Randomness from Borel through Turing and into the 21st Century

Rod Downey
Victoria University
Wellington
New Zealand

Tokyo, May, 2013

Plan

- First I will talk about the general theory of algorithmic randomness.
- Second I plan to relate this to work of Turing, not published in his lifetime, and see how he anticipates ideas from today.

Plan-randomness

- Since this is a general talk, I will give a basic lecture in this area, hopefully
- concentrating on recent themes.
- For more ... there are nice books (I can think of at least two!).
- Apologies to the experts.

Randomness

- How should we understand randomness?
- Can we generate randomness?
- What does this mean anyway?
- Can we *quantify* the amount of randomness?
- What does randomness do as a *computational resource*?

The great men

- Turing 1950:

  > *" An interesting variant on the idea of a digital computer is a "digital computer with a random element." These have instructions involving the throwing of a die or some equivalent electronic process; one such instruction might for instance be, "Throw the die and put the-resulting number into store 1000." Sometimes such a machine is described as having free will (though I would not use this phrase myself)."*

- von Neumann 1951:

  > *"Any one who considers arithmetical methods of producing random digits is, of course, in a state of sin."*

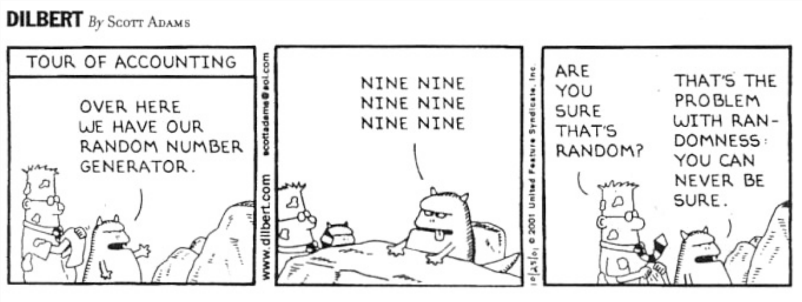
- It is fair to say that both had the idea of "pseudo-random" numbers with no formalization.

But even earlier

"How dare we speak of the laws of chance?
Is not chance the antithesis of all law?"

— Joseph Bertrand, Calcul des Probabilités, 1889

Intuitive Randomness



Intuitive Randomness

Which of the following binary sequences seem random?

A 000000000000000000000000000000000000000000000000000000000000000000

B 001101001101001101001101001101001101001101001101001101001101001101

C 010001101100000101001110010111011100000001001000110100010101

D 001001101101100010001111010100111011001001100000001011010100

E 010101110110111101110010011010110111001101101000011011110111

F 0111011111001101100110100100001111100110110000001101101010 1

G 000001100010111000100000000101000010110101000000100000000100

H 010100110111101101110101010000010111100000010101110101010001

Intuitive Randomness

Non-randomness: increasingly complex patterns.

A 000000000000000000000000000000000000000000000000000000000000

B 001101001101001101001101001101001101001101001101001101001101

C 010001101100000101001110010111011100000001001000110100010101

D 001001101101100010001111010100111011001001100000001011010100

E 010101110110111101110010011010110111001101101000011011110111

F 011101111100110110011010010000111111001101100000011011010101

G 000001100010111000100000000101000010110101000000010000000100

H 010100110111101101110101010000010111100000010101110101010001

Intuitive Randomness

Randomness: bits coming from atmospheric patterns.

A 0000000000000000000000000000000000000000000000000000000000000000

B 001101001101001101001101001101001101001101001101001101001101

C 0100011011000001010011100101110111011100000001001000110100010101

D 00100110110110001000111101010100111011001001100000001011010100

E 01010111011011110111001001101010110111001101101000011011110111

F 01110111110011011001101001000001111111001101100000011011010101

G 000001100010111000100000000101000001011010100000001000000000100

H 010100110111101101110101010000010111100000010101110101010001

Intuitive Randomness

Partial Randomness: mixing random and nonrandom sequences.

A 000000000000000000000000000000000000000000000000000000000000

B 001101001101001101001101001101001101001101001101001101001101

C 010001101100000101001110010111011100000001001000110100010101

D 001001101101100010001111010100111011001001100000001011010100

E 010101110110111101110010011010110111001101101000011011110111

F 011101111100110110011010010000111111001101100000011011010101

G 000001100010111000100000000101000010110101000000010000000100

H 010100110111101101110101010000010111100000001010111010101000 1

Intuitive Randomness

Randomness relative to other measures: biased coins.

