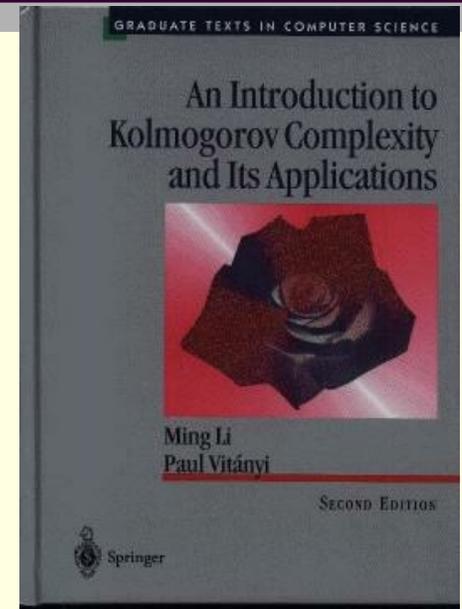


Kolmogorov complexity and its applications



Ming Li
University of Waterloo

We live in an information society. Information science is our profession.

Fundamental Questions:

- What is “**information**”, mathematically, and how to use it to prove theorems?
- What is a computable “**random number**”...what properties does it have ?
- What is an “**incompressible string**”...what properties does it have ?

Motivation:

A case of Dr. Samuel Johnson

(1709-1784)



... Dr. Beattie observed, as something remarkable which had happened to him, that he chanced to see both No.1 and No.1000 hackney-coaches. “Why sir,” said Johnson “there is an equal chance for one’ s seeing those two numbers as any other two.”

***Boswell’ s Life of
Johnson***

Further Motivation:

Alice goes to the court

- Alice complains: T^{100} is not random.
- Bob asks Alice to produce a random coin flip sequence.
- Alice flipped her coin 100 times and got
THTTHTHTHTHHHTTTTH ...
- But Bob claims Alice's sequence has probability 2^{-100} , and so does his.
- How do we define randomness?



Further Motivation, Cont

Alice goes to the court

Bob proposes to flip a coin with Alice:

- Alice wins a dollar if Heads;
- Bob wins a dollar if Tails

Result: TTTTTT 100 Tails in a roll.



Alice lost \$100. She feels being cheated.



History: What is the Information in Individual String?

- What is the information content of an individual string?
 - 111 1 (n 1' s)
 - $\pi = 3.1415926 \dots$
 - $n = 2^{1024}$
 - Champernowne' s number:
0.1234567891011121314 ...
is normal in scale 10 (every block has same frequency)
 - All these numbers share one commonality: **there are “small” programs to generate them.**
- Popular youtube explanation:
<http://www.youtube.com/watch?v=KyB13PD-UME>

History: What is the Information in Individual String?

(1) **Information Theory:** Shannon-Weaver theory is on an **ensemble**. But what is information in an **individual object**?

Shannon's information theory does not seem to help here.

(2) **Inductive inference:** Bayesian approach using **universal prior distribution**

(3) **Kolmogorov Theory:** TM state size

Andrey Nikolaevich Kolmogorov

(1903-1987, Tambov, Russia)



- Measure Theory
- Probability
- Analysis
- Intuitionistic Logic
- Cohomology
- Dynamical Systems
- Hydrodynamics
- Kolmogorov complexity

Preliminaries and Notations

- Binary Strings: x, y, z .
- $x = x_1x_2 \dots$ an infinite binary sequence
 - Finite subsequence $x_{i:j} = x_i x_{i+1} \dots x_j$
 - $|x|$ is number of bits in x .
- Sets, $A, B, C \dots$
 - $|A|$, number of elements in set A .
- Fix an **effective enumeration of all Turing machines (TMs)**: M_1, M_2, M_3, \dots
- $\langle M_n \rangle$ is description of TM M_n
- **Universal Turing machine U** :
 - $U(0^n1x) = M_n(x)$ = gives output of TM M_n with input x

Kolmogorov Theory

Let U be a universal TM that takes as input the description $p = \langle M \rangle$ of a TM M and produces as output $U(p)$.

