A weak randomness notion for probability measures

André Nies\textsuperscript{1} and Frank Stephen\textsuperscript{2}

1 Department of Computer Science, The University of Auckland, Private Bag 92019, Auckland, New Zealand, andre@cs.auckland.ac.nz
2 Department of Mathematics and Department of Computer Science, National University of Singapore, 10 Lower Kent Ridge Road, Block S17, Singapore 119076, Republic of Singapore, fstephan@comp.nus.edu.sg

Abstract

We study probability measures on Cantor space, thinking of them as statistical superpositions of bit sequences. We say that a measure $\mu$ on the space of infinite bit sequences is Martin-Löf absolutely continuous if the non-Martin-Löf random bit sequences form a $\mu$ null set. We analyse this notion as a weak randomness notion for measures. We begin with examples and a robustness property related to Solovay test. Then we study the growth of initial segment complexity for measures (defined as a $\mu$-average over the complexity of strings of the same length) and relate it to our weak randomness property. We introduce K-triviality for measures. We seek an appropriate effective version of the Shannon-McMillan-Breiman theorem where the trajectories are replaced by measures.

1998 ACM Subject Classification 03D32,68Q30

Keywords and phrases algorithmic randomness, probability measure on Cantor space, Kolmogorov complexity, statistical superposition, quantum states

Digital Object Identifier 10.4230/LIPIcs...

1 Introduction

The theory of algorithmic randomness is usually developed for bit sequences. A central randomness notion based on algorithmic tests is the one due to Martin-Löf\textsuperscript{[10]}. Let $\{0,1\}^N$ denote the topological space of infinite bit sequences. A probability measure $\mu$ on $\{0,1\}^N$ can be seen as a statistical superposition of bit sequences. The bit sequences $Z$ form an extreme case: the corresponding measure $\mu$ is the Dirac measure $\delta_Z$, i.e., $\mu$ is concentrated on $\{Z\}$. The opposite extreme is the uniform measure $\lambda$ which independently gives each bit value the probability $1/2$. The uniform measure represents the maximum disorder as no bit sequence is preferred over any other.

We introduce an algorithmic randomness notion for probability measures. The notion is weak enough so that $\lambda$ qualifies as random. Thus, randomness in this weak sense is compatible with, from another point of view, being computable. Nothing changes for bit sequences: $\delta_Z$ is random iff $Z$ is Martin-Löf random.

Recall that a measure $\mu$ on $\{0,1\}^N$ is called absolutely continuous if each $\lambda$-null set is a $\mu$-null set. Our notion is a weakening of absolute continuity, requiring the $\lambda$-null set in the hypothesis to be effective in the sense of Martin-Löf. Given that there is a universal Martin-Löf test, and hence a largest effective null set, all we have to require is that $\mu(C) = 0$ where $C$ is the class of bit sequences that are not Martin-Löf random.
We provide a brief overview of the paper. Some background and formal definitions will be given in Section 2. In Section 3 we extend the well-known equivalence of Martin-Löf test and Solovay tests to measures. In Section 4 we show that our weak notion of randomness is implied by a notion of Martin-Löf randomness for measures that is obtained when viewing the measures as points in a canonical computable probability space, in the sense of [7]. That stronger randomness notion forces the measure to be atomless.

Randomness of infinite bit sequences is linked to the descriptive complexity of their initial segments via the Levin-Schnorr theorem, which intuitively says that randomness of Z means incompressibility, up to the same constant b, of all the initial segments x. Formally, one requires that $K(x) \geq |x| - b$ for each initial segment x of Z, where $K(x)$ is the prefix free version of Kolmogorov complexity of a string x. The “n-th initial segment” of a measure $\mu$ is given by its values $\mu[x]$ for all strings x of length n, where [x] denotes the set of infinite sequences extending x. It is natural to define the initial segment complexity $K(\mu \upharpoonright_n)$ of this initial segment as the $\mu$-average of the individual complexities of those strings. With this definition, in Section 3 we show that both implications of the analog of the Levin-Schnorr theorem fail. However, we also show in Proposition 24 that for measures that are random in our weak sense, $K(\mu \upharpoonright_n)/n$ converges to 1. Thus, such measures have effective dimension 1; see, for example, Downey and Hirschfeldt [1, Section 12.3].

Opposite to random bit sequences are the K-trivial sequences, where the initial segment complexity grows no faster than that of a computable set; for background see e.g. Nies [14, Section 5.3]. In Section 6 we extend this notion to statistical superpositions of bit sequences: we introduce K-trivial measures and show that they all have countable support.

The Shannon-McMillan-Breiman Theorem from the 1950s (see [19], where it is called the Entropy Theorem) says informally that for an ergodic measure $\rho$ on $\{0,1\}^\mathbb{N}$, outside a null set every bit sequence Z reflects the entropy of the measure by the limiting weighted information content on its sufficiently large initial segments. In the final Section 7 we study what happens when Z is replaced by a measure $\mu$ that is Martin-Löf a.c. with respect to $\rho$ and we take the $\mu$-average of the information contents at the same length.

We note that our research was partly motivated by a recent attempt to define Martin-Löf randomness for quantum states corresponding to infinitely many qubits [15]. Using the terminology of [15], probability measures correspond to the quantum states $\rho$ where the matrix $\rho \upharpoonright_{M_n}$ is diagonal for each n, where $M_n$ is the algebra of $2^m \times 2^m$ complex matrices.

For general background on recursion theory and algorithmic randomness we refer the readers to the textbooks of Calude [1], Downey and Hirschfeldt [4], Li and Vitányi [9], Nies [14], Odifreddi [10, 17], and Soare [20]. There are also lecture notes on recursion theory available online [21].

\section{Measures and Randomness}

We begin with some background for randomness for bit sequences. We use standard notation: letters $Z, X, \ldots$ denote elements of the space of infinite bit sequences $\{0,1\}^\mathbb{N}$, $\sigma, \tau$ denote finite bit strings, and $[\sigma] = \{Z : Z \succ \sigma\}$ is the set of infinite bit strings extending $\sigma$. $Z \upharpoonright_n$ denotes the string consisting of the first n bits of Z. For quantities $r, s$ depending on the same parameters, we write $r \leq^+ s$ for $r \leq s + O(1)$.