A 000000000000000000000000000000000000000000000000

B 001101001101001101001101001101001101001101001101

C 010001101100000101001110010111011100000001001000110100010101

D 001001101101100010001111010100111011001001100000001011010100

E 010101110110111101110010011010110111001101101000011011110111

F 011101111100110110011010010000111111001101100000011011010101

G 000001100010111000100000000010100001011010100000001000000000100

H 010100110111101101110101010000010111100000001010111010101010001

Algorithmic

We need a way to talk about laws: “computable statistical tests” or “effective statistical tests” so that we can make sense of our intuition.

(So that, the expected behaviour of a distribution aligns itself to the behaviour of a particular input.)
The idea is to use the methods of computability and complexity theory.

Our Setting

- We will work with strings, members of Σ^*, typically $\{0,1\}^* = 2^{<\omega}$.
- Reals considered as infinite (binary) sequences, living in Cantor space 2^ω.
- Possible to have other settings, notably Levin-Gács, etc.
- For a string σ the uniform=Lebesgue measure of the collection of all infinite extensions of σ, $[\sigma]$, have $\mu([\sigma]) = 2^{-|\sigma|}$.

Three Approaches to Randomness at an Intuitive Level

- The statistician's approach: Deal directly with rare patterns using measure theory. Random sequences should not have effectively rare properties. (von Mises, 1919, finally Martin-Löf 1966)
- Computably generated null sets represent effective statistical tests.
- The coder's approach: Rare patterns can be used to compress information. Random sequences should not be compressible (i.e., easily describable) (Kolmogorov, Levin, Chaitin 1960-1970's).
- Kolmogorov complexity; the complexity of σ is the length of the shortest description of σ.
- The gambler's approach: A betting strategy can exploit rare patterns. Random sequences should be unpredictable. (Solomonoff, 1961, Schnorr, 1975, Levin 1970)
- No effective martingale (betting) can make an infinite amount betting on the bits.

The statisticians approach

- von Mises, 1919. A random sequence should have as many 0's as 1's. But what about 101010101010101010.....
- von Mises idea: If you select a subsequence $\{a_{f(1)}, a_{f(2)}, \ldots\}$ (e.g. $f(1) = 3, f(2) = 10, f(3) = 29,000$, so the 3rd, the 10th, the 29,000 th etc) then the number of 0's and 1's divided by the number of elements selected should tend to $\frac{1}{2}$. (Law of Large Numbers)
- But what selection functions should be allowed?
- Church, 1940 suggested (partial) computable selections.
- Ville, 1939 showed no countable selection possible. Essentially not enough statistical tests.

Turing

- As we later see, Turing was very interested in normality and absolute normality.
- Numbers are normal base d if they obey the frequency considerations : the number of i's that occur is $\frac{1}{d}$. Alsolutely normal if normal to any base.
- These concepts go back to Borel and his contemporaries.
- Schmidt (1960) was the first to construct absolutely normal numbers.
- Normality is a primitive form of randomness related to automata, more later.
- Interestingly, many recent advances in additive number theory based on supposing the primes are "random" but we don't understand,....returing to our story.

von Mises

- Recapping, the first person to look seriously at the notion of a random individual sequence was Richard von Mises (1919).
- Let $f : \omega \to \omega$ be an increasing injection, a selection function.
- Then a random X should satisfy the following.

$$\lim_{n\to\infty} \frac{\{m \mid m \leq n \wedge X(f(m)) = 1\}}{n} = \frac{1}{2}.$$

- Later Church 1940 said use (partial) computable f.

Ville's Theorem

Let $S(\alpha, n)$ denote the number of 1's in the first n bits of α and similarly S_f for the selected places.

Theorem (Ville's Theorem 1939)

Let E be any countable collection of selection functions. Then there is a sequence $\alpha = \alpha_0\alpha_1 \dots$ such that the following hold.

1. $\lim_n \frac{S(\alpha,n)}{n} = \frac{1}{2}$.
2. *For every $f \in E$ that selects infinitely many bits of α, we have* $\lim_n \frac{S_f(\alpha,n)}{n} = \frac{1}{2}$.
3. *For all n, we have* $\frac{S(\alpha,n)}{n} \leq \frac{1}{2}$.

