A short walk into randomness

Silvio Capobianco[1]

[1]Institute of Cybernetics at TUT

Institute of Cybernetics at TUT
October 18, 2012

Revision: October 25, 2012

Introduction

- Classical probability theory is concerned with randomness of selections of specific items from given sets.
- But it cannot express the notion of randomness of single objects.
- In the case of strings, this is done by algorithmic information theory, originated independently by Andrei Kolmogorov, Gregory Chaitin, and Ray Solomonoff.
- A very nice contribution comes from Per Martin-Löf.
- An approach by Peter Hertling and Klaus Weihrauch allows extension to more general cases.

What is randomness?

$$00000000000000000000000000000000\ldots$$

$$01010101010101010101010101010101\ldots$$

$$01000110110000010100111001011101\ldots$$

$$00110110101101011000010110101111\ldots$$

Disclaimer

Any one who considers arithmetic methods of producing random digits is, of course, in a state of sin. For, as has been pointed out several times, there is no such thing as a random number—there are only methods to produce random numbers, and a strict arithmetical procedure is of course not such a method.

John von Neumann

von Mises' definition

Given an infinite binary sequence $a = a_0 a_1 a_2 \ldots$, we will say that a is random if the following two conditions are satisfied:

1. The following limit exists:

$$\lim_{n\to\infty} \frac{\{i < n \mid a_i = 1\}}{n} = p$$

2. For every admissible place selection rule $\phi : \{0,1\}^* \to \{0,1\}$, chosen to select those indices for which $\phi(a_0 \ldots a_{n-1}) = 1$, we also have

$$\lim_{n\to\infty} \frac{\{i < n \mid a_{n_i} = 1\}}{n} = p$$

But what is "admissible" supposed to mean?

Notation

Let A be a Q-ary alphabet.

- A^n is the set of strings or words of length n over A. $A^* = \bigcup_{n \geq 0} A^n$.
  For $n = 0$ we set $A^0 = \{\lambda\}$ where λ is the empty string.
  For $i \geq 1$ and $j \leq |x|$ we set $x_{[i..j]} = x_i x_{i+1} \dots x_{j-1} x_j$.
- A^ω is the set of sequences or infinite words.
  We have indices start from 1, so $x = x_1 x_2 \dots x_n \dots$
- The product topology on A^ω has a subbase formed by the cylinders $wA^\omega = \{x \in A^\omega \mid x_{[1..|w|]} = w\}$
- The product measure μ_Π is defined on the Borel σ-algebra generated by the cylinders as the unique extension of $\mu_\Pi(wA^\omega) = Q^{-|w|}$
- The prefix encoding of $x = x_1 x_2 \dots x_n$ is $\overline{x} = 0x_1 0x_2 \dots 0x_n 1$
- $\mathtt{str} : \mathbb{N} \to A^*$ is the Smullyan encoding of n as a Q-ary string, *e.g.*, $0 \to \lambda$, $1 \to 0$, $2 \to 1$, $3 \to 00$, $4 \to 01$, etc.
- $\langle \cdot, \cdot \rangle : A^* \times A^* \to A^*$ is a pairing function for strings.

Computers

A computer is a partial function

$$\phi : A^* \times A^* \to A^*$$

$\phi(u, y)$ is the output of the computer ϕ with program u and input y.

A computer is prefix-free, or a Chaitin computer if, for every $w \in A^*$, the function

$$C_w(x) = \phi(x, w)$$

has a prefix-free domain.
This reflects the idea of self-delimiting computations: the length of a program is embedded in the program itself.

The Invariance Theorem

There exists a (prefix-free) computer Φ with the following property:

for every (prefix-free) computer ϕ there exists a constant c such that, if $\phi(x, w)$ is defined, then there exists $x' \in A^*$ such that $\Phi(x', w) = \phi(x, w)$ and $|x'| \leq |x| + c$.

Such computers are called universal.

For the rest of this talk we fix a universal computer ψ and a universal Chaitin computer U.

Kolmogorov complexity

The Kolmogorov complexity of $x \in A^*$ conditional to $y \in A^*$ associated with the computer ϕ on the alphabet Q is the partial function $K_\phi : A^* \times A^* \to \mathbb{N}$ defined by

$$K_\phi(x \mid y) = \min\{n \in \mathbb{N} \mid \exists u \in A^n \mid \phi(u, y) = x\}$$

If ϕ is a Chaitin computer we speak of prefix(-free) Kolmogorov complexity and write H_ϕ instead of K_ϕ.

