On Analogues of the Church -Turing Thesis in Algorithmic Randomness

Christopher P. Porter
Université Paris 7
LIAFA

MCMP
Munich, Germany
4 December 2013

Our starting point: The Church -Turing thesis

(CTT) The informal notion of an effectively calculable number-theoretic function has the same extension as the formal notion of a Turing computable number-theoretic function.

Gödel: "[W]ith this concept one has for the first time succeeded in giving an absolute definition of an interesting epistemological notion, i.e., one not depending on the formalism chosen."

An alternative scenario to consider

One key datum that lends credence to the Church-Turing thesis is the lack of alternative definitions that seriously contend to capture the intuitive notion of effective calculability.

But what if matters had been otherwise, and there were multiple, non-equivalent definitions of computable function, each of which had a reasonable claim to capturing the intuitive notion of effectively calculable function?

In such a scenario, how would our understanding of the concept of computability have differed from how it is currently understood?

A multiplicity of definitions of randomness

In theory of algorithmic randomness, we are confronted with precisely such a scenario.

In the first place, there are multiple, non-equivalent definitions of random sequence.

To further complicate matters, several of these definitions have even been claimed to capture the intuitive conception of randomness.

That is, multiple non-equivalent randomness-theoretic theses have appeared in the algorithmic randomness literature.

The Martin-Löf-Chaitin Thesis

(MLCT): The informal notion of an intuitively random sequence has the same extension as the formal notion of a Martin-Löf random sequence.

The Martin-Löf-Chaitin thesis is articulated as analogous to the Church-Turing thesis:

> *The Church -Turing Thesis and the Martin-Löf-Chaitin Thesis are really similar: each of them is a statement of identification of a mathematical notion with an intuitive metamathematical one. (J.-P. Delahaye, "Randomness, Unpredictability, and Absence of Order", pg. 152)*

Two Alternative Theses

Schnorr's Thesis (ST): The informal notion of an intuitively random sequence has the same extension as the formal notion of a Schnorr random sequence.

The Weak 2-Random Thesis (W2RT): The informal notion of an intuitively random sequence has the same extension as the formal notion of a weakly 2-random sequence.

Given that

$$\mathsf{W2R} \subsetneq \mathsf{MLR} \subsetneq \mathsf{SR},$$

clearly at least two of the MLCT, ST, and the W2RT must be false.

One caveat

The search for a single “correct” definition of algorithmic randomness is not a serious concern in current research in algorithmic randomness.

However, as several of the previously mentioned theses have appeared in recent technical and philosophical discussions of algorithmic randomness, the status of these three theses merits our attention.

My goal for today

For reasons that cannot be discussed here, in the 1960s there was a concern with finding a single adequate definition of randomness, one that would, in Martin-Löf's words, satisfy "all intuitive requirements."

However, in recent years there has been a shift in the algorithmic randomness community away from judging definitions of randomness to be adequate or inadequate solely on the basis of how well they capture certain pre-theoretic intuitions about the concept of randomness.

Here I try to articulate why such a shift has taken place and to provide an account of a more reasonable criterion of adequacy for definitions of randomness that is more in line with the current situation in the field.

Outline of the remainder of the talk

1. Background on algorithmic randomness
2. Randomness-theoretic theses
3. Almost everywhere typicality

1. Background on algorithmic randomness

Extensional definitions of randomness

A distinctive feature of definitions of algorithmic randomness is that they are what one might call *extensional*.

Roughly, a definition of randomness is extensional if it counts an object as random whenever it satisfies certain properties that the typical outcomes of some paradigmatically random process satisfy.

More precisely, an extensional definition of randomness $\mathscr{D}$ can formulated in terms of a countable collection of properties $\{\Phi_i\}_{i\in\omega}$, such that

- for each $i \in \omega$, the set $\{X \in 2^\omega : \Phi_i(X)\}$ has measure one, and
- $X \in 2^\omega$ is $\mathscr{D}$-random if and only if $\Phi_i(X)$ for every $i \in \omega$.