A subset $G$ of $\{0,1\}^\mathbb{N}$ is called effectively open if $G = \bigcup [\sigma_i]$ for a computable sequences $\langle \sigma_i \rangle_{i \in \mathbb{N}}$ of strings.

\begin{definition}
A Martin-Löf test (ML-test, for short) is a sequence $\langle G_m \rangle$ of uniformly effectively open sets such that $\lambda G_m \leq 2^{-m}$ for each m. A bit sequence Z fails the test if
\( Z \in \bigcap_m G_m \), otherwise it passes the test. \( Z \) is Martin-Löf random (ML-random) if \( Z \) passes each ML-test.

Let \( K(x) \) denote the prefix free version of descriptive (i.e., Kolmogorov) complexity of a bit string \( x \).

\section*{Theorem 2 (Levin [8], Schnorr [18])} \( Z \) is ML-random \( \iff \exists \forall n K(Z|_n) \geq n - b. \)

Using the notation of [14] Ch. 3], let \( \mathcal{R}_b \) denote the set of bit sequences \( Z \) such that \( K(Z|_n) < n - b \) for some \( n \). It is easy to see that \( \langle \mathcal{R}_b \rangle_{b \in \mathbb{N}} \) forms a Martin-Löf test. The Levin-Schnorr theorem says that this test is universal: \( Z \) is ML-random iff it passes the test.

Unless otherwise stated, all measures will be probability measures. We use the letters \( \mu, \nu, \rho \) for probability measures; \( \lambda \) denotes the uniform measure. So \( \lambda[\sigma] = 2^{-|\sigma|} \). We now provide the formal definition of our weak randomness notion for measures.

\section*{Definition 3} A measure \( \mu \) is called Martin-Löf absolutely continuous (Martin-Löf a.c., for short) if \( \inf m \mu(G_m) = 0 \) for each Martin-Löf test \( \langle G_m \rangle_{m \in \mathbb{N}} \).

If \( \inf_m \mu(G_m) = 0 \) we say that \( \mu \) passes the test. If \( \inf_m \mu(G_m) \geq \delta \) where \( \delta > 0 \) we say \( \mu \) fails the test at level \( \delta \).

In the definition it suffices to consider Martin-Löf tests \( \langle G_m \rangle \) such that \( G_m \uparrow G_{m+1} \) for each \( m \), because we can replace \( \langle G_m \rangle \) by the Martin-Löf test \( G_m = \bigcup_{k > m} G_k \), and of course \( \inf_m \mu(G_m) = 0 \) implies \( \inf_m \mu(G_m) = 0 \). So we can change the definition above, replacing the condition \( \inf_m G_m = 0 \) by the only apparently stronger condition \( \lim_m G_m = 0 \).

The intersection of a universal ML-test consists of the non-Martin-Löf random sequences. Since such a test exists, we have:

\section*{Fact 4} A measure \( \mu \) is Martin-Löf a.c. iff the sequences which are not Martin-Löf random form a \( \mu \)-null set.

We already mentioned the diametrically opposite examples, which give an impression of the range of the notion.

\section*{Example 5} The uniform measure \( \lambda \) is a Martin-Löf a.c. measure.

A Dirac measure \( \delta_Z \) is Martin-Löf a.c. iff \( Z \) is Martin-Löf random.

Each ML-random sequence \( Z \) satisfies the law of large numbers

\[ \lim_n \left| \{i < n: Z(i) = 1\} \right| = 1/2; \]

see e.g. [14] Prop. 3.2.19]. So if \( \mu \) is Martin-Löf a.c., then \( \mu \) almost surely, \( Z \) satisfies the law of large numbers. Thus, for \( p \neq 1/2 \), a Bernoulli measure on \( \{0,1\}^\mathbb{N} \), that independently gives probability \( p \) to a 0 in each position, is not a Martin-Löf a.c. measure.

For a measure \( \nu \) and string \( \sigma \) with \( \nu[\sigma] > 0 \) let \( \nu_\sigma \) be the localisation: \( \nu_\sigma(A) = 2^{-|\sigma|} \nu(A \cap [\sigma]) \). Clearly if \( \nu \) is Martin-Löf a.c. then so is \( \nu_\sigma \).

A set \( S \) of probability measures is called convex if \( \mu_i \in S \) for \( i \leq k \) implies that the convex combination \( \mu = \sum \alpha_i \mu_i \) is in \( S \), where the \( \alpha_i \) are reals in \( [0,1] \) and \( \sum \alpha_i = 1 \). The extreme points of \( S \) are the ones that can only be written as convex combinations of length 1 of elements of \( S \).

\section*{Proposition 6} The Martin-Löf a.c. probability measures form a convex set. Its extreme points are the Martin-Löf a.c. Dirac measures, i.e. the measures \( \delta_Z \) where \( Z \) is a ML-random bit sequence.
A weak randomness notion for probability measures

Proof. Convexity is trivial: let \( \mu = \sum_i c_i \mu_i \) as above where the \( \mu_i \) are Martin-Löf a.c. measures. Suppose \( (G_m) \) is a Martin-Löf test. Then \( \lim_m \mu_i(G_m) = 0 \) for each \( i \), and hence \( \lim_m \mu(G_m) = 0 \).

If \( \mu \) is a Dirac measure then it is an extreme point of the Martin-Löf a.c. measures. Conversely, if \( \mu \) is not Dirac there is a least number \( t \) such that the decomposition

\[
\mu = \sum_{|\sigma| = t, \mu[\sigma] > 0} \mu[\sigma] \cdot \mu_{\sigma}
\]

is nontrivial. Hence \( \mu \) is not an extreme point.

