Martin-Löf randomness

March 1, 2023

1 Randomness of individual infinite sequences

One of the surprising consequences of the theory of random finite strings is that it leads to a consistent theory of the randomness of *individual* infinite sequences. For example, it is possible to say that a sequence of zeroes is non-random, and the Chaitin's Omega which we defined in Homework 1 is random.

This definition is moreover *mathematically robust* - multiple, very different approaches, agree on which sequences are random and which are not. We will see two approaches in this class. The various approaches which lead to equivalent definitions are:

1. Computational: we can define a sequence to be random if, all its long enough prefixes are Kolmogorov incompressible. This is the *incompressibility* definition.
2. Measure-theoretic: Random sequences are those which "pass" every effective probability 1 property. This was the definition which was originally studied by Martin-Löf.
3. Unpredictability: Given some prefix of the sequence, no algorithm should be able to predict the next bit of the sequence.

In this course, we will study the computational and the unpredictability approaches, and show their equivalence.

As a historical note: the notion of a martingale is a very important modern tool in mathematical finance and stochastic processes. It was in the context of algorithmic randomness that Ville first defined the modern notion of martingales in 1939. Later, the American mathematician Doob developed the modern theory of stochastic processes based on martingales.

In this chapter, we establish that the random sequences obey the usual "randomness notions". We include van Lambalgen's theorem which establishes the symmetry of relative randomness of one sequence with respect to another (this is an effective version of Fubini's theorem in analysis). In the end, we will discuss a very counterintuitive result which emerges when we study which sequences

can be computed from a random sequence. (Intuitively, what patterns may emerge when we look at a random sequence?)

2 The incompressibility approach

This definition of a random sequence is immediately accessible based on what we have so far used for finite strings - basically, the definition says that an infinite sequence is random if all its prefixes are incompressible. We do allow for a slight relaxation: we allow finitely many prefixes to be compressible. But after a point, all further prefixes must be incompressible.

Notation: Throughout this chapter, we will use capital letters like X and Y to denote infinite binary sequences. If m is a positive integer, then X_m denotes the m^{th} bit of X. If m and n are natural numbers, then $X[m:n]$ will denote the finite substring $X_m \dots X_n$ if $m \leq n$ and the empty string if $m < n$. Even though it is possible to use $X[1:n]$ to denote the n-length prefix of X, we adopt a shorter notation for this particular case - $X \upharpoonright n$.

We would like to say that if all prefixes of an infinite sequence are incompressible, then the sequence is random. There are two things to note.

1. Suppose we start by saying that an infinite sequence X is random if $C(X \upharpoonright 1)$, $C(X \upharpoonright 2)$, ... are the maximum possible values for those lengths. However, Martin-Löf in 1969 showed that *no* sequence satisfies this - every sequence has infinitely many compressible prefixes, if we use C. So this approach fails.

2. We now show that K works for the above idea. However, we will relax the condition slightly - finitely many prefixes of a random sequence X may be compressible. However, for all sufficiently large n, $K(X \upharpoonright n)$ must be maximal.

Definition 2.1 (Levin, Chaitin)**.** An infinite binary sequence A is *random* if for all n, $K(A \upharpoonright n) \geq n - O(1)$.

Note that, because of the constant, the above definition is equivalent to saying that for all sufficiently large n, $K(A \upharpoonright n) \geq n - O(1)$.

What we need to show is that with probability 1, a binary sequence selected at random has this property. But before we do this, we show one particular example of a random sequence. Let

$$\omega = \sum_{\substack{i \in \mathbb{N} \\ M_i(i)\downarrow}} \frac{1}{2^i}. \tag{1}$$

Theorem 2.2 (Chaitin)**.** *ω defined in (1) is random.*

Proof. We know that ω is lower semicomputable, and that it is irrational. Let $f : \mathbb{N} \to \mathbb{Q}$ be a total computable function such that for all m, $f(m) < f(m+1) < \omega$ and $\lim_{m\to\infty} f(m) = \omega$.

We define a program P that given this m, prints the first string s whose complexity is greater than n, *i.e.* $K(s) > n$. Since $f(m)$ is computable from m, then we have $n < K(s) \leq K(f(m)) + O(1)$.

How do we find this first string? Consider the set of prefix-free programs $\{p_k \mid k \leq m\}$. This is clearly computable given m.

