算法随机性的前世今生

方楠

Regularity in Chaos: Probability

- Although a lot of processes seem to exhibit a rather chaotic behavior locally, many of them show statistical regularities when observed over a longer time.

- Probability theory tries to provide a mathematical framework for these regularities.

Interpretations of Probability

Richard von Mises

- Frequentist's (Von Mises) Approach:
  - Define probabilities in terms of frequencies, in fact, probabilities are relative frequencies.
  - This means in order to define probabilities, one has to define the objects from which the frequencies are derived.
  - We will study this approach with respect to infinite sequences of 0 and 1.

Probability = Global Regularities

- Let $\xi = \xi_0\xi_1\ldots$ be an infinite binary sequence.
- The *relative frequency* of "1" at n is defined as

$$F(\xi, n) = \frac{\xi_0 + \xi_1 + \cdots + \xi_{n-1}}{n}$$

- We expect from a random sequence of coin tosses that

$$\lim_{n\to\infty} F(\xi, n) = \frac{1}{2}$$

- If the limit exists, we call it the *limiting frequency* of 1.

Von Mises: Kollektivs

- Von Mises, *Grundlagen der Wahrscheinlichkeitsrechung* (1919)
- A Kollektiv has to satisfy two requirements:
  - (1) The limiting frequency exists. (global regularity)
  - (2) The limiting frequency persists for any subsequence selected by an admissible selection rule. (local irregularity)
- Question: What is an admissible selection rule?

Selection Rules

- A selection rule is a function that selects from a given sequence an infinite subsequence.
- Von Mises gives as examples of admissible rules
  - select every other bit
  - select all bits whose index is a prime number
- We can think of more complicated rules:
  - select all bits that follow a sequence of 25 ones.
  - Select all bits with index n such that up at n we have seen the same amount of zeros and ones.

Admissibility

- Von Mises in *Probability, Statistics, and Truth* (1936):

  "The only essential condition is that the question whether or not a certain member of the original sequence belongs to the selected subsequence should be settled ***independently of the result*** *of the corresponding observation, i.e., before anything is known about this result."*

- not admissible: select all indices n with $\xi_n = 1$.

- According to von Mises, if a sequence of events is truly random, then it does not exhibit any patterns that one could exploit to devise a successful gambling system against it.

Criticism

- Kamke [1932], Tornier [1933]:
  - The notion of Kollektiv is inconsistent.
  - Suppose ξ is a Kollektiv.
  - Consider the set S of all increasing functions on integers. This set can be formed independently of ξ.
  - However, one function in S selects a subsequence 1111111111.... from ξ.
- One might want to require admissible selection rules to be countable.

Ville's Theorem

- **Ville's Theorem (1937):** *Let S be any countable collection of selection functions. Then there is a sequence $\xi = \xi_0\xi_1\ldots$ such that the following hold.*

(i) $\lim_{n\to\infty} F(\xi, n) = \frac{1}{2}$

(ii) *For every $f \in S$ that selects infinitely many bits of ξ, we have*

$\lim_{n\to\infty} F(\xi_f, n) = \frac{1}{2}$, *here ξ_f is the subsequence selected by f.*

(iii) *For all n, we have* $F(\xi, n) \leq \frac{1}{2}$.

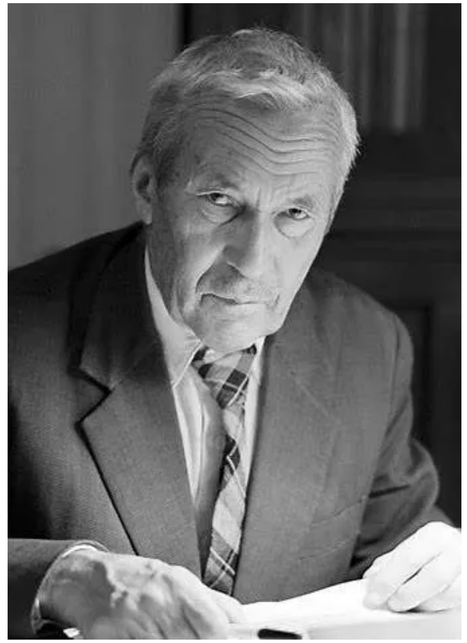
Kolmogorov's Theory

- In 1933, Kolmogorov gave his formulation of probability theory.
- The fundamental notion in Kolmogorov's theory is that of a **probability space**, which consists of
  - the sample space Q; the space of all events F (a σ-algebra of Q); a measure function μ on F
  - and should satisfy the following axioms:

