The Algorithmic Approach to Randomness

Christopher P. Porter
Université Paris 7
LIAFA

Institut de Mathématiques de Toulouse
Université Paul Sabatier
Toulouse, France
20 March 2014

Introduction

Over roughly the past fifteen years, the theory of algorithmic randomness has grown into a mature and fruitful sub-branch of computability theory.

Despite the many developments in this area of research, one can make the case that the conceptual foundations of algorithmic randomness are in need of clarification.

In particular, we might ask:

- What is the relationship between notions of algorithmic randomness and the notions of randomness that occur in classical mathematics?
- Does the theory of algorithmic randomness provide a conceptual analysis of the notion of randomness?

My approach

The approach I will take today is to frame the discussion in terms of three general problems for the theory of algorithmic randomness:

1. the undergeneration problem;
2. the overgeneration problem; and
3. the disconnect problem.

In fact, these problems are faced by what I call *logical definitions of randomness*, of which the definitions of algorithmic randomness are instances.

Logical definitions of randomness

The main ingredients

The main definitions of algorithmic randomness are given in terms of the following ingredients:

- a collection of objects $\mathcal{O}$;
- a probability measure μ on $\mathcal{O}$; and
- a collection of properties $\{\Phi_i\}_{i \in I}$, expressible in some language $\mathcal{L}$ and satisfiable by objects in $\mathcal{O}$, such that for each $i \in I$,

$$\mu(\{x \in \mathcal{O} : \Phi_i(x)\}) = 1.$$

Hereafter, I will refer to the properties $\{\Phi_i\}_{i \in I}$ as *measure one properties* (where the measure is understood to be the relevant underlying measure μ).

Putting the ingredients together

From a triple $(\mathcal{O}, \mu, \{\Phi_i\}_{i \in I})$ satisfying the conditions from the previous slide, we get a definition $\mathscr{D}$ of μ-randomness for objects in $\mathcal{O}$ by stipulating that

$$x \in \mathcal{O} \text{ is } \mathscr{D}\text{-random if and only if } \Phi_i(x) \text{ for every } i \in I.$$

One immediate consequence of this definitional framework is that, assuming that there is some $x \in \mathcal{O}$ and some $i \in I$ such that x does not satisfy Φ_i, we can partition $\mathcal{O}$ into

- a non-empty collection of $\mathscr{D}$-random objects, and
- a non-empty collection of non-$\mathscr{D}$-random objects.

Valuative randomness

Note that this approach differs radically from one frequently occurring notion of randomness in classical mathematics, which I refer to as *valuative randomness*.

Roughly speaking, the idea behind valuative randomness is this: to be random is to be the value of a random variable.

Recall that a random variable is simply a measurable function from a sample space Ω to some space, usually $\mathbb{R}$.

The usage of ‘random’ is not exact here; randomness is usually attributed to the function itself, but sometimes it is also attributed to individual outputs of the function.

ϕ-valued random variables

However, it is important to emphasize that in practice, the range of a random variable can be any collection of mathematical objects:

- complex numbers
- vectors
- matrices
- functions
- graphs
- closed sets
- Banach spaces
- and so on...

Let ϕ be a mathematical object such as one from any of the collections listed above.

Then "the random ϕ" is simply a ϕ-valued random variable.

Where exactly is the randomness?

We are supposed to think of a random variable as yielding the values of some 'random' experiment (such as some measurement of some randomly selected individual).

Thus, a ϕ-valued random variable is can be understood as yielding as output a randomly chosen ϕ from the relevant collection of objects.

Note that this random experiment/choice isn't technically part of the definition of a random variable, but in applications, such experiments or choices are associated to random variables.

Almost sure events

Random variables can take values that appear to be "non-random," at least informally speaking.

For instance, a real-valued random variable can take the value $0.111\ldots$.

However, there is a sense in which such outcomes are atypical.

In particular, one can associate a probability distribution to a random variable, and by means of such a probability distribution, one can define events that happen almost surely (i.e. with probability one).

