

18. Find the entropy from the given data set.

The entropy of a random variable is the average level of information, surprise, or uncertainty inherent to the variable's possible outcomes. Given a discrete random variable X which takes value in the alphabet x and distributed according to the $P(x)[0,1]$. The entropy is $H[X]$

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x)$$

The choice of base for log varies for different applications.

Base 2 gives the unit of bits while base e gives **natural units**.

Base e gives the units of $H(X)$.

An equivalent definition of entropy is the expected value of the self information of a variable.

