Algorithmic Randomness of Closed Sets*

GEORGE BARMPALIAS, *School of Mathematics, University of Leeds, Leeds LS2 9JT, UK.*
E-mail: georgeb@math.leeds.ac.uk

PAUL BRODHEAD, DOUGLAS CENZER and SEYYED DASHTI, *Department of Mathematics, University of Florida, P.O. Box 118105, Gainesville, Florida 32611, USA.*
E-mail: brodhead@math.ufl.edu; cenzer@math.ufl.edu; mashadeo@math.ufl.edu

REBECCA WEBER, *Department of Mathematics, Dartmouth College, Hanover, NH 03755-3551, UK.*
E-mail: rebecca.weber@dartmouth.edu

Abstract

We investigate notions of randomness in the space $\mathcal{C}[2^{\mathbb{N}}]$ of non-empty closed subsets of $\{0,1\}^{\mathbb{N}}$. A probability measure is given and a version of the Martin-Löf test for randomness is defined. Π^0_2 random closed sets exist but there are no random Π^0_1 closed sets. It is shown that any random closed set is perfect, has measure 0, and has box dimension $\log_2(4/3)$. A random closed set has no n-c.e. elements. A closed subset of $2^{\mathbb{N}}$ may be defined as the set of infinite paths through a tree and so the problem of compressibility of trees is explored. If $T_n = T \cap \{0,1\}^n$, then for any random closed set $[T]$ where T has no dead ends, $K(T_n) \geq n - O(1)$ but for any k, $K(T_n) \leq 2^{n-k} + O(1)$, where $K(\sigma)$ is the prefix-free complexity of $\sigma \in \{0,1\}^*$.

Keywords: Computability, randomness, Π^0_1 classes.

1 Introduction

The study of algorithmic randomness has been of great interest in recent years. The basic problem is to quantify the randomness of a single real number; here we will extend this problem to the randomness of the set of paths through a finitely branching tree. Early in the last century, von Mises [31] suggested that a random real should obey reasonable statistical tests, such as having a roughly equal number of zeroes and ones of the first n bits, in the limit. Thus a random real would be *stochastic* in modern parlance. If one considers only *computable* tests, then there are countably many and one can construct a real satisfying all tests.

An early approach to randomness was through betting. Effective betting on a random sequence should not allow one's capital to grow unboundedly. The betting strategies used are constructive martingales, introduced by Ville [30] and implicit in the work of Lévy [22], which represent fair double-or-nothing gambling.

Martin-Löf [24] observed that stochastic properties could be viewed as special kinds of measure zero sets and defined a random real as one which avoids certain effectively presented measure 0 sets. That is, a real $x \in 2^{\mathbb{N}}$ is Martin-Löf random if for every effective

*A preliminary version of this article appeared in the Proceedings of CIE 2006 [2].

Vol. 17 No. 6,
doi:10.1093/logcom/exm033

sequence $S_1, S_2, \ldots$ of c.e. open sets with $\mu(S_n) \leq 2^{-n}$, $x \notin \bigcap_n S_n$. It is easy to see that this is equivalent to the condition that we get if we replace 2^{-n} above with q_n for a computable sequence (q_i) of rationals such that $\lim_i q_i = 0$.

At the same time Kolmogorov [18] defined a notion of randomness for finite strings based on the concept of *incompressibility*. The stronger notion of prefix-free complexity was developed by Levin [21], Gács [16] and Chaitin [9] and extended to infinite words. Schnorr [27] later proved that the notions of constructive martingale randomness, Martin-Löf randomness and prefix-free randomness are equivalent.

In this article we want to consider algorithmic randomness on the space $\mathcal{C}$ of non-empty closed subsets P of $2^{\mathbb{N}}$. Some definitions are needed. Fix a finite alphabet $A = \{0, 1, \ldots, k-1\} = k$; we will make use of the alphabets $\{0, 1\}$ and $\{0, 1, 2\}$. For a finite string $\sigma \in A^n$, let $|\sigma| = n$. Let λ denote the empty string, which has length 0. A word (a) of length 1 is may be identified with the symbol a. For two strings σ, τ, say that τ extends σ and write $\sigma \sqsubseteq \tau$ if $|\sigma| \leq |\tau|$ and $\sigma(i) = \tau(i)$ for $i < |\sigma|$. Similarly $\sigma \sqsubset x$ for $x \in 2^{\mathbb{N}}$ means that $\sigma(i) = x(i)$ for $i < |\sigma|$. Let $\sigma^\frown\tau$ denote the concatenation of σ and τ. Let $x\lceil n = (x(0), \ldots, x(n-1))$. Now a non-empty closed set P may be identified with a tree $T_P \subseteq A^*$ as follows. For a finite string σ, let $I(\sigma)$ denote $\{x \in 2^{\mathbb{N}} : \sigma \subset x\}$. Then $T_P = \{\sigma : P \cap I(\sigma) \neq \emptyset\}$. Note that T_P has no dead ends, that is if $\sigma \in T_P$ then either $\sigma^\frown 0 \in T_P$ or $\sigma^\frown 1 \in T_P$.

For an arbitrary tree $T \subseteq A^*$, let $[T]$ denote the set of infinite paths through T, that is,

$$x \in [T] \iff (\forall n) x\lceil n \in T.$$

It is well known that $P \subseteq 2^{\mathbb{N}}$ is a closed set if and only if $P = [T]$ for some tree T. P is a Π^0_1 class, or effectively closed set, if $P = [T]$ for some computable tree T. Note that if P is a Π^0_1 class, then T_P is a Π^0_1 set, but not in general computable. P is said to be a *decidable* Π^0_1 class if T_P is computable. P is said to be a *strong* Π^0_2 class, if T_P is a Π^0_2 set, or equivalently if $P = [T]$ for some Δ^0_2 tree; P is said to be a *strong* Δ^0_2 class if T_P is Δ^0_2. Thus any Π^0_1 class is also a strong Δ^0_2 class. Any decidable Π^0_1 class contains a computable element (in particular the leftmost and rightmost paths) and similarly any strong Δ^0_2 class contains a Δ^0_2 element. On the other hand, there exist Π^0_1 classes with no computable elements and strong Π^0_2 classes with no Δ^0_2 elements. The complement of a Π^0_1 class is sometimes called a c.e. open set.

There is a natural effective enumeration $P_0, P_1, \ldots$ of the Π^0_1 classes and thus an enumeration of the c.e. open sets. Thus we can say that a sequence $S_0, S_1, \ldots$ of c.e. open sets is *effective* if there is a computable function, f, such that $S_n = 2^{\mathbb{N}} - P_{f(n)}$ for all n. For a detailed development of Π^0_1 classes, see [7] or [8].

For background and terminology on computable functions and computably enumerable sets, see [28].

The betting approach to randomness is formalized as follows.

Definition 1.1 (Ville[30])

(i) A *martingale* is a function $m : k^{<\omega} \to [0, \infty)$ such that for all $\sigma \in k^{<\omega}$,

$$m(\sigma) = \frac{1}{k} \sum_{i=0}^{k-1} m(\sigma^\frown i).$$

(ii) A *martingale* m succeeds on $X \in k^{\mathbb{N}}$ if

$$\limsup_{n \to \infty} m(X\lceil n) = \infty.$$

That is, the betting strategy results in an unbounded amount of money made on the k-ary infinite sequence X.

(iii) The success set of m is the set $S^\infty[m]$ of all sequences on which m succeeds.

That is, a martingale on $2^{<\omega}$ is the capital function of a fair double-or-nothing betting strategy. When working on $3^{<\omega}$ the strategy is triple-or-nothing.

Definition 1.2

A martingale m is *constructive* (*effective*, c.e.) if it is *lower semi-computable*; that is, if there is a computable function $\hat{m} : k^{<\omega} \times \mathbb{N} \to \mathbb{Q}$ such that

(i) for all σ and t, $\hat{m}(\sigma, t) \leq \hat{m}(\sigma, t+1) < m(\sigma)$, and
(ii) for all σ, $\lim_{t\to\infty} \hat{m}(\sigma, t) = m(\sigma)$.

In other words, $m(w)$ is approximated from below by rationals uniformly in w. A sequence in $k^{\mathbb{N}}$ is constructive martingale random if no constructive martingale succeeds on it.

Some flexibility may be gained by also considering *non-monotonic* martingales; i.e. martingales which bet on the bits of a sequence out of order. While for a monotonic martingale only the *amount* of the next bet is determined from the bits seen previously, for a non-monotonic martingale both the amount and the *location* of the next bet are determined from the bits seen previously (the next bit may precede them, follow them, or lie in the middle). These martingales must obey two rules: the standard fair-betting rule that monotonic martingales obey, and the rule that they never bet on the same bit twice. We refer the reader to Downey and Hirschfeldt [11] for the formal definition.

Although *a priori* allowing non-monotonic martingales strengthens the notion of randomness, since more strategies must be defeated, in fact in the c.e. case they are equivalent. Muchnik *et al.* [25] (Theorem 8.9) show that ML-random sequences defeat all *computable* non-monotonic martingales (in fact they show this with respect to general measures, not just the coin-toss measure). The proof does not depend on the computability of the martingale, however; the martingale is used to define a Martin-Löf test which may be enumerated equally well alongside the enumeration of the martingale. Therefore, as defeating all c.e. non-monotonic martingales is clearly sufficient to be ML-random, the two are equivalent.