Thus, if some property Θ occurs almost surely with respect to the probability distribution associated to a ϕ-valued random variable, we say, "the random ϕ has Θ almost surely."

Comparing the logical and valuative approaches

The key distinction between the logical and valuative approaches is the former is *discriminative* while the latter is not.

That is, on the logical approach, one discriminates between the random and the non-random objects.

In fact, one first specifies the non-random elements, which form a set of measure zero (with respect to the given measure), and the remaining elements are taken to be the random elements.

By contrast, on the valuative approach, *any* object in the relevant domain of objects can be the value of a random variable (and thus can be random).

For the most part, on the valuative approach, we never attribute non-randomness to any objects.

The undergeneration problem

Motivating the undergeneration problem

Recall that a logical definition is formulated in terms of a collection of “measure one properties” $\{\Phi_i\}_{i\in I}$.

The undergeneration problem concerns the choice of the properties $\{\Phi_i\}_{i\in I}$.

For each object $x \in \mathcal{O}$, such we define the formula $\Phi_x(y)$ to be

$$y \neq x.$$

Then assuming that $\mu(\{x\}) = 0$, we will have

$$\mu(\{y \in \mathcal{O} : \Phi_x(y)\} = 1.$$

If μ is continuous (i.e., $\mu(\{y\}) = 0$ for every $y \in \mathcal{O}$), then each of the properties in $\{\Phi_x\}_{x\in\mathcal{O}}$ will be a measure one property, but together they will yield an empty definition of randomness.

Addressing the undergeneration problem: a first step

We can at least avoid this problem by restricting to countably many measure one properties.

The countable intersection of sets of measure one is a set of measure one, so the resulting definition of randomness will be non-empty.

But this raises a further problem: Which countable collection of properties do we choose?

The challenge: provide a *principled* restriction of the properties $\{\Phi_i\}_{i\in\omega}$.

Von Mises and undergeneration

The undergeneration problem was first raised for von Mises' definition of random sequences (which can be seen as a prototype of the definitions of randomness formulated according to the logical approach).

Roughly, von Mises' definition required that a random sequence have relative limiting frequencies that are invariant under selection of subsequences by "admissible place selections."

Von Mises' contemporaries objected that any place selection could be counted as admissible according to von Mises' definition of admissibility.

Thus, they claimed, von Mises' definition of randomness yields an empty collection of random sequences.

Responses to undergeneration

In response to this objection, Abraham Wald proved in 1937 that for any choice of countably many place selections, the definition of randomness given in terms of this countable collection is non-empty.

In fact, Wald showed it has size continuum, while Doob showed that it has measure one (with respect to the induced product measure).

In 1940, Church suggested that von Mises' definition should be restricted to all *computable* place selection rules.

This definition was still found to be inadequate, as Ville proved that not every sequence that is random according to Church's definition satisfies the law of the iterated logarithm.

Towards Martin-Löf's definition of randomness

Given a finite string $\sigma \in 2^{<\omega}$, we'd like to test whether it is of 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 Martin-Löf's 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*: $\alpha = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots, \frac{1}{2^n}, \dots$

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, $X\restriction k \notin U_i$.

$X \in 2^\omega$ 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

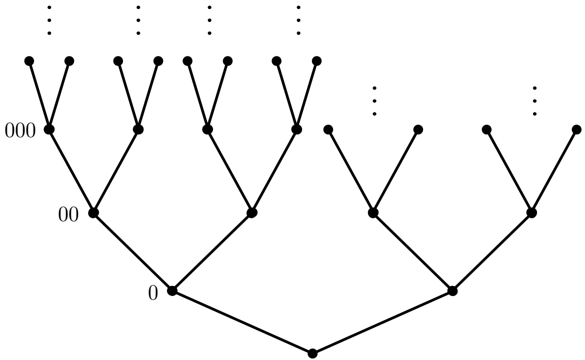
$$\lambda([\![\sigma]\!]) = 2^{-|\sigma|}.$$

