Entropy and Coding

Nicola Prezza

References and Reading

[1] Chapter 2 of: Navarro, Gonzalo. Compact data structures: A practical approach. Cambridge University Press, 2016.

Exercises

1 Entropy (1) A binary *de Bruijn* sequence of order n is a *circular* string of length $N = 2^n$ over the alphabet $\{0,1\}$ containing all combinations of n bits as substrings (example: 11101000). It is well known that there are $\frac{2^{2^{n-1}}}{2^n}$ distinct binary de Bruijn sequences.

1.1 What is the worst-case entropy (as a function of n) of the set of all de Bruijn sequences of order n?

1.2 What is the worst-case entropy *per symbol* (i.e. divide by the sequence length) of the set of all de Bruijn sequences of length $N = 2^n$? how does this compare with the set of all binary sequences of length N?

1.3 Let S be a binary de Bruijn sequence. What is the zero-order empirical entropy of S?

1.4 Let S be a binary de Bruijn sequence of order n. What is the $(n-1)$-th empirical entropy of S? (note: take into account the circularity of S to compute symbols' contexts)

2 Entropy (2) A *Fibonacci word* is a binary string recursively defined as follows:

- $S_0 = 0$
- $S_1 = 1$
- $S_n = S_{n-1}S_{n-2}$

Let F_n be the n-th Fibonacci number. It is known that $\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \varphi$, where $\varphi = \frac{1+\sqrt{5}}{2}$ is the golden ratio. Using this fact, compute $\lim_{n\to\infty} H_0(S_n)$, i.e. the zero-order empirical entropy of S_n for n that tends to infinity.

3 Huffman As seen in Section 2.6.3 of [1], a text encoded with a canonical Huffman code can be decoded in $O(n \log\log n)$ time and $|\Sigma|\log|\Sigma| + O(\log^2 n)$ bits of space. Devise an optimal $O(n)$-time decoding algorithm using $O(n\sqrt{n})$ bits of space.