\( \blacktriangleleft \)

3 Solovay tests

Recall that a Solovay test is a sequence \( (S_k)_{k \in \mathbb{N}} \) of uniformly \( \Sigma^0_1 \) sets such that \( \sum_k S_k < \infty \). A bit sequence \( Z \) passes such a test if \( Z \not\in S_k \) for almost every \( k \). (Each ML-test is a Solovay test, but the passing condition is stronger for Solovay tests). A basic fact from the theory of algorithmic randomness (e.g. [14, 3.2.9]) states that a bit sequence (e.g. [14, Prop. 3.2.19]): Let \( Z \) pass such a test if \( \sum_k S_k < \infty \). Then \( Z \) is ML-random iff \( Z \) passes each Solovay test. We say that a measure \( \mu \) passes a Solovay test \( (S_k)_{k \in \mathbb{N}} \) if \( \lim_k \mu(S_k) = 0 \).

The fact that passing all Martin-Löf tests is equivalent to passing all Solovay tests generalises to measures. We use the following variant for measures of a result by Tejas Bhojraj (in preparation) that he proved in the setting of quantum states.

Fact 7. Let \( \mu = \sum_k c_k \delta_{S_k} \) where \( \forall k \geq 0, 0 \leq c_k \leq 1 \) and \( \sum_k c_k = 1 \). Then \( \mu \) is Martin-Löf a.c. iff all \( Z_k \) are Martin-Löf random.

Proof. The implication from left to right is immediate. For the converse implication, given a Martin-Löf test \( (G_m) \), note that the \( Z_k \) pass this test as a Solovay test. Hence for each \( k \), there is \( M \) such that \( Z_k \not\in G_m \) for each \( k \leq r \) and \( m \geq M \). This implies that \( \mu(G_m) \leq \sum_{k > r} c_k \). So \( \lim_m \mu(G_m) = 0 \).

The fact that passing all Martin-Löf tests is equivalent to passing all Solovay tests generalises to measures. We use the following variant for measures of a result by Tejas Bhojraj (in preparation) that he proved in the setting of quantum states.

Proposition 8. A measure \( \mu \) is Martin-Löf a.c. iff \( \mu \) passes each Solovay test.

Proof. Each Martin-Löf test is a Solovay test. So the implication from right to left is immediate.

For the implication from left to right, suppose that \( (S_k) \) is a Solovay test that \( \mu \) fails. So there is \( \delta > 0 \) such that \( \mu(S_k) > \delta > 0 \) for infinitely many \( k \). We will define a Martin-Löf test \( (G_m)_{m \in \mathbb{N}} \) that \( \mu \) fails at level \( \delta/2 \), i.e. \( \inf_m \mu(G_m) \geq \delta/2 \).

Let \( S_{k,t} \) denote the clopen set generated by the strings \( \sigma \) of length \( t \) such that \( [\sigma] \subseteq S_k \).

First we use a minor modification of the proof that the two test notions are equivalent for bit sequences (e.g. [14, Prop. 3.2.19]): Let \( G_{m,t} \) be the clopen set generated by strings \( \sigma \) of length \( t \) such that

\[
[\sigma] \subseteq S_{k,t} \text{ for } \delta^{2^m-1} \text{ many } k \leq t.
\]

Similar to the proof of [14, Prop. 3.2.19] one shows that \( \lambda G_{m,t} \leq 2^{-m}/\delta \). Let \( G_m = \bigcup_t G_{m,t} \) and note that \( G_m \) is effectively open uniformly in \( m \). So, fixing a \( k \in \mathbb{N} \) such that \( k > \log_2(1/\delta) \), \( (G_k)_{m \in \mathbb{N}} \) forms a Martin-Löf test.

Given \( m \), we pick \( t \in \mathbb{N} \) sufficiently large so that for some set \( M \subseteq \{0, \ldots, t-1\} \) of size \( 2^m \), we have \( \mu(S_{k,t}) > \delta \) for each \( k \in M \). We show that \( \mu(G_{m,t}) > \delta/2 \).
Let $\sigma$ range over strings of length $t$. If $[\sigma] \not\subseteq G_{m,t}$ then $\sum_{k \leq t} \mu([\sigma] \cap S_{k,t}) \leq 2^{m-1}\delta \mu[\sigma]$ by definition of $G_m$. Therefore

$$\sum_{[\sigma] \not\subseteq G_{m,t}} \sum_{k \leq t} \mu([\sigma] \cap S_{k,t}) \leq 2^{m-1}\delta.$$  

Since $2^m \delta \leq \sum_{k \in M} \mu(S_{k,t})$, this implies

$$\sum_{[\sigma] \subseteq G_m} \sum_{k \in M} \mu[\sigma] > 2^{m-1}\delta.$$  

Since $|M| = 2^m$ this shows $\mu G_{m,t} > \delta/2$ as required. \hfill \blacklozenge

### 4 Full Martin-Löf randomness of measures

Let $M(\{0,1\}^N)$ be the space of probability measures on Cantor space (which is canonically a compact topological space). A probability measure $P$ on this space has been introduced implicitly in Mauldin and Monticino [14]. Culver’s thesis [3] shows that this measure is computable. So the framework of [7] yields a definition of Martin-Löf randomness for points in the space $M(\{0,1\}^N)$.

To define $P$, first let $R$ be the closed set of representations of probability measures; namely, $R$ consists of the functions $X: \{0,1\}^* \to [0,1]$ such that $X_\emptyset = 1$ and $X_{\sigma} = X_{\sigma 0} + X_{\sigma 1}$ for each string $\sigma$. $P$ is the unique measure on $R$ such that for each string $\sigma$ and $r, s \in [0,1]$, we have

$$P(X_\emptyset \leq r \mid X_{\sigma} = s) = \min(1, r/s).$$

Intuitively, we choose $X_{\sigma 0}$ at random w.r.t. the uniformly distribution on the interval $[0, X_{\sigma}]$, and the choices made at different strings are independent.

► **Proposition 9.** Every probability measure $\mu$ that is Martin-Löf random wrt to $P$ is Martin-Löf absolutely continuous.

For the duration of this proof let $\mu$ range over $M(\{0,1\}^N)$. For an open set $G \subseteq \{0,1\}^N$, let

$$r_G = \int \mu(G)dP(\mu).$$

Our proof of Prop. 9 is based on two facts. For open $G \subseteq \{0,1\}^N$, let

$$r_G = \int \mu(G)dP(\mu).$$

► **Fact 10.** $r_G = \lambda(G)$.

