S&M
Split and Merge Compression Algorithm

By

Abdullah Hashim

Lempel & Ziv algorithm

Lempel & Ziv algorithm

To study the practical behaviour of the ***L&Z*** algorithm in the context of predicting the properties of the environment Θ, a dictionary of 512 words is used the first 256 words are the single ASCII character of the alphabet ***A***, the second 256 words are the multi-cheater words added to the dictionary over the 256 intervals. The initial probability distribution at interval n=0 of the 256 characters were assigned ten different sets of values, by fixing the initial probability of a leading character (c_1) to one of the ten values in the range $\{0.001 \leq p(c_1) \leq 1\}$, other character initial probabilities where made to be equal to $\{[(1- p(c_1)] / 255\}$. At each interval n, a new word added to the dictionary according to the initial probability distribution of the environment Θ. All words assumed to be equiprobable and the probability of each characters were recomputed. Let at interval n, the words w_1, w_2, … ,and w_k contains z_1, z_2, … , and z_k of the character α, the probability of character α at the n-th interval is given by:

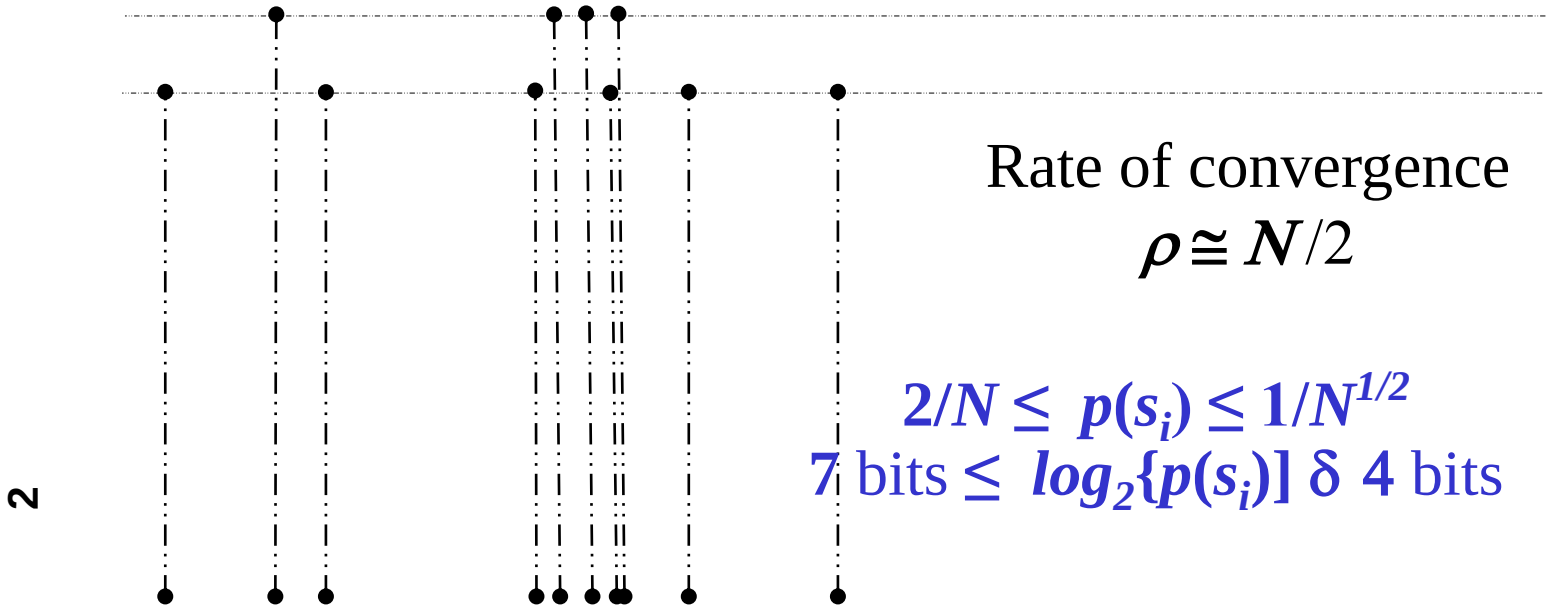
$$p(\alpha) = (1/n) \sum_{i=1}^{k} (z_i / |w_i|)$$

The behaviour of the algorithm is determined by plotting the average $Q(n)$ value over hundred (100) trials, for each of the ten different set of initial probabilities, against the intervals n, where $Q(n)$ is given by the expression:

$$Q(n) = \sum_{i=1}^{256} p^2(s_i) = p^2(s_1) + p^2(s_2) + \ldots + p^2(s_{256})$$

Lempel & Ziv algorithm

2 Results of Practical Simulation



Rate of convergence

$\rho \cong N/2$

$2/N \leq p(s_i) \leq 1/N^{1/2}$

7 bits $\leq log_2\{p(s_i)] \leq 4$ bits

3 db points