CS860, Winter, 2010

Kolmogorov complexity and its applications

Ming Li
School of Computer Science
University of Waterloo
http://www.cs.uwaterloo.ca/~mli/cs860.html

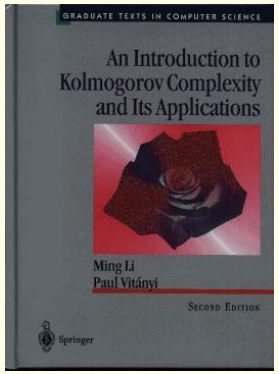
We live in an information society. Information science is our profession. But do you know what is "information", mathematically, and how to use it to prove theorems?

Examples

- Average case analysis of Shellsort.
- Lovasz Local Lemma
- What is the distance between two pieces of information carrying entities? For example, distance from an internet query to an answer.

Lecture 1. History and Definitions

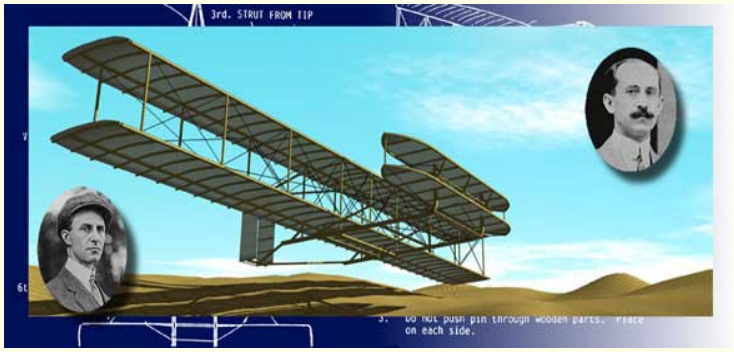
- History
  - Intuition and ideas in the past
  - Inventors
- Basic mathematical theory
- Textbook: Li-Vitanyi: An introduction to Kolmogorov complexity and its applications. You may use any edition (1st , 2nd , 3rd) except that the page numbers are from the 2nd edition.



1. Intuition & history

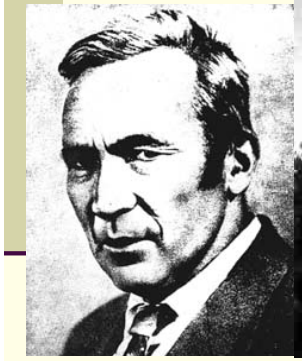
- What is the information content of an individual string?
  - 111 …. 1 (n 1's)
  - π = 3.1415926 …
  - n = 2^{1024}
  - Champernowne's number:
    0.1234567891011121314 …
    is normal in scale 10 (every block has same frequency)
  - All these numbers share one commonality: there are "small" programs to generate them.
- Shannon's information theory does not help here.
- Popular youtube explanation:
  http://www.youtube.com/watch?v=KyB13PD-UME

1903: An interesting year



This and the next two pages were taken from Lance Fortnow

1903: An interesting year



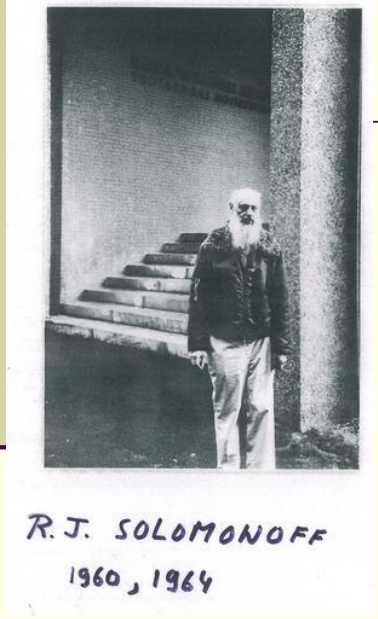
Kolmogorov

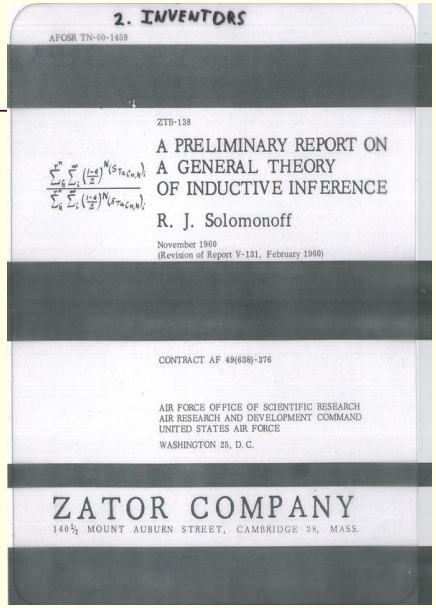
Church

von Neumann

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

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




2. INVENTORS

AFOSR TN-60-1459

ZTB-138

A PRELIMINARY REPORT ON
A GENERAL THEORY
OF INDUCTIVE INFERENCE

R. J. Solomonoff

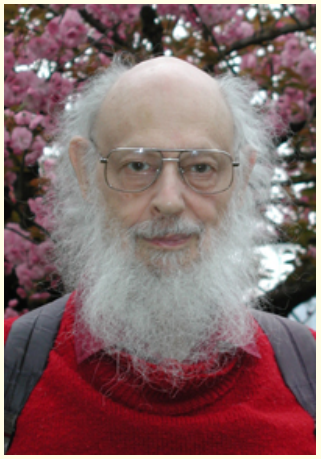
November 1960
(Revision of Report V-131, February 1960)