Solomonoff (1960)-Kolmogorov (1963)-Chaitin (1965):

The amount of information in a string x is the size of the smallest description $\langle M \rangle$ of any TM M generating x .

$K_U(x) = \min_n \{ |\langle M_n \rangle| : U \text{ simulates TM } M_n \text{ with no input, which gives output } x \}$

Invariance Theorem: It does not matter which universal Turing machine U we choose. I.e. all “encoding methods” are ok.

Proof of the Invariance theorem

- For a fixed effective enumeration of all Turing machines (TM' s): M_1, M_2, \dots
- U is a universal TM such that with no input to n th TM M_n produces x

$$U(0^n1) = M_n() = x$$

- Then for all x : $K_U(x) < K_n(x) + O(1)$
 - Note: The constant $O(1)$ depends on n , but not x .
- Fixing U , we write $K(x)$ instead of $K_U(x)$. **QED**

Formal statement of Invariance Theorem:

There exists a computable function f_0 such that for all computable functions f , there is a constant c_f such that for all strings $x \in \{0, 1\}^*$

$$K_{f_0}(x) \leq K_f(x) + c_f$$

Kolmogorov Theory **Properties**

- Intuitively: **$K(x)$ = length of shortest description of x**
- Properties of $K(x)$:
 - $K(xy) \leq K(x) + K(y) + O(\log(\min\{K(x), K(y)\}))$
 - $K(xx) = K(x) + O(1)$ **Why?**
 - $K(1^n) \leq O(\log n)$ **Why?**
 - $K(n!) \leq O(\log n)$ **Why?**
 - For all x , $K(x) \leq |x| + O(1)$ **Why?**

Kolmogorov Theory **Properties**

- Intuitively: $K(x)$ = length of shortest description of x
- Properties of $K(x)$:
 - $K(xy) \leq K(x) + K(y) + O(\log(\min\{K(x), K(y)\}))$
 - $K(xx) = K(x) + O(1)$ **since just need TM generating x**
 - $K(1^n) \leq O(\log n)$ **since can use binary encoding of n**
 - $K(\pi_{1:n}) \leq O(\log n)$ **since can use binary encoding of n**
 - For all x , $K(x) \leq |x| + O(1)$ **since can encode x in TM**

Kolmogorov Theory **Conditional** **Properties**

- Recall $K(x)$ = length of shortest description of x
- Define conditional Kolmogorov complexity similarly,
 $K(x|y)$ = length of shortest description of x given y .
- Properties of $K(x|y)$:
 - $K(x|\epsilon) = K(x)$ **Why?**
 - $K(x|x) = O(1)$ **Why?**

Kolmogorov Theory **Conditional** **Properties**

- $K(x)$ = length of shortest description of x
- Define conditional Kolmogorov complexity similarly,
 $K(x|y)$ = length of shortest description of x , given y as *input*
- Properties of $K(x|y)$:
 - $K(x|\varepsilon) = K(x)$ **since empty string ε provides no additional info on x**
 - $K(x|x) = O(1)$ **since just need TM that generates x**

Basics of Kolmogorov Theory

- **Incompressibility:** For constant $c > 0$, a string $x \in \{0, 1\}^*$ is **c-incompressible** if $K(x) \geq |x| - c$.
- For constant c , we often simply say that x is **incompressible**.
- Incompressible strings have properties similar to **random** strings.

Lemma. There are at least $2^n - 2^{n-c} + 1$ c-incompressible strings of length n .

Proof. The number of programs with length $< n - c$ is $\sum_{k=0, \dots, n-c-1} 2^k = 2^{n-c} - 1$.