Prefix-free randomness for reals is defined as follows. A Turing machine M which takes inputs from A^*, where A is a finite alphabet, is called prefix-free if it has prefix-free domain dom(M); that is, if $\sigma \sqsubseteq \tau$ are strings in dom(M), then σ must equal τ. For any finite string τ, the *prefix-free complexity of τ with respect to M* is

$$K_M(\tau) = \min\{|\sigma|, \infty : M(\sigma) = \tau\}.$$

There is a *universal* prefix-free function U such that, for any prefix-free M, there is a constant c such that for all τ

$$K_U(\tau) \leq K_M(\tau) + c.$$

We let $K(\tau) = K_U(\tau)$ and call it the *prefix-free complexity of τ*. Then x is called *prefix-free random* if there is a constant c such that $K(x\lceil n) \geq n - c$ for all n. This means that the initial segments of x are not *compressible*.

The equivalence of these three notions of randomness (via tests, betting or incompressibility) is a result of Schnorr [27] and is a fundamental result in the theory of algorithmic randomness. While these definitions and results are usually given for binary strings and sequences, they carry over to k-ary strings and sequences as well. See for example Calude [5, 6]. The following lemma will be needed.

LEMMA 1.3
If P is a Π^0_1 class of measure 0, then P has no random elements.

PROOF. Let T be a computable tree such that $P=[T]$, and for each n, let $P_n = \bigcup\{I(\sigma) : \sigma \in T \cap \{0,1\}^n\}$. Then $\{P_n\}_{n\in\mathbb{N}}$ is an effective sequence of clopen sets with $P=\bigcap_n P_n$ and $\lim_n \mu(P_n) = \mu(P) = 0$. Furthermore,

$$\mu(P_n) = 2^{-n}|T \cap \{0,1\}^n|$$

and is therefore a computable sequence. Thus $\{P_n\}_{n\in\mathbb{N}}$ is a Martin-Löf test, showing that P has no random elements.

We will want to use the following result from the literature [31]. ■

THEOREM 1.4 (Von-Mises-Church-Wald Computable Selection Theorem)
For any random sequence x and any computable 1-1 function g, the sequence $z(n) = x(g(n))$ is random.

2 Martin-Löf randomness of closed sets

In this section, we define a measure on the space $\mathcal{C}$ of non-empty closed subsets of $2^{\mathbb{N}}$ and use this to define the notion of randomness for closed sets. We then obtain several properties of random closed sets.

An effective one-to-one correspondence between the space $\mathcal{C}$ and the space $3^{\mathbb{N}}$ is defined as follows. Let a closed set Q be given and let $T=T_Q$ be the tree without dead ends such that $Q=[T]$.

Define the code $x = x_Q \in \{0,1,2\}^{\mathbb{N}}$ for Q as follows. Let $\lambda = \sigma_0, \sigma_1, \sigma_2, \ldots$ enumerate the elements of T in order, first by length and then lexicographically. We now define $x = x_Q = x_T$ by recursion as follows. For each n, $x(n)=2$ if $\sigma_n{}^\frown 0$ and $\sigma_n{}^\frown 1$ are both in T, $x(n)=1$ if $\sigma_n{}^\frown 0 \notin T$ and $\sigma_n{}^\frown 1 \in T$ and $x(n)=0$ if $\sigma_n{}^\frown 0 \in T$ and $\sigma_n{}^\frown 1 \notin T$. For example, if $Q=\{0,1\}^{\mathbb{N}}$, then $x_Q = (2,2,\ldots)$ and if $Q=\{y\}$, then $x_Q = y$. Let Q_x denote the unique closed set Q such that $x_Q = x$.

Now define the measure μ^* on $\mathcal{C}$ by

$$\mu^*(\mathcal{X}) = \mu(\{x_Q \neg \mathcal{Q} \in \mathcal{X}\}).$$

Informally this means that given $\sigma \in T_Q$, there is probability 1/3 that both $\sigma{}^\frown 0 \in T_Q$ and $\sigma{}^\frown 1 \in T_Q$ and, for $i=0,1$, there is probability 1/3 that only $\sigma{}^\frown i \in T_Q$. In particular, this means that $Q \cap I(\sigma) \neq \emptyset$ implies that for $i=0,1$, $Q \cap I(\sigma{}^\frown i) \neq \emptyset$ with probability 2/3.

Let us comment briefly on why some other natural representations were rejected. Suppose first that we simply enumerate all strings in $\{0,1\}^*$ as $\sigma_0, \sigma_1, \ldots$ and then represent T by its characteristic function so that $x_T(n) = 1 \iff \sigma_n \in T$. Then in general a code x might not represent a tree. That is, once we have $(01) \notin T$ we cannot later decide that $(011) \in T$. Suppose then that we allow the empty closed set by using codes $x \in \{0,1,2,3\}^*$ and modify our original definition as follows. Let $x(n) = i$ have the same definition as above for $i \leq 2$ but let $x(n) = 3$ mean that neither $\sigma^\frown_n 0$ nor $\sigma^\frown 1$ is in T. Informally, this would mean that for $i = 0, 1$, $\sigma \in T$ implies that $\sigma^\frown i \in T$ with probability 1/2. The advantage here is that we can now represent all trees. But this is also a disadvantage, since for a given closed set P, there are many different trees T with $P = [T]$. The second problem with this approach is that we would have $[T] = \emptyset$ with positive probability. We briefly return to this subject in Section 6.

Now we will say that a closed set Q is (Martin-Löf) random if the code x_Q is Martin-Löf random. This definition clearly relativizes to any oracle in accordance with the definitions of relative randomness in the Cantor space. Since random reals exist, it follows that random closed sets exists. Furthermore, there are Δ^0_2 random reals, so we have the following.

Theorem 2.1
There exists a random closed set Q such that T_Q is Δ^0_2.

Note that if T_Q is Δ^0_2, then Q must contain Δ^0_2 elements (in particular the leftmost path). Since there exist strong Π^0_2 classes with no Δ^0_2 elements, there are strong Π^0_2 classes Q such that T_Q is not Δ^0_2.

The following lemma will be needed throughout.

Lemma 2.2
For any $Q \subseteq 2^{\mathbb{N}}$ which is either closed or open,

$$\mu^*(\{P : P \subseteq Q\}) \leq \mu(Q).$$

Proof. Let $\mathcal{P}_C(Q)$ denote $\{P : P \subseteq Q\}$. We first prove the result for non-empty clopen sets U in place of Q by the following induction. Suppose $U = \bigcup_{\sigma \in S} I(\sigma)$, where $S \subseteq \{0,1\}^n$. For $n = 1$, either $\mu(U) = 1 = \mu^*(\mathcal{P}_C(U))$ or $\mu(U) = 1/2$ and $\mu^*(\mathcal{P}_C(Q)) = 1/3$. For the induction step, let $S_i = \{\sigma : i^\frown\sigma \in S\}$, let $U_i = \bigcup_{\sigma \in S_i} I(\sigma)$, let $u_i = \mu(U_i)$ and let $v_i = \mu^*(\mathcal{P}_C(U_i))$, for $i = 0, 1$. Then considering the three cases in which S includes both initial branches or just one, we calculate that

$$\mu^*(\mathcal{P}_C(U)) = \frac{1}{3}(v_0 + v_1 + v_0 v_1).$$

Thus, by induction we have

$$\mu^*(\mathcal{P}_C(U)) \leq \frac{1}{3}(u_0 + u_1 + u_0 u_1).$$

Now

$$2u_0 u_1 \leq u_0^2 + u_1^2 \leq u_0 + u_1,$$

and therefore

$$\mu^*(\mathcal{P}_C(U)) \leq \frac{1}{3}(u_0 + u_1 + u_0 u_1) \leq \frac{1}{2}(u_0 + u_1) = \mu(U).$$

For a closed set Q, let $Q = \bigcap_n U_n$, where U_n is clopen and $U_{n+1} \subseteq U_n$ for all n. Then $P \subset Q$ if and only if $P \subseteq U_n$ for all n. Thus

$$\mathcal{P}_C(Q) = \bigcap_n \mathcal{P}_C(U_n),$$

so that

$$\mu^*(\mathcal{P}_C(Q)) = \lim_{n \to \infty} \mu^*(\mathcal{P}_C(U_n)) \leq \lim_{n \to \infty} \mu(U_n) = \mu(Q).$$

Finally, for an open set Q, let $Q = \bigcup_n U_n$ be the union of an increasing sequence of clopen sets U_n. Then, by compactness,

$$\mathcal{P}_C(Q) = \bigcup_n \mathcal{P}_C(U_n),$$

so that

$$\mu^*(\mathcal{P}_C(Q)) = \lim_{n\to\infty} \mu^*(\mathcal{P}_C(U_n)) \le \lim_{n\to\infty} \mu(U_n) = \mu(Q).$$

This completes the proof of the lemma. ■

Next we will consider the intersection of a random closed set with an interval $I(\sigma)$ and the disjoint union of random closed sets.

First recall van Lambalgen's theorem.

THEOREM 2.3 (Van Lambalgen [29])
The following are equivalent.