Randomness as typicality

The definitions that I will discuss today are all very similar in form.

Each such definition can be seen as a formalization of the idea that random sequences are statistically typical.

Not all definitions of algorithmic randomness take this form

- Some definitions are given in terms of unpredictability.
- Others are given in terms of incompressibility.

Towards a statistical definition of randomness, 1

Given a finite string $\sigma \in 2^{<\omega}$, we'd like to test whether it is random.

Null hypothesis: σ is random.

How do we test this hypothesis?

We employ a statistical test $\mathcal{T}$ that has a critical region U corresponding to the significance level α.

If our string is contained in the critical region U, we reject the hypothesis of randomness at level α (say, $\alpha = 0.05$ or $\alpha = 0.01$).

Towards a statistical definition of randomness, 2

Given an infinite sequence $X \in 2^\omega$, we'd like to test whether it is random.

Null hypothesis: X is random.

How do we test this hypothesis?

Towards a statistical definition of randomness, 2

Given an infinite sequence $X \in 2^\omega$, we'd like to test whether it is random.

Null hypothesis: X is random.

How do we test this hypothesis?

We test initial segments of X at *every level of signficance*: $\alpha = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots, \frac{1}{2^n}, \ldots$

A test for 2^ω is now given by an infinite collection $(\mathcal{T}_i)_{i\in\omega}$ of tests for $2^{<\omega}$, where the critical region U_i of $\mathcal{T}_i$ corresponds to the significance level $\alpha = 2^{-i}$.

Formally...

A *Martin-Löf test* is a sequence $(U_i)_{i\in\omega}$ of uniformly computably enumerable sets of strings such that for each i,

$$\sum_{\sigma\in U_i} 2^{-|\sigma|} \leq 2^{-i}.$$

(Think of each U_i as the critical region for a statistical test $\mathcal{T}_i$ at significance level $\alpha = 2^{-i}$.)

A sequence $X \in 2^\omega$ *passes a Martin-Löf test* $(U_i)_{i\in\omega}$ if there is some i such that for every k, $X{\upharpoonright}k \notin U_i$.

$X \in 2^\omega$ is *Martin-Löf random*, denoted $X \in \mathrm{MLR}$, if X passes *every* Martin-Löf test.

The measure-theoretic formulation

Given $\sigma \in 2^{<\omega}$,

$$[\![\sigma]\!] := \{X \in 2^\omega : \sigma \prec X\}.$$

These are the basic open sets of 2^ω.

The Lebesgue measure on 2^ω is defined by

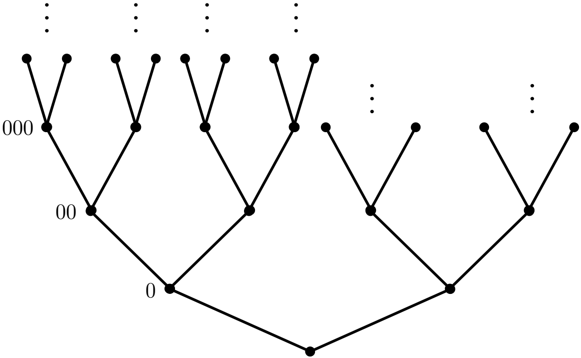
$$\lambda([\![\sigma]\!]) = 2^{-|\sigma|}.$$

