Subject : Intro. to Info. Theory

No. : 1
Date : 12 / 17 / 2019

* Kolmogorov

$K(\underline{s}) = \min_{\underline{p}:\, U(\underline{p}) = \underline{s}} |\underline{p}|$

$\Downarrow$

$K(\underline{X}) = \underbrace{\sum_{\underline{s}} p(\underline{s})\, K(\underline{s})}_{\text{expectation}}$

Shannon

$\underbrace{\text{"}H(\underline{s})\text{"}}_{\text{self-info.}} = \log \frac{1}{p(\underline{s})}$

$\Downarrow$

$\neq$ in general

$H(\underline{X}) = \sum_{\underline{s}} p(\underline{s}) \log \frac{1}{p(\underline{s})}$

("irregular" ensemble)

eg. $K(\underline{s}) \gtrsim |\underline{s}|$

w/ $p(\underline{s}) = 1$

$K(\underline{X}) \gtrsim \underset{\underset{N}{\shortparallel}}{|\underline{s}|}$

$H(\underline{X}) = 0$

* what if "regular" ensemble

$\underline{X} = \underbrace{\underline{B}}_{\text{binary ensemble}}{}^{N \leftarrow \text{iid}}$

w/ $P_0 = 1-p$ & $P_1 = p$

known :

$$\underbrace{\underset{\underset{\text{similar lower bound}}{\leq}}{H(\underline{X}) \leq} (\underset{\underset{\mathcal{H}(p)}{\shortparallel}}{H(\underline{B})} + \epsilon)\, N \quad \text{when } N \text{ large enough}}_{\text{Shannon's 1st Theorem}}$$

todo :

$\underset{\text{similar lower bound (see notes)}}{\leq}$ $K(\underline{X}) \leq (\mathcal{H}(p) + \epsilon)\, N$ when N large enough

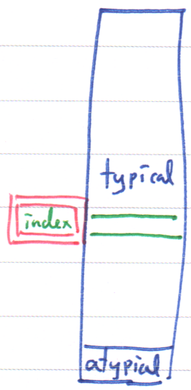
$\Downarrow$

$\frac{1}{N} K(\underline{X}) \approx \frac{1}{N} H(\underline{X}) \approx \mathcal{H}(p)$ when N large enough

Double A

* proof outline

$\underline{X} = \underline{B}^N$



typical ⇒ $\boxed{\text{a program}}$ that generates $\underline{s}$ w/ length $\approx H(p)$ — actually, $\le (H(p) + \frac{1}{2}\epsilon) N$

$1-\frac{\epsilon}{4} \le \underbrace{\sum_{\underline{s}}}_{\text{typical}} p(\underline{s}) \le 1$

} contribution $\le (H(p)+\frac{\epsilon}{2}) \cdot N$

atypical ⇒ $K(\underline{s}) \le N + C \le 2N$ for $N \ge C$

$\underbrace{\sum_{\underline{s}}}_{\text{atypical}} p(\underline{s}) \le \frac{\epsilon}{4}$ for N large enough

} contribution $\le \frac{1}{2}\epsilon \cdot N$

total $\le (H(p)+\epsilon) N$

* $\boxed{\text{the program}}$

for $\boxed{\text{each } \underline{s} \text{ in typical set}}$

each $\underline{s}$ with #1's between N_1 and N_2, lexicographically

count until $\boxed{\text{index}}$ and output the $\underline{s}$

input: $\underbrace{N}_{\log N \text{ bits}}$, $\underbrace{N_1 , N_2}_{2 \log N \text{ bits}}$ known from typicality , $\boxed{\text{index}}$

$\le (H(p) + \frac{\epsilon}{4}) N$ bits when N large enough

$\le \frac{\epsilon}{4} N$ when N large enough

const $\boxed{\text{program}}$ size

contribution $\le (H(p)+\frac{\epsilon}{2}) N$

* consider monkeys typing "uniform" programs

$\underline{p}$ of length M : $\boxed{\frac{1}{2^M} \quad \text{probability}}$

if uniform on program length M and uniform in $\underline{p}$ probability

$$p(\underline{s}) = \sum_{M=0}^{\infty} \sum_{\substack{|\underline{p}|=M \\ U(\underline{p})=\underline{s}}} 2^{-|\underline{p}|}$$

halts w/ output

can then infer s_t after observing $\underbrace{s_1, s_2, \cdots, s_{t-1}}_{\underline{s}_{t-1}}$

by comparing $P(\underline{s}_{t-1} 0 *)$ and $P(\underline{s}_{t-1} 1 *)$

called Solomonoff Inference, will discuss technical details next time