A Brief History of Randomness

Jan Reimann

Describing Chaos

Ovid, Metamorphoses:

“Before there was earth or sea or the sky that covers everything, Nature appeared the same throughout the whole world: what we call chaos: a raw confused mass, nothing but inert matter, badly combined discordant atoms of things, confused in the one place.”

The Fundamental Question

Can chaos be described? Defined?

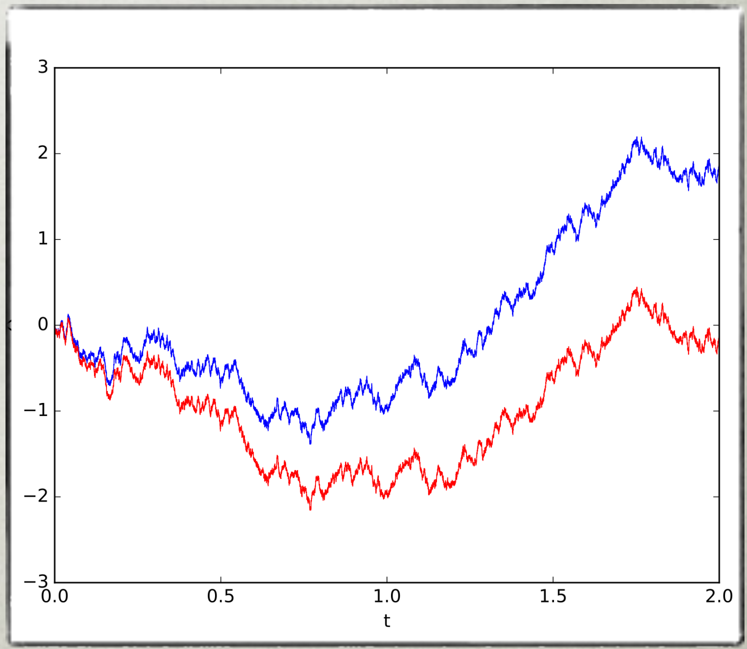
Does not language impose an order on things that opposes chaotic behavior?

As soon as you can define chaos, does it not cease to be true chaos?

- We will describe a certain aspect of these problems - the search for a mathematical definition of randomness.

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:
  - Rolling dice
  - Coin tosses
  - Brownian motion
- Probability theory tries to provide a mathematical framework for these regularities.



Interpretations of Probability

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

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

- We expect from a random sequence of coin tosses that

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

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

Von Mises: Kollektivs

Von Mises, *Grundlagen der Wahrscheinlichkeitsrechung* (1919)

- A Kollektiv has to satisfy two requirements:

  (1) The limiting frequency exists.

  (2) The limiting frequency persists for any subsequence selected by an *admissible selection rule*.

- 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.
- **not** admissible: select all indices n with $\xi_n = 1$.
- How about: toss a coin and select the n-th bit if the n-th toss is Heads?

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."*

In German: *"ohne Benützung der Merkmalunterschiede"*

Criticism

- Kamke [1932], Tornier [1933]:

  The notion of Kollektiv is **inconsistent**.

  - Suppose ξ is a Kollektiv.
  - Consider the set S of all increasing sequences of integers.
  - This set can be formed independently of ξ.
  - However, one sequence in S selects a subsequence 1111111111.... from ξ.

Kolmogorov’s Theory

In 1933, Kolmogorov gave his formulation of probability theory:

- *“The theory of probability, as a mathematical theory, can and should be developed from axioms... all further exposition must be based exclusively on these axioms, independent of the usual concrete meaning...”*

- *“...the concept of a* ***field of probabilities*** *is defined as a system of sets which satisfies certain conditions. What the elements of this set represent is of no importance in the purely mathematical development...”*

Kolmogorov’s Theory

The fundamental notion in Kolmogorov’s theory is that of a **probability space**, which consists of

- a set *E*,
- a non-negative measure function μ on a certain family of subsets of E (a σ-algebra) that satisfies $\mu(\emptyset) = 0$, $\mu(E) = 1$, and

  if (A_n) is a sequence of mutually disjoint sets, then

$$\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$$

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.

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.

Gödel’s Theroems

In 1931, Gödel’s famous proofs of the incompleteness theorems appeared.