Now the trick: Recall that given any approximation of ω to within 2^{-n}, it is possible to decide whether $p_k(\lambda)$ halts, $1 \leq k \leq n$. Equivalently, we can determine whether $U(p_k, \lambda)$ is defined, $1 \leq k \leq n$.

Now, consider $S_m = \{U(p_k, \lambda) \mid k \leq m, |p_k| \leq n\}$. This set contains all strings x such that $K(x) \leq n$. This set is decidable, given m (since from m, we can compute an approximation of ω to within 2^{-n}.)

Hence, there is a program P such that given m, it outputs the first string s in the standard enumeration which is outside S_m. By the definition of S_m, we have $K(s) > n$.

Since s is computable from $f(m)$, we have $K(s) \leq K(f(m)) + O(1)$. Hence we have $n < K(f(m)) + O(1)$. Since the first n bits of $f(m)$ coincide with the first n bits of ω, we have $n < K(\omega \upharpoonright n) + O(1)$.

This completes the proof. □

Chaitin's omega is a random that is somewhat strange. Intuitively, a random does not have any useful, extractable information. But Chaitin's omega contains information about the Halting problem in a way that can be easily extracted. Later, we will deal with the Kučera-Gács theorem, which also talks about what information can be embedded into a random.

These, however are not the norm: a random, intuitively, should not contain easily extractable information, just as, looking at grainy black and white spots on a television should not suggest any cogent image. There are more advanced notions of randomness than what we cover in this course, where information cannot usefully be embedded, appealing more to our intuition of what randomness looks like.

We have thus shown that there is a random. What we have not shown is that the set of randoms has probability 1.

Going from C to K is what made ω well-defined. K also has the following nice property - we need not check for the incompressibility of *all* prefixes. It is sufficient to check an infinite, computable set of prefix lengths.

Lemma 2.3 (Fortnow, and Nies, Stephan, Terwijn)**.** *If there is an infinite computable set $S \subseteq \mathbb{N}$ such that for every $n \in S$, $K(X \upharpoonright n) \geq n + O(1)$, then X is random.*

Proof. Suppose X is not random. Consider a computable set $M = \{m_0 < m_1 < \dots\}$. Fix $c \in \mathbb{N}$.

We need to show a prefix $X \upharpoonright m_i$, $m_i \in M$, which is compressible.

Since X is not random, there is a length n such that $K(X \upharpoonright n) \leq n - c - d$, where d is the constant in the invariance theorem.

The issue is that n itself need not be in M.

But, consider $K(X \upharpoonright m_n)$. Without loss of generality, let $m_n > n$. Then $X \upharpoonright m_n = (X \upharpoonright n)z$ for some string z. If we have a prefix code for $X \upharpoonright n$, then its concatenation with z will form a prefix-free code for $X \upharpoonright m_n$. [1]

Hence

$$\begin{aligned} K(X \upharpoonright m_n) &\leq K(X \upharpoonright n) + |z| + d \\ &\leq (n - c - d) + (m_n - n) + d \\ &= m_n - c, \end{aligned}$$

as required.

□

The next result characterizes random sequences using a criterion related to the sum of exponentials of randomness deficiencies.

Theorem 2.4. *X is Martin-Löf random if and only if*

$$\sum_{n=1}^{\infty} 2^{n - K(X[0 \ldots n-1])} < \infty. \tag{2}$$

Proof. If X is non-random, then there are infinitely many n such that $K(X[0 \ldots n-1]) < n$, hence $n - K(X[0 \ldots n-1]) > 0$, implying

$$\sum_{n=1}^{\infty} 2^{n - K(X[0 \ldots n-1])} = \infty.$$

Conversely, suppose that for some X, we have

$$\sum_{n=1}^{\infty} 2^{n - K(X[0 \ldots n-1])} = \infty.$$

We form a Martin-Löf test containing X. For any $c \in \mathbb{N}$, define

$$U_c = \left\{ A \in \Sigma^{\infty} \mid \sum_{n=1}^{\infty} 2^{n - K(A[0 \ldots n-1])} \geq 2^c \right\}.$$

[1] This follows since z has length equal to $m_n - n$, and there is a unique string of that length which forms the suffix of X.

Clearly, $X \in \cap_{c\in\mathbb{N}} U_c$, and U_cs are uniformly computably enumerable in c. It suffices to show that $\mu(U_c) < 2^{-c}$.