- If $y = \lambda$ is the empty string we write $K_\phi(x)$ and $H_\phi(x)$.
- We omit ϕ if $\phi = \psi$ (complexity) or $\phi = U$ (prefix complexity).
- The canonical program of a string x is the smallest string (in lexicographic order) x^* such that $U(x^*) = x$.
- The invariance theorem ensures that $|x^*|$ is defined up to $O(1)$.

Basic estimates

$K(x) \leq |x| + O(1)$

- Consider the computer $\phi(u, y) = u$.

$H(x) \leq |x| + 2\log|x| + O(1)$.

- Consider the Chaitin computer $C(\overline{u}, y) = u$.

If $f : A^* \to A^*$ is a computable bijection then $H(f(x)) = H(x) + O(1)$.

- Consider the Chaitin computer $C(x) = f(U(x))$.
- In particular, $H(\langle x, y\rangle) = H(\langle y, x\rangle) + O(1)$.

For fixed y, $K(x|y) \leq K(x) + O(1)$ and $H(x|y) \leq H(x) + O(1)$.

- Consider the Chaitin computer $C(u, y) = U(u, \lambda)$.

There are less than $Q^{n-t}/(Q-1)$ strings of length n with $K(x) < n - t$.

- There are $(Q^{n-t} - 1)/(Q - 1)$ Q-ary strings of length $< n - t$.

Kolmogorov complexity is not computable!

The set $CP = \{x^* \mid x \in A^*\}$ of canonical programs is immune, *i.e.*, it is infinite and has no infinite recursively enumerable subset.

- For every infinite r.e. S there exists a total computable g s.t. $S' = g(\mathbb{N}_+) \subseteq S$, and if $g(i) \in CP$ then $i - c \leq 3\log i + k$ for suitable constants c, k.

The function $f : A^* \to A^*$, $f(x) = x^*$ is not computable.

- The range of f is precisely CP.

The prefix Kolmogorov complexity H is not computable.

- If $H|_{\mathrm{dom}\,\phi} = \phi$ for some partial recursive $\phi : A^* \to \mathbb{N}$ with infinite domain, then we might construct recursive $B \subseteq \mathrm{dom}\phi$ s.t. $f(0^i1) = \min\{x \in B \mid H(x) \geq Q^i\}$ satisfies $Q^i \leq H(f(0^i1))$ i.o.

However, H is semicomputable from above.

- $H(x) < n$ if and only if, for suitable y and t, $|y| < n$ and $U(y, \lambda) = x$ in at most t steps.

Randomness according to Chaitin

For $n \geq 0$ let

$$\Sigma(n) = \max_{x \in A^n} H(x) = n + H(\mathtt{str}(n)) + O(1)$$

We say that x is Chaitin m-random if $H(x) \geq \Sigma(|x|) - m$.
For $m = 0$ we say that x is Chaitin random.

Chaitin random strings are those with maximal prefix Kolmogorov complexity for their own length.

Call RAND_m^C the set of Chaitin m-random strings. Omit m if $m = 0$.

Theorem. For a suitable constant $c > 0$,

$$\gamma(n) = |\{x \in A^n \mid H(x) = \Sigma(n)\}| \geq Q^{n-c} \quad \forall n \in \mathbb{N}$$

Relating H with K

For all $x \in A^*$ and $t \geq 0$, if $K(x) < |x| - t$ then

$$H(x) < |x| + H(\mathtt{str}(|x|)) - t + O(\log_Q t)$$

- As K is upper semicomputable, given n and t, we only need $n - t$ Q-ary digits to extract $x \in A^n$ with $K(x) < n - t$.
- But there are at most $Q^{n-t}/(Q-1)$ such strings, and those also satisfy

$$H(x \mid \langle \mathtt{str}(n), \mathtt{str}(t) \rangle) < n - t + O(1)$$

- Then

$$\begin{aligned} H(x) &< n - t + H(\langle \mathtt{str}(n), \mathtt{str}(t) \rangle) + O(1) \\ &< n - t + H(\mathtt{str}(n)) + O(\log_Q t) \end{aligned}$$

As a consequence,

for every $x \in \mathrm{RAND}_t^C$ and every T s.t. $T - O(\log_Q T) \geq t$
one has $K(x) < |x| - T$

Martin-Löf tests

A Martin-Löf test is a recursively enumerable set $V \subseteq A^* \times \mathbb{N}_+$ such that:

1. The level sets $V_m = \{x \in A^* \mid (x, m) \in V\}$ form a nonincreasing sequence, *i.e.*, $V_{m+1} \subseteq V_m$ for every $m \geq 1$.
2. For every $n \geq m \geq 1$, $|A^n \cap V_m| \leq Q^{n-m}/(Q-1)$.