The first two say that α is random insofaras all legal selections.

The problem is 3 since it says you always get more 0's than 1's in the first n bits.

Martin-Löf randomness

- Martin-Löf, 1966 suggests using shrinking effective null sets as representing effective tests-abstract tests. Basis of modern effective randomness theory.

Definition (Martin-Löf)

1. A *Martin-Löf test* is a sequence $\{U_n\}_{n\in\omega}$ of uniformly Σ^0_1 classes such that $\mu(U_n) \leq 2^{-n}$ for all n. ($U_n = \{[\sigma] \mid \sigma \in W_{f(n)}\}$)
2. Equivalently for ML, pass all Solovay tests meaning $A \in U_i$ for only finitely many i, so don't need nesting.
3. For example, consider eliminating all reals with the Ville Property of more 0's than 1's. We could let $U_1 = \{[0]\}$, $U_2 = \{[001], [010], [100]\}$, etc.
4. It is okay for α to be in some of these but it passes the test if it leaves them from some point onwards. Form some point onwards, it avoids having more 0's than 1's.
5. α is Martin-Löf random iff *alpha* passes all Martin-Löf tests.

The computational paradigm

- Can think of a machine $U(\tau) = \sigma$ as the information of the bits of τ describing σ.
- The length of the shortest τ is the U-Kolmogorov complexity of σ, $C_U(\sigma)$.
- σ is random if $C_U(\sigma) \geq |\sigma|$.
- Have universal machines and can define an optimal C up to a constant.
- Intentional meaning is not quite right as τ provides τ plus $|\tau|$ bits of information.
- This is avoided using telephone numbers, prefix-free complexity K.

K-randomness

- Prefix freeness gets rid of the use of length as extra information:
- Notice that prefix-freeness means that the domain of the machines has measure.
- The Coding Theorem (Levin-Gács) says that "Occam's Razor=Bayes' Theorem" in that if $Q(\sigma) = -\log(\mu\{\tau \mid U(\tau) = \sigma\}$, then $Q(\sigma) = K(\sigma)$.
- Roughly, the a priori probability of a string corresponds to its shortest description.

K-randomness

- (Levin, Chaitin) α is K- random if there is a c s.t.

$$\forall n(K(\alpha \upharpoonright n) > n - c).$$

Theorem (Schnorr)

X is K-random iff X is ML-random.

- I remark that other forms are possible, such as process complexity, which acts continuously.
- Eg. $U(\sigma) = \tau$ and $U(\sigma') = \tau'$ and $\sigma \prec \sigma'$ implies $\tau \prec \tau'$. (process)

Theorem (Levin, then Schnorr)

X is K-random iff for all n, $K_P(X \upharpoonright n) =^+ K_M(X \upharpoonright n) =^+ n$.

- The Coding Theorem fails for both monotone and process (Gács, then Day).

Martingales

- von Mises again. This time think about predicting the next bit of a sequence. Then you bet on the outcome. You should not win!
- (Levy) A martingale is a function $f : 2^{<\omega} \mapsto \mathbb{R}^+ \cup \{0\}$ such that for all σ,
$$f(\sigma) = \frac{f(\sigma 0) + f(\sigma 1)}{2}.$$
- the martingale *succeeds* on a real α, if $\limsup_n F(\alpha \upharpoonright n) \to \infty$.

- Think of betting on a sequence where you know that every second bit is 1. Then every second bit you could double your stake. This martingale exhibits exponential growth and that can be used to characterize computable reals.
- Ville proved that null sets correspond to success sets for martingales. They were used extensively by Doob in the study of stochastic processes.

- A supermartingale is a function $f : 2^{<\omega} \mapsto \mathbb{R}^+ \cup \{0\}$ such that for all σ,
$$f(\sigma) \geq \frac{f(\sigma 0) + f(\sigma 1)}{2}.$$
- Schnorr showed that Martin-Löf randomness corresponded to effective (super-)martingales failing to succeed.
- f as being effective or computably enumerable if $f(\sigma)$ is a c.e. real, and at every stage we have effective approximations to f in the sense that $f(\sigma) = \lim_s f_s(\sigma)$, with $f_s(\sigma)$ a computable increasing sequence of rationals.

