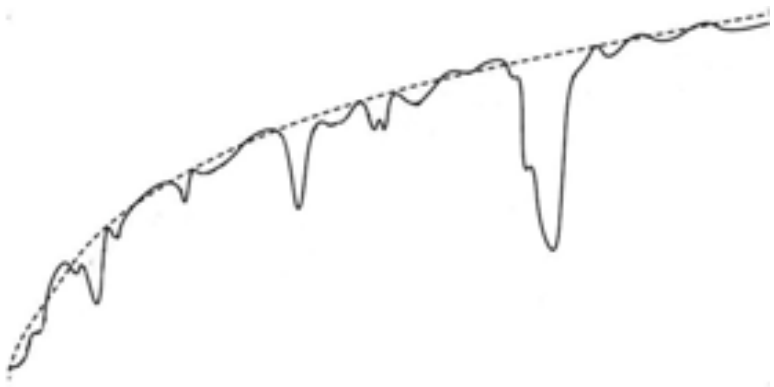
Kolmogorov complexity

Lecture II

George Barmpalias

Chinese Academy of Sciences

UESTC - Chengdu, December 5, 2023



Contents of Lecture II

- Instanteneous codes and Kraft inequality
- Construction of prefix-free codes
- proof of coding theorem
- Counting theorems
- Symmetry of information
- incompressibility method (LLL example)

Links in Red

Theorem. The following hold:

- $C(\sigma\tau) \leq^+ C(\sigma,\tau) \leq^+ C(\sigma) + C(\tau) + 2\log C(\sigma)$.
- $\forall d\ \exists\ \sigma,\tau: \quad C(\sigma\tau) > C(\sigma) + C(\tau) + d$.

Conservation of information ? This is weird...

How can $\sigma\tau$ have more information than $C(\sigma) + C(\tau)$?

Explanation: A program τ for σ

- carries information in its digits, but also in its length $|\tau|$
- this can make $C(\sigma)$ smaller than it should be.

By restricting the underlying machines to:

- Self-delimiting (one-way reading of the the input-tape)
- or exuivalently, prefix-free

we obtain a refined complexity $K(\sigma)$.

Plain versus prefix-free complexity

Theorem. $K(\sigma\tau) \stackrel{+}{\leq} K(\sigma,\tau) \stackrel{+}{\leq} K(\sigma) + K(\tau)$

Theorem.

- $C(\sigma) \stackrel{+}{=} \min\{n : K(\sigma \mid n) \leq n\}$
- $C(\sigma) \stackrel{+}{=} K(\sigma \mid C(\sigma))$.