**Proof.** Clearly, for each $n$ we have

$$\sum_{|\sigma| = n} \rho[\sigma] = \int \sum_{|\sigma| = n} \mu([\sigma])dP(\mu) = 1.$$  

Furthermore, $\rho_{\sigma} = \rho_{\eta}$ whenever $|\sigma| = |\eta| = n$ because there is a $P$-preserving transformation $T$ of $M(\{0,1\}^N)$ such that $\mu([\sigma]) = T(\mu([\eta]))$. Therefore $\rho_{\sigma} = 2^{-|\sigma|}$.

If $\sigma, \eta$ are incompatible then $r_{\sigma \cup [\eta]} = r_{\sigma} + r_{[\eta]}$. Now it suffices to write $G = \bigcup_{i} [\sigma_i]$ where the strings $\sigma_i$ are incompatible, so that $\lambda G = \sum_i 2^{-|\sigma_i|}$. \hfill \blacklozenge

► **Fact 11.** Let $\mu \in M(\{0,1\}^N)$ and let $(G_m)_{m \in \mathbb{N}}$ be a ML-test such that there is $\delta \in \mathbb{Q}^+$ with $\forall m \mu(G_m) > \delta$. Then $\mu$ is not ML-random w.r.t. $P$. 

A weak randomness notion for probability measures

Proof. Observe that by the foregoing fact
\[ \delta \cdot \mathbb{P}([\mu: \mu(G_m) \geq \delta]) \leq \int \mu(G_m)d\mathbb{P}(\mu) = \lambda(G_m) \leq 2^{-m}. \]

Let \( G_m = \{ \mu: \mu(G_m) > \delta \} \) which is uniformly effectively open in the space of measures \( \mathcal{M}([0, 1]^N) \). Fix \( k \) such that \( 2^{-k} \leq \delta \); then \( (G_{m+k})_{m\in\mathbb{N}} \) is a ML-test w.r.t. \( \mathbb{P} \) that succeeds on \( \mu \).

Culver [3] shows that each measure \( \mu \) that is Martin-Löf random w.r.t. \( \mathbb{P} \) is non-atomic. So because of the measures \( \delta_Z \) for Martin-Löf random bit sequences \( Z \), the converse of Prop. 9 fails: not every Martin-Löf a.c. measure is Martin-Löf random with respect to \( \mathbb{P} \).

5 Initial segment complexity of a measure \( \mu \)

Let \( K(\mu | n) = \sum_{|x|=n} K(x)\mu[x] \) be the \( \mu \)-average of all the \( K(x) \) over all strings \( x \) of length \( n \). In a similar way we define \( C(\mu | n) \). Note that for a Dirac measure \( \delta_Z \), we have \( K(\delta_Z | n) = K(Z | n) \).

In this section we use standard inequalities such as \( C(x) \leq^+ K(x) \), \( K(x) \leq^+ |x| + 2 \log |x| \) and \( K(0^n) \leq^+ 2 \log n \). We also use that for each \( r \) there are at most \( 2^r - 1 \) strings such that \( C(x) < r \). See e.g. [14, Ch. 2]. Recall that \( \lambda \) denotes the uniform measure on \( \{0, 1\}^N \).

fact 12.

(a) \( C(\lambda | n) \geq^+ n \).
(b) \( K(\lambda | n) \geq^+ n + K(n) \).

Proof. Chaitin [2] showed that there is a constant \( c \) such that, for all \( d \), there are at most \( 2^{n+c-d} \) strings \( x \in \{0, 1\}^n \) with \( C(x) \leq n - d \). Similarly, among the strings of length \( n \), there are at most \( 2^{n+c-d} \) strings with \( K(x) \leq n + K(n) - d \). In other words, the fraction of strings of length \( n \) where, for (a), \( C(x) \leq n - d \), and, for (b), \( K(x) \leq n + K(n) - d \), respectively, is in each case at most \( 2^{c-d} \). Now for each \( d \), from the estimated lower bound \( n \) and \( n + K(n) \), respectively, one subtracts the fraction of the strings of length \( n \) for which the Kolmogorov complexity is at least \( d \) below the average in order to correct the lower bound.

For, if the complexity of a string \( x \) below the lower bound, it has to be considered \( r \) times, for \( d = 1, \ldots, r \). Let \( c_d \) is the fraction of strings of length \( n \) with \( C(x) \leq n - d \) and \( k_d \) is the fraction of strings with \( K(x) \leq n + K(n) - d \). Then
\[ C(\lambda | n) \geq n - \sum_{d \geq 0} c_d \text{ and } K(\lambda | n) \geq n + K(n) - \sum_{d \geq 0} k_d. \]

Using Chaitin’s bounds gives then the corrected estimates on the averages of
\[ C(\lambda | n) \geq n - \sum_{d \geq 0} 2^{c-d} \text{ and } K(\lambda | n) \geq n + K(n) - \sum_{d \geq 0} 2^{c-d}. \]

Now one uses that \( \sum_{d \geq 0} 2^{c-d} \leq 2^{c+1} \) and that \( 2^{c+1} \) is a constant independent of \( n \) and only dependent on the universal machine in order to get that \( C(\lambda | n) \geq^+ n \) and \( K(\lambda | n) \geq^+ n + K(n) \).

We say that a measure \( \mu \) has complex initial segments if \( K(\mu | n) \geq^+ n \). We will show that both implications of the analog of the Levin-Schnorr Theorem 2 fail for measures. One implication, that a Martin-Löf a.c. measure cannot have initial segment complexity growing slower than \( n - O(1) \), is disproved by a simple example of a measure with countable support.
Example 13. There is a Martin-Löf a.c. measure $\mu$ such that $K(\mu \upharpoonright n) \leq^+ n - \log n$.