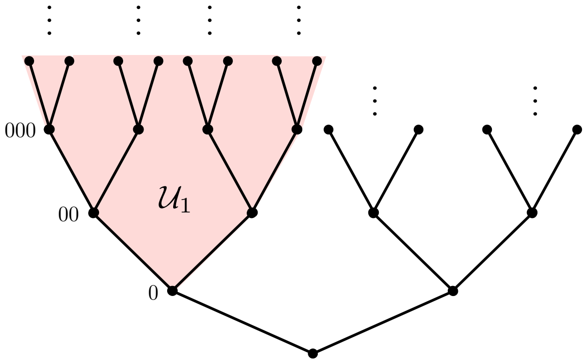
Thus we can consider a Martin-Löf test to be a collection $(\mathcal{U}_i)_{i \in \omega}$ of uniformly effectively open subsets of 2^ω such that

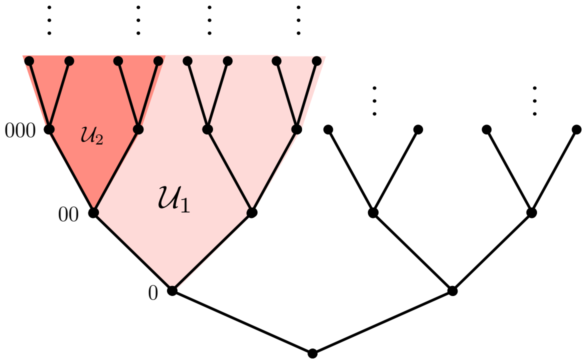
$$\lambda(\mathcal{U}_i) \leq 2^{-i}$$

for every i.

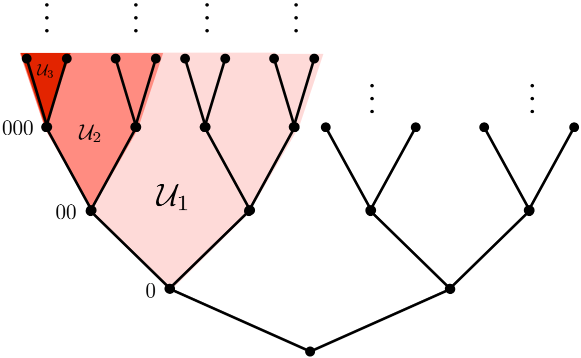
Moreover, X passes the test $(\mathcal{U}_i)_{i \in \omega}$ if $X \notin \bigcap_i \mathcal{U}_i$.


000
00
0

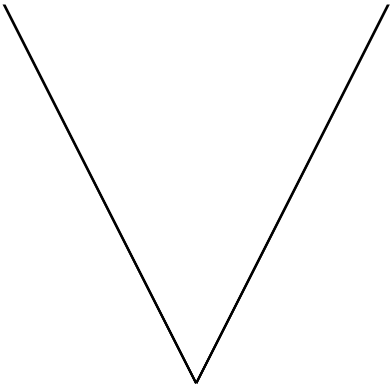

000
00
0
$\mathcal{U}_1$

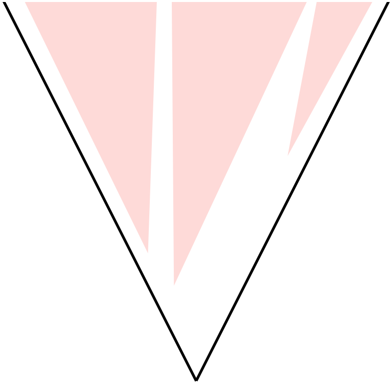


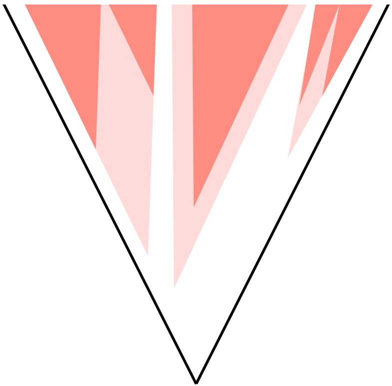
000
$\mathcal{U}_2$
$\mathcal{U}_1$
00
0