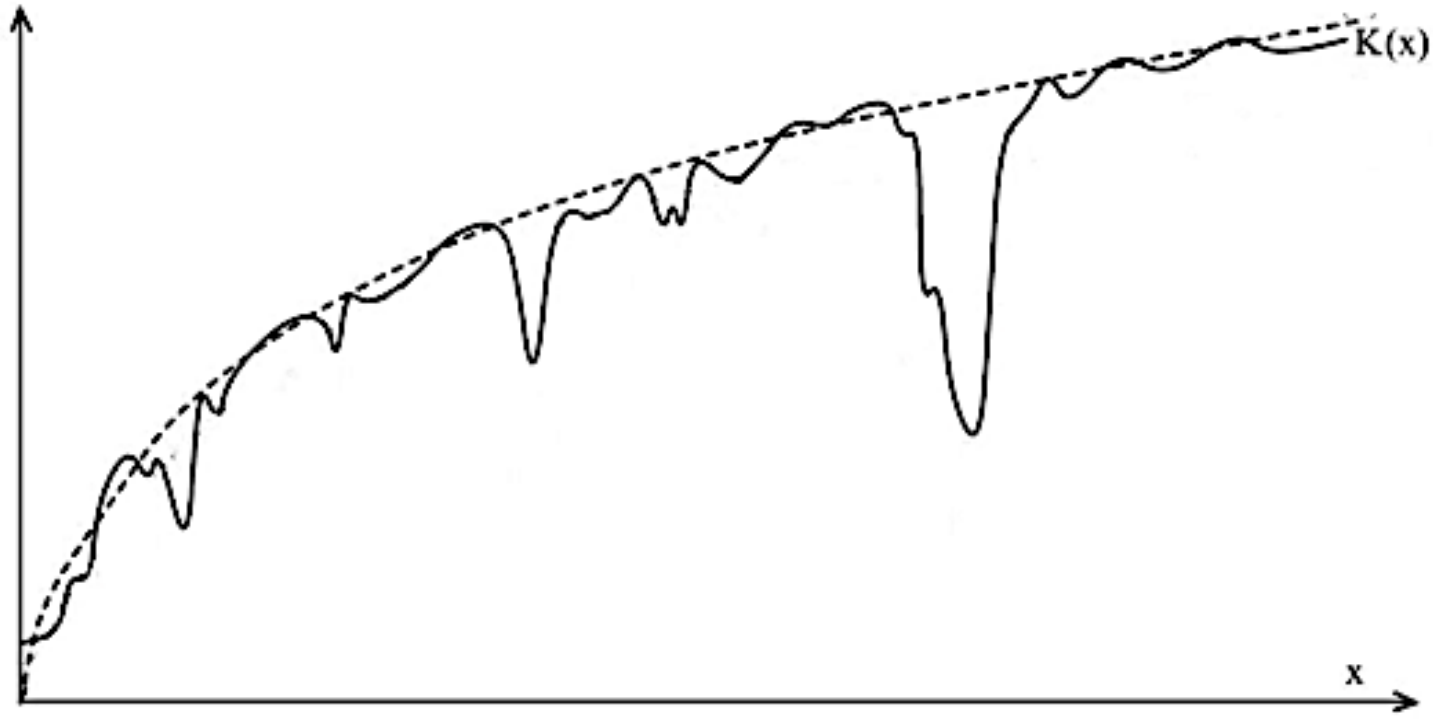
Theorem.

$$\max_{i \leq n} K(i) \stackrel{+}{=} \log n + K(\log n)$$

$$C(\sigma) \stackrel{+}{\leq} |\sigma| \ \wedge \ \ C(n) \stackrel{+}{\leq} \log n \ \wedge \ \ K(n) \stackrel{+}{\leq} 2 \log n$$

$$K(|\sigma|) \stackrel{+}{\leq} K(\sigma) \stackrel{+}{\leq} |\sigma| + K(|\sigma|)$$

$$K(\log n) \stackrel{+}{\leq} K(n) \stackrel{+}{\leq} \log n + K(\log n)$$



Optimal descriptions are incompressible (same for K):

$$C(\sigma^*) \stackrel{+}{=} C(\sigma) = |\sigma^*|$$

The existence of σ with a certain property can be shown by:

- *Explicit construction* and verification of the required property
- **Probabilistic method**: show that the required property occurs with high or non-zero probability
- **Incompressibility method**: show that the negation of the required property allows the compression of σ.

A good example: algorithmic Lovasz Local Lemma

Fortnow (Moser's proof via Kolmogorov complexity)

Moser-Tardos and Messner-Thierauf

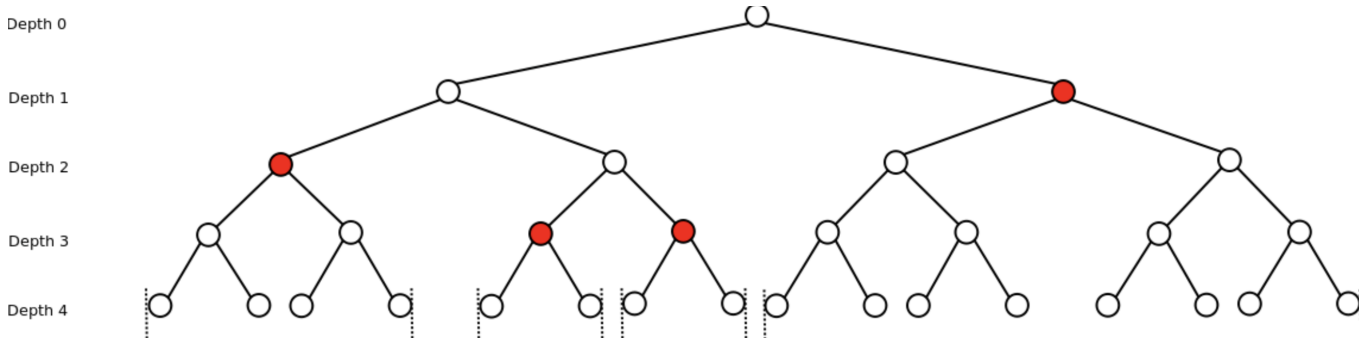
Prefix-free codes

A set of strings is *prefix-free* if no member of it is a strict prefix of another member of it.

