A comparison of growths of simple and prefix entropies

Alexei Semenov Andrej Muchnik

Institute of New Technologies
Moscow

Plain (simple) entropy $KS(x)$ of a binary word x is the length of the shortest code of x in a certain optimal encoding system.

Prefix entropy $KP(x)$ of a binary word x is the length of the shortest code of x in a certain optimal encoding system such that any code does not extend any other code.

Clearly, prefix encoding is *uniquely decodable:*

no binary word can be divided
into codes in two different ways

A uniquely decodable but not prefix code:

$$\{0, 010\}$$

McMillan–Kraft Theorem
For any uniquely decodable encoding system there exists a prefix encoding with the same code-word lengths.

$w_1, \ldots, w_k$ are code words, $L = \max_i \ell(w_i)$

$$(w_1 + \cdots + w_k)^n \leq \sum_{l \leq nL} \sum_{\ell(v)=l} v$$

$$\left(2^{-\ell(w_1)} + \cdots + 2^{-\ell(w_k)}\right)^n \leq nL \cdot 1$$

Consequently,

$$2^{-\ell(w_1)} + \cdots + 2^{-\ell(w_k)} \leq 1$$

Passage to the limit proves the theorem for the countable number of code words.

Given the code-word lengths, a prefix encoding is constructed easily step by step beginning with the smallest length.

Obviously, for any x

$$KS(x) \leq KP(x) + O(1)$$

$$KS(x) \leq \ell(x) + O(1)$$

$$\sum_x 2^{-KS(x)} \geq \sum_x 2^{-\ell(x)-O(1)} = \infty$$

$$\sum_x 2^{-KP(x)} \leq 1$$

Therefore, $KP(x) - KS(x)$ is unbounded.

Does $KP(x) - KS(x)$ tend to infinity?

Theorem (R. Solovay, unpublished)

$$KP(x) - KS(x) \ \geq$$
$$KP(KP(x)) - 2KP(KP(KP(x))) - O(1)$$

Proof

$$S_{n,i} \subseteq \{0,1\}^n$$
$$i \neq k \Rightarrow S_{n,i} \cap S_{n,k} = \emptyset$$
$$\left|S_{n,i}\right| = \left\lfloor 2^{n-KP(i)} \right\rfloor$$

Define an auxiliary coding function f: enumerating from above the function KP, for $y \in S_{n,i}$ and x such that $KP(x) < n + i$, put $f(y) = x$ (if $f(y)$ was not defined before).

Let $KP(x) = m$, $KP(m) = j$.
Let $n = m - j + 2KP(j)$, $i = j - 2KP(j) + 1$.

It is sufficient to prove that $\forall^{\infty} i \, \forall n$

$$\left|S_{n,i}\right| \geq \left|\{z \mid KP(z) < n + i\}\right|$$

$KP(x) = m$, $KP(m) = j$,
$n = m - j + 2KP(j)$, $i = j - 2KP(j) + 1$.

$$\alpha(q) = \sum_{KP(z)<q} 2^{-q}$$

$\alpha(q)$ is enumerable from below semimeasure

$$\sum_q \alpha(q) = \sum_z 2^{-KP(z)} \leq 1$$

Hence, $\alpha(q) \leq 2^{-KP(q)+O(1)}$,

$$\Big|\{z \mid KP(z) < m + 1\}\Big| \leq 2^{m-KP(m)+O(1)} =$$

$$2^{n-2KP(j)+O(1)} = \Big|S_{n,i}\Big| \cdot 2^{KP(i)-2KP(j)+O(1)}$$

$$KP(i) \leq KP(j) + KP(KP(j)) + O(1)$$

□

$$KP(x) \geq$$
$$KS(x) + KP(KP(x)) - O(\log\log \ell(x))$$

This bound is sharp:

$$KP(x) \leq KS(x) + KP(KS(x)) + O(1) \leq$$
$$KS(x) + KP(KP(x)) +$$
$$KP(KP(x) - KS(x)) + O(1) \leq$$
$$KS(x) + KP(KP(x)) + O(\log\log \ell(x))$$

The graphs of KS and KP are close:

$$KS(y) \gg KS(x) \quad \Leftrightarrow \quad KP(y) \gg KP(x)$$

$\big(K(y) \gg K(x)$ means $K(y) - K(x) \geq \Omega(\log\max\{\ell(x), \ell(y)\})\big)$

Theorem

(*Theor. Comp. Sci., 2002, 271, pp.15–35*)

$\forall C \; \exists x, y$

$$KS(y) > KS(x) + C$$

and

$$KP(y) < KP(x) - C$$