The first incompleteness theorem states that any consistent and recursively axiomatizable theory *T* extending Peano arithmetic (more precisely, a weaker subsystem of PA) is incomplete, i.e. there exists a sentence ϕ in the language of arithmetic such that neither ϕ nor $\neg\phi$ is provable from *T.*

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.*

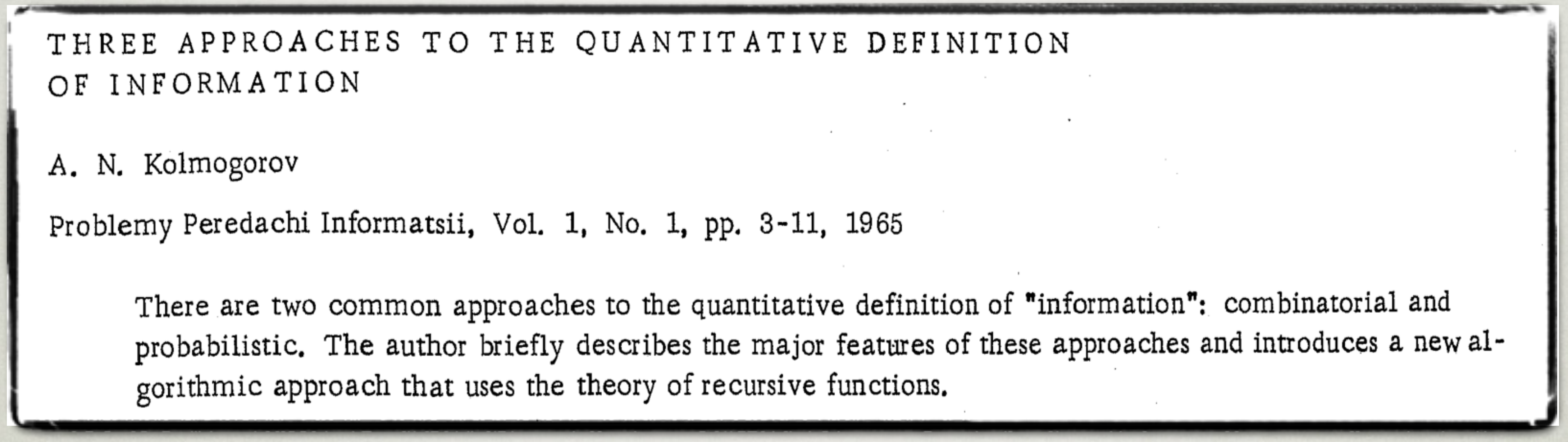
Arithmetic and Computability

- Primitive recursive functions are defined in terms of closure under simple arithmetic functions.
- However, they do not cover all intuitively computable functions (e.g. Ackermann function)
- Several extension have been proposed (Gödel-Herbrand, Kleene, Church).
- 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*).

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*).
- In particular, it showed that undecidable problems exist, e.g. the *halting problem*.
- 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



THREE APPROACHES TO THE QUANTITATIVE DEFINITION OF INFORMATION

A. N. Kolmogorov

Problemy Peredachi Informatsii, Vol. 1, No. 1, pp. 3-11, 1965

There are two common approaches to the quantitative definition of "information": combinatorial and probabilistic. The author briefly describes the major features of these approaches and introduces a new algorithmic approach that uses the theory of recursive functions.

- 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 sequences of the same length have the same probability, $1/2^{\text{length}}$.
- Yet we would call 0000000000000...000 less random than 0100010101111010010…010
- Basic idea: the first sequence has a short algorithm, the second not.
- Paradigm: 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

$$C_M(x) = \min\{\text{length}(\sigma) \colon M(\sigma) = x\}$$

- 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 developed independently by Solomonoff.

Properties of Complexity

- The machine U is essentially a universal Turing machine, as constructed by Turing.
- 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 description of less than thirteen words.

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

Random Finite Strings

- Call a string x *random* (or *incompressible)* if $C(x) = C_U(x) \geq \text{length}(x)$.
- *Notation*: $|x|$ denotes the length of 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 such that for all x,

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

(the *"copy-machine"*)

Random Infinite Strings?

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

$$C(X|_n) \geq n - c.$$

- *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, if a string x is sufficiently long, then there is an initial segment y of x such that $C(y) < |y| - c$.

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*.

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).