0	0	0	000
10	100	1000	001
110	101	1001	010
1110	11000	10100	011
11110	11001	10101	1000
111110	11010	10110	1001
1111110	11011	10111	1010
11111110	1110000	11000000	1011
111111110	1110001	11000001	11000
1111111110	1110010	11000010	11001

- also called instanteneous codes or self-delimiting codes
- form a subset of the uniquely decodable codes.

Kraft inequality



Prefix-free sets as antichains.

Theorem (Kraft). Any prefix code with the codeword lengths $\ell_i, i < n$ satisfies $\sum_{i<n} 2^{-\ell_i} \leq 1$.

Proof. Map $\sigma \mapsto [0.\sigma, 0.\sigma + 2^{-|\sigma|})$.

Then prefix-free sets correspond to pairwise disjoint intervals. ◀

About Kraft inequality and Github code

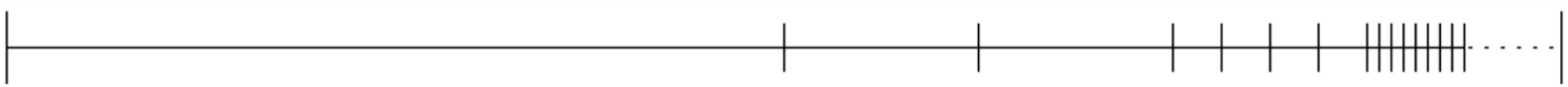
Online generation of prefix-free codes

Given requests (ℓ_i) with $\sum_i 2^{-\ell_i} \leq 1$, monitor the space via the *trace*:

$$t_s := 1 - \sum_{i<s} 2^{-\ell_i}$$

.

Greedy code assignment (intervals)



The lexicographically greedy assignment of intervals produces a prefix-free code, as long as Kraft's inequality holds.

The 1s in t_s indicate the available intervals: $\sigma \mapsto [0.\sigma, 0.\sigma + 2^{-|\sigma|})$.

Information content measures

Definition. A function I from strings to integers ≥ 0 is an information content measure if

- it is effectively approximable from below
- $\sum_\sigma 2^{-I(\sigma)} \leq 1$.

Theorem. K is a $O(1)$-minimal information content measure.

Information content measures

- are produced via online prefix-free codes
- define a distribution (left semimeasure) on the strings.

The optimal non-decreasing information content measure is

$$K^*(n) := \max_{i \leq n} K(i)$$

The probability that the universal machine prints σ on random input is

$$P(\sigma) \;:=\; \sum_{U(\tau)=\sigma} 2^{-|\tau|}$$

Analogue (finite, algorithmic) of Shannon's source-coding theorem:

Coding Theorem. $P(\sigma) \stackrel{\times}{=} 2^{-K(\sigma)}$ and $K(\sigma) \stackrel{+}{=} -\log P(\sigma)$

Proof. Clearly $2^{-K(\sigma)} = 2^{-|\sigma^*|} \stackrel{+}{\leq} P(\sigma)$

Note that $\sigma \mapsto -\log P(\sigma)$ is an information content measure.

So $K(\sigma) \stackrel{+}{\leq} -\log P(\sigma)$ and $P(\sigma) \stackrel{\times}{\leq} 2^{-K(\sigma)}$. ◀

The probability of n-bit output is

$$P(n) := \sum_{|U(\tau)|=n} 2^{-|\tau|} \sim 2^{-K(n)}$$

Complexity of complexity

Sometimes Kolmogorov complexity is complex (computational depth).

Theorem. For every n there exists an n-bit string σ with

$$K(K(\sigma) \mid \sigma) \stackrel{+}{=} \log n$$

Proof by *allocation game*:

- Player 1 picks string of high complexity
- Player 2 describes its current complexity
- Player 1 compresses it and picks another random string
- Player 2 can insist with a new description of the updated complexity or describe the new string

Player 2 runs out of descritions: he fails compressing some $K(\sigma)$.