Let $m \in \mathbb{N}$.

$$\sum_{w\in\Sigma^m} \sum_{n=1}^{m} 2^{n-K(X[0\ldots n-1])} = \sum_{w\in\Sigma^m} \sum_{v\sqsubseteq w} 2^{|v|-K(v)}.$$

We see that[2] a string of length $m-k$ will appear 2^k times in the sum (since that string will have 2^k extensions of length m). Hence, we the above sum is equal to

$$\begin{aligned}\sum_{v\in\Sigma^{\leq m}} 2^{m-|v|} \times -^{|v|-K(v)} &= \sum_{v\in\Sigma^{\leq m}} 2^m 2^{-K(v)} \\ &= 2^m \sum_{v\in\Sigma^{\leq m}} 2^{-K(v)} \\ &< 2^m,\end{aligned}$$

by Kraft's inequality. □

3 The unpredictability approach

The following theorem says that a martingale cannot succeed very much with very high probability. This is a version of Markov inequality for martingales.

Theorem 3.1 (Kolmogorov inequality, Ville)**.** *Let d be a martingale.*

1. *For any string w and any prefix-free set S of extensions of w, we have*

$$\sum_{v\in S} \frac{d(v)}{2^{|v|}} \leq \frac{d(w)}{2^{|w|}}.$$

2. $$P\left(\{w \ : \ d(w) \geq k\}\right) \leq \frac{d(\lambda)}{k}.$$

Proof. (1) Suppose S is finite. We prove this by induction on the number of elements in S. Suppose $n = 1$. Then $d(v)2^{-|v|} \leq d(w)2^{-|w|}$ follows from the definition of the martingale.

Now suppose the inequality holds for every set with at most n elements. Now, consider a set S with $n+1$ elements. Let y be the longest string such that every element of S extends y (y may be distinct from w).

[2]Thanks to Aditi Goyal for pointing out an error in the earlier presentation.

Then the extensions of $y0$ and $y1$ in S have both fewer than n elements. Let the set of extensions of $y0$ be S_0 and $y1$ be S_1. By the inductive hypothesis, for $i \in \{0,1\}$, we have

$$\sum_{z \in S_i} \frac{d(z)}{2^{|z|}} = \frac{d(yi)}{2^{|yi|}}.$$

Hence,

$$\sum_{y \in S} \frac{d(y)}{2^{|y|}} \leq \frac{d(y0) + d(y1)}{2^{|y|+1}} \leq \frac{d(y)}{2^{|y|}} \leq \frac{d(w)}{2^{|w|}}.$$

(2) Let $P \leq R_k$ be a prefix-free such that the set of extensions of P be R_k. Then by part (1), with $w = \lambda$,

$$P([[P]]) = \sum_{w \in P} \frac{1}{2^{|w|}} \leq \sum_{w \in P} \frac{d(w)}{k} \frac{1}{2^{|w|}} \leq \frac{d(\lambda)}{k}.$$

□

This implies that the success set of a martingale has probability 0.

Theorem 3.2. *Let $d : \Sigma^* \to [0, \infty)$ be a martingale. Then $P(S^\infty[d]) = 0$.*

Proof. We know that

$$S^\infty[d] = \bigcap_N \bigcup_n \{X \in \Sigma^\infty \ : \ d(X[0 \ldots n-1]) > N\}.$$

By the Kolmogorov inequality,

$$P\left[\bigcup_n \{X \in \Sigma^\infty \ : \ d(X[0 \ldots n-1]) > N\}\right] \leq \frac{1}{N}.$$

Hence,

$$P\left[\bigcap_N \bigcup_n \{X \in \Sigma^\infty \ : \ d(X[0 \ldots n-1]) > N\}\right] \leq \lim_{N \to \infty} \frac{1}{N} = 0.$$

□

Definition 3.3. The set of sequences $X \in \Sigma^\infty$ on which the universal martingale fails is called *Martin-Löf random*.

Since every martingale succeeds only on a probability 0 set, the set of Martin-Löf randoms has probability 1.

4 Equivalence between the two approaches

5 Van Lambalgen's theorem

6 the Kučera-Gács Theorem

6.1 Languages and infinite binary sequences

Every language L is a subset of Σ^*. An equivalent representation of languages is using the "characteristic sequence", which can be seen as a "bit-map" representation of the language.

Definition 6.1 (Characteristic Sequence)**.** The *characteristic sequence* of a language $L \subseteq \Sigma^*$ is the infinite binary sequence χ_L where for each $i \in \mathbb{N}$, $\chi_L = 1$ if $s_i \in L$ and $\chi_L = 0$ otherwise.

