CS860, Winter, 2010

Kolmogorov complexity and its applications

Ming Li
University of Waterloo

1

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

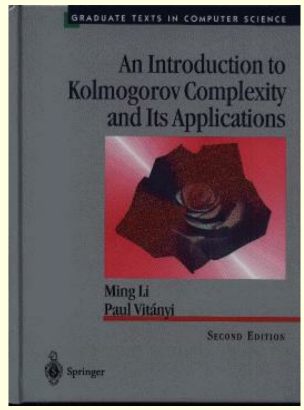
Fundamental Questions:

- 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 ?

2

Lecture 1. History and Definitions

- History
  - Intuition and ideas in the past
  - Inventors
- Basic mathematical theory



- For more see book:

Li-Vitanyi: An introduction to Kolmogorov complexity and its applications.

3

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*

4

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
  THTTHHTHTHHHTTTTH …
- But Bob claims Alice's sequence has probability 2^{-100}, and so does his.
- How do we define randomness?

5

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.

6

History: What is the Information in Individual String?

- What is the information content of an individual string?
  - 111 …. 1 (n 1's)
  - π = 3.1415926 …
  - n = 2^{1024}
  - Champernowne's number:
    0.12345678910111213 14 …
    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

7

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**

8

Andrey Nikolaevich Kolmogorov *(1903-1987, Tambov, Russia)*

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

9

Preliminaries and Notations

- Binary Strings: x, y, z.
- $x=x_1x_2$ ... an infinite binary sequence
  - Finite subsequence $x_{i:j} = x_i\, x_{i+1} \ldots x_j$
  - |x| is number of bits in x.
- Sets, A, B, C …
  - |A|, number of elements in set A.
- Fix an **effective enumeration of all Turing machines (TMs)**: M_1, M_2, M_3, …
- **<** M_n> 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

10

3. Kolmogorov Theory

Let U be a universal TM that takes as input the description p=<M> 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 <M> of any TM M generating x.

$K_U(x) = \min_n \{ |\langle M_n \rangle| :$ U simulates TM M_n 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.

11

Proof of the Invariance theorem

- For a fixed effective enumeration of all Turing machines (TM's): M_1, M_2, …
- U is a universal TM such that with no input to nth 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 the 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 ε $\{0,1\}^*$

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

12

Kolmogorov Theory continued…

- Intuitively*: 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)* and K*(x|y):*
  - K(xx) = K(x) + O(1) since just need TM that generates x
  - K(xy) ≤ K(x) + K(y) + O(log(min{K(x), K(y)})
  - $K(1^n) \le O(\log n)$ since can use binary encoding of n
  - $K(\pi_{1:n}) \le O(\log n)$ since can use binary encoding of n
  - For all x, K(x) ≤ |x|+O(1) since can encode x in TM
  - K(x|x) = O(1) since just need TM that generates x
  - K(x|ε) = K(x) since empty string ε provides no additional info on x

13

3.1 Basics

- Incompressibility: For constant c>0, a string $x \in \{0,1\}^*$ is c-incompressible if K(x) ≥ |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. There are only $\sum_{k=0,\ldots,n-c-1} 2^k = 2^{n-c} - 1$ programs with length less than n-c. Hence only that many strings (out of total 2^n strings of length n) can have shorter programs (descriptions) than n-c.
QED.

14

Facts

- 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 M is the shortest TM description for x, then
  - $K(M) \geq |M| - O(1)$ and
    $K(x|M) = O(1)$.
- A is recursively enumerable (r.e.) if the elements of A can be listed by a Turing machine.
- A is sparse if the set of all length n strings of A is $\leq p(n)$ for some polynomial p. If a subset A of {0,1}* is recursively enumerable (r.e.), and A is sparse, then for all x in A, $|x|=n$,
- $K(x) \leq O(\log p(n)) + O(K(n)) = O(\log n)$.

15

3.3 Properties

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 >> |M'|=length of description of M'.
Let M' simulate M on input (x,n), for all |x|=n 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 $|x|=n >> |M'|$, a contradiction. QED

16

3.4 Godel's Theorem

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 $|x| >> K$, output (first) such x. Then $K(x) < K + O(1)$. But the proof for "x is random" implies that $K(x) \geq |x| >> K$, a contradiction. QED

17

3.5 Barzdin's Lemma

- A characteristic sequence of set A is an infinite binary sequence $\chi=\chi_1\chi_2 \ldots$, where $\chi_i=1$ iff i ε 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 $\chi=\chi_1\chi_2 \ldots$, by $\chi_i=1$ if U(i-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

18

Kolmogorov Theory Applications to 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.

19

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

20