- *Definition:* A ***Martin-Löf test*** is an algorithm T such that, for input n, effectively lists a sequence $x_1^{(n)}, x_2^{(n)}, x_3^{(n)}, \ldots$ such that

$$\sum_i 2^{-|x_i|} \leq 2^{-n}\,.$$

- A sequence X is *covered by T* if for every n there exists an i such that

$$x_i^{(n)} \text{ is an initial segment of X.}$$

- X is ***Martin-Löf random*** if it is not covered by any test.

Martin-Löf Randomness

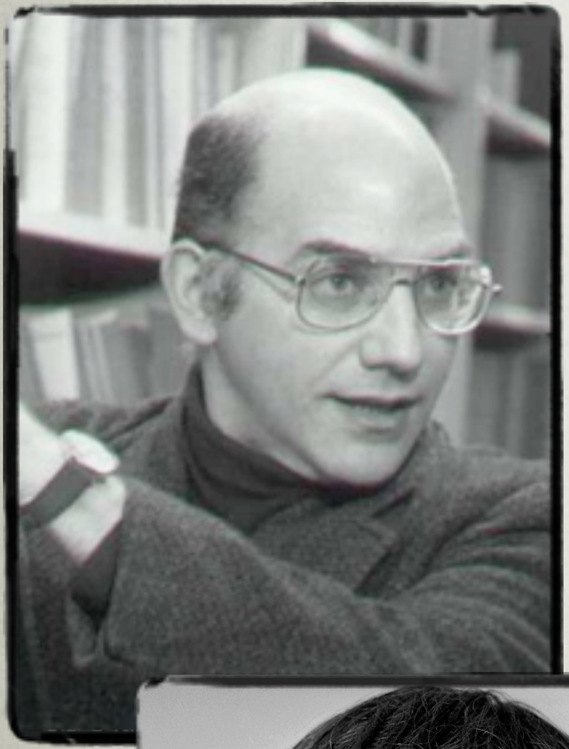
- Since there are only countably many Martin-Löf tests (there are only countably many algorithms), almost every sequence is random.

- Most probabilistic laws, such as the Law of Large Numbers, can be captured by tests, that is, *every* Martin-Löf random sequence X satisfies

$$\lim_{n\to\infty} \frac{\#\text{ 1's in } X|_n}{n} = \frac{1}{2}.$$

- On the other hand, no computable sequence can be random.
  - For example, π is not random in this framework. (It fails the test of “being π”.)

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.*

Prefix-Free Complexity

- A Turing machine is *prefix-free* if no two halting inputs are prefixes of one another (example: phone numbers)
- **Prefix-free complexity *K*** is defined as the Kolmogorov complexity with respect to a machine that is universal among prefix-free Turing machines.
- The resulting complexity measure is rather robust. In many ways, it resembles Shannon entropy, just that it applies to single strings instead of probability distributions over them:
  - $K(xy) \leq^{+} K(x) + K(y)$
  - If $\{X_n\}_{n \in \mathbb{N}}$ is an i.i.d. process of random variables over a finite alphabet, and we let $X^n = X_1 X_2 \ldots X_n$, then

$$\lim_{n \to \infty} \mathbb{E} \frac{1}{n} K(X^n) = H(X)$$

Randomness as Incompressibility

- In 1973, Schnorr proved that for prefix-free complexity, randomness and incompressibility coincide:

  A sequence is Martin-Löf random iff there exists a c such that for all n,

$$K(X|_n) \geq n - c$$

A Random Sequence

- Let *U* be a universal prefix-free machine.
- *Theorem* [Chaitin]:

The real number

$$\Omega = \sum_{M \text{ halts on } x} 2^{-|x|}$$

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.
- Applications range from set theory (Solovay models) to model selection (Minimum Description Length principle).

References

- M. Li and P. Vitanyi, *An Introduction to Kolmogorov Complexity and Its Applications*, Springer, 1997.

  A standard reference for Kolmogorov Complexity.

- V. Uspenskij, A. Semenov,A. Shen, *Can an (individual) sequence of zeros and ones be random?* Russian Math. Surveys 45 (1990), no. 1, 121–189.

- R. Downey and D. Hirschfeldt, *Computability and Randomness*, Notices of the AMS, August 2019.

  Survey articles on Martin-Löf randomness and its equivalent characterizations.

- M. van Lambalgen, *Random Sequences*, PhD-Thesis, University of Amsterdam, 1987.

  A detailed mathematical and philosophical discussion of von Mises' ideas. Available online http://staff.science.uva.nl/~michiell/docs/