All coincide

Theorem (Schnorr)

A real α is Martin-Löf random iff no effective (super-)martingale. succeeds on α.

Recent Major Themes

- Computational power of randoms
- Information theory and characterizing computability.
- Reflections in analysis, ergodic theory etc.
- Calibrating randomness.

Randoms should be computationally weak

- We now know that there are two kinds of randoms, those which resemble Chaitin's $\Omega = \sum_\sigma 2^{-K(\sigma)}$ and more typical ones.
- There has been a lot of popular press about the "number of knowledge" etc, which is random, but has high computational power.
- We would theorize randoms to be stupid: computationally weak.
- But, for all X there is a random Y with $X \leq_T Y$. (Kučera-Gács)

Stephan's Theorem

Theorem (Stephan)

If X is random and X has enough computational power to compute a $\{0,1\}$-valued function f such that for all e, $f(e) \neq \varphi(e)$, (ie X is called PA) then X computes the halting problem.

- Stupidity Tests
- There are two ways to convince someone you are stupid:
- The first people pass the stupidity test as they are so smart that they know how to be stupid, the second really are stupid.
- That is, with sufficient randomness, randomness begins to resemble order. This is kind of remarkable. We are still trying to understand it.
- In music it is quite difficult to distinguish between aleatoric (or chance) and totally serial (based on a pattern) music.

How Chaos Resembles Order

Highly random objects can resemble highly patterned ones.

A musical example.

Excerpt A: from *Music of Changes* by John Cage

Excerpt B: from *Structures for Two Pianos* by Pierre Boulez

Cage's piece is an example of aleatory music.

Boulez's piece is an example of total serialism.

- What this means is that if X is $\emptyset'$-random (ie random relative to the halting problem) then it is already computationally weak.
- Recent work by Bienvenu and others look at adding statements asserting certain strings are random to logical systems. Again, as expected, this is not a way around the incompleteness phenomenon. (Except in the resource bounded case.)
- Of course, there is always the Chaitin-Muchnik incompleteness theorem.

Theorem (Chaitin, Muchnik)

There exists a constant d (which only depends on the particular axiomatic system S and the choice of description language) such that there does not exist a string σ for which the statement $K(\sigma) \geq d$ (as formalized in S) can be proven within the axiomatic system S.

- Lots more work relating (calibrations of) randomness and fixed point free functions. For example.

Theorem (Barmpalias, Lewis, and Ng)

Each PA degree is the join of two randoms.

- This theme had realizations as to aligning randomness with weaker notions of computing fixed point free functions, and things like $K(X \upharpoonright h(n)) \geq n$. and "autocomplex" degrees.
- That is somewhat random in terms of initial segment complexity is fixed point free.

Calibrating randomness

- When is X more random than Y? When is X somewhat random?
- One way is to vary the tests or gales. Stronger tests mean stronger randomness.
- Examples : Schnorr randomness (means that $\mu(V_e) = 2^{-e}$), computable randomness (means that computable martingales).
- Intricate dance with Turing degrees, Sample theorem: if $\mathbf{a}$ is not computationally powerful in terms of its jump ($\mathbf{a}$ is not high) (Nies, Stephan, Terwijn) then in $\mathbf{a}$ these randomness notions all coincide. That is A is MLR iff Schnorr random iff computably random.
- Varying oracles. $n+1$-randomness equals randomness relative to $\emptyset^{(n)}$. (Miller-Yu) if $A \leq_T B$ are random and B is n-random, so is A.

- Many reducibilities and measures of relative randomness. Eg $Y \leq_K X$ means $K(Y \upharpoonright n) \leq K(X \upharpoonright n) + c$ for all n. $Y \leq_{LR} X$ means every real Y can derandomize X can also.
- Sample theorem. $\Omega = \sum_{U(\sigma)\downarrow} 2^{-|\sigma|}$ is Chaitins' Omega. Seems to depend on the machine, but in the same way as for the halting problem.

Theorem (Slaman-Kučera)

A left-c.e. real (the halting probability of a prefix free machine) is random iff it is Solovay complete.

