

Entropia empirica

Data la stringa s , la sua entropia empirica di ordine 0 è:

$$H_0(s) = \sum_{i=1}^m \frac{n_i}{n} \log \frac{n}{n_i},$$

dove n_i è il numero di occorrenze di σ_i in s . Comprimendo s con Huffman o codifica aritmetica si ottiene un messaggio di lunghezza vicina a $|s| H_0(s)$.

Definiamo anche l'entropia empirica di ordine k :

$$H_k(s) = \frac{1}{|s|} \sum_{w \in \Sigma^k} |w_s| H_0(w_s).$$

dove w_s è una stringa formata dai simboli che seguono le occorrenze di w in s (con ripetizioni). Fornisce un limite inferiore per un compressore con memoria k .