1. $A \oplus B$ is n-random.
2. A is n-random and B is n-A-random.
3. B is n-random and A is n-B-random.
4. A is n-B-random and B is n-A-random.

Let us call the coding of a closed set Q by the nodes of its representative tree with no dead ends the *canonical code* of Q. We wish now to introduce a second method of coding, the *ghost code*. A ghost code of Q is an infinite ternary string whose terms correspond to all nodes of $2^{<\omega}$ in lexicographical order. The terms corresponding to the nodes of Q's tree (the 'canonical nodes') agree with the corresponding terms in the canonical code; the remaining 'ghost nodes' may hold any values. Ghost codes are non-unique, and every closed set has a non-random ghost code (if the closed set itself is random take the code with ghost nodes all equal to zero, say). This method of coding is more convenient for some purposes; for example, we will use it to show that if Q_0, Q_1 are closed sets and $Q = \{0^\frown x : x \in Q_0\} \cup \{1^\frown x : x \in Q_1\}$, Q is random if and only if the Q_i are random relative to each other. The utility of the ghost codes rests on the following correspondence.

THEOREM 2.4
The canonical code of a closed set $Q \subseteq 2^{\mathbb{N}}$ is random if and only if Q has some random ghost code. Furthermore, for any y, the canonical code r is y-random if and only if Q has a ghost code which is y-random.

PROOF. ($\Leftarrow$) Suppose the canonical code of Q is non-random. Then there is a c.e. martingale m that succeeds on it. From any initial segment σ of a ghost code g for Q, the subsequence $\hat{\sigma}$ of exactly the canonical nodes of σ is computable. Therefore it is computable whether the bit of g after σ is canonical or ghost. From m, define the martingale m' which bets as follows:

$$m'(\sigma^\frown i) = \begin{cases} m(\hat{\sigma}^\frown \mathrm{i}) & \text{next bit is a canonical node} \\ m'(\sigma) & \text{next bit is a ghost node.} \end{cases}$$

That is, m' holds its money on ghost nodes and bets identically to m on canonical nodes. It is clear that m' succeeds on the ghost code g and thus g is non-random.

($\Rightarrow$) Now suppose the canonical code r for Q is random, and let q be an infinite ternary string that is random relative to r (and so by Theorem 2.3 $r \oplus q$ is random). We claim the ghost code g obtained by using the bits of r as the canonical nodes and the bits of q in their original order as the ghost nodes is random. It is clear that g is a ghost code for Q.

Suppose m is a c.e. martingale that bets on g. From m it is straightforward to define a non-monotonic martingale m' which mimics m's bets exactly but performs them on $r \oplus q$, succeeding whenever m succeeds. As r and q were chosen to be relatively random, this will show g is random.

As discussed previously, from $g\lceil n$ it is computable whether $g(n)$ will be a ghost node or a canonical node, and which position in g or r it occupies in either case. Therefore, assuming the bits seen so far may be assembled into an initial segment σ of g, m' takes the values $m(\sigma^\frown i)$, $i < 3$, as its bets on the corresponding bit of r or g, whichever is appropriate. Having seen that bit, then, it can assemble a $(|\sigma| + 1)$-length initial segment of g and repeat the process. As m' makes identical bets to m and has identical outcomes, since it cannot succeed on $r \oplus g$, m cannot succeed on g and g is random.

To relativize ($\Rightarrow$), suppose that r is y-random, so that $r \oplus y$ is random by Van Lambalgen's Theorem 2.3. Then in the proof simply choose q to be random relative to $r \oplus y$, and then g will be random relative to y. The other direction relativizes in a straightforward way. ■

The primary purpose of the ghost codes is to remove the dependence on the particular closed set under discussion when interpreting bits of the code as nodes of the tree. This is especially useful when subdividing the tree, as in the following definition.

Definition 2.5
The tree join of closed sets P_0 and P_1 is the closed set

$$Q = \{0^\frown x : x \in P_0\} \cup \{1^\frown x : x \in P_1\}.$$

Given ghost codes r_0, r_1 for the P_i, their tree join $r_0 \boxplus r_1$ is the code for Q with the corresponding ghost node values.

The standard *recursion–theoretic* join is defined by

$$r_0 \oplus r_1 = (r_0(0), r_1(0), r_0(1), r_1(1), \ldots).$$

We wish to relate the recursion–theoretic join and the tree join.

Lemma 2.6
Given two ghost codes r_0, r_1, the tree join $r_0 \boxplus r_1$ is random if and only if the recursion theoretic join $r_0 \oplus r_1$ is random.

Proof. It is clear that there is a computable permutation π which uniformly maps any tree join $r_0 \boxplus r_1$ to the recursion–theoretic join $r_0 \oplus r_1$. That is, in $r_0 \oplus r_1$, the entries of r_0 and r_1 alternate, whereas $r_0 \boxplus r_1$ starts with a 2, followed by blocks from r_0 and r_1, as follows. First $r_0(0)$, $r_1(0)$, then $r_0(1)$, $r_0(2)$, $r_1(1)$, $r_1(2)$, and continuing with pairs of blocks of size 4, 8 and so on. The result now follows from the Von-Mises–Church–Wald Computable Selection Theorem 1.4. ■

We now obtain the following corollary of Theorems 2.3 and 2.4, and Lemma 2.6.

COROLLARY 2.7
Suppose P_i, $i=0,1$, are closed sets with canonical codes r_i and let P be the tree join of P_0, P_1. Then P is random if and only if $r_0 \oplus r_1$ is random.

PROOF. ($\Leftarrow$) Suppose that $r_0 \oplus r_1$ is random. Then by Theorem 2.3, r_0 and r_1 are mutually relatively random. By Theorem 2.4, P_0 has a ghost code g_0 which is random relative to r_1, and so also vice versa, and then P_1 has a ghost code g_1 which is random relative to g_0. Again by Theorem 2.3, the recursion–theoretic join $g_0 \oplus g_1$ is random, so by Theorem 2.6 the tree join $g_0 \boxplus g_1$ is also random, and hence P possesses a random ghost code and is random.
($\Rightarrow$) Suppose now that P is random, and therefore possesses a random ghost code g. The code g may be thought of as a tree join $g_0 \boxplus g_1$, which is therefore random, and so by Theorem 2.6, $g_0 \oplus g_1$ is random. By Theorem 2.3, the individual codes g_0, g_1 are therefore mutually relatively random. Now by the relatived version of Theorem 2.4, r_0 is random relative to g_1. But r_1 is computable from g_1 and hence r_0 is random relative to r_1 as well. Similarly, r_1 is r_0-random and thus, again by 2.3, $r_0 \oplus r_1$ is random. ∎

3 Members of random closed sets

For any finite string σ of length n, the probability that a closed set Q meets $I(\sigma)$ is $(2/3)^n$. For a computable real y, the sequence $\{Q : Q \cap I(y\lceil n) \neq \emptyset\}$ thus forms a Martin-Löf test in the space $\mathcal{C}$ of closed sets, which shows that y does not belong to any Martin-Löf random closed set. That is, for each n, $\{x : Q_x \cap I(y\lceil n) \neq \emptyset\}$ is a c.e. open set and has measure $(2/3)^n$ in $\{0,1,2\}^{\mathbb{N}}$, where Q_x is the closed set with code x. We omit the details, since we will now prove a stronger result.

For any computable, non-decreasing function f, we say that a real $\beta \in \{0,1\}^{\mathbb{N}}$ is *f-c.e.* if there exists a computable approximating function ϕ such that, for all $i \in \mathbb{N}$,

(i) $\phi(i,0)=0$;
(ii) $\lim_s \phi(i,s) = \beta(i)$;
(iii) $\{s : \phi(i,s+1) \neq \phi(i,s)\}$ has cardinality $\leq f(i)$.

The reals which are f-c.e. for some computable function f are part of the well-known Ershov hierarchy [14, 28].

THEOREM 3.1
Suppose that f is computable and bounded by a polynomial. Then no random closed set has any f-c.e. paths.

PROOF. Let f be as above, β an f-c.e. real and P a closed set containing β. Let ϕ be the f-approximating function for β. Also let $M_n \subseteq \{0,1\}^n$ be the set of different ϕ-approximations to $\beta\lceil n$ during the stages.

A priori, $|M_n|$ is exponential. However, for a fixed n, $\beta\lceil n$ can change at most $\sum_{i<n} f(i)$ times, so $|M_n|$ is also bounded by a polynomial, i.e. there is $k \in \mathbb{N}$ such that for almost all n, $|M_n| < n^k$. Now let

$$S_n = \bigcup_{\sigma \in M_n} \{P \mid P \in \mathcal{C} \ \&\ P \cap I(\sigma) \neq \emptyset\}. \tag{1}$$

Then (S_n) is a uniformly c.e. sequence of open sets in the space $\mathcal{C}$ of closed sets of $2^{\mathbb{N}}$ and for all n, $P \in S_n$. Also for almost all n,

$$\mu^*(S_n) \leq \sum_{\sigma \in M_n} \mu^*(\{P \mid P \in \mathcal{C} \ \&\ P \cap I(\sigma) \neq \emptyset\}) = |M_n| \cdot \left(\frac{2}{3}\right)^n \leq n^k \cdot \left(\frac{2}{3}\right)^n.$$

Since $\lim_n[n^k \cdot (2/3)^n] = 0$ there is a computable subsequence of (S_n) which is a Martin-Löf test and so P is not random. ■

For any K-trivial real A and any unbounded non-decreasing computable function h, A is h-c.e. [26]. Thus it follows from Theorem 3.1 that a random closed set can have no K-trivial paths. We observe that Theorem 3.1 cannot be extended to ω-c.e. in general, because there are left-c.e. (and hence ω-c.e.) random reals, and by Theorem 3.13 each of these belongs to a random closed set.