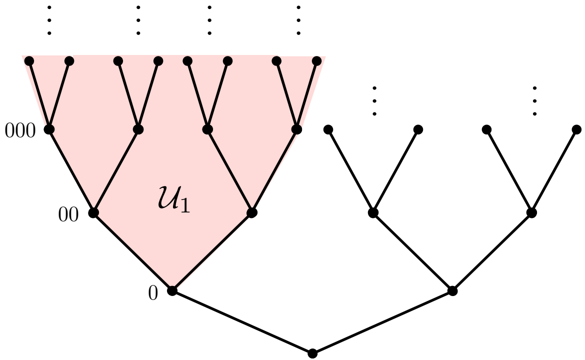
Thus we can consider a Martin-Löf test to be a collection $(\mathcal{U}_i)_{i\in\omega}$ of uniformly effectively open subsets of 2^ω such that

$$\lambda(\mathcal{U}_i) \leq 2^{-i}$$

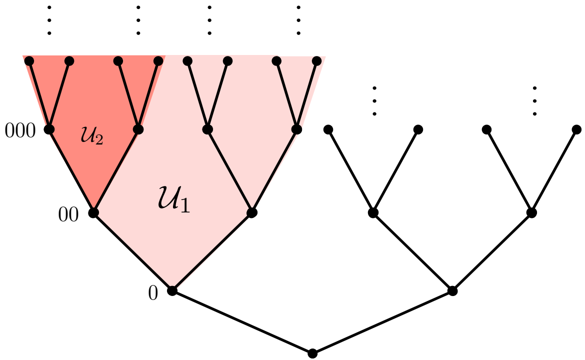
for every i.

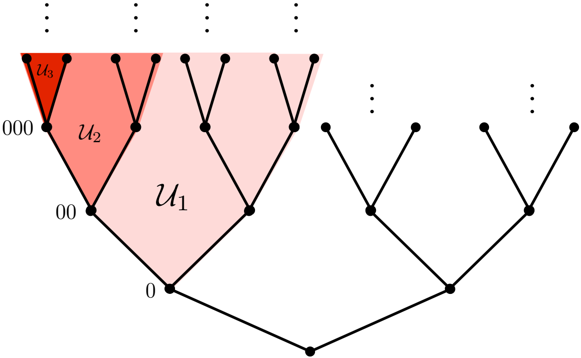
Moreover, X passes the test $(\mathcal{U}_i)_{i\in\omega}$ if $X \notin \bigcap_i \mathcal{U}_i$.