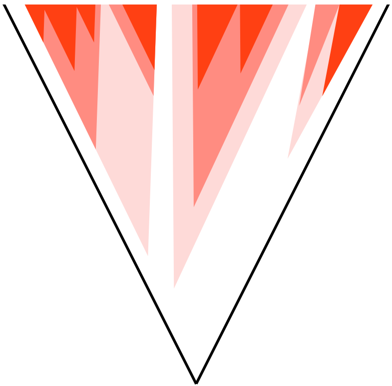


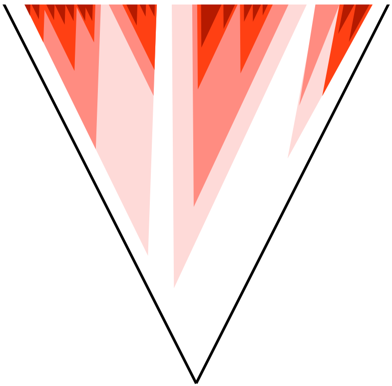
$\mathcal{U}_3$
000
$\mathcal{U}_2$
$\mathcal{U}_1$
00
0











Is the undergeneration problem answered?

For each Martin-Löf test $(U_i)_{i\in\omega}$, the set

$$2^\omega \setminus \bigcap_{i\in\omega} [\![U_i]\!]$$

corresponds to one of the measure one properties in the description of the logical approach to randomness.

Does this choice of measure one properties successfully answer the undergeneration problem?

Alternative definitions of randomness complicate matters.

Schnorr randomness

Schnorr presented a more constructive definition of randomness as an alternative to Martin-Löf randomness.

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 $\mathsf{MLR} \subseteq \mathsf{SR}$.

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

As $\mathsf{MLR} \subsetneq \mathsf{SR}$, 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 $\mathsf{W2R} \subseteq \mathsf{MLR}$.

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

In sum, we have

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

Shortly, we will see that there are good reasons to hold that each of these definitions is provides a legitimate response to the undergeneration problem.

The overgeneration problem

What is the overgeneration problem?

The overgeneration problem concerns the choice of the measure μ in logical definitions of randomness.

For any object $x \in \mathcal{O}$, there is a measure μ_x such that $\mu(\{x\}) = 1$ (i.e., the *Dirac measure* concentrated on x).

Then the only μ-random element of $\mathcal{O}$ is x.

If we are too general in our approach to defining randomness, we run the risk of counting every object as random with respect to some definition.

Randomness with respect to non-computable measures

One need not appeal to Dirac measures to formulate the overgeneration problem.

If we consider, say, Martin-Löf randomness with respect to non-computable measures on 2^ω, one can prove the following:

Theorem (Reimann-Slaman)

For every sequence X, X is non-computable if and only if there is some measure μ such that

(i) $\mu(\{X\}) = 0$ *and*

(ii) X *is Martin-Löf random with respect to* μ.

Surprisingly, this fact can be witnessed by a *single* measure!

Artifactual definitions of randomness

The measures in Reimann-Slaman theorem are admittedly exotic (for instance, it is necessary that they give *some* points positive measure, i.e. they are necessarily discontinuous).

A case can be made that the Reimann-Slaman theorem and related results are artifacts of the computational framework used to define randomness (particularly when we consider non-computable measures).

Of course, we can ask: which notions of randomness are merely artifactual and which ones are not?

Restricting to computable measures?

However, the measures considered in mathematical practice are computable measures (the Lebesgue measure, Bernoulli measures with rational parameter p, etc.).

In fact, it is quite difficult to produce an example of a non-computable measure, especially without appealing to the standard tricks from computability theory.

The stability of randomness w.r.t. computable measures

Further, from the point of view of algorithmic randomness, there is a high degree of stability among the sequences random with respect to some computable measure:

Theorem (Levin-Kautz)

For every non-computable sequence X, if X is Martin-Löf random with respect to some computable measure μ, then X is Turing equivalent to a Martin-Löf random sequence with respect to the Lebesgue measure.

The Turing equivalence of two sequences means that they can be effectively transformed into one another.