Recall that a real x is 1-generic if for any Π^0_1 class $Q = [T]$, either $x \notin Q$ or there exists $\sigma \in T$ such that $I(\sigma) \subseteq Q$. Note that 1-genericity is closed downwards under Turing reductions below $\mathbf{0}'$ [17].

Theorem 3.2
No random closed set has any 1-generic Δ^0_2 path.

Proof. Let x be a 1-generic Δ^0_2 real with an effective approximation $x(n) = lim_s f(n, s)$, and let P a closed set containing x. Let $x_s = (f(0, s), f(1, s), \ldots$ and let $M_n = \{x_s \lceil s : s > n\}$ and observe that by the 1-genericity of x every M_n contains an initial segment of x. That is, if $T = \{\sigma : (\forall s > n)$ not $x_s \lceil s \sqsubseteq \sigma\}$, then we must have $x \notin [T]$, so that some $x_s \lceil s \sqsubseteq x$, or $x \in [T]$, This is because if $x \in [T]$, then for some k, $I(x \lceil k) \subseteq [T]$. But for any k and n, there certainly exists $s > n$ such that $x \lceil k \sqsubseteq x_s \lceil k$, which would put $x \notin [T]$ by the definition of T.

Now if (S_n) is as in (1) we have $\mu(S_n) \leq \sum_{s>n} \left(\frac{2}{3}\right)^s$ and $P \in \cap_n S_n$. Since the sequence $\sum_{s>n} \left(\frac{2}{3}\right)^n$ is uniformly computable and converges to zero, there is a computable subsequence of (S_n) which is a Martin-Löf test and so P is not random. ■

It is a plausible conjecture that this can be extended to 1-generic paths in general.

Theorem 3.3
If Q is a random closed set, then Q has no isolated elements.

Proof. Let $Q = [T]$ and suppose by way of contradiction that Q contains an isolated path x. Then there is some node $\sigma \in T$ such that $Q \cap I(\sigma) = \{x\}$. For each n, let

$$S_n = \{P \in \mathrm{C} : |\{\tau \in \{0, 1\}^n : P \cap I(\sigma^\frown \tau) \neq \emptyset\}| = 1\}.$$

That is, $P \in S_n$ if and only if the tree T_P has exactly one extension of σ of length $n + |\sigma|$. It follows that

$$|P \cap I(\sigma)| = 1 \iff (\forall n) P \in S_n$$

Now for each n, S_n is a clopen set in $\mathcal{C}$ and again by induction, S_n has measure $(2/3)^n$. Thus the sequence $S_0, S_1, \ldots$ is a Martin-Löf test. It follows that for some n, $Q \notin S_n$. Thus there are at least two extensions in T_Q of σ of length $n+|\sigma|$, contradicting the assumption that x was the unique element of $Q \cap I(\sigma)$. ■

COROLLARY 3.4
If Q is a random closed set, then Q is perfect and hence has continuum many elements.

THEOREM 3.5
Every random closed set contains a random element.

PROOF. Suppose that a closed set Q has no random element and consider the following Martin-Löf test on the space $\mathcal{C}$:

$$U_i = \{P \mid P \in \mathcal{C} \text{ and } P \subseteq V_i\}$$

where (V_i) is a universal Martin-Löf test on the Cantor space. By Lemma 2.2, $\mu^*(U_i) \leq \mu(V_i) \leq 2^{-i}$ so that (U_i) is a Martin-Löf test on $\mathcal{C}$. But $Q \in \cap_i U_i$, so Q is not random. ■

The previous results might suggest that every element of a random closed set is a random real. However, it turns out that every random closed set contains a non-random real.

We need the following classic result of Chernoff [10] (a version of Bernoulli's *Weak Law of Large Numbers*) here and also for another theorem to follow. See [23] for an exposition.

LEMMA 3.6 (Chernoff)
Let E be an event which we will refer to as 'success'. If E occurs with probability p, then for any natural numbers n and any ε with $0 \leq \varepsilon \leq 1$, the probability that out of n mutually independent trials, the number of successes differs from pn by $> \varepsilon pn$ is $\leq 2^{-\varepsilon^2 pn/3}$.

THEOREM 3.7
Not every element of a random closed set is random; in particular, the leftmost and rightmost paths in a random closed set are not random reals.

PROOF. We will show that, for a random closed set Q, the leftmost path is not stochastically random, that is, the asymptotic frequency of 0's is 2/3. Since an effectively random real in $2^{\mathbb{N}}$ must have asymptotic frequence of 1/2 for 0's and 1's, this will suffice to prove that the leftmost path is not random. We define a Martin-Löf test as follows. Fix a rational ε such that $0 < \varepsilon < 1$. For each n, let S_n be the family of closed sets (that is, codes for closed sets) such that the first n bits of the leftmost path have either $<(2/3)(1-\varepsilon)n$, or $>(2/3)(1+\varepsilon)n$ occurrences of 0. By the definition of our probability measure, we have

$$\mu^*(S_n) = \sum_{|m-\frac{2}{3}n|>\frac{2}{3}\varepsilon n} \binom{n}{m} \left(\frac{2}{3}\right)^m \left(\frac{1}{3}\right)^{n-m}.$$

It now follows from Chernoff's Lemma 3.9 that

$$\mu^*(S_n) \leq 2e^{-\varepsilon^2 2n/9}.$$

Thus the measures of the test sets S_n have effective limit zero. It is easy to see that the sequence $\{S_n\}$ is computably enumerable. For each n, S_n is a clopen set and in fact the union of the finite family of intervals $I(\sigma)$ in $\mathcal{C}$ such that σ codes a tree up to level n in which the leftmost path has either $<(2/3)(1-\varepsilon)n$, or $>(2/3)(1+\varepsilon)n$ occurrences of 0.

Furthermore, $S'_n = \bigcup_{p \geq n} S_p$ is also a Martin-Löf test. It follows that for any random closed set Q, and any $\varepsilon > 0$, there is an n such that for all $m \geq n$, the frequency of 0's in the first m bits of the leftmost path is always within ε of 2/3. Thus the leftmost path is not effectively random. ∎

Recall that the leftmost and rightmost elements of any strong Δ^0_2 closed set are Δ^0_2. Given Theorems 3.8 and 3.10, we ask: Does a Δ^0_2 random closed set contain a Δ^0_2 random path?

Theorem 3.8
Every random strong Δ^0_2 closed set contains a random Δ^0_2 real.

Proof. Let Q be a random strong Δ^0_2 class. By Theorem 3.8, Q contains a random real x. Let P be a Π^0_1 class in the Cantor space which contains only randoms and contains x (this exists since the class of random reals is an effective union of Π^0_1 classes). Note that $P \cap Q$ is a non-empty strong Δ^0_2 class and it follows that the leftmost path of $P \cap Q$ is a Δ^0_2 real which must be random since it belongs to P. ∎

Note that the above theorem does not combine with the low basis theorem to establish the existence of a low random real in any random strong Δ^0_2 class. Thus we pose the question of whether for any random closed set Q, if T_Q is low, then Q has a low random element.

Next we want to find a random closed set which does not contain a Δ^0_2 path. Now it is easy [7, 8] to construct a strong Π^0_2 class P of positive measure which contains no Δ^0_2 elements; of course P must contain a random real since it has measure 1. The difficult problem is to construct a *random* strong Π^0_2 class with no Δ^0_2 elements. We have the following result in this direction, which yields a random strong Δ^0_3 closed set with no Δ^0_2 elements.

Theorem 3.9
For any set A there is an A-random closed set Q such that $T_Q \leq_T A''$ but Q has no elements $\leq_T A'$.

Proof. It is enough if we prove the claim for $A = \emptyset$ because the argument relativises to any oracle A in a straightforward way. For $A = \emptyset$ we use a finite injury construction over $\emptyset'$ to construct Q with the above properties. In the construction we will $\emptyset'$-approximate the canonical code of a tree T which has no Δ^0_2 paths. To make sure that the tree T is random we fix a Π^0_1 class P of positive measure in the space $3^{\mathbb{N}}$ (where the code for T lies) which contains only randoms, and we make sure that at every stage our approximation (as a finite ternary string) to T's canonical code can be extended to a path in P. Then by compactness the canonical code of our tree will be in P and so the tree will be random. The changes in the approximations are motivated by the requirements:

$$R_e : \text{if } \Phi_e^{\emptyset'} \text{is total then the real it defines is not in } [T].$$

Let α_s be a finite string approximation of the canonical code α we are building. We will have $|\alpha_s| = s$. Strategy R_e will come into power after stage e and will restrain

α up to some $r_e \geq e$ (the default value is $r_e[0] = e$). Also it might request some changes in α after the e-th bit. We start with $\alpha_0 = \emptyset$ and at stage $s+1$, assuming inductively that $\alpha_s \downarrow$ and $[\alpha_s] \cap P \neq \emptyset$ we ask for the least $i < s$ such that R_i *requires attention*. This happens if

(i) The longest defined initial segment τ of $\Phi^{\emptyset'}_{i,s+1}$ is larger than ever before;
(ii) there exists $\sigma \in \{0,1,2\}^*$ such that $\alpha_s \lceil (\max_{j<i} r_j[s]) \sqsubseteq \sigma$, $I(\sigma) \cap P \neq \emptyset$, $|\sigma| = s+1$, and τ is not consistent with the finite tree with code σ.