000
00
0

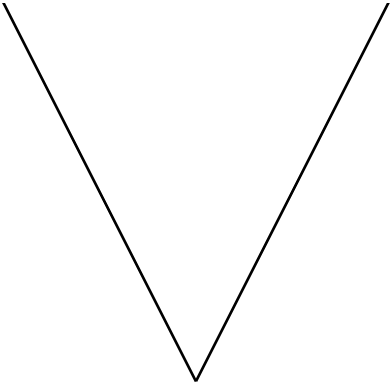


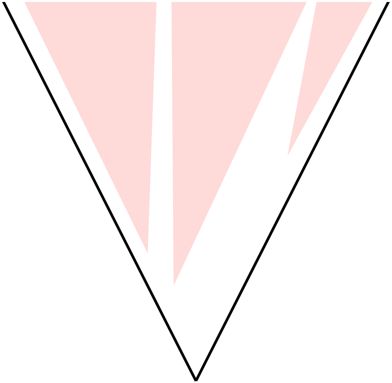
000
00
$\mathcal{U}_1$
0

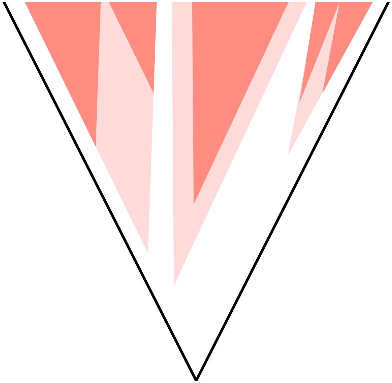

000
$\mathcal{U}_2$
$\mathcal{U}_1$
00
0

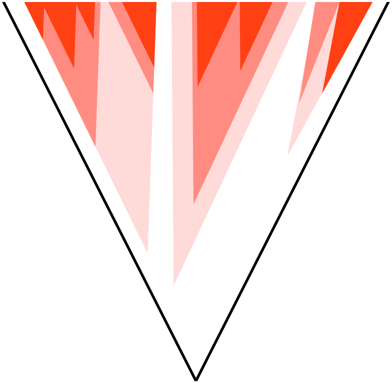


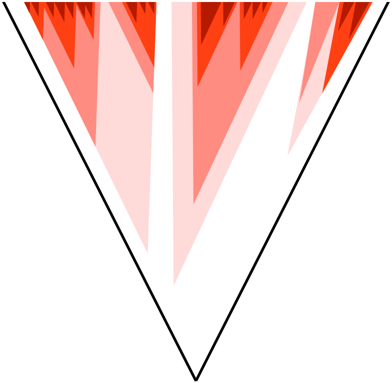
000
$\mathcal{U}_3$
$\mathcal{U}_2$
$\mathcal{U}_1$
00
0











Schnorr randomness

Schnorr presented a more constructive definition of randomness as an alternative to Martin-Löf randomness (for reasons I'll discuss shortly).

A *Schnorr test* is a collection $(\mathcal{U}_i)_{i\in\omega}$ of uniformly effectively open subsets of 2^ω such that

$$\lambda(\mathcal{U}_i) = 2^{-i}$$

for every i.

$X \in 2^\omega$ is *Schnorr random*, denoted $X \in \mathrm{SR}$, if it passes every Schnorr test.

MLR vs SR

Since every Schnorr test is a Martin-Löf test, a sequence that passes every Martin-Löf test thus passes every Schnorr test.

Consequently, we have $\mathrm{MLR} \subseteq \mathrm{SR}$.

With some work, one can show there is some $X \in \mathrm{SR} \setminus \mathrm{MLR}$.

As $\mathrm{MLR} \subsetneq \mathrm{SR}$, we say that Schnorr randomness is *weaker* than Martin-Löf randomness (or that Martin-Löf randomness is *stronger* than Schnorr randomness).

Weak 2-randomness

Another alternative to Martin-Löf randomness is known as weak 2-randomness.

Instead of strengthening the notion of a test (as in the definition of a Schnorr test), we can weaken it.

A *generalized Martin-Löf test* is a collection $(\mathcal{U}_i)_{i\in\omega}$ of uniformly effectively open subsets of 2^ω such that

$$\lim_{i\to\infty} \lambda(\mathcal{U}_i) = 0.$$

$X \in 2^\omega$ is *weakly 2-random*, denoted $X \in \mathrm{W2R}$, if it passes every generalized Martin-Löf test.

MLR vs W2R

Every Martin-Löf test is a generalized Martin-Löf test, and thus we have $\mathrm{W2R} \subseteq \mathrm{MLR}$.

Further, there is some $X \in \mathrm{MLR} \setminus \mathrm{W2R}$.

In sum, we have

$$\mathrm{W2R} \subsetneq \mathrm{MLR} \subsetneq \mathrm{SR}.$$

There are many more definitions of algorithmic randomness, but these are the three definitions that have been the subject of a randomness-theoretic thesis.

2. Randomness-theoretic theses

Several motivating questions

- What grounds have been offered for accepting the various randomness-theoretic theses?
- Are there any good reasons to prefer one thesis over the others?
- On what grounds should we reject a given thesis as unsatisfactory?
- More specifically, what should count as a counterexample to a given thesis?

In support of the MLCT

Martin-Löf does not explicitly state the MLCT, but he makes several remarks that come close.

First, he states that his notion of a test includes any sequential test "of present or future use in statistics".

He adds that his definition "seems to satisfy all intuitive requirements."

- The law of large numbers.
- The law of the iterated logarithm.

It had been open for almost thirty years whether there is a definition of randomness satisfying these properties.

"Such a definition has so far not been obtained by other methods."

The confluence of definitions as evidence for the MLCT

The primary piece of evidence one finds cited in the support of the MLCT is that Martin-Löf randomness is extensionally equivalent to a number of intensionally distinct definitions of randomness.