Proof. We let $\mu = \sum c_k \delta_{Z_k}$ where $Z_k$ is Martin-Löf random and $0^{n_k} \prec Z_k$ for a sequence $(c_k)$ of reals in $[0,1]$ that add up to 1, and a sufficiently fast growing sequence $n_k$. Then $\mu$ is Martin-Löf a.c. by Fact 7. For $n$ such that $n_k \leq n < n_{k+1}$ we have

\[
K(\mu \upharpoonright n) \leq^+ \sum_{i=0}^{k} c_i \cdot (n + 2 \log n) + \sum_{i=k+1}^{\infty} c_i \cdot 2 \log n
\]

\[
\leq^+ (1 - c_{k+1})n + 2 \log n.
\]

Hence, to achieve $K(\mu \upharpoonright n) \leq^+ n - \log n$ it suffices to ensure that $3 \log(n_{k+1}) < c_{k+1} n_k$. For instance, we can let $c_k = 1/(3(k+1)(k+2))$ and $n_k = 2^{k+4}$. ▶

To falsify the converse implication, we need to provide a measure $\mu$ such that $K(\mu \upharpoonright n) \geq^+ n$ yet $\mu$ is not a Martin-Löf a.c. measure. This will be immediate from the following fact on the growth of initial segment complexity for bit sequences.

Theorem 14. There are a Martin-Löf random $X$ and a not Martin-Löf random $Y$ such that, for all $n$, $K(X \upharpoonright n) + K(Y \upharpoonright n) \geq^+ 2n$.

Proof. Let $X$ be a low Martin-Löf random set (i.e., $X' \equiv_T \emptyset'$). We claim that there is a strictly increasing function $f$ such that the complement of the range of $f$ is a recursively enumerable set $E$, and $K(X \upharpoonright m) \geq m + 3n$ for all $m \geq f(n)$. To see this, recall that $\lim_n K(X \upharpoonright n) = \infty$. Since $X$ is low there is a computable function $p$ such that for all $n$, $\lim_n p(n,s)$ is the maximal $m$ such that $K(X \upharpoonright m) \leq m + 3n$. Define $f(n,s)$ for $n \leq s$ as follows. $f(0,0) = 0$; for $s > 0$ let $n$ be least such that $p(n,s-1) \neq p(n,s)$ or $n = s$. Let $f(n,s) = r$ be larger than all previous values assigned by $f$ to a pair, and let $f(m,s) = r + m - n$ for $n < m \leq s$. Let $f(m,s) = f(m,s-1)$ for $m < n$. Then $f(n) = \lim_n f(n,s)$ is a function as required, which verifies the claim.

Now let $g(n) = \max \{ m : f(m) \leq n \}$ (with the convention that $\max(\emptyset) = 0$). Since $g$ is unbounded, by a result of Miller and Yu [13, Cor. 3.2] there is a Martin-Löf random $Z$ such that there exist infinitely many $n$ with $K(Z \upharpoonright n) \leq n + g(n)/2$. Let

\[
Y = \{ n + f(n) : n \in Z \}.
\]

Note that $K(Z \upharpoonright n) \leq K(Y \upharpoonright n) + g(n) + K(g(n))$, as one can enumerate the set $E$ until there are, up to $n$, only $g(n)$ many places not enumerated and then one can reconstruct $Z \upharpoonright n$ from $Y \upharpoonright n$ and $g(n)$ and the last $g(n)$ bits of $Z$. As $Z$ is Martin-Löf random, $K(Z \upharpoonright n) \geq^+ n$ and so,

\[
K(Y \upharpoonright n) \geq^+ n - g(n) - K(g(n)) \geq^+ n - 2g(n).
\]

The definitions of $X,f,g$ give $K(X \upharpoonright n) \geq n + 3g(n)$. This shows that $K(X \upharpoonright n) + K(Y \upharpoonright n) \geq 2n$ for almost all $n$.

However, the set $Y$ is not Martin-Löf random, as there are infinitely many $n$ with $K(Z \upharpoonright n) \leq^+ n + g(n)/2$. Now $Y \upharpoonright n + g(n)$ can be computed from $Z \upharpoonright n$ and $g(n)$, as one needs only to enumerate $E$ until the $g(n)$ nonelements of $E$ below $n$ are found and they allow to see where the zeroes have to be inserted into the string $Z \upharpoonright n$ in order to obtain $Y \upharpoonright n + g(n)$. Note furthermore, that $K(g(n))/4$ for almost all $n$ and thus $K(Y \upharpoonright n + g(n)) \leq^+ n + 3/4 \cdot g(n)$ for infinitely many $n$, so $Y$ cannot be Martin-Löf random. ▶

Corollary 15. There is a measure $\mu$ with complex initial segments which is not Martin-Löf a.c.
A weak randomness notion for probability measures

The measure \( \mu = (\delta x + \delta y)/2 \) has only two equal-weighted atoms and one of these atoms is not Martin-Löf random. So every component of a universal Martin-Löf test has at least \( \mu \)-measure 1/2. On the other hand, \( K(\mu \upharpoonright n) \geq n \) for almost all \( n \) by the preceding result.

A bit sequence \( Z \in \{0,1\}^N \) is called strongly Chaitin random if there is \( d \) such that 
\[
K(Z \upharpoonright n) \geq n + K(n) - d \text{ for infinitely many } n. 
\]
(This is equivalent to 2-randomness by [12]; for detail see e.g. [14, 8.1.14] or [4]).

We may extend this notion to measures. Each such measure is Martin-Löf a.c.:

**Theorem 16.** Suppose that \( \mu \) is a measure such that \( K(\mu \upharpoonright n) \geq n + K(n) - r \) for infinitely many \( n \). Then \( \mu \) is a Martin-Löf a.c. measure.

**Proof.** Suppose that \( \mu \) is not a Martin-Löf a.c. measure. So there is a Martin-Löf test \( \langle G_d \rangle_{d \in \mathbb{N}} \) and \( \epsilon > 0 \) such that \( \mu(G_d) > \epsilon \) for each \( d \). We view \( G_d \) as given by an enumeration of strings, uniformly in \( d \); thus \( G_d = \bigcup \{\sigma_i \} \) for an effectively computable sequences \( \langle \sigma_i \rangle \).