What is the advantage of this representation? This helps us to use the tools in the theory of languages, e.g. reducibility, to the study of infinite binary sequences. This is what we will do now. (It also allows the use of tools related to infinite binary sequences, e.g. *real analysis* to study languages, but this is beyond the scope of this course.)

6.2 The Kučera-Gács theorem

The following is a deep and unexpected connection between Martin-Löf randomness and Turing reducibility. We show a result that is counterintuitive at first: every infinite sequence can be computed from some random sequence. In order to make it precise, we need to define what is means when we say "computed from". This is done using the notion of reducibility between languages, equivalently, between infinite binary sequences.

6.3 Reducibility between languages

In computability theory, we usually use the notion of reducibility in the following sense: Suppose we have to show that a language A is undecidable. To do this, we find a B which we already know to be undecidable. Then we show that B is Turing-reducible to A, denoted $B \leq_T A$. If B is Turing reducible to A and B is undecidable, then A is undecidable as well. We recall the definition of Turing-reducibility.

(The notation $\leq_T$ is a mnemonic to remember the direction of the implication: B is less hard than A. So if B is undecidable, then A is undecidable as well.)

Definition 6.2 (Turing reducibility)**.** A language B is said to be *Turing reducible* to A, denoted $B \leq_T A$, if there is an oracle Turing machine M such that $B = M^A$.

The operation of the oracle Turing machine can be visualized as follows. Imagine that the machine M in addition to its work-tapes, is provided with an infinite *read-only* tape which is provided with the characteristic sequence of A. This is given for free, and M can query bits of χ_A while computing membership of x in B. These answers may be used to determine the final answer for $x \stackrel{?}{\in} B$.

In this section, we use another notion of reducibility. Instead of Turing reducibility, we use a weaker form of reducibility, called *weak truth-table reducibility*.

Note: When the reducibility becomes weaker (less general), then the relationship between the languages becomes *stronger*.

Definition 6.3 (weak truth-table reducibility)**.** A language B is *weak truth-table reducible* to A, denoted $B \leq_{wtt} A$, if there is a total computable function $f : \Sigma^* \to \Sigma^*$ and an oracle Turing machine M such that $B = M^A$ and for every x, the use $\phi(x) \leq f(x)$.

To understand *wtt*-reducibility, contrast with Turing reducibility: In Turing reducibility, on input x, at some point M has to stop asking queries to the oracle, and do the final computation and halt. But in general, it may be impossible to predict or precompute how many queries M makes to A for a given input x.

In contrast, in *wtt*-reducibility, we can precompute the bound on index of the furthest query to A that M makes on x: it will be at most $f(x)$.

6.4 The Theorem

The presentation follows Section 8.3 from Downey and Hirschfeldt. We first show a technical lemma.

Lemma 6.4 (Space Lemma, see Merkle and Mihalović)**.** *Given a rational q and a positive integer k, we can compute a length $\ell(q,k)$ such that for any martingale d and any $\sigma \in \Sigma^*$,*

$$|\{\tau \in \Sigma^{2^{\ell(q,k)}} \ : \ d(\sigma\tau) \leq qd(\sigma)\}| \geq k.$$

This is not surprising, since it is a consequence of Kolmogorov's inequality saying that all martingales will fail to make a lot of money on most extensions of σ. So there are many extensions where the martingale's success is limited. We now prove the lemma.

Proof. IOU. □

We now want to code a given language L into a Martin-Löf random R. Of course, we will not be able to do too much: it is impossible to, for example, encode a decidable language L into the even bits of R - if we do so, we can simply bet on the even bits of R and succeed on R. This would

contradict the assumption that R is Martin-Löf random. However, we can use the space lemma to encode L into R in a *wtt*-reducible manner.

What we accomplish is the following: for each bit in L we have a rough idea of the *stretch* in R to embed it. But we cannot precompute the precise position, since that would contradict the randomness of R.

Theorem 6.5 (Kučera-Gács Theorem)**.** *For every $L \subseteq \Sigma^*$, there is a Martin-Löf random R such that $\chi_L \leq_T R$.*

Proof. Suppose d is the universal c.e. martinagle. We assume that we have done the savings trick, so that if $\liminf_{n\to\infty} d(R[0\ldots n-1]) < \infty$, then R is non-Martin-Löf random.

Now we prepare the stretches to embed χ_L: Let $q_0 > q_1 > \ldots$ be a collection of positive rationals such that their partial product sequence $\prod_{i=1}^{n} q_i$ converges.