We say that $x \in A^n$ passes V at level $m < n$ if $x \notin V_m$.

If ϕ is a (not necessarily prefix-free!) computer, then

$$V = V(\phi) = \{(x, m) \mid K_\phi(x) < |x| - m\}$$

is a Martin-Löf test. Such tests are called representable.

A non-representable test

Let $x_0, x_1, x_2 \in \{0,1\}^3$ and $V = \{(x_0, 1), (x_1, 1), (x_2, 1)\}$.

- By contradiction, assume $V = V(\phi)$.
- Then there exist $y_0, y_1, y_2 \in \{0,1\}^*$ s.t. $|y_i| \leq 1$ and $\phi(y_i) = x_i$.
- Then necessarily $\{y_0, y_1, y_2\} = \{\lambda, 0, 1\}$.
- But then, $K_\phi(\phi(\lambda)) = 0 < 1 = |\phi(\lambda)| - 2$.
- Then $(\phi(\lambda), 2) \in V(\phi)$—contradiction.

Critical levels

The critical level function of a M-L test V is

$$m_V(x) = \begin{cases} \max\{m \mid x \in V_m\}, & \text{if } x \in V_1\,, \\ 0\,, & \text{otherwise}\,. \end{cases}$$

If $x \neq V_q$ for some $q < |x|$ we say that x is q-random.
If, in addition, $V = V(\phi)$ is representable, then:

- If $m_V(x) > 0$ then $m_V(x) = |x| - K_\phi(x) - 1$.
- $m_V(x) = 0$ if and only if $K_\phi(x) \geq |x| - 1$.

On the other hand, if

- $|A^n \cap V_m| \leq Q^{n-m-1}$ for every $n \geq m \geq 1$, and
- there is at most one $(x, m) \in V$ with $|x| = m + 1$,

then V is representable.

Universal Martin-Löf tests

A M-L test $\mathcal{U}$ is universal if for every M-L test V there exists a constant c such that

$$V_{m+c} \subseteq \mathcal{U}_m \;\; \forall m \geq 1$$

that is, if $\mathcal{U}$ refines all M-L tests at once.

For a computer ψ the following are equivalent:

1. ψ is a universal computer.
2. For every M-L test V there exists a constant c s.t.

$$m_V(x) \leq |x| - K_\psi(x) + c \;\; \forall x \in A^*$$

3. $V(\psi)$ is a universal M-L test and in addition there exists c s.t.

$$K_\psi(x) \leq |x| + c \;\; \forall x \in A^*$$

Martin-Löf asymptotic formula

Let ψ be a universal computer and let $\mathcal{U}$ be a universal M-L test. Then there exists a constant $c = c(\psi, \mathcal{U})$ such that

$$\left|\,|x| - K_\psi(x) - m_{\mathcal{U}}(x)\,\right| \leq c \quad \forall x \in A^*$$

As a consequence,

for fixed $t \geq 0$,
almost all $x \in \mathrm{RAND}_t^C$ are declared eventually random
by every Martin-Löf test V

Randomness for sequences

An intuitive definition might be:

a sequence is random if and only if all its finite prefixes are

However:

- Given $x \in \{0,1\}^\omega$ and $n \in \mathbb{N}$, let $N_0(x;n)$ be the numbers of consecutive 0s from position n.
- It is well known that $\limsup_{n\to\infty} N_0(x;n)/\log_2 n = 1$ for almost all x.
- Thus, for almost all x there are infinitely many n s.t. $x_{[1..n]} = x_{[1..n-\log_2 n]} 0^{\log_2 n}$.
- For those n we have $K(x_{[1..n]}) \approx n - \log_2 n$.