Let \( G_d^\leq_n \) denote the clopen set generated by the strings in this enumeration of length at most \( n \). (Note that this set is not effectively given as a clopen set, but we effectively have a description of it as a \( \Sigma_1^0 \) set). Let \( c \) be a constant such that \( K(x) \leq n + K(n) + c \) for each \( x \) of length \( n \).

If \( x \) is a string of length \( n \) such that \( [x] \subseteq G_d^\leq_n \) then
\[
K(x \upharpoonright n, d) \leq n - d.
\]
To see this let \( M \) be the fixed machine that, on a pair of auxiliary inputs \( n, d \), waits for \( [x] \subseteq G_d^\leq_n \) and once that happens provides a description of length \( n - d \) for \( x \) (so the descriptions for different \( x \) are prefix free). It follows that for such \( x \) (and after increasing \( c \) if necessary)
\[
K(x) \leq n + K(n) - d + 2 \log d + c.
\]
For each \( d, n \), letting \( x \) range over strings of length \( n \), we have
\[
K(\mu \upharpoonright n) = \sum K(x) \mu[x] = \sum_{[x] \subseteq G_d^\leq_n} K(x) \mu[x] + \sum_{[x] \not\subseteq G_d^\leq_n} K(x) \mu[x]
\]
The first summand is bounded above by \( \mu(G_d^\leq_n)(n + K(n) - d + 2 \log d + c) \), the second by \( (1 - \mu(G_d^\leq_n))(n + K(n) + c) \). We obtain
\[
K(\mu \upharpoonright n) \leq n + K(n) + c - \mu(G_d^\leq_n) d/2.
\]
Now for each \( d \), for sufficiently large \( n \) we have \( \mu(G_d^\leq_n) > \epsilon \). So given \( r \) let \( d = 2r/\epsilon \); then for large enough \( n \) we have \( K(\mu \upharpoonright n) \leq n + K(n) + c - r \). So \( \mu \) is not strongly Chaitin random.

We ask whether a known fact for bit sequences lifts to measures.

**Question 17.** Is every strongly Chaitin random measure Martin-Löf a.c. relative to the halting problem \( \emptyset' \)?
6 \textbf{K-triviality for measures}

\textbf{Definition 18.} A measure $\mu$ is called $K$-trivial if $K(\mu \upharpoonright n) \leq^+ K(n)$ for each $n$.

For Dirac measures $\delta_A$ this is the same as saying that $A$ is $K$-trivial in the usual sense. More generally, any finite convex combination of such Dirac measures is $K$-trivial.

\textbf{Proposition 19.} Suppose $\mu$ is $K$-trivial. Then $\mu$ is supported by its set of atoms.

Thus, if $\mu$ is $K$-trivial for constant $b$ then $\mu$ has the form $\sum_{r<N} \alpha_r \delta_{A_r}$ where $N \leq \infty$ and each $\alpha_r$ is positive and $\sum_{r<N} \alpha_r = 1$. Clearly each $A_r$ is $K$-trivial for constant $b/\alpha_r$.

\textbf{Proof.} Assume for a contradiction that $\mu$ gives a measure of $\epsilon > 0$ to the set of its non-atoms. Note that there is a constant $b$ such that $K(x) \geq K(|x|) - b$ for each $x$. Fix $c$ arbitrary with the goal of showing that $K(\mu \upharpoonright n) \geq K(n) - b + \epsilon c/2$ for large enough $n$.

There is $d$ (in fact $d = O(2^n)$) such that for each $n$ there are at most $d$ strings $x$ of length $n$ with $K(x) \leq K(n) + c$ (see e.g. [11, 2.2.26]).

Let $S_n = \{x : |x| = n \land \mu[x] \leq \epsilon/2d\}$. By hypothesis we have $\mu[S_n] \geq \epsilon$ for large enough $n$. Therefore by choice of $d$ we have $\mu[S_n \cap \{x : K(x) > |x| + c\}] \geq \epsilon/2$. Now we can give a lower bound for the $\mu$ average of $K(x)$ over all strings $x$ of length $n$:

$$\sum_{|x|=n} K(x) \mu[x] \geq (1 - \epsilon/2)(K(n) - b) + (\epsilon/2)(K(n) + c) \geq K(n) - b + \epsilon c/2,$$

as required.

Notice that we only used the weaker hypothesis that $\liminf_n [K(\mu \upharpoonright n) - K(n)]$ is finite.

It would be interesting to characterise the countable convex combinations of $K$-trivials that yield $K$-trivial measures. The following is easily checked.

\textbf{Fact 20.} Suppose that $\alpha_r$ is $K$-trivial with constant $b_r$, and $\sum_r \alpha_r b_r \leq c < \infty$ where each $\alpha_r$ is positive and $\sum_r \alpha_r = 1$. Then $\mu = \sum_r \alpha_r \delta_{A_r}$ is $K$-trivial with constant $c$.

For instance, we can build a computable $K$-trivial measure with infinitely many atoms as follows. Let $A_r = 0^{r+1}1^\infty$, so that $K(A_r \upharpoonright n) \leq^+ K(n) + 2 \log r$. Let $\mu = \sum 2^{-r+1} A_r$. By the above fact $\mu$ is $K$-trivial. If we vary the construction by letting $A_r = 0^{r+1}1B$ where $B$ is $K$-trivial but non-recursive, we obtain a $K$-trivial $\mu$ with infinitely many atoms, and none of them recursive.

On the other hand, the following example shows that not every infinite convex combination of $K$-trivial Dirac measures yields a $K$-trivial measure. Let $\mu = \sum_k \alpha_k \delta_{A_k}$, where $A_k = \{\ell : \ell \in \Omega \land \ell < k\}$, and $\alpha_k = (k+1)^{-1/2} - (k+2)^{-1/2}$. All sets $A_k$ are finite and thus $K$-trivial. Furthermore, the sum of all $\alpha_k$ is 1. We have

$$K(\mu \upharpoonright n) = \sum_{x \in \{0,1\}^n} K(x) \mu(x) \geq (\sum_{m \geq n} \alpha_m) \cdot K(\Omega \upharpoonright n) \geq (n+2)^{-1/2} \cdot (n+2) = \sqrt{n+2}$$

for almost all $n$ and thus the average grows faster than $K(n) + c$. So the measure is not $K$-trivial.