    I. $\mu(E) \geq 0$ for all $E \in F$

    II. $\mu(Q) = 1$

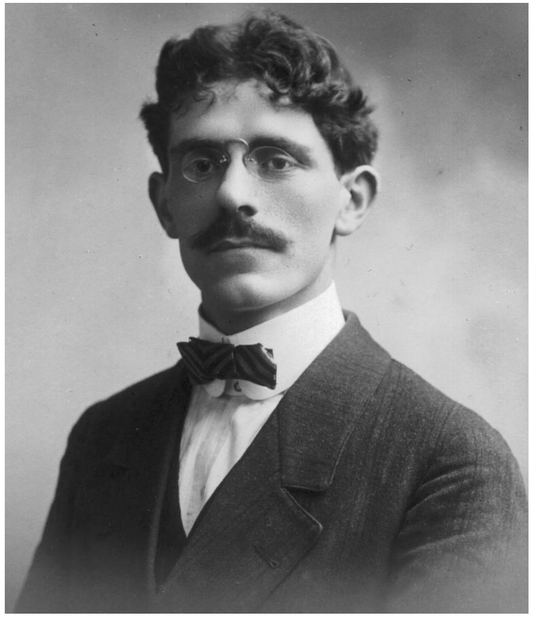
    III. if $\{A_n\}$ is a sequence of mutually disjoint sets in F, then $\mu(\bigcup_n A_n) = \sum \mu(A_n)$



Andrey Kolmogorov

The Geneva Conference

- In 1937, the University of Geneva hosted a conference on probability theory.
- Part of the conference was devoted to foundational questions, in particular a comparison of Kolmogorov's recently published axiomatization and von Mises' theory.
- Frechet gave a lecture that compiled a long list of objections to von Mises.
- The majority of researchers preferred Kolmogorov's approach.



Maurice Fréchet

Von Mises Revisited

- The Geneva conference, together with the subsequent publication of results by Ville, all but established the view that von Mises' approach had failed.
- Over the years, Kolmogorov's theory became the widely accepted framework that we know today.
- It was Kolmogorov himself that revived interest in von Mises' ideas.
- This revitalization would have been impossible without the dramatic developments in mathematical logic that had taken place since the 1930's.

Arithmetization and Computability

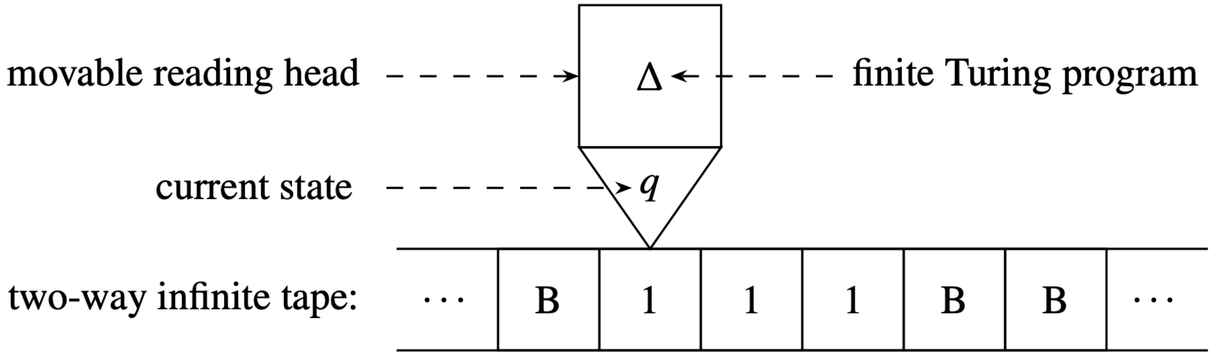
Kurt Gödel

Alan Turing

- In 1931, Gödel's famous proofs of the incompleteness theorems appeared. The critical feature of the proof is arithmetization, assigning a code number to every element of the syntax.
- Gödel showed that this coding can be done using primitive recursive functions. This idea was extended by others to cover all intuitively computable functions.
- In 1936, Turing gave another equivalent characterization, now known as Turing machines, differing greatly from the more mathematical models using definability in arithmetic.
- The equivalence of the various models indicated that an adequate formalization of the notion of computability had been found (Church-Turing Thesis).

Turing Machine

- A Turing machine M consists of
  - A two-way infinite tape divided into cells, with each cell marked as B or 1
  - A reading head which scans one cell of the tape at a time
  - A finite set of internal states
    $Q = \{q_0, q_1, \ldots, q_n\}, n \geq 1$
  - A finite set of instructions
    $\Delta = \{\delta_0, \delta_1, \ldots, \delta_m\}, m \geq 0$



Each instruction $\delta \in \Delta$ is a 5-tuple of the form $\delta = (q_i, s, q_j, s', D), q_i, q_j \in Q$where are states, $s, s' \in S = \{B,1\}$ are symbols, and $D \in \{L, R\}$ is a direction (left or right).

