximation more and more precise of another Cantor space element. It is basically the same here, the approximation being the set of open sets containing a given element.

Lemma VIII.1. *Let* $\mathbf{X}$ *be an effectively compact computable topological space. Let* $\{\mathcal{Q}_e\}$ *be a standard enumeration of the* Π^0_1 *subsets of* $\mathbf{X}$*. There exists a partial computable function* $S : \mathbb{N} \to \mathbf{X}$ *such that, for all* e*, if* $\mathcal{Q}_e$ *is a singleton* $\{x_e\}$*, then* $S(e) = x_e$. ∎

Proof: As $\mathbf{X}$ is effectively compact, each Π^0_1 subset of $\mathbf{X}$ is also effectively compact. So the set of finite sequences of basic open sets of X covering each $\mathcal{Q}_e$ is effectively enumerable uniformly e. If a specific $\mathcal{Q}_e$ contains only one point, then a single basic open set contains this point iff it covers $\mathcal{Q}_e$. So when enumerating finite sequences of basic open sets of $\mathbf{X}$ covering $\mathcal{Q}_e$, we decide to keep only the sequences containing one element. This way we obtain an enumeration of all basic open sets containing the point, which makes it a computable point. ∎