If there is no such i then we extend α_s by one bit such that $[\alpha_{s+1}] \cap P \neq \emptyset$. Otherwise we let $\alpha_{s+1} = \sigma$ and $r_i[s+1] = s+1$. The construction proceeds in a straightforward way and we can prove inductively that for every e, R_e is satisfied, stops requiring attention and r_e reaches a limit. Then the limit $\alpha = \lim_s \alpha_s$ exists and we also have that α is random by compactness. The satisfaction of the requirements comes from a measure-theoretic fact. Consider R_e and inductively assume that after stage s_e no R_i with $i < e$ requires attention. Then $r = \max_{i<e} r_i$ will remain constant. Since P contains only randoms and $[\alpha \lceil \max_{i<e} r_i] \cap P \neq \emptyset$,

$$\mu([\alpha \lceil r] \cap P) > 0$$

and on the other hand, if $\beta = \Phi^{\emptyset'}_e$ we have seen that

$$\mu\{\gamma \mid \gamma \in 3^{\mathbb{N}} \text{ and } \gamma \text{ is the canonical code of a tree which has } \beta \text{ as a path}\} = 0.$$

This means that if at stage s_e the requirement R_e is not yet satisfied, it will receive attention at a later stage and get satisfied permanently. ∎

As a converse to Theorem 3.8 we have the following.

Theorem 3.10
For any random $r \in 2^{\mathbb{N}}$, there exists a random closed set containing r as a path.

The proof of this theorem was originally given by Joe Miller and Antonio Montalbán and has been subsequently improved thanks to the anonymous referee.

Proof. Let r be a random real and let x be the canonical code of an r-random closed set. We alter x to the code x' of a closed set guaranteed to contain r but changed as little as possible to achieve that.

To determine $x'(n)$, assume $x' \lceil n$ has been defined. If $x(n) = 2$ or $x(n)$ corresponds to a node not along r, set $x'(n) = x(n)$. If $x(n) \in \{0,1\}$ corresponds to $r(k)$, set $x'(n) = r(k)$.

The closed set defined by x' will clearly contain r. For a contradiction, assume x' is non-random and let m' be a c.e. martingale that succeeds on it. We build a non-monotonic martingale m to bet on $x \oplus r$. On bits of x, m will be a triple-or-nothing martingale; on r, it will be double-or-nothing.

First note that from initial segments of x and r we may reconstruct an initial segment of x' computably, and we always know from an initial segment of x' whether the next bit is along r or not, and which bit of r it is. We will construct m so that after every stage of betting (which will be one bet by m' and one or two bets by m), the value of m is equal to the value of m'. At every stage it will be clear we have revealed enough bits of x and r to reconstruct x' to the needed length.

Suppose inductively m and m' hold equal capital after the stage of betting on the last node of $\sigma \sqsubset x'$. If the bit $x'(n)$ following σ is not on r, m bets identically to m'; i.e. $m(x(n)=i)=m'(\sigma^\frown i)$ for $i<3$. In that case $x(n)=x'(n)$ so our inductive hypothesis holds. If $x'(n)$ is on r, set $m(x(n)=2)=m'(\sigma^\frown 2)$ and for $i=0,1$, set $m(x(n)=i)=(1/2)[m'(\sigma^\frown 0)+m'(\sigma^\frown 1)]$. If $x'(n)=2$, then the capital for both m and m' is $m'(\sigma^\frown 2)$, so the inductive hypothesis holds and we proceed to the next stage. Otherwise m bets on $r(k)$ for the appropriate k, setting $m(r(k)=i)=m'(\sigma^\frown i)$ for $i=0,1$. On $r(k)$, the sum of m's capital on each of the two outcomes must average to the previous capital; as the previous capital was $(1/2)[m'(\sigma^\frown 0)+m'(\sigma^\frown 1)]$ this clearly holds. By construction $r(k)=x'(n)=i$, so both m and m' now have capital $m'(\sigma^\frown i)$ and the inductive hypothesis holds. As m' is c.e., m will also be.

As the values of m' along x' are a subsequence of the values of m along $x \oplus r$, if m' succeeds so does m, contradicting our assumption on $x \oplus r$. Therefore x' is the code of a random closed set containing the given random path r. ■

4 Measure and dimension

Theorem 4.1

If Q is a random closed set, then $\mu(Q)=0$.

Proof. We will show that in the space $\mathcal{C}$ of closed sets, the μ^*-probability that a closed set P has Lebesgue measure 0, is 1. This is proved by showing that for each m, $\mu(P)\geq 2^{-m}$ with μ^*-probability 0. For each m, let

$$S_m=\{P:\mu(P)\geq 2^{-m}\}.$$

We claim that for each m, $\mu^*(S_m)=0$. The proof is by induction on m.

For $m=0$, we have $\mu(P)\geq 1$ if and only if $P=2^{\mathbb{N}}$, which is if and only if $x_P=(2,2,\ldots)$, so that S_0 is a singleton and thus has measure 0.

Now assume by induction that S_m has measure 0. Then the probability that a closed set $P=[T]$ has measure $\geq 2^{-m-1}$ can be calculated in two parts.

(i) If T does not branch at the first level, say $T_0=\{(0)\}$ without loss of generality. Now consider the closed set $P_0=\{y:0^\frown y\in P\}$. Then $\mu(P)\geq 2^{-m-1}$ if and only if $\mu(P_0)\geq 2^{-m}$, which has probability 0 by induction, so we can discount this case.

(ii) If T does branch at the first level, let $P_i=\{y:i^\frown y\in P\}$ for $i=0,1$. Then $\mu(P)=(1/2)(\mu(P_0)+\mu(P_1))$, so that $\mu(P)\geq 2^{-m-1}$ implies that at least one of $\mu(P_i)\geq 2^{-m-1}$. (Note that the reverse implication is not always true.) Let $p=\mu^*(S_{m+1})$. The observations above imply that

$$p\leq \frac{1}{3}(1-(1-p)^2)=\frac{2}{3}\,p-\frac{1}{3}p^2,$$

and therefore $p=0$.

To see that a random closed set Q must have measure 0, fix m and let $S=S_m$. Then S is the intersection of an effective sequence of clopen sets V_ℓ, where for $P=[T]$,

$$P\in V_\ell \iff \mu([T_\ell])\geq 2^{-m}.$$

Since these sets are uniformly clopen, the sequence $m_\ell = \mu^*(V_\ell)$ is computable. Since $\lim_\ell\ m_\ell = 0$, it follows that this is a Martin-Löf test and therefore no random set Q belongs to $\bigcap_\ell V_\ell$. Then in general, no random set can have measure $\geq 2^{-m}$ for any m. ■

Recall that a Π^0_1 class P is decidable if T_P is decidable. It follows that a non-empty decidable Π^0_1 class must contain a computable element (for example, the leftmost path). No computable real can be random and it follows that no decidable Π^0_1 class can be random. We will extend this to arbitrary Π^0_1 classes in Corollary 4.3.

Theorem 4.2
Let Q be a Π^0_1 class with measure 0. Then no subset of Q is random.

Proof. Let T be a computable tree (possibly with dead ends) and let $Q=[T]$. Then $Q=\bigcap_n U_n$, where $U_n=[T_n]$. Since $\mu(Q)=0$, it follows from Lemma 2.2 that $\lim_n\ \mu^*(\mathcal{P}_C(U_n))=0$. But $\mathcal{P}_C(U_n)$ is a computable sequence of clopen sets in $\mathcal{C}$ and $\mu^*(\mathcal{P}_C(U_n))$ is a computable sequence of rationals with limit 0. Thus $\mathcal{P}_C(U_n)$ is a Martin-Löf test, so that for any random closed set, there exists n such that $P \notin \mathcal{P}_C(U_n)$ and hence P is not a subset of U_n. ■

Since any random class has measure 0, we have the following immediate corollary.

Corollary 4.3
No Π^0_1 class can be random.

Surprisingly, we can compute the (Kolmogorov) box dimension of a random closed set, and in fact it turns out that all random closed sets have the same dimension. The intuition for this comes from the following lemma. For any function F mapping the space $\mathcal{C}$ of closed sets into $\Re$, the *expected* value of F on $\mathcal{C}$ is the integral $\int F(P)$ with respect to the probability measure μ^*.

Lemma 4.4
In the space $\mathcal{C}$ of closed sets, the expected cardinality of $\{\sigma \in \{0,1\}^n : Q \cap I(\sigma) \neq \emptyset\}$ is exactly $(4/3)^n$ for every n, where Q is chosen uniformly at random according to μ^*.

Proof. Let $S_n = \{\sigma \in \{0,1\}^n : Q \cap I(\sigma) \neq \emptyset\}$, for a randomly chosen Q from $\mathcal{C}$.

