Combinatorial Information Distance*

Joel Ratsaby

Abstract

Let $|A|$ denote the cardinality of a finite set A. For any real number x define $t(x) = x$ if $x \geq 1$ and 1 otherwise. For any finite sets A, B let $\delta(A, B) = \log_2\left(t\left(\left|B \cap \overline{A}\right||A|\right)\right)$. We define a new cobinatorial distance $d(A, B) = \max\left\{\delta\left(A, B\right), \delta\left(B, A\right)\right\}$ which may be applied to measure the distance between binary strings of different lengths. The distance is based on a classical combinatorial notion of information introduced by Kolmogorov.

Keywords: *Set distance, Lempel-Ziv complexity, Combinatorial entropy, Set entropy, Binary sequences*