- $A \leq_S B$ roughly means that effectively approximating B allows us to B-tightly effectively approximate A.
- Another: $\emptyset^{(n)}$-randomness is definable in terms of K. (Bienvenu, Muchnick, Shen, Vereshchagin)

Effective Dimensions

- Fractional dimension: Caratheordory, Hausdorff etc.
- Based on the idea of weakening Lebesgue measure.
- We know that if we are not random, then there is a martingale that wins on A. Let's make it harder to win, measuring "how close" the real is to being random.
- (Lutz) An s-gale is a function $F : 2^{<\omega} \mapsto \mathbb{R}$ such that

$$F(\sigma) = 2^{-s}(F(\sigma 0) + F(\sigma 1)).$$

  (Here $0 \leq s \leq 1$, $s = 1$ corresponds to Lebesgue measure.)
- The basic idea here is that not betting on one outcome or the other is bad.
- Usually, decide that we are not prepared to favour one side or the other in our bet. Thus we make $F(\sigma i) = F(\sigma)$ at some node σ.In the case of an s-gale, then we will be unable to do this, without automatically losing money due to inflation.

- Lutz has shown that effective Hausdorff dimension can be characterized using these notions.
- It is not important exactly what the definition is but we get the following.
- (Lutz, Hitchcock) For a class X the following are equivalent:
  (i) $\dim(X) = s$.
  (ii) $s = \inf\{s \in \mathbb{Q} : X \subseteq S[d]$ for some s-gale $F\}$.
  (iii) $s = \inf\{s \in \mathbb{Q} : X \subseteq S_{2^{(1-s)n}}[d]$ for some martingale $d\}$.
- An equivalent characterization due to Lutz is $\liminf_{n\to\infty} \frac{K(X\upharpoonright n)}{n}$.

- Lutz comment:
- "Informally speaking, the above theorem says the the dimension of a set is the most hostile environment (i.e. most unfavorable payoff schedule, i.e. the infimum s) in which a single betting strategy can achieve infinite winnings on every element of the set."
- While Schnorr did not do any of this, he did look at exponential orders. He comments:
- "To our opinion the important statistical laws correspond to null sets with fast growing orders. Here the exponentially growing orders are of special significance."

Themes

- Can be used for aperiodic tiling (Levin, Shen, Vereshchagin etc)
- Can have for all m, n, $K(X[m, m+n])$ is high (That is $\frac{K(X[m,m+n])}{n} \geq 1 - \epsilon$).
- Simpson recently used effective dimension for new results in Symbolic dynamics namely, classical dimension equals the entropy (generalizing a difficult result of Furstenburg 1967).
- Very close relationship between ergodic theory and randomness e.g.

Theorem (Hochman and Meyerovitch)

The values of entropies of subshifts of finite type over $\mathbb{Z}^d$ for $d \geq 2$ are exactly the complements of halting probabilities.

- Lutz, Mayordomo and others: use resource bounded versions to measure things like NP.

- The easiest way to make something of Hausdorff dimension $\frac{1}{2}$ is to take a random and "thin it."
- Is this the only way?

Theorem (Joe Miller)

There is a real X of (effective) Hausdorff dimension $\frac{1}{2}$ such that every $Y \equiv_T X$ has Hausdorff dimension $\leq \frac{1}{2}$.

- Extracting randomness is hard.
- However, with two independent sources, it is possible to get a Y computable from both of them to within ϵ of random (Zimand).
- 0,1 law for effective packing dimension.
- (Mayordomo) Packing dimension $\limsup_{n\to\infty} \frac{K(X\upharpoonright n)}{n}$.

Speculations

- Use randomness for understanding quantum physics.
- Can already buy it over the counter (Quandis) (see Calude and Svozil).
- Program is to figure out what is needed to make physics work.
- Is the universe granular? Is computability emergent?
- Can the universe manufacture randomness, computability, incomputability etc?
- Also left out applications in biology, music, etc.

Normality again-Turing

- Borel 1909. X is normal base b if the base b expansion has for all $0 \leq i \leq b-1$
$$\lim_{n\to\infty} \frac{|\{j \leq n \mid X(j) = i\}|}{n} = \frac{1}{b}.$$
- Absolutely normal if it is normal to all bases $b \geq 2$.

Theorem (Borel)

Almost all numbers are absolutely normal.

- *Question*(Borel) Give an explicit construction of an absolutely normal number.
- Sierpinski and Lebesgue 1917 gave an intricate limiting construction of absolutely normal numbers. Unclear whether the construction is effective.
- *Question*(Steihhaus) Can a number be normal to one base and not to another?
- Yes: Cassels 1959, Schnidt 1960.