In a sense, an atomless measure can come arbitrarily close to being $K$-trivial.

\textbf{Proposition 21.} For each nondecreasing unbounded function $f$ which is computably approximable from above there is a non-atomic measure $\mu$ such that $K(\mu \upharpoonright n) \leq^+ K(n) + f(n)$.

\textbf{Proof.} There is a recursively enumerable set $A$ such that, for all $n$, $A$ has up to $n$ and up to a constant $f(n)/2$ non-elements. One let $\mu$ be the measure such that $\mu(x) = 2^{-m}$ in the case that all ones in $x$ are not in $A$ and $\mu(x) = 0$ otherwise, here $m$ is the number of non-elements
of \( A \) below \(|x|\). One can see that when \( \mu(x) = 2^{-m} \) then \( x \) can be computed from \(|x|\) and the string \( b_0b_1 \ldots b_{m-1} \) which describes the bits at the non-elements of \( A \). Thus \( K(x) \leq K(|x|) + K(b_0b_1 \ldots b_{m-1}) \leq K(|x|) + 2m \). It follows that \( K(\mu \upharpoonright n) \leq K(n) + f(n) \), as the \( \mu \)-average of strings \( x \in \{0,1\}^n \) with \( K(x) \leq K(n) + f(n) \) is at most \( K(n) + f(n) \) plus a constant.

7 Towards effective Shannon-McMillan-Breiman for measures

We review some notions from the field of symbolic dynamics, a mathematical area closely related to Shannon information theory. It is useful to admit alphabets other than the binary one. Let \( \mathcal{A}^\infty \) denote the topological space of one-sided infinite sequences of symbols in an alphabet \( \mathcal{A} \). Randomness notions etc. carry over from the case of \( \mathcal{A} = \{0,1\} \). A dynamics on \( \mathcal{A}^\infty \) is given by the shift operator \( T \), which erases the first symbol of a sequence. A measure \( \rho \) on \( \mathcal{A}^\infty \) is called shift invariant if \( \rho(G) = \rho(T^{-1}(G)) \) for each open (and hence each measurable) set \( G \). The empirical entropy of a measure \( \rho \) along \( Z \in \mathcal{A}^\infty \) is given by the sequence of random variables

\[
h_n^\rho(Z) = -\frac{1}{n} \log |\mathcal{A}| \rho[Z|n].
\]

A shift invariant measure \( \rho \) on \( \mathcal{A}^\infty \) is called ergodic if every \( \rho \) integrable function \( f \) with \( f \circ T = f \) is constant \( \rho \)-almost surely. An equivalent condition that is easier to check is the following: for \( u, v \in \mathcal{A}^* \),

\[
\lim_n \frac{1}{N} \sum_{k=0}^{n-1} \rho([u] \cap T^{-k}[v]) = \rho[u]\rho[v].
\]

For ergodic \( \rho \), the entropy \( H(\rho) \) is defined as \( \lim_n H_n(\rho) \), where

\[
H_n(\rho) = -\frac{1}{n} \sum_{|w|=n} \rho[w] \log \rho[w].
\]

Thus, \( H_n(\rho) = E_\rho h_n^\rho \) is the expected value with respect to \( \rho \). One notes that \( H_{n+1}(\rho) \leq H_n(\rho) \leq 1 \) so that the limit exists.

A well-known result from the 1950s due to Shannon, McMillan and Breiman (see, e.g., [19]) states that for an ergodic measure \( \rho \), for \( \rho \)-a.e. \( Z \) the empirical entropy along \( Z \) converges to the entropy of the measure.

Theorem 22 (SMB theorem). Let \( \rho \) be an ergodic measure on the space \( \mathcal{A}^\infty \). For \( \rho \)-a.e. \( Z \) we have \( \lim_n h_n^\rho(Z) = H(\rho) \).

A measure \( \rho \) on \( \mathcal{A}^\infty \) is called computable if the real \( \rho[x] \) is computable, uniformly in \( x \in \mathcal{A}^* \). For such a measure we can define Martin-Löf tests and Martin-Löf randomness with respect to \( \rho \) (\( \rho \)-ML randomness for short) as above; in Definition 1 we simply replace the uniform measure \( \lambda \) by \( \rho \). The basic theory of \( \rho \)-ML randomness is developed in a way similar to the uniform case. In particular, there is a universal \( \rho \)-ML test. Recalling Fact 4, we say that a measure \( \mu \) is Martin-Löf a.e. with respect to \( \rho \) if \( \mu(C) = 0 \) where \( C \) is the class of sequences in \( \mathcal{A}^\infty \) that are not ML-random with respect to \( \rho \).

If a computable measure \( \rho \) is shift invariant, then \( \lim_n h_n^\rho(Z) \) exists for each \( \rho \)-ML-random \( Z \) by a result of Hochman [5]. Hoyrup [6, Thm. 1.2] gave an alternative proof for ergodic \( \rho \), and also showed that in that case we have \( \lim_n h_n^\rho(Z) = H(\rho) \) for each \( \rho \)-ML.
random $Z$. We extend this result to measures $\mu$ that are Martin-Löf a.c. with respect to $\rho$.

We require as a hypothesis that the $h^\rho_n$ are uniformly bounded. This holds e.g. for Bernoulli measures and the measures given by a Markov process. On the other hand, using a renewal process it is not hard to construct an ergodic computable measure $\rho$ where this hypothesis fails. For instance, over the binary alphabet let $\rho$ be the shift invariant measure such that $\rho[10^k1] = 2^{-k-1}$ for each $k \in \mathbb{N}$.

**Theorem 23.** Let $\rho$ be a computable ergodic measure on the space $\mathcal{K}_\infty$ such that for some constant $D$, each $h^\rho_n$ is bounded above by $D$. Suppose the measure $\mu$ is Martin-Löf a.c. with respect to $\rho$. Then $\lim_n E_\mu h^\rho_n = H(\rho)$.

