Basic information theory was viewed earlier in the course. Entropy in context of a discrete distribution is “a measure of the randomness or unpredictability of a sequence of symbols drawn” [1, 630] from such a distribution. The units of entropy depend on the number system used, but otherwise is a unit-less value. Entropy depends on the probabilities of the discrete items in the distribution, and not on the items themselves.

$$H = -\sum_{i=1}^{m} P_i \log_2 P_i = E[\log \frac{1}{P}] \tag{2}$$

The relative entropy also known as Kullback-Leibler distance is a measure between two probabilities over the same variable.

$$D_{KL}(p(x), q(x)) = \sum_{x} q(x) \ln \frac{q(x)}{p(x)} \tag{3}$$

$$D_{KL}(p(x), q(x)) = \int_{-\infty}^{\infty} q(x) \ln \frac{q(x)}{p(x)} \tag{4}$$

If there are two distributions, then there is a possibility of the distributions have information in common. A few exceptions arise due to the mutual information. Mutual information is the reduction of uncertainty about one variable due to information about another.