Let $\ell(q_s, 2)$ be as in the space lemma (Lemma **??**) so that for any σ, there are at least two extensions τ at length $|\sigma| + \ell(q_s, 2)$ where $d(\sigma\tau) \leq q_s d(\sigma)$.

Partition $\mathbb{N}$ into consecutive intervals $\{I_s\}_{s=1}^{\infty}$ such that $|I_s| = \ell(q_s, 2)$.

Now we embed χ_L into a specific R which is Martin-Löf random. This R will *wtt*-compute χ_L. At stage s, we specify the bits of R in the interval I_s. Assume that the prefix of R we determine before stage s by σ_{s-1}. If $s > 0$, then assume by induction that $d(\sigma_{s-1}) \leq \prod_{i=1}^{s-1} q_s$.

□

7 Ville's Theorem

Recall from our discussion of the history of the subject that von Mises wanted to define a *Kollektiv* as a sequence whose asymptotic frequency of 1s is 1/2. Moreover, the asymptotic frequency of 1s in any *admissible* subsequence is also 1/2.

The following theorem by Ville shows that a sequence can be in some sense non-random even when the above properties are satisfied. Thus von Mises' approach cannot work as stated (even though there may be modifications which work).

A *selection function* is a function $f : \Sigma^* \to \{Y, N\}$. It makes sense in the context of selecting on prefixes of an infinite sequence. When we input a finite prefix $\alpha \upharpoonright n$ of an infinite sequence α, it outputs whether to select the next position - Y indicates that we must select the next position, and N indicates that we must not select the next position. The selected positions define a subsequence of α.

Let $S(\alpha, n)$ be the number of ones in $\alpha \upharpoonright n$ and $S_f(\alpha, n)$ be the number of ones in the subsequence of $\alpha \upharpoonright n$ which are selected by f.

Theorem 7.1. *Let E be any finite collection of selection functions. Then there is a sequence α such that the following hold.*

1. $\lim_{n\to\infty} \frac{S(\alpha,n)}{n} = \frac{1}{2}$.
2. *For every $f \in E$ that selects infinitely many indices,* $\lim_{n\to\infty} \frac{S_f(\alpha,n)}{n} = \frac{1}{2}$.
3. *For all n, we have* $\frac{S(\alpha,n)}{n} \leq \frac{1}{2}$.

What is counterintuitive about the theorem is item (3) - if a sequence is truly random, then the averages must converge to 1/2 in a two-sided manner - at some times, the average should be greater than 1/2. But Ville's theorem says that there is a sequence where the oscillation about 1/2 is one sided. A clever betting algorithm can utilize (3) to bet and succeed on α - for example, if at some prefix length, the number of zeroes and ones are equal, then it is certain that the next bit will be a 0, so the algorithm can put its entire money on 0. In general, it can always bet more on 0 than on 1.

Note. Ville's theorem was more general. It works against countably infinitely many selection functions. However, the central probabilistic idea of the proof is contained in the above theorem.

We now prove the result.

Proof. Without loss of generality, we consider that the function which always outputs Y is in E. Thus, it suffices to construct α to satisfy (2) and (3).

Suppose we have constructed $\alpha[0 \dots n-1]$.

Let

$$C(n) = \{f \in E \mid f(\alpha \upharpoonright n) = Y\}.$$

This is the subcollection of all functions which select the $(n+1)^{\text{st}}$ position. By our additional assumption, $C(n)$ is non-empty.

We now extend α by one bit. We let $\alpha[n]$ to be 1 if the same set of functions $C(n)$ has appeared an even number of times among $C(0)$, $C(1)$, $\dots$, $C(n)$. Otherwise, set $\alpha[n]$ to be 0.

It is clear that the construction of α satisfies property (3). This is because for every 1 in α, there is a unique 0 in α that has appeared before. Thus the number of ones in α on any prefix is at most the number of zeroes in the prefix.

Suppose that for any $f \in E$ the selected indices are $n_0 < n_1 < \dots$. We now analyze the situation considering the different subsets of E that f can be part of.

For any C containing f, let $n_{i_0} < n_{i_1} < \dots$ be the (possibly finite) subsequence of $n_0 < n_1 < \dots$, where $C(n_i) = C$, we have $\alpha[n_{i_j}] = 0$ for every even j and 1 for every odd j. So for all n, the

number of 1s among the first n many bits of a selected by f differs from the number of 0s by at most 1. Taking limits as $n \to \infty$, we get (3).

□