In the case that μ is computable and continuous, the transformations in the Levin-Kautz theorem are effectively uniformly continuous.

Taking stock

We've seen some possible responses to the undergeneration problem, in the form of three specific definitions of algorithmic randomness.

Further, we've seen a possible line of response to the overgeneration problem, albeit one in need of further development.

The most pressing of the three problems, however, is the disconnect problem.

The disconnect problem

What is the disconnect problem?

As we've seen, each definition $\mathscr{D}$ of algorithmic randomness partitions the domain in $\mathcal{O}$ into two non-empty collections:

- the $\mathscr{D}$-random objects, and
- the non-$\mathscr{D}$-random objects.

This is *not* a feature shared by the uses of randomness in classical mathematics (such as the examples we've seen).

Without any clear connection to the classical uses of randomness in mathematics, let alone to commonly held intuitions about the concept of randomness, are we even justified in referring to logical definitions of randomness as definitions *of* randomness?

Response one: denial

One way to respond to the disconnect problem is to simply deny that it's a problem and simply embrace the disconnect.

Why should the definitions of algorithmic randomness answer to the uses of randomness in classical mathematics, or to commonly held intuitions of randomness?

As we will see, this response concedes too much.

Response two: a unique solution

Another way to respond to the disconnect problem, which would yield a very strong solution to the problem, is to identify a single "correct" definition of randomness.

Just as the notion of Turing computable function captures the intuitive conception of effectively calculable function, we could hope to isolate a single definition of randomness that captures the intuitive conception of randomness.

A worry about the "unique solution" response

Although some have held that there is a single such correct definition, such a view has always been articulated for definitions of random sequence with respect to the Lebesgue measure.

For instance, each of the definitions of randomness introduced earlier have been held to capture the intuitive conception of randomness.

But what about definitions of randomness for other objects, and with respect to different measures?

Should we hope for one general definition of randomness that is correct for each choice of objects and underlying measure?

Response three: "almost everywhere" typicality

A promising response to the disconnect problem is to appeal to recent results in algorithmic randomness that reveal deep connections between various notions of randomness and certain "almost everywhere" theorems from classical mathematics.

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

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.

A.e. typicality in classical analysis

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

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

A.e. typicality in ergodic theory

$x \in \text{MLR}$	$\Leftrightarrow$	Birkhoff's ergodic theorem holds at x for all computable ergodic transformations with respect to every lower semi-computable function.
$x \in \text{SR}$	$\Leftrightarrow$	Birkhoff's ergodic theorem holds at x for all computable ergodic transformations with respect to every computable function.
$x \in \text{W2R}$	$\Rightarrow$	A weak version of Birkhoff's ergodic theorem holds at x for all computable measure-preserving transformations with respect to every lower semi-computable function.

The emerging picture

There are other promising developments along a slightly different lines:

- Martin-Löf random closed sets;
- Martin-Löf random Brownian motion;
- effective notions of Hausdorff and packing dimension;
- connections to information theory.

Taken together, these developments provide good grounds for dismissing the worry that there is a disconnect between algorithmic definitions of randomness and notions of randomness from classical mathematics.

References

Textbooks on algorithmic randomness:

- Downey, Hirschfeldt, *Algorithmic randomness and complexity*, Springer, 2010.
- Nies, *Computability and randomness*, Oxford, 2009.

Algorithmic randomness and differentiability

- Brattka, Miller, Nies, “Randomness and differentiability,” forthcoming.
- Pathak, Rojas, Simpson, “Schnorr randomness and the Lebesgue differentiation theorem,” *Proceedings of the AMS*, 2014.

References (continued)

Algorithmic randomness and ergodic theory

- Bienvenu, Day, Hoyrup, Mezhirov, Shen, "A constructive version of Birkhoff's ergodic theorem for Martin-Löf random points," *Information and Control*, 2012.
- Gács, Hoyrup, Rojas, "Randomness on computable probability spaces—a dynamical point of view," *Theory of Computing Systems*, 2011.
- Franklin, Towsner, "Randomness and nonergodic systems," forthcoming.