The proof is by induction on n. For $n=1$, we have two cases. With probability 2/3, $\mathrm{card}(S_1)=1$ and with probability 1/3, $\mathrm{card}(S_1)=2$. Thus the expected value is exactly 4/3. For $n+1$, there are again two cases. With probability 2/3, $\mathrm{card}(S_1)=1$, so that the expected $\mathrm{card}(S_{n+1})$ equals the expected $\mathrm{card}(S_n)$, which is $(4/3)^n$ by induction. With probability 1/3, $\mathrm{card}(S_1)=2$, in which case the expected $\mathrm{card}(S_{n+1})$ is twice the expected $\mathrm{card}(S_n)$, that is, $2(4/3)^n$. Thus we have the expected value

$$\mathrm{card}(\mathrm{S}_{n+1}) = \frac{2}{3}\left(\frac{4}{3}\right)^n + \frac{1}{3}\cdot 2\left(\frac{4}{3}\right)^n = \left(\frac{4}{3}\right)^{n+1}. \qquad ■$$

The box dimension of a closed set in the Cantor space, if it exists, is given by the following limit:

$$\dim_B F(Q) = \lim_{n\to\infty} \frac{\log_2(\mathrm{card}(\mathrm{T_Q} \cap \{0,1\}^n)).}{n}$$

(See [1] for this formulation of the box dimension in $\{0,1\}^{\mathbb{N}}$.) Now by Lemma 4.4, the expected value of $(T_Q \cap \{0,1\}^n)$ for a random closed set Q is $(4/3)^n$, which suggests that the box dimension of Q should be $\log_2(4/3)$.

Lemma 4.5
Let Q be a random closed set. Then for any $\varepsilon > 0$, there exists a $m \in \mathbb{N}$ such that, for all $n > m$, $(4/3)^n(1-\varepsilon)^n <$ card $(T_Q \cap \{0,1\}^n) < (4/3)^n(1+\varepsilon)^n$.

Proof. For each n, let $c_n(Q)$, or just c_n, denote card $(T_Q \cap \{0,1\}^n)$. We will use three applications of Chernoff's Lemma 3.9. First we show that there exists m such that for all $n > m$, $c_{6n} \geq n$. Since the tree $T_Q \cap \{0,1\}^{\leq 6n-1}$ has at least $6n$ nodes, it follows from Chernoff's Lemma that the number of branching nodes is less than n with probability $\leq 2^{-n/6}$. Thus $c_{6n} < n$ with probability $< 2^{-n/6}$. Then the probability that $c_{6n} < n$ for any $n \geq m$ is less than

$$\sum_{n=m}^{\infty} 2^{-n/6} = \frac{2^{-m/6}}{1-2^{-1/6}}.$$

This provides a computable sequence of clopen sets with measures bounded by a computable sequence with limit zero and hence a Martin-Löf test. It follows that for any random closed set Q, there exists m_0 such that $c_{6n} \geq n$ for all $n \geq m_0$. Now for $n > m_0$, there are at least $6n^2$ nodes in $T_Q \cap \{0,1\}^{\leq 12n-1} - \{0,1\}^{\leq 6n-1}$, so that again by Chernoff's Lemma, the probability that $< n^2$ of these are branching nodes is $\leq 2^{-n^2/6}$. It follows as above that there exists $m_1 > 3$ such that $c_{12n} \geq n^2$ for all $n \geq m_1$. Now suppose that $m \geq 12m_1$ and that $12n \leq m < 12(n+1) < 16n$. Then $n \geq m_1$, so that

$$c_m \geq c_{12n} \geq n^2 > (m/16)^2.$$

Again by Chernoff's Lemma, the probability that the number of branching nodes from $T_Q \cap \{0,1\}^n$ differs from $(1/3)c_n$ by $> (1/3)c_n^{-1/4}c_n$ is $< 2^{-\sqrt{c_n}/9}$. But this is exactly the probability that c_{n+1} differs from $(4/3)c_n$ by $> (1/3)c_n^{-1/4}c_n$. For $n > m_1$, we know that $c_n \geq (n/16)^2$, so that $\sqrt{c_n} \geq (n/16)$ and $c_n^{-1/4} \leq (4/\sqrt{n})$ and hence $2^{-\sqrt{c_n}/9} \leq 2^{-n/144}$. Thus the probability p_n that c_{n+1} differs from $(4/3)c_n$ by more than $(c_n/9)\sqrt{n}$ is $< 2^{-n/144}$. Then the probability that for any $n \geq m_1$, c_{n+1} differs from $(4/3)c_n$ by more than $(4/3)\sqrt{n}c_n$ is bounded by

$$\sum_{n=m}^{\infty} p_n = \sum_{n=m}^{\infty} 2^{-n/144} = \frac{2^{-m/144}}{1-2^{-144}}.$$

This again provides a Martin-Löf test which shows that for any random closed set Q, there exists m_2 so that for $n > m_2$,

$$(*) \qquad \frac{4}{3}\left(1-\frac{1}{\sqrt{n}}\right)c_n \leq c_{n+1} \leq \frac{4}{3}\left(1+\frac{1}{\sqrt{n}}\right)c_n.$$

Now given ε, choose $m \geq m_2$ so that $(1+\frac{1}{\sqrt{m}})^2 < 1+\varepsilon$ and $1-\varepsilon < (1-\frac{1}{\sqrt{m}})^2$.
Then for any k,

$$c_m\left(\frac{4}{3}\right)^{2k}(1-\epsilon)^k < c_m\left(\frac{4}{3}\right)^{2k}\left(1-\frac{1}{\sqrt{m}}\right)^{2k} < c_{m+2k} < c_m\left(\frac{4}{3}\right)^{2k}(1+\frac{1}{\sqrt{m}})^{2k} < c_m\left(\frac{4}{3}\right)^{2k}(1+\varepsilon)^k.$$

Now let k be large enough so that

$$(1-\varepsilon)^{m+k} \leq c_m \leq \left(\frac{4}{3}\right)^m (1+\varepsilon)^{m+k}.$$

Then the desired inequality

$$\left(\frac{4}{3}\right)^n (1-\varepsilon)^n < c_n < \left(\frac{4}{3}\right)^n (1+\varepsilon)^n.$$

will hold for even $n \geq m+2k$. For odd n, this inequality will hold by the inequality $(*)$ above. ■

THEOREM 4.6
For any random closed set Q, the box dimension of Q is $\log_2(4/3)$.

PROOF. Given $\varepsilon > 0$, let m be given by Lemma 4.5. Then for $n > m$, we have

$$n\log_2\frac{4}{3} + n\log_2(1-\varepsilon) \leq \log_2(\mathrm{card}(T_Q \cap \{0,1\}^n) \leq n\log_2\frac{4}{3} + n\log_2(1+\varepsilon),$$

so that

$$\log_2\frac{4}{3} + \log(1-\varepsilon) \leq \frac{\log_2(\mathrm{card}(T_Q \cap \{0,1\}^n)}{n} \leq \log_2\frac{4}{3} + \log_2(1+\varepsilon),$$

and therefore $\dim_B(Q) = \lim_n(\log_2(\mathrm{card}(\mathrm{T_Q} \cap \{0,1\}^n)))/n = \log_2(4/3)$. ■

5 Prefix-free complexity of closed sets

In this section, we consider randomness for closed sets in terms of incompressibility of trees. Of course, Schnorr's theorem tells us that P is random if and only if the code $x_P \in \{0,1,2\}^{\mathbb{N}}$ for P is prefix-free random, that is, $K_3(x_P\lceil n) \geq n - O(1)$. (Schnorr's theorem for arbitrary finite alphabets is shown in [6].) Here we write K_3 to indicate that we would be using a universal prefix-free function $U: \{0,1,2\}^* \to \{0,1,2\}^*$. However, many properties of trees and closed sets depend on the levels $T_n = T \cap \{0,1\}^n$ of the tree. For example, if $[T_n] = \cup\{I(\sigma) : \sigma \in T_n\}$, then $[T] = \bigcap_n [T_n]$ and $\mu([T]) = \lim_{n\to\infty} \mu([T_n])$.

So we want to consider the compressibility of a tree in terms of $K(T_n)$. Now there is a natural representation of T_n as a string of length 2^n. That is, list $\{0,1\}^n$ in lexicographic order as $\sigma_1, \ldots, \sigma_{2^n}$ and represent T_n by the string $e_1, \ldots, e_{2^n}$ where $e_i = 1$ if $\sigma_i \in T$ and $e_i = 0$ otherwise. Henceforth we identify T_n with this natural representation.

It is interesting to note that the code for T_n will have a shorter length than the natural representation. For example, if $[T] = \{y\}$ is a singleton, then $x = y$ and for each n, the code for T_n is $x\lceil n$. If x is the code for the full tree $\{0,1\}^*$, then $x = (2,2,\ldots)$ and the code for T_n is a string of $(2^n - 1)$ 2's, those labels attached to nodes of length $< n$. For the remainder of this section, we will use T_n to mean the natural representation and x_n to mean the code.

One question here is whether there is a formulation of randomness in terms of the incompressibility of T_n. We will give some partial answers. It seems plausible that $P=[T]$ is random if and only if there is a constant c such that $K(T_n) \geq 2^n - c$ for all n. We will see that this is not possible for any tree. First we give a lower bound for the prefix-free complexity of a random tree.

