# Information Theory - Basics

Sanjeev Narayanan

29th April 2018

## 1 Introduction

Information theory provides a constructive criterion for setting up probability distributions on the basis of partial knowledge, and leads to statistical inference, also answers two fundamental concepts in communication theory - Entropy (H) and Channel capacity (C). However, Information theory is not restricted to merely communication theory. The figure below gives a broad relationship with other fields [1].

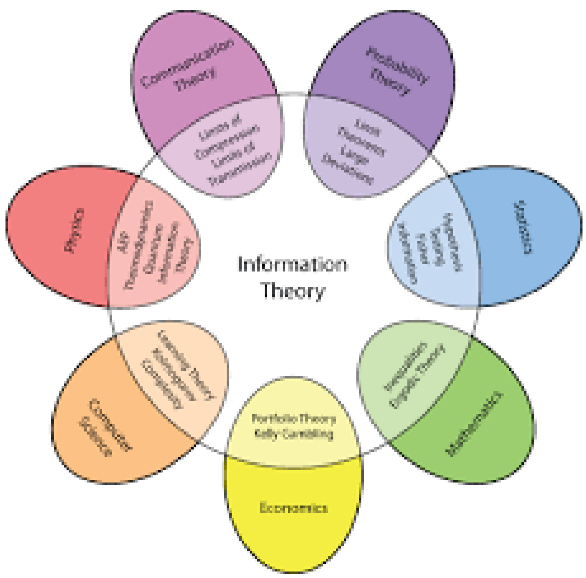


Figure 1: The relationship of information theory with other fields

## 2 Entropy, Relative Entropy and Mutual Information

### 2.1 