Examples?

- Widely conjectured $\pi, e, \sqrt{2}, \zeta(3)$, etc.
- every irrational algebraic number
- Far from what is provable. None of these have been proven to be normal to any base. (But Bailey and Crandall 2001, via an unproven hypothesis.)
- What about explicit examples
- We can easily see that any Martin-Löf random real will be...

Turing

> *Although it is known that almost all numbers are [absolutely] normal no example of [an absolutely] normal number has ever been given. I propose to show how [absolutely] normal numbers may be constructed and to prove that almost all numbers are [absolutely] normal constructively.*

Notice how Turing anticipates the computational paradigm for algorithmic randomness.

- What Turing did was to show that almost all numbers are absolutely normal constructively, and then
- Derive the computable construction of such a number trivially.
- As analysed by Figueira, Becher and Picci (2007) that Turing's unpublished note shows is that The set of non-normal numbers has computable measure 0, and hence, for instance, all computably random reals will be absolutely normal.

Turing's vision

- Jack Lutz, lecture of CCR 2012:

  > *Placing computability constraints on a nonconstructive theory like Lebesgue measure seems a priori to weaken the theory, but it may strengthen the theory for some purposes*
  > *This vision is crucial for present-day investigations of*
  > - *individual random sequences,*
  > - *dimensions of individual sequences,*
  > - *measure and category in complexity classes, etc.*

- Turing's work is 75 years old.

From a modern perspective

- Use polynomial martingales, rather than computable ones to get computable examples.

Theorem (Strauss-97)

Almost every polynomial time computable real is absolutely normal.

Theorem (Lutz and Mayordomo, Becher and Slaman, Figueira and Nies)

Absolutely normal numbers can be constructed in time $O(n^2)$.

Finite state compressors

- All of this is related to finite state dimension. I don't have time, but the idea going back to Schnorr and Stimm is to use finite state gamblers for the martingales. (See the work of Lutz and Mayordomo for more.

Theorem (Schnorr and Stimm 1972; Bourke, Hitchcock, and Vinodchandran 2005)

. *A real X is normal in base b if and only if $dim^b_{FS}(X) = 1$.*

- For more on this and relationship with additive number theory, look at Jack Lutz' lecture in CCR 2012 (online from the Newton Institute).

- For example the Champernowne number 01234567891011.. is normal base 10. (1933).
- The Copeland-Erdos number $C_b(A)$ is the infinite contatenation of elements of A in order.

Theorem (Copeland and Erdos, 1946)

If A is sufficiently dense base b, namely

$$\liminf_{n\to\infty} \frac{\log |A \cap \{1, \ldots, n\}|}{\log n} = 1,$$

(e.g the primes) then $C_b(A)$ is normal base b.

- Interestingly there has been a lot of recent work, after Demuth, looking at the relationship between almost everywhere differentiation of (computable) functions, levels of randomness, and things like the Lebesgue Theorem on monotone functions. I am sure Andre will talk on this.
- And work on Ergodic Theorems classifying the randomness needed for them.
- Terry Tao blog suggestes that there is a deep connection.

Turing musings

- Turing had the elements to be able to develop Martn-Löf randomness but did not.
- Turing understood the computational uses of algorithmic pseudo-randomness. For example: (1950)

  > *"It is probably wise to include a random element in a learning machine.... A random element is rather useful when searching for the solution of some problem."*

- Turing 1950 gives an example of search for the solution to some numerical problem, pointing out that if we do this systematically, we will often have a lot of overhead corresponding to our previous searches.
- He says if problem has solutions reasonably densely in the sample space random methods should succeed.
- Our example Polynomial Identity Testing.
- Turing (1951) even speculates that randomness is necessary for intelligence.

Want to learn more?

- Calibrating randomness (BSL) Downey, Hirschfeldt, Nies Terwijn.
- Computability and Randomness, Nies OUP
- V. Becher, Turings note on normal numbers, 2012 CIE
- V. Becher, S. Figueira, and R. Picchi, Turings unpublished algorithm for normal numbers, Theoretical Computer Science (2007).
- Algorithmic randomness and complexity, Downey and Hirchfeldt.

Thank You