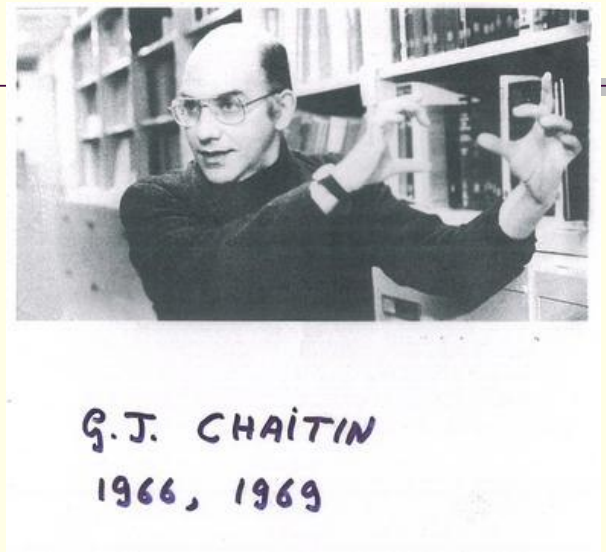
CONTRACT AF 49(638)-376

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
AIR RESEARCH AND DEVELOPMENT COMMAND
UNITED STATES AIR FORCE
WASHINGTON 25, D.C.

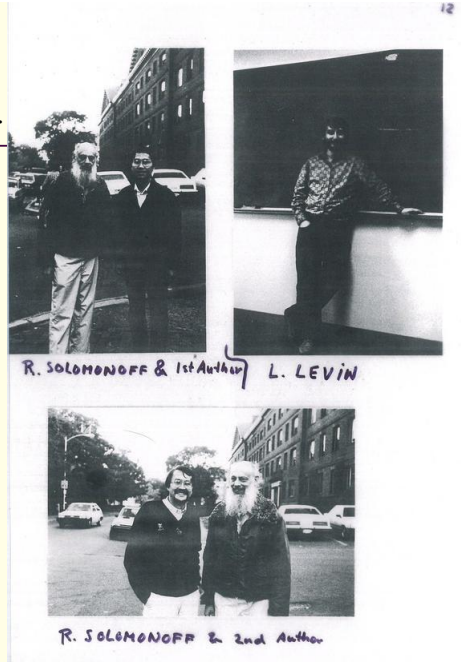
ZATOR COMPANY
140½ MOUNT AUBURN STREET, CAMBRIDGE 38, MASS.

Ray Solomonoff: 1926 -- 2009





When there were no digital cameras (1987).



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*

The case of cheating casino

Bob proposes to flip a coin with Alice:

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

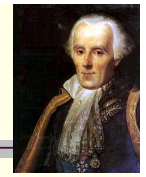
Result: TTTTTT …. 100 Tails in a roll.

Alice lost $100. She feels being cheated.

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?

2. Roots of Kolmogorov complexity and preliminaries



Laplace, 1749-1827

(1) Foundations of Probability

- P. Laplace: … a sequence is extraordinary (nonrandom) because it contains rare regularity.
- 1919. von Mises' notion of a random sequence S:
  - $\lim_{n\to\infty}\{$ #(1) in n-prefix of S$\}/n = p$, $0<p<1$
  - The above holds for any subsequence of S selected by an "admissible" function.
- But if you take any partial function, then there is no random sequence a la von Mises.
- A. Wald: countably many. Then there are "random sequences.
- A. Church: recursive selection functions
- J. Ville: von Mises-Wald-Church random sequence does not satisfy all laws of randomness.

Roots …

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

(3) Inductive inference. Bayesian approach using universal prior distribution

(4) Shannon's State x Symbol (Turing machine) complexity.

Preliminaries and Notations

- Strings: x, y, z. Usually binary.
  - $x=x_1x_2$ ... an infinite binary sequence
  - $x_{i:j} = x_i\, x_{i+1} \ldots x_j$
  - $|x|$ is number of bits in x. Textbook uses l(x).
- Sets, A, B, C …
  - $|A|$, number of elements in set A. Textbook uses d(A).
- K-complexity vs C-complexity, names etc.
- I assume you know Turing machines, universal TM's, basic facts from CS360.