As a side effect,

there is no such thing as a random string
in the sense stated above

Testing sequentially

A Martin-Löf test V is sequential if it satisfies the following property:

$$\forall m \geq 1 \, \forall x, y \in A^* \; : \;\; x \in V_m, y \in xA^* \Rightarrow y \in V_m$$

- The family of sequential M-L tests is r.e.
- There exists a universal sequential M-L test U such that, for every sequential M-L test V, there exists a constant $c = c(V)$ such that $V_{m+c} \subseteq U_m$ for every $m \geq 1$.
- A sequential M-L test U is universal if and only if, for every sequential M-L test V, there exists a constant $c = c(V)$ such that $m_V(x) \leq m_U(x) + c$ for every $x \in A^*$.
- If U and W are universal sequential M-L tests, then for every $x \in A^*$

$$\lim_{n \to \infty} m_U(x_{[1..n]}) < \infty \Leftrightarrow \lim_{n \to \infty} m_W(x_{[1..n]}) < \infty$$

Randomness for sequences

We say that $x \in A^\omega$ fails a sequential M-L test V if

$$x \in \bigcap_{m \geq 1} V_m A^\omega$$

This is actually equivalent to saying that

$$\lim_{n \to \infty} m_U(x_{[1..n]}) = \infty$$

We call $\mathbf{rand}(V)$ the set of sequences that do **not** fail V. Then

$$\mathbf{rand} = \bigcap_{V \text{ sequential}} \mathbf{rand}(V) = \mathbf{rand}(U)$$

Characterizations of $\mathbf{rand}$

- $A^\omega \setminus \mathbf{rand}$ is the union of all the constructible μ_Π-null subsets of A^ω. (Observe that non-random sequences are those that fail the universal test.)
- $x \in \mathbf{rand}$ if and only if, for every r.e. $C \subseteq A^* \times \mathbb{N}_+$ such that $\mu_\Pi(C_j A^\omega) < Q^{-j}/(Q-1)$ for all $j \geq 1$, there exists $i \geq 1$ s.t. $x \notin C_i A^\omega$.
  (This is because such C's can easily be turned into M-L tests.)
- **Chaitin:** $x \in \mathbf{rand}$ if and only if there exists $c > 0$ s.t. $H(x_{[1..n]}) \geq n - c$ for every $n \geq 1$.
- **Solovay:** $x \in \mathbf{rand}$ if and only if, for every r.e. $X \subseteq A^* \times \mathbb{N}_+$ such that $\sum_{i \geq 1} \mu_\Pi(X_i A^\omega) < \infty$, there exists $N \in \mathbb{N}$ s.t $x \notin X_i A^\omega$ for every $i > N$.
- **Chaitin:** $x \in \mathbf{rand}$ iff $\lim_{n \to \infty}(H(x_{[1..n]}) - n) = \infty$.
- If $\phi : \mathbb{N} \to \mathbb{N}$ is a computable bijection, then $x \in \mathbf{rand}$ if and only if $x \circ \phi \in \mathbf{rand}$.

Is there a simpler characterization?

Martin-Löf theory formalizes the intuitive concept:

a random sequence passes all computable statistical tests

We ask if we can say something as such:

a random sequence satisfies every property
which is true for μ_Π-almost every string

However:

- Given $x \in A^\omega$, say that $y \in A^\omega$ satisfies $P(x)$ if for every $n \geq 1$ there exists $m \geq n$ such that $y_i \neq x_i$.
- Then $P(x)$ is satisfied by μ_Π-almost all $y \in A^\omega$, but not by x.

Once again: there ain't no such thing as a free lunch.

Normal sequences

Given $x \in A^\omega$ and $w \in A^* \cup A^n$, set

$$\mathrm{occ}(w, x) = \{i \geq 1 \mid x_{[i..i+n-1]} = w\}$$

We say that x is n-normal if

$$\lim_{i \to \infty} \frac{|\mathrm{occ}(w, x) \cap [1, i]|}{i} = \frac{1}{Q^n} \;\; \forall w \in A^n$$

A string which is n-normal for every $n \geq 1$ is said to be normal.

Observe that n-normality is the same as

$$\liminf_{i \to \infty} \frac{|\mathrm{occ}(w, x) \cap [1, i]|}{i} \geq \frac{1}{Q^n} \;\; \forall w \in A^n$$

Random sequences are 1-normal

By contradiction, suppose $\liminf_i |\mathrm{occ}(a, x) \cap [1, i]|/i < Q^{-1} - k^{-1}$.