THEOREM 5.1
If P is a random closed set and $T = T_P$, then there is a constant c such that $K(T_n) \geq \left(\frac{7}{6}\right)^n - c$ for all n.

PROOF. Let $P = [T]$ be a random closed set. Let m be given by Lemma 4.5, for $\varepsilon = (7/6)$, so that for $n > m$,

$$\mathrm{card}(\mathrm{T}_n) \geq \left(\frac{7}{6}\right)^n.$$

It follows that the code x_n for T_n has length $\geq (7/6)^n$. Since x is random, we know that, for $n \geq m$,

$$K_3(x_n) \geq \left(\frac{7}{6}\right)^n - a,$$

for some constant a. Now we can compute x_n from T_n, so that

$$K(T_n) \geq K_3(x_n) - b,$$

for some constant b. The result now follows.

That is, let U (mapping $\{0,1\}^*$ to $\{0,1\}^*$) be a universal prefix-free Turing machine and let $K(T_n) = \min\{|\sigma| : U(\sigma) = T_n\}$. Let M be a prefix-free machine M (mapping $\{0,1\}^*$ to $\{0,1,2\}^*$) such that $M(T_n) = x_n$. Then define V by

$$V(\sigma) = M(U(\sigma)).$$

Then $K_V(x\lceil n) \leq K(T_n)$, so that for some constant e, $K_3(x_n) \leq K(T_n) + e$ and hence

$$K(T_n) \geq K_3(x_n) - e \geq \left(\frac{7}{6}\right)^n - b - e. \qquad \blacksquare$$

Going in the other direction, we can compute T_n uniformly from $x\lceil 2^n$, so that as above, $K_3(x\lceil 2^n) \geq K(T_n) - b$ for some b. Thus in order to conclude that P is random, we would need to know that $K(T_n) \geq 2^n - c$ for some c. The next result shows that this is not possible, since trees are naturally compressible.

THEOREM 5.2
For any tree $T \subseteq \{0,1\}^*$, there are constants $k > 0$ and c such that $K(T_\ell) \leq 2^\ell - 2^{\ell-k} + c$ for all ℓ.

Proof. For the full tree $\{0,1\}^*$, this is clear so suppose that $\sigma \notin T$ for some $\sigma \in \{0,1\}^m$. Then for any level $\ell > m$, there are $2^{\ell-m}$ possible nodes for T which extend σ and T_ℓ may be uniformly computed from σ and from the characteristic function of T_ℓ restricted to the remaining set of nodes. That is, fix σ of length m and define a prefix-free computer M as follows. The domain of M is strings of the form $0^\ell 1\tau$ where $|\tau| = 2^\ell - 2^{\ell-m}$. M outputs the standard representation of a tree T_ℓ such that no extension of σ is in T_ℓ and such that τ tells us whether strings not extending σ are in T_ℓ. It is clear that M is prefix-free and we have $K_M(T_\ell) = \ell + 1 + 2^\ell - 2^{\ell-m}$. Thus $K(T_\ell) \leq \ell + 1 + 2^\ell - 2^{\ell-m} + c$ for some constant c. Now $\ell + 1 < 2^{\ell-m-1}$ for sufficiently large ℓ and thus by adjusting the constant c, we can obtain c' so that

$$K(T_\ell) \leq 2^\ell - 2^{\ell-m-1} + c'. \qquad \blacksquare$$

We might next conjecture that $K(T_\ell) > 2^{\ell-c}$ is the right notion of prefix-free randomness. However, classes with small measure are more compressible.

Theorem 5.3
If $\mu([T]) < 2^{-k}$, then there exists c such that, for all ℓ,

$$K(T_\ell) \leq 2^{\ell-k+1} + c.$$

Proof. Suppose that $\mu([T]) < 2^{-k}$. Then for some level n, T_n has $< 2^{n-k}$ nodes $\sigma_1, \ldots, \sigma_t$. Now for any $\ell > n$, T_ℓ can be computed from the fixed list $\sigma_1, \ldots, \sigma_t$ and the list of nodes of T_ℓ taken from the at most $2^{\ell-k}$ extensions of $\sigma_1, \ldots, \sigma_t$. It follows as in the proof of Theorem 5.2 that for some constant c and all ℓ,

$$K(T_\ell) \leq 2^{\ell-k} + \ell + 1 + c.$$

Thus for large enough so that $\ell + 1 \leq 2^{\ell-k}$, we have

$$K(T_\ell) \leq 2^{\ell-k+1} + c,$$

as desired. $\blacksquare$

Note that if $\mu([T]) = 0$, then for any k, there is a constant c such that $K(T_\ell) \leq 2^{\ell-k} + c$. But by Theorem 4, random closed sets have measure zero. Thus if P is random, then it is not the case that $K(T_n) \geq 2^{n-k}$.

Finally, we will construct an effectively closed set with not too much compressibility. The standard example of a random real, Chaitin's Ω [9], is a c.e. real and therefore Δ^0_2. Thus there exists a Δ^0_2 random tree T and by Theorem 5.1, $K(T_\ell) \geq \left(\frac{7}{6}\right)^n - c$ for some c. We have a more modest result for Π^0_1 classes.

Theorem 5.4
There is a Π^0_1 class $P = [T]$ such that $K(T_n) \geq n$ for all n.

Proof. Recall the universal prefix-free machine U and let $S = \{\sigma \in \mathrm{Dom}(U) : |U(\sigma)| \geq 2^{|\sigma|}\}$. Then S is a c.e. set and can be enumerated as $\sigma_1, \sigma_2, \ldots$. The tree $T = \bigcap_s T^s$ where T^s is defined at stage s. Initially we have $T^0 = \{0,1\}^*$. We say that σ_t *requires attention* at stage $s \geq t$ when $\tau = U(\sigma_t) = T^s_n$ for some n (so that $|\tau| = 2^n$) and $n \geq |\sigma_t|$. Action is taken by selecting some path $\rho_t \in T_s$ of length n and defining T^{s+1} to contain all nodes of T^s which do not extend ρ_t. Then $\tau \neq T^{s+1}_n$ and furthermore $\tau \neq T^r_n$ for any $r \geq s+1$ since future action will only remove more nodes from T_n.

At stage $s+1$, look for the least $t \leq s+1$ such that σ_t requires action and take the action described if there is such a t. Otherwise, let $T^{s+1} = T^s$.

Let A be the set of t such that action is ever taken on σ_t. Recall from the Kraft Inequality that $\sum_t 2^{-|\sigma_t|} < 1$. Since $|\rho_t| \geq |\sigma_t|$, it follows that $\sum_{t \in A} 2^{-|\rho_t|} < 1$ as well. Now $\mu([T]) = 1 - \sum_t 2^{-|\rho_t|} > 0$ and therefore $[T]$ is non-empty.

It follows from the construction that for each t, action is taken for σ_t at most once.

Now suppose by way of contradiction that $U(\sigma) = T_n$ for some σ_t with $|\sigma| \leq n$. There must be some stage $r \geq t$ such that for all $s \geq r$, $T^s_n = T_n$ and such that action is never taken on any $t' < t$ after stage r. Then σ_t will require action at stage $r+1$ which makes $T^{r+1}_n \neq T^r_n$, a contradiction.

6 Conclusions and future research

In this article we have proposed a notion of randomness for closed sets and derived several interesting properties of random closed sets. Random strong Π^0_2 classes exist but no Π^0_1 class is random. A random closed set has measure zero and box dimension $\log_2 \frac{4}{3}$; it is perfect and hence uncountable. Results on members of random closed sets include the following. A random closed set contains no f-c.e. elements, if f is polynomially bounded. Every random closed set Q contains a random real, not every element of a random closed set is random and every random real belongs to some random closed set. On the other hand we do not know the answer to the following.

Problem 6.1
Does every random closed set with Δ^0_2 canonical code contain a low random element?

We conjecture a negative answer. It is a well-known fact that every real is computed by a random real. The corresponding question for trees is as follows.

Problem 6.2
Let A by an incomputable set. Is there a random closed set such that all of its elements compute A?

We have examined the notion of compressibility for trees based on the prefix-free complexity of the nth level T_n of a tree. We showed that for any random closed set (and hence for some strong Π^0_2 class), there exists c such that $K(T_n) \geq (7/6)^n - c$ for all n. We constructed a Π^0_1 class $P = [T]$ such that $K(T_n) \geq n$ for all n. It seems a reasonable conjecture that if $K(T_n) \geq (4/3)^n - c$ for all n, then the closed set $[T]$ is random. We would like to explore the notion that Π^0_1 classes are more compressible than arbitrary closed sets.