Hence only that many strings (out of total 2^n strings of length n) can have shorter programs (descriptions) than $n - c$. **QED.**

Basics of Kolmogorov Theory

- Recall: a finite string x is incompressible if $K(x) \geq |x| - c$ for a constant c .
- If $x = uvw$ is incompressible, then $K(v) \geq |v| - O(\log |x|)$.
- If $\langle M \rangle$ is the shortest TM description for x , then
 - $K(\langle M \rangle) \geq |\langle M \rangle| - O(1)$ and
 $K(x|\langle M \rangle) = O(1)$.
- A is **recursively enumerable (r.e.)** if the elements of A can be listed by a Turing machine.

Properties of Kolmogorov Theory

Theorem (Kolmogorov) $K(x)$ is not partially recursive.
(That is, there is no Turing machine M such that M accepts (x,m) if $K(x) \geq m$ and undefined otherwise.)

Proof. If such M exists, then design M' as follows:
Choose $n \gg |\langle M' \rangle| = \text{length of description of } M'$.
Choose a sufficiently large constant $c > 0$.
Let M' simulate M on input (x,n) , for all $|x| = n + c$ in “parallel” (one step each), and then output the first x such that M says yes.

Thus we have a contradiction:

- $K(x) \geq n$ by M ,
- but M' outputs x .

Hence $|M'| \geq K(x) \geq n$, but by choice $|x| = n \gg |\langle M' \rangle|$, a **contradiction**. **QED**

Kolmogorov Theory Applications to Complexity Theory

Kolmogorov Theory can give elegant proofs in Complexity Theory:

- Proofs that certain sets are **not regular**
- Complexity **Lower Bounds for 1 Tape TMs**
- **Communication Lower Bounds:** What is the distance between two pieces of information carrying entities? For example, distance from an internet query to an answer.

Other Kolmogorov Theory Applications

- **Mathematics:** probability theory, logic.
- **Physics:** chaos, thermodynamics.
- **Computer Science:** average case analysis, inductive inference and learning, shared information between documents, data mining and clustering, incompressibility method -- examples:
 - Lower bounds on Turing machines, formal languages
 - Shellsort average case
 - Heapsort average case
 - Circuit complexity
 - Combinatorics: Lovazs local lemma and related proofs.
 - Distributed protocols
- **Philosophy, biology etc:** randomness, inference, complex systems, sequence similarity
- **Information theory:** information in individual objects, information distance
 - Classifying objects: documents, genomes
 - Query Answering systems



Kolmogorov Theory: Further Results



Kolmogorov Theory: Further Results

Theorem. The statement “ x is random” (x is incompressible) is not provable.

Proof (G. Chaitin). Let F be an axiomatic theory. Let $K(F) = K$ be the size of the compressed encoding of F . If the theorem is false and statement “ x is random” is provable in F , then we can enumerate all proofs in F to find a proof of “ x is random” and $|x| \gg K$, output (first) such x . Then $K(x) < K + O(1)$. But the proof for “ x is random” implies that $K(x) \geq |x| \gg K$, a contradiction. QED

Kolmogorov Theory: Further Results:

Barzdin's Lemma

- A **characteristic sequence** of set A is an infinite binary sequence $\chi = \chi_1\chi_2 \dots$, where $\chi_i = 1$ iff $i \in A$.

Theorem. (i) The characteristic sequence χ of an r.e. set A satisfies $K(\chi_{1:n}|n) \leq \log n + c_A$ for all n .

(ii) There is an r.e. set, $K(\chi_{1:n}|n) \geq \log n$ for all n .

Proof.

Proof of (i): Use the number 1's in the prefix $\chi_{1:n}$ as a termination condition, implies $K(\chi_{1:n}|n) \leq \log n + c_A$

Proof of (ii): By diagonalization: Let U be the universal TM. Define $\chi = \chi_1\chi_2 \dots$, by $\chi_i = 1$ if $U(i\text{-th program}, i) = 0$, otherwise $\chi_i = 0$. χ defines an r.e. set. And, for each n , we have $K(\chi_{1:n}|n) \geq \log n$ since the first n programs of length $< \log n$ are all different from $\chi_{1:n}$ by definition.
QED