> *Different intuitive starting points have generated the same set of random sequences. This has been taken to be evidence that ML-randomness or equivalently (prefix-free) Kolmogorov randomness is really the intuitive notion of randomness, in much the same way as the coincidence of Turing machines, Post machines, and recursive functions was taken to be evidence for Church's Thesis, the claim that any one of these notions captures the intuitive notion of effective computability.*

-Eagle, "Chance vs. Randomness"
Stanford Encyclopedia of Philosophy

Evidence for Schnorr's Thesis

Schnorr argues that Martin-Löf randomness is too strong a notion of randomness, as it counts too many sequences as random.

> *The acceptable definition of random sequences cannot be any formulation of recursive function theory which contains all relevant statistical properties of randomness, but it has to be precisely a characterization of all those properties of randomness that have a physical meaning. These are intuitively those properties that can be established by statistical experience.*

He provides a criterion of physical meaning for a property of randomness and proves that Schnorr randomness satisfies this criterion.

Schnorr also appeals to the confluence of definitions in support of his own thesis.

Evidence for the W2RT

In "Recognizing Strongly Random Reals", Osherson and Weinstein provide a learning-theoretic characterization of weak 2-randomness.

They also argue that Martin-Löf randomness is not a strong enough definition of randomness.

- Some Martin-Löf random sequences are Δ^0_2 (or equivalently, decidable in the limit).
- But no intuitively random sequence is decidable in the limit.
- Thus there are Martin-Löf random sequences that are not intuitively random.

Weak 2-randomness does not face this problem.

- For every Δ^0_2 $X \in 2^\omega$, $\{X\}$ is a Π^0_2 singleton.

A challenge

Note that the advocate of the MLCT faces a challenge, namely

(1) to provide reasons for rejecting all $X \in \mathrm{SR} \setminus \mathrm{MLR}$ as not intuitively random, and

(2) to do so in such a way that it does not undermine the claim that every sequence in $\mathrm{MLR} \setminus \mathrm{W2R}$ is intuitively random.

For instance, every $X \in \mathrm{SR} \setminus \mathrm{MLR}$ computes a function that dominates all computable functions, a property one might argue is incompatible with being intuitively random.

However, the advocate of the MLCT cannot appeal to this property, as there are also Martin-Löf random sequences that satisfy this property.

The other definitions of randomness face a similar problem.

The upshot

The evidence offered in support of the various theses is at best indecisive.

In particular, the so-called "intuitive conception of randomness" is in need of further clarification so as to underwrite, for instance, the claim that a given formally random sequence is not intuitively random.

The situation is more dire for advocates of the various randomness-theoretic theses in light of the results on almost everywhere typicality that I will now discuss.

3. Almost everywhere typicality

Almost everywhere theorems

In classical analysis, it is very common to encounter theorems that hold of almost every member of some fixed domain of objects, usually some subset of the real numbers.

A number of these results involve some collection $\mathscr{C}$ of real-valued functions $f : [0,1] \to \mathbb{R}$ and have the form

$$(\forall f \in \mathscr{C})(\forall^{\text{a.e.}} x \in [0,1])\ \Phi(x, f),$$

where

- $\forall^{\text{a.e.}}$ is the almost everywhere quantifier (so that $(\forall^{\text{a.e}} x \in [0,1])\ \Phi(x)$ means that the set $\{x : \Phi(x)\}$ has Lebesgue measure one), and
- $\Phi(x, f)$ is some predicate such as "f is differentiable at x."

An informal gloss

Such results are commonly glossed as follows:

If we choose a point $x \in [0, 1]$ at random, then with probability one, the property $\Phi(\cdot, f)$ will hold at x.

Alternatively, we might say that it is the *typical* behavior of points $x \in [0, 1]$ for the each of the above properties $\Phi(\cdot, f)$ to hold at x, or that these properties hold of the random member of $[0, 1]$.