Other notions of randomness might also be considered. A general probability measure ν_f may be defined on $3^{\mathbb{N}}$ from a function $f\colon \{0,1,2\}^* \to [0,1]$ such that $\sum_{i=0,1,2} f(\sigma^\frown i) = 1$ for all σ. The interval $I(\sigma)$ then has ν_f-measure $\prod_{n<|\sigma|} f(\sigma\lceil(n+1))$. We will say that ν_f is a *computable* measure if f is computable. The probability measure ν is *non-atomic* if for any $x \in 3^{\mathbb{N}}$, $\nu(\{x\}) = 0$. The function f (and the corresponding measure ν_f) is *bounded* if there is an upper bound $b < 1$ such that $f(\sigma) < b$ for all $\sigma \in \{0,1,2\}^*$. It is easy to see that any bounded measure is non-atomic. If there exist constants b_0, b_1, b_2 strictly between 0 and 1, such that for all σ, $f(\sigma^\frown i) = b_i$, then we will say that ν_f is *regular*. For any regular measure, we can define the notion of a ν-Martin-Löf test and the resulting notion of a ν-Martin-Löf-random (or just ν-random) real. It is easy to see that ν-random reals exist for any ν and hence ν-random closed sets exist. The results on ghost codes and joins will hold for any regular measure. The corresponding version of Lemma 2.2 will hold if ν is regular with b_0 and $b_1 \le (1/2)$. The proofs of Theorem 4.2 and Corollary 4.3, that no subset of a measure-zero Π^0_1 class is random, also go through under this assumption.

Some of the results in this article may also be obtained for ν_f where $f(\sigma^\frown i) \le (1/2)$ for $i=0,1$. For example with respect to ν_f a random closed set will have no isolated elements and it will always contain a random element. For any regular measure, either the leftmost or the rightmost path will be non-random, since either $b_0 + b_2 > (1/2)$ or $b_1 + b_2 > (1/2)$. The proof of Theorem 3.6 that every random closed set has measure 0 seems to require, for ν_f-randomness, that $f(\sigma^\frown 2) \le (1/2)$ for all σ.

Returning to the notion of randomness which allows trees with dead ends, let b_3 now be the probability that a given node has no extensions and let the probability be regular as above. Then a simple recursion shows the probability p of a given closed set being empty satisfies the equation

$$p = b_3 + (b_0 + b_1)p + b_2 p^2.$$

Solving for p, we obtain

$$(p-1)(b_2 p - b_3) = 0.$$

Thus either $p=1$ or $p=(b_3/b_2)$. It follows that if $b_2 \le b_3$, then $p=1$, that is, almost every closed set is empty. Suppose now that $b_3 < b_2$ and let p_n be the probability that a given tree T has no paths of length n. Then it can be seen by induction that $p_n \le (b_3/b_2)$ for all n. That is, $p_1 = b_3 \le (b_3/b_2)$ and then

$$p_{n+1} = b_3 + (1 - b_2 - b_3)p_n + b_2 p_n^2 \le \frac{b_3}{b_2}.$$

Hence in this case, the probability that a given closed set is empty is $b_3/b_2 < 1$. In this case, one could presumably develop a notion of a random tree and a random closed set and explore the properties of random closed sets.

A real x is said to be *K-trivial* if $K(x\lceil n) \le K(n) + c$ for some c. Much interesting work has been done on the K-trivial reals. Chaitin showed that if A is K-trivial, then $A \le_T \mathbf{0}'$. Solovay constructed a non-computable K-trivial real. Downey *et al.* [12] showed that no K-trivial real is c.e. complete. The notion of a K-trivial closed set was introduced in [4]. It was shown in particular that every K-trivial class contains a K-trivial member, but there exist K-trivial Π^0_1 classes with no computable members.

The related notion of a random continuous function was introduced in [3]. It was shown that a random continuous function F on $2^{\mathbb{N}}$ cannot be computable, so that the graph of F cannot be Π^0_1 class. For any random F and computable x, $F(x)$ is a random real, however the image of F need not be a random closed set. The authors can now show that the set of zeroes of a random continuous function is a random closed set. Random Brownian motions have been studied by Fouche [15] and are a special case of random continuous functions on the real line, which is another area of interest for further research.

Acknowledgements

The authors wish to thank Jack Lutz and Joe Miller for helpful discussions, and the referee for comments that greatly improved the paper. Much of the contents of this article was discussed during the AIM workshop on Effective Randomness in August, 2006. Research partially supported by National Science Foundation grants DMS 0532644, 0554841 and 0652732.

References

[1] K. B. Athreya, J. M. Hitchcock, J. Lutz, and E. Mayordomo. Effective strong dimension, algorithmic information and computational complexity. *SIAM Journal of Computing* (in press).

[2] P. Brodhead, D. Cenzer, and S. Dashti. Random closed sets. In *Logical Approaches to Computational Barriers, Proc. CIE 2006*, A. Beckmann, U. Berger, B. Löwe, and J. Tucker (eds), pp. 55–64. Vol. 3988 of *Springer Lecture Notes in Computer Science*, Springer-Verlag, Berlin, 2006.

[3] P. Brodhead, D. Cenzer, and J. B. Remmel. Random continuous functions. In *Proceedings CCA 2006*, pp. 275–287. Vol. 167 of *Springer Electronic Notes in Computer Science*, 2007.

[4] G. Barmpalias, D. Cenzer, J. B. Remmel and R. Weber. K-trivial closed sets and continuous functions. In *Proceedings of CIE 2007*, S. B. Cooper, B. Löwe and A. Sorbi (eds), pp. 135–145. Vol. 4497 of *Springer Lecture Notes in Computer Science*, Springer-Verlag, Berlin, 2007.

[5] C. Calude. *Theories of computational complexity*. Vol. 35 of *Annals of Discrete Mathematics*, Elsevier, North-Holland, Amsterdam, 1988.

[6] C. Calude. *Information and Randomness: An Algorithmic Perspective*, Springer-Verlag, Berlin, 1994.

[7] D. Cenzer and J. B. Remmel. Π^0_1 classes. *ASL Lecture Notes in Logic*, in press.

[8] D. Cenzer and J. B. Remmel. Π^0_1 classes. *In Handbook of Recursive Mathematics,* Vol. **2**: *Recursive Algebra, Analysis and Combinatorics*, Y. Ersov, S. Goncharov, V. Marek, A. Nerode, J. Remmel (eds), pp. 623–821. Vol. 139 of *Elsevier Studies in Logic and the Foundations of Mathematics*, 1998.

[9] G. Chaitin. Information-theoretical characterizations of recursive infinite strings. *Theoretical Computer Science*, **2**, 45–48, 1976.

[10] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sums of observations. *Annals of Mathematical Statistics*, **23**, 493–509, 1952.

[11] R. Downey and D. Hirschfeldt. *Algorithmic Randomness and Complexity*, in preparation. Current draft available online at http://www.mcs.vuw.ac.nz/~downey/

[12] R. Downey, D. Hirschfeldt, A. Nies, and F. Stephan. Trivial reals. In *Proceedings of the 7th and 8th Asian Logic Conferences*, pp. 101–131. World Scientific Press, Singapore, 2003.

[13] R. Downey and L. Yu. Arithmetical sacks forcing. *Archive for Mathematical Logic*, **45**, 715–720, 2006.

[14] Y. Ershov. A hierarchy of sets, Part I. *Algebra and Logic*, **7**, 24–43, 1968.

[15] W. Fouche. Arithmetical representations of Brownian motion. *Journal of Symbolic Logic*, **65**, 421–442, 2000.

[16] P. Gács. On the symmetry of algorithmic information. *Soviet Mathematics* (*Doklady*), **15**, 1477–1480, 1974.

[17] C. Haught. The degrees below a 1-generic degree below $\mathbf{0}'$. *Journal of Symbolic Logic*, **51**, 770–777, 1996.

[18] A. N. Kolmogorov. Three approaches to the quantitative definition of information. In *Problems of Information Transmission*, **1**, 1–7, 1965.

[19] A. Kučera. Measure, Π^0_1 classes and complete extensions of *PA*. In *Recursion Theory Week (Oberwolfach, 1984)*, pp. 245–259. Vol. 1141 of *Lecture Notes in Mathematics*, Springer, Berlin, 1985.

[20] A. Kučera. On the use of diagonally nonrecursive functions. In *Logic Colloquium '87*, Ebbinghaus *et al*. (eds), pp. 219–239. North Holland, 1989.

[21] L. Levin. On the notion of a random sequence. *Soviet Mathematics* (*Doklady*), **14**, 1413–1416, 1973.

[22] P. Lévy. *Théorie de l'Addition des Variables Aleatoires*, Gauthier-Villars, Paris, 1937 (2nd edn, 1954).

[23] M. Li and P. Vitanyi. *An introduction to Kolmogorov Complexity and Its Applications*, 2nd edn. Springer, Berlin, 1997.

[24] P. Martin-Löf. The definition of random sequences. *Information and Control*, **9**, 602–619, 1966.

[25] A. A. Muchnik, A. L. Semenov, and V. A. Uspensky. Mathematical metaphysics of randomness. *Theoretical Computer Science*, **207**, 263–271, 1998.

[26] A. Nies. *Computability and Randomness*, in preparation. Current draft available at http://www.cs.auckland.ac.nz/˜nies

[27] C. P. Schnorr. A unified approach to the definition of random sequences. *Mathematical Systems Theory*, **5**, 246–258, 1971.

[28] R. Soare. *Recursively Enumerable Sets and Degrees*. Springer-Verlag, Berlin, 1987.

[29] M. van Lambalgen. *Random Sequences*, PhD Dissertation, University of Amsterdam, The Netherlands, 1987.

[30] J. Ville. *Étude Critique de la Notion de Collectif*. Gauthier-Villars, Paris, 1939.

[31] R. von Mises, Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, **5**, 52–99, 1919.

Received 20 October 2006