Lecture 14

Tuesday, June 11, 2019 10:25 PM

Kolmogorov Complexity,

or how to measure randomness of a given individual binary string

"Random" strings :

0101010101 01 ⇐: "repeat 01 6 times"

0011101010 : the string itself

Randomness measure for distributions of strings.

$\mathcal{D} \subseteq \{0,1\}^n$ $\quad 0 \le d_x \le 1$: probability of string x

$$\sum_{x \in \{0,1\}^n} d_x = 1$$

Shannon's Entropy of $\mathcal{D}$:

$$\sum_{x \in \{0,1\}^n} d_x \cdot \log \frac{1}{d_x}$$

$= n$ for Uniform

$=n$ for UNIF

$$= \underset{x \sim \mathcal{D}}{\mathrm{Exp}} \left[\log_2 \frac{1}{d_x} \right]$$

average amount of information in a string x sampled from $\mathcal{D}$

$\mathcal{D} = \text{uniform}, \ \forall \ d_x = \frac{1}{2^n}$

Kolmogorov wanted the "worst-case" measure of randomness defined for individual strings (as opposed to distribs).

Kolmogorov's Idea: Use Computability Theory!

Def descriptive complexity of a string $x \in \{0,1\}^n$ is the length of shortest pair

$$d(x) = \langle M, w \rangle$$

s.t. TM M on input w will output string x

will output string x.

$$\langle M, w\rangle = \underline{\langle M\rangle w}$$

$\langle M\rangle$: repetition code

$$\langle 0100\rangle \xrightarrow{\text{Rep. Code}} 00\ 11\ 00\ 00$$

$$\Big\downarrow \text{ Decoding (easy)}$$

$$0\ 1\ 0\ 0$$

$$|\langle M\rangle w| = 2\cdot|\underset{\substack{\text{original}\\ \langle M\rangle}}{\langle M\rangle}| + |w|$$

$$\langle M\rangle w = \text{Rep. Code}(M)\ \underline{01}\ w$$

$$|\text{Encoding length}(M,w)| = 2\cdot|\langle M\rangle| + 2 + |w|$$

$$|d(x)| = \text{min over all } \langle M,w\rangle \text{ s.t. } M \text{ on } w \text{ outputs } x$$

s.t. M on w outputs x

of $|\langle M, w\rangle|$

Def: A string x is called K.-random if $|d(x)| \geq |x|$.

Observe: $\forall x \in \{0,1\}^n$,

$|d(x)| \leq n + O(1)$

Define M = "On input w,

output w."

x can be described as $\langle M, x\rangle$.

---

why do K.-random strings exist?

$\{0,1\}^n$: 2^n strings

Suppose each of them is not K.-random

i.e.,

i.e., "

$x \longmapsto$ encoded using $\leq n-1$ bits

$$\left.\begin{array}{l} x_1 \longmapsto e_1 \\ x_2 \longmapsto e_2 \\ \vdots \\ x_{2^n} \longmapsto e_{2^n} \end{array}\right\}$$ all encodings must be <u>distinct</u>!

$$2^0 + 2^1 + 2^2 + \ldots + 2^{n-1} = 2^n - 1 < 2^n$$

<u>Def:</u> $K(x) = |d(x)|$ Kolmogorov fn

<u>Claim:</u> (1) $\exists c$ constant

$\forall x, \quad K(xx) \leq K(x) + c$

(2) $\forall x, y, \quad K(xy) \leq 2 \cdot K(x) + K(y) + c$

Given computable fn p,

$d_p(x) =$ lex. shortest string s

s.t. $p(s) = x$

s.t. $p(s) = x$

<u>Claim:</u> $\forall$ comp. p, $\exists$ const. c s.t. $\forall x$, $K(x) \leq K_p(x) + c$.

<u>Thm:</u> (1) $K : \{0,1\}^* \to \{0,1\}^*$ is not computable.

(2) $\{ x \in \{0,1\}^* \mid K(x) \geq |x| \}$ is not semi-decidable.

Computability

K-random $x \in \{0,1\}^n$

Complexity

x that has no small "circuit"