Kolmogorov complexity is rarely complex (random strings).

Counting theorems I

Recall $K(|\sigma|) \stackrel{+}{\leq} K(\sigma) \stackrel{+}{\leq} |\sigma| + K(|\sigma|)$

Theorem. The following hold:

- $\max\{K(\sigma) : |\sigma| = n\} \stackrel{+}{=} |\sigma| + K(|\sigma|)$
- $|\{\sigma : |\sigma| = n \ \wedge \ \ K(\sigma) \leq |\sigma| + K(n) - m\}| = O(2^{n-m})$

The probability of compressing by c bits decreases exponentially in c.

Since n, m are independent by substitution:

$$|\{\sigma : |\sigma| = n \ \wedge \ \ K(\sigma) \leq |\sigma| - c\}| = O(2^{n-K(n)-c})$$

Counting theorems II

Theorem. Let $D_n := \{\sigma : |\sigma| = n \ \wedge \ \ U(\sigma)\downarrow\}$ and

- $p_n := |\{\sigma : |\sigma| = n \ \wedge \ \ U(\sigma)\downarrow\}|$
- $P_n := |\{\sigma : |\sigma| \leq n \ \wedge \ \ U(\sigma)\downarrow\}|$
- $d_n := \{\sigma : K(\sigma) \leq n\}$.

Then $p_n \sim P_n \sim d_n \sim 2^{n-K(n)}$ and

$$K(D_n) \stackrel{+}{=} n$$

Corollary. The number of shortest descriptions of any object is bounded by a universal constant.

Mutual Information and Symmetry

Definition. The *mutual information* of σ, τ is

$$I(\tau : \sigma) := K(\sigma) - K(\sigma \mid \tau^*)$$

Theorem. $I(\tau : \sigma) \stackrel{+}{=} I(\sigma : \tau)$

Follows from:

$$K(\sigma, \tau) \stackrel{+}{=} K(\sigma) + K(\tau \mid \sigma^*) \stackrel{+}{=} K(\tau) + K(\sigma \mid \tau^*)$$

Note: $K(\tau \mid \sigma^*) \stackrel{+}{=} K(\tau \mid \sigma, K(\sigma))$.

Similarly: $C(\sigma, \tau) \stackrel{+}{=} C(\sigma) + C(\tau \mid \sigma^*) + O(\log(C(\sigma, \tau)))$.

Theorem. $K(\sigma,\tau) \stackrel{+}{=} K(\sigma) + K(\tau \mid \sigma^*)$.

Proof. Clearly $K(\sigma,\tau) \stackrel{+}{\leq} K(\sigma) + K(\tau \mid \sigma^*)$. It remains to show

$$K(\tau \mid \sigma^*) \stackrel{+}{\leq} K(\sigma,\tau) - K(\sigma)$$

We enumerate online prefix-free code and the weight is:

$$2^{K(\sigma)} \cdot \sum_{\tau} 2^{-K(\sigma,\tau)}$$

But $I(\sigma) := \sum_{\tau} 2^{-K(\sigma,\tau)}$ is an information content measure.

So $I(\sigma) \stackrel{+}{\leq} P(\sigma) \stackrel{+}{=} 2^{-K(\sigma)}$ and

$$\sum_{\tau} 2^{-K(\sigma,\tau)} \stackrel{+}{\leq} 2^{-K(\sigma)}$$

so the required code exists. ◀

Resource-bounded Kolmogorov complexity

Non-equivalent formalizations (non-robustness).

Many equalities and results:

- fail to transfer, or need additive factors
- depend on strong computational complexity hypotheses.

We did not cover

- Solomonoff's theory of inductive inference
- compressibility of infinite binary sequences
- Resource-bounded versions of Kolmogorov complexity
- Computability-theoretic aspects of Kolmogorov complexity

Take home - Online slides and content

Remember:

- Incompressibility vs probabilistic: counting arguments
- Classical versus algorithmic information theory
- Incompressibility and Undecidability
- Coding theorem
- Universal distribution (on strings)
- Counting theorems
- Symmetry and conservation of information

http://barmpalias.net/teaching/courses/UESTC_KolmLect.html