Computability as Descriptive Context

- The work of Gödel, Turing, and others paved the way for a profound analysis of the complexity of subsets of the natural numbers (arithmetical hierarchy).
- It suggests itself to adopt this context of logical analysis as a framework for admissible selection rules.
- Church (1940) proposed to admit only computable selection rules.

A New Approach: Kolmogorov Complexity

- In 1963, Kolmogorov came to the conclusion that the frequency interpretation needed a precise formulation after all.
- He wanted to show that the concept of a finite random sequence can be rigorously defined.
- His idea was impressively simple, elegant, yet profound.

Kolmogorov Complexity

- The problem: How should we define a finite random sequence?
- All finite strings of the same length n have the same probability $1/2^n$.
- Yet we would call 0000000000000...000 less random than 01000101011111010010...010
- Observation: the first sequence has a short algorithm, the second not.
- Basic idea: A sequence is random if it has no shorter algorithmic description than itself.

Kolmogorov Complexity: Formal Definition

- Let M be a Turing machine. M computes a partial recursive function mapping strings to strings.
- We define the M-complexity of a string x as

$$C_M(x) = \min\{\,|\sigma| : M(\sigma) = x\}$$

- The complexity of x depends on the choice of M. Can we choose M so that it reflects the "true" complexity of x?
- Theorem [Kolmogorov]: There exists a machine U such that for all other machines M,

$$C_U(x) \leq C_M(x) + e_M$$

  where e_M is a constant depending on M but not on x.
- The theory was also developed independently by Solomonoff.

Properties of Complexity

- The machine U is essentially a universal Turing machine, as constructed by Turing.
- Fix a universal Turing machine U, define the (plain) Kolmogorov Complexity function C as $C(x) = C_U(x)$.
- The function C is not computable. This follows from the undecidability of the halting problem, but can also be obtained as an application of *Berry's paradox*:
  - If C were computable, then the following would be a computable description of some string x:

    The shortest string that has no computable description of less than fourteen words.

  - However, we just gave a computable description of less than fourteen words!

Random Finite Strings

- Call a string x *random* (or *incompressible*) if $C(x) \geq |x|$.
- More generally, given a constant $c > 0$, we say x is *c-incompressible* if

$$C(x) \geq |x| - c.$$

- A simple counting argument shows that for each n, most strings of length n are rather incompressible:

$$|\{x : |x| = n \ \& \ C(x) \geq n - c\}| \geq 2^n(1 - 2^{-c}) + 1$$

- On the other hand, there exists a constant d (code of the “copy-machine”) such that for all x,

$$C(x) \leq |x| + d$$

Random Infinite Strings?

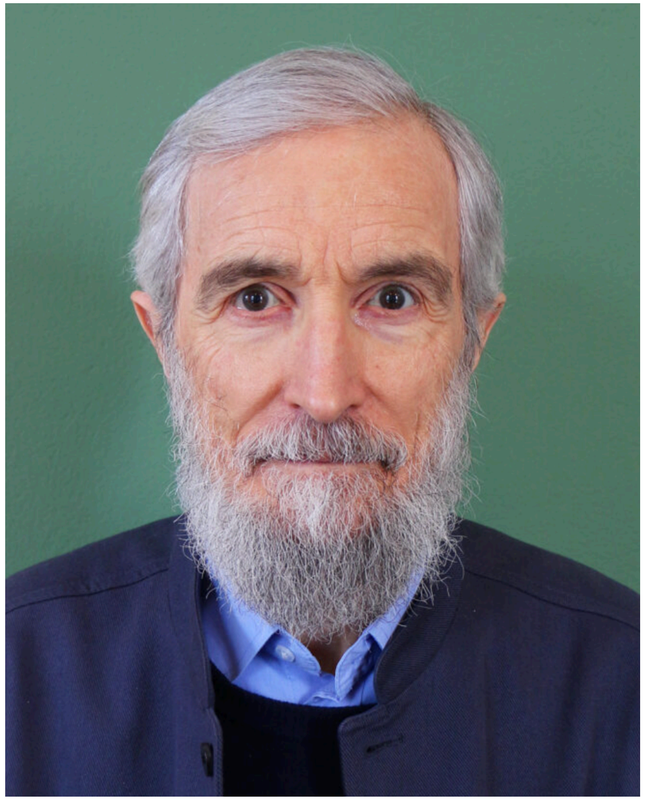
- Now let X be an infinite binary sequence. Let $X \upharpoonright n$ denotes the first n bits of X.
- One could define: X is random iff there exists a $c > 0$ such that

$$C(X \upharpoonright n) \geq n - c.$$

- Problem: such X do not exist. [Martin-Löf]
  - The reason for this that the length of a string can be used to code (compress) information, too.
  - For every $c > 0$, if a string σ is sufficiently long, then there is an initial segment τ of σ such that $C(\tau) < |\tau| - c$.