Hereafter, such typical behavior will be referred to as *a.e. typicality*.

A theorem involving a.e. typicality

Consider the following example of a.e. typicality:

Theorem: For every real-valued function $f : [0, 1] \to \mathbb{R}$ of bounded variation, f is differentiable almost everywhere.

A few observations:

- The function quantifier in this theorem ranges over sets of size $2^{\mathfrak{c}}$, the size of the power set of the continuum.
- The properties "being a point of differentiability of some real-valued function of bounded variation" and "being a point of non-differentiability of some real-valued function of bounded variation" are satisfied by every point in [0,1].

A restricted version of the theorem

Now consider:

For every *computable* non-decreasing real-valued function $f : [0,1] \to \mathbb{R}$, f is differentiable almost everywhere.

A few observations:

- The function quantifier in this theorem now ranges over countably many functions.
- Thus the property "being a point of differentiability of every computable real-valued function of bounded variation" is the intersection of countably many sets of measure one, which is itself a set of measure one.

The connection to randomness

It follows that the collection of points x such that some computable real-valued function f is not differentiable at x is a null set.

In fact, it is an effective null set:

Theorem (Brattka, Miller, Nies)

$z \in [0,1]$ is Martin-Löf random if and only if every computable, real-valued function $f : [0,1] \to \mathbb{R}$ of bounded variation is differentiable at z.

That is, Martin-Löf randomness is necessary and sufficient for this particular instance of a.e. typicality.

More examples

In fact, each of the definitions we've considered here is necessary and sufficient for some notion of a.e. typicality.

$x \in \mathsf{MLR}$	$\Leftrightarrow$	Every computable real-valued function of bounded variation is differentiable at x.
$x \in \mathsf{SR}$	$\Leftrightarrow$	For every L_1-computable real-valued function f, the Lebesgue differentiation theorem holds for f at x.
$x \in \mathsf{W2R}$	$\Leftrightarrow$	Every computable real-valued a.e.-differentiable function is differentiable at x.

There are a number of other examples, some involving definitions of randomness that we have not considered here.

The challenge for the randomness-theoretic theses

Let us consider the status of the MLCT in light of our data.

I claim that the advocate of the MLCT faces a dilemma:

(i) explain why it is that the points of differentiability of computable functions of bounded variation pick out precisely the intuitively random points while these other instances of a.e. typicality do not, or

(ii) concede that there is nothing more intuitively random about points in the first collection than the points in the other two collections.

A similar dilemma is faced by advocates of the other two theses.

An alternative approach

Whereas on the approach taken by the advocates of the various randomness theses, only one definition can be correct, on the alternative approach outlined here, multiple extensionally non-equivalent definitions of randomness can be, in a certain sense, correct:

With respect to one notion of a.e. typicality, one definition of randomness $\mathscr{D}$ captures precisely that notion of typicality, and all definitions that differ from $\mathscr{D}$ extensionally fail in this respect.

However, with respect to other notions of a.e. typicality, $\mathscr{D}$ fails to pick out the capture notion of typicality, but some non-equivalent notion $\mathscr{D}'$ does the trick.

A classificatory role for definitions of randomness

The picture that emerges is that

(i) the various definitions of randomness each correspond to a specific *degree* of randomness,

(ii) these degrees of randomness correspond to mathematically significant instances of a.e. typicality, but

(iii) there is no single degree of randomness that can capture all mathematically significant instances of a.e. typicality.

This suggests a research program (which is already under way), namely to classify various notions of a.e. typicality in classical mathematics according to the corresponding definition of randomness.

In Closing

With much foresight, Kolmogorov wrote,

> *The notions of information theory in their application to infinite sequences make possible some very interesting research that, although it is not necessary from the point of view of the foundations of probability, may have a certain significance in the study of the algorithmic aspect of mathematics as a whole.*