**Proof.** For each pair of rationals $s < t$ and each $k \in \mathbb{N}$, let $U_k^{s,t}$ be the effectively open set of those $Z \in \mathcal{K}_\infty$ such that the sequence $h^\rho_n(Z)$ has $k$ upcrossings of $[s,t]$. By Hochman [5, Thm. 1.3] we have $\rho(U_k^{s,t}) < c \cdot \alpha^k$ for positive constants $c < 1$ and $\alpha$ that depend only on $s,t$. Therefore a subsequence of $\bigcap_{k} U_k^{s,t}$ is a Martin-Löf test w.r.t. $\rho$, and hence $\mu(\bigcap_{k} U_k^{s,t}) = 0$. Since $s < t$ are arbitrary and $h^\rho_n$ is bounded by a fixed constant $D$ this implies that $h^\rho_n(Z)$ converges to a value $g(Z)$, $\mu$-almost surely.

By Hoyrup’s result, $\lim_n h^\rho_n(Z) = H(\rho)$ for each $\rho$-ML random $Z$. Since the sequences that are not ML-random w.r.t. $\rho$ form a null set w.r.t. $\mu$, we infer that $g(Z) = H(\rho)$ for $\mu$-a.e. $Z$. The Dominated Convergence Theorem now shows that $\lim_n E_\mu h^\rho_n = E_\mu g = H(\rho)$, as required.

**Proposition 24.** Let $\rho$ be a computable ergodic invariant measure, and suppose $\mu$ is a Martin-Löf a.c. measure with respect to $\rho$. Then

$$\lim_n \frac{1}{n} K(\mu|_{n}) = \lim_n \frac{1}{n} C(\mu|_{n}) = H(\rho).$$

**Proof.** We can use $K$ and $C$ interchangeably because $C(x) \leq K(x) \leq K(C(x)) + K(\{x\})$ [14, 2.4.1]. In analogy to the functions $h_n$, let $k_n(Z) = K(Z|n)/n$. The argument for $K$, say, is very similar to the one in the theorem above, replacing the functions $h_n$ by $k_n$. Note that $k_n$ is bounded above by a constant because $K(x) \leq |x| + 2 \log |x|$. With the $U_k^{s,t}$ defined analogously, $\rho(U_k^{s,t}) < c \cdot \alpha^k$ for some positive constants $c, \alpha$ with $\alpha < 1$ by Hochman [5, Thm. 1.4]. Hoyrup’s result [6, Thm. 1.2] states that $\lim_n k_n(Z) = H(\rho)$ for each $\rho$-ML random $Z$. Now we can apply the Dominated Convergence Theorem as before.

**Acknowledgment.** A. Nies was supported in part by the Marsden Fund of the Royal Society of New Zealand, UoA 13-184. F. Stephan is supported in part by the Singapore Ministry of Education Academic Research Fund Tier 2 grant MOE2016-T2-1-019 / R146-000-234-112.

A preliminary version was summarised in the Logic Blog 2018 on the first author’s homepage at [https://www.cs.auckland.ac.nz/~nies/](https://www.cs.auckland.ac.nz/~nies/)

**References**

1. Cristian S. Calude. *Information and Randomness - An Algorithmic Perspective*. Second Edition, Springer, Heidelberg, 2002.
2. Gregory J. Chaitin. A theory of program size formally identical to information theory. *Journal of the Association for Computing Machinery*, 22:329-340, 1975.
3. Quinn Culver. *Topics in Algorithmic Randomness and Effective Probability*. PhD thesis, University of Notre Dame, 2015.
4. Rodney G. Downey and Denis Hirschfeldt. *Algorithmic Randomness and Complexity*. Springer-Verlag, Berlin, 2010. 855 pages.
A weak randomness notion for probability measures

1. Michael Hochman. Upcrossing inequalities for stationary sequences and applications. *Annals of Probability*, 37(6):2135–2149, 2009.
2. Mathieu Hoyrup. The dimension of ergodic random sequences. In Christoph Dürre and Thomas Wilke, editors, *STACS 2012*, pages 567–576, 2012.
3. Mathieu Hoyrup and Cristóbal Rojas. Computability of probability measures and Martin-Löf randomness over metric spaces. *Information and Computation*, 207(7):830–847, 2009.
4. Leonid A. Levin. The concept of a random sequence. *Doklady Akademii Nauk SSSR*, 212:548–550, 1973.
5. Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and its Applications*. Texts in Computer Science. Springer, New York, third edition, 2008.
6. Per Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966.
7. Daniel Mauldin and Michael Monticino. Randomly generated distributions. *Israel Journal of Mathematics*, 91(1-3):215–237, 1995.
8. Joseph S. Miller. The $K$-degrees, low for $K$-degrees, and weakly low for $K$ sets. *Notre Dame Journal of Formal Logic*, 50(4):381–391, 2009.
9. Joseph S. Miller and Liang Yu. Oscillation in the initial segment complexity of random reals. *Advances in Mathematics*, 226(6):4816–4840, 2011.
10. André Nies. *Computability and Randomness*, volume 51 of *Oxford Logic Guides*. Oxford University Press, Oxford, 2009. 444 pages. Paperback version 2011.
11. André Nies and Volker Scholz. Martin-Löf random quantum states. Technical Report on http://www.arxiv.org no 1709.08422, 2017, new version 2018.
12. Piergiorgio Odifreddi. *Classical Recursion Theory*. North-Holland, Amsterdam, 1989.
13. Piergiorgio Odifreddi. *Classical Recursion Theory, Volume II*. Elsevier, Amsterdam, 1999.
14. Claus-Peter Schnorr. Process complexity and effective random tests. *Journal on Computer and System Sciences*, 7:376–388, 1973.
15. Paul C. Shields. *The Ergodic Theory of Discrete Sample Paths*. Graduate Studies in Mathematics 13. American Mathematical Society, 1996.
16. Robert I. Soare. *Recursively Enumerable Sets and Degrees. A Study of Computable Functions and Computably Generated Sets*. Springer, Heidelberg, 1987.
17. Frank Stephan. *Recursion Theory*. Lecture notes, School of Computing, National University of Singapore, Technical Report TR10/12, 2012.