A New Approach to Infinite Sequences

- In 1965, Martin-Löf studied under Kolmogorov in Moscow.
- He was able to give a definition of randomness for infinite sequences that many consider as adequate.
- His approach combines Kolmogorov's framework of probability theory and computability.
- Instead of stability under selection rules, Martin-Löf's fundamental property is passing *effective statistical tests*.



Per Martin-Löf

Martin-Löf Randomness

- An *effective statistical test* is essentially a nullset (i.e. a set of probability zero) that can be represented effectively (i.e. is definable at a certain level of the arithmetical hierarchy).

- Formal definition: [Martin-Löf]

  - A *Martin-Löf test* is an computable algorithm M such that, for input n, outputs a set U_n of finite strings such that

$$\mu(U_n) \leq 2^{-n}.$$

  - A sequence X fails a Martin-Löf test $\{U_n\}$ if for every n there exists a string $\sigma \in U_n$ such that σ is an initial segment of X.

  - X is *Martin-Löf random* if it passes all Martin-Löf test.

Martin-Löf Randomness

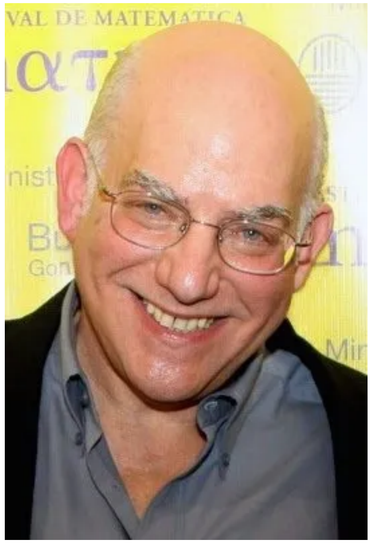
- Since there are only countably many Martin-Löf tests (there are only countably many Turing machines), almost every sequence is random.
- Most probabilistic laws, such as the Law of Large Numbers, can be captured by tests, that is, every Martin-Löf random sequence X satisfies

$$\lim_{n\to\infty} F(X,n) = \frac{1}{2}.$$

- On the other hand, no computable sequence can be random.

Prefix-Free Complexity

- In the 1970's, Kolmogorov's student Levin and independently Chaitin worked on a variation of Kolmogorov complexity to overcome the deficiencies of (plain) complexity noted by Martin-Löf.
- The result was the prefix-free complexity K.



Gregory Chaitin

Leonid Levin

Prefix-Free Complexity

- A Turing machine is *prefix-free* if no two halting inputs are prefixes of one another.
- There exists also a universal prefix-free machine.
- *Prefix-free Kolmogorov complexity* K is defined as the Kolmogorov complexity with respect to a universal prefix-free machine.
- X is defined as *1-random* iff there exists a $c > 0$ such that

$$K(X \upharpoonright n) \geq n - c.$$

Randomness by martingales

- A *supermartingale* is a function d maps finite strings to nonnegative reals, such that for all σ

$$d(\sigma) \geq \frac{d(\sigma 0) + d(\sigma 1)}{2}.$$

- It is a martingale if only "=" holds.

- A (super)martingale d succeeds on a set X if $\limsup_n d(A \upharpoonright n) = \infty$.

- This notion can be used to formulate Von Mises' original idea about the unpredictability aspect of randomness.

Randomness Notions Coincide

- In several papers appeared in 1971 and 1973, Schnorr proved the following are equivalent:

  i) X is 1-random.

  ii) X is Martin-Löf random.

  iii) There is no left-c.e. (super)martingale succeeds on X.

Claus P. Schnorr

- Here a function f is left-c.e. if there are a family of computable functions f_i such that $\lim_n f_n(x) = f(x)$ and $f_n(x) \leq f_{n+1}(x)$ for all n and x.

Chaitin's Constant

- Let U be a universal prefix-free machine.
- Let Ω denote the *halting probability* of U, defined as

$$\Omega = \sum_{U(\sigma)\downarrow} 2^{-|\sigma|}.$$

- Theorem [Chaitin]: Ω is Martin-Löf random.

A Hierarchy of Randomness

- While Martin-Löf randomness is very robust (equivalence with respect to different paradigms), it is not the only notion of randomness one can consider in this framework.

- By giving tests more (e.g. access to an undecidable problem) or less (e.g. finite automata) computational power, one obtains different versions of algorithmic randomness.

- Different modification of the martingales could also lead to different notions of randomness. (such as nonmonotonic randomness)