3. Mathematical Theory

Solomonoff (1960)-Kolmogorov (1963)-Chaitin (1965):
The amount of information in a string is the size of the smallest program generating that string.

$$C_U(x) = \min_p \{|p| : U(p) = x\}$$

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

- Fix an effective enumeration of all Turing machines (TM's): $T_1, T_2, \ldots$
- Let U be a universal TM such that (p produces x)
  $U(0^n1p) = T_n(p)$
- Then for all x: $C_U(x) < C_{Tn}(x) + O(1)$ --- O(1) depends on n, but not x.
- Fixing U, we write C(x) instead of $C_U(x)$. QED

Formal statement of the Invariance Theorem: There exists a computable function S_0 such that for all computable functions S, there is a constant c_S such that for all strings $x \in \{0,1\}^*$

$C_{S0}(x) \le C_S(x) + c_S$

It has many 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:
  - Shellsort average case
  - Heapsort average case
  - Circuit complexity
  - Lower bounds on Turing machines, formal languages
  - Combinatorics: Lovazs local lemma and related proofs.
- Philosophy, biology etc – randomness, inference, complex systems, sequence similarity
- Information theory – information in individual objects, information distance
  - Classifying objects: documents, genomes
  - Query Answering systems

Mathematical Theory cont.

- Intuitively: $C(x)$= length of shortest description of x
- Define conditional Kolmogorov complexity similarly, $C(x|y)$=length of shortest description of x given y.
- Examples
  - $C(xx) = C(x) + O(1)$
  - $C(xy) \leq C(x) + C(y) + O(\log(\min\{C(x),C(y)\})$
  - $C(1^n) \leq O(\log n)$
  - $C(\pi_{1:n}) \leq O(\log n)$
  - For all x, $C(x) \leq |x|+O(1)$
  - $C(x|x) = O(1)$
  - $C(x|\varepsilon) = C(x)$

3.1 Basics

- Incompressibility: For constant $c>0$, a string $x \in \{0,1\}^*$ is c-incompressible if $C(x) \geq |x|-c$. For constant c, we often simply say that x is incompressible. (We will call incompressible strings 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.

Facts

- If x=uvw is incompressible, then
  $C(v) \geq |v| - O(\log |x|)$.
- If p is the shortest program for x, then
  $C(p) \geq |p| - O(1)$
- $C(x|p) = O(1)$
- If a subset of $\{0,1\}^*$ A is recursively enumerable (r.e.) (the elements of A can be listed by a Turing machine), and A is sparse ($|A^{=n}| \leq p(n)$ for some polynomial p), then for all x in A, $|x|=n$,
  $C(x) \leq O(\log p(n)) + O(C(n)) = O(\log n)$.

3.2 Asymptotics

- Enumeration of binary strings: 0,1,00,01,10, mapping to natural numbers 0, 1, 2, 3, …
- C(x) →∞ as x →∞
- Define **m**(x) to be the monotonic lower bound of C(x) curve (as natural number x →∞). Then

  m(x) →∞, as x →∞
- **m**(x) < Q(x) for all unbounded computable Q.
- Nonmonotonicity: for x=yz, it does not imply that C(y)≤C(x)+O(1).

m(x) graph

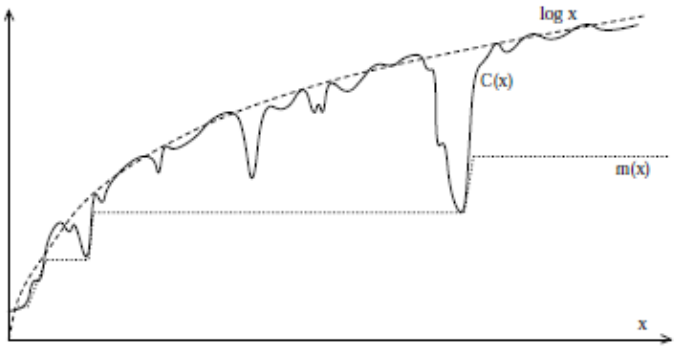


FIGURE 2.1. The graph of the integer function $C(x)$

3.3 Properties

Theorem (Kolmogorov) C(x) is not partially recursive. That is, there is no Turing machine M s.t. M accepts (x,k) if C(x)≥k and undefined otherwise. However, there is H(t,x) such that

$\lim_{t\to\infty} H(t,x)=C(x)$

where H(t,x), for each fixed t, is total recursive.

Proof. If such M exists, then design M' as follows. Choose n >> |M'|. M' simulates M on input (x,n), for all |x|=n in "parallel" (one step each), and outputs the first x such that M says yes. Thus we have a contradiction: C(x)≥n by M, but |M'| outputs x hence |x|=n >> |M'| ≥ C(x) ≥ n. QED

3.4 Godel's Theorem

Theorem. The statement "x is random" is not provable.

Proof (G. Chaitin). Let F be an axiomatic theory. C(F)= C. 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| >> C, output (first) such x. Then C(x) < C +O(1) But the proof for "x is random" implies that C(x) ≥ |x| >> C. Contradiction. QED

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 $i \varepsilon A$.

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

Proof.

(i) Using number 1's in the prefix $\chi_{1:n}$ as termination condition (hence logn)

(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 $C(\chi_{1:n}|n) \geq \log n$ since the first n programs (i.e. any program of length < logn) are all different from $\chi_{1:n}$ by definition. QED