- Then, for infinitely many values of j, $x \in S_i A^\omega$ where

$$S = \left\{ (y, i) \mid y \in A^i, \frac{|\mathrm{occ}(a, y) \cap [1, i]|}{i} < \frac{1}{Q} - \frac{1}{k} \right\}$$

- The random variables $Y_j = [y_j = a]$ are independent, and

$$S_i A^\omega = \left\{ \sum_{j=1}^{i} Y_j < \frac{i}{Q} \left(1 - \frac{Q}{k} \right) \right\}$$

- By the Chernoff bound, $\mu_\Pi(S_i A^\omega) < e^{-\frac{Q}{k^2} i}$.
- By Solovay's criterion, $x \notin \mathbf{rand}$.

. . . in fact, random sequences are normal *tout court*

Given $n \geq 1$ and $x \in A^\omega$, define $x^{(n)} \in (A^n)^\omega$ by

$$x_i^{(n)} = x_{(i-1)n+1} x_{(i-1)n+2} \cdots x_{in}$$

Then $x \in \mathbf{rand}$ if and only if $x^{(n)} \in \mathbf{rand}$.
The thesis then follows from the following theorem by Niven and Zuckerman:

$$x \text{ is } n\text{-normal if and only if } x^{(n)} \text{ is 1-normal}$$

General randomness spaces

A randomness space is a triple (X, B, μ) where:

- X is a topological space (*e.g.*, A^ω).
- B is a total numbering of a subbase for X (*e.g.*, $B_i = w_i A^\omega$).
- μ is a probability measure on the Borel σ-algebra of X (*e.g.*, μ_Π).

Given two sequences $V = \{V_n\}_{n \geq 0}$, $W = \{W_m\}_{m \geq 0}$ of open subsets of X, we say that V is W-computable if there exists a r.e. $A \subseteq \mathbb{N}$ such that

$$V_n = \bigcup_{\pi(n,m) \in A} W_m \;\; \forall n \geq 0 \,,$$

where $\pi(x, y) = (x + y)(x + y + 1)/2 + x$ is the standard pairing function for natural numbers.

We define $D : \mathbb{N} \to \mathcal{PF}(\mathbb{N})$ as the inverse of $E : \mathcal{PF}(\mathbb{N}) \to \mathbb{N}$ defined by

$$E(S) = \sum_{i \in S} 2^i$$

Given $V = \{V_n\}$ we define $V' = \{V'_n\}$ as $V'_n = \bigcap_{m \in D(n+1)} V_n$.

A general framework for randomness

Let (X, B, μ) be a randomness space.

- A randomness test on X is a B'-computable family $V = \{V_n\}$ of open subsets of X such that $\mu(V_n) < 2^{-n}$ for every $n \geq 0$.
- An object $x \in X$ fails a randomness test V if $x \in \bigcap_{n \geq 0} V_n$.
- $x \in X$ is random if it does not fail any randomness test on X.

Theorem. (Hertling and Weihrauch)
Let $x \in A^\omega$ and let $B_i = \mathtt{str}(i) A^\omega$. The following are equivalent.

1. $x \in \mathbf{rand}$.
2. x is random as an element of the randomness space (A^ω, B, μ_Π).

An application to cellular automata theory

Let G be a discrete group and let $\phi : \mathbb{N} \to G$ be a computable bijection such that $m : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ satisfying $\phi(m(i,j)) = \phi(i) \cdot \phi(j)$ for every i and j is a computable function. Let A be a Q-ary alphabet.

- Set the product topology on A^G.
- Define $B : \mathbb{N} \to A^G$ as $B_{Qi+j} = \{c : G \to A \mid c(\phi(i)) = a_j\}$.
- Define the product measure on A^G as the only probability measure μ_Π that extends $\mu_\Pi(\{c(g) = a\}) = Q^{-1}$ to the Borel σ-algebra.

Then (A^G, B, μ_Π) is a randomness space.

- In addition, $c \in A^G$ is random if and only if $c \circ \phi \in \mathbf{rand}$.
- Thus, the notion of randomness does not depend on the choice of ϕ.

Theorem (Calude, Hertling, Jürgensen and Weihrauch, 2001)

Let F be the global law of a d-dimensional CA. The following are equivalent.

1. F is surjective.
2. $F(c)$ is random for every c which is itself random.

Conclusions

- Chaitin's approach to randomness: program-size complexity.
- Martin-Löf's approach: computable statistical tests.
- In some, very precise sense, there **is** such thing as a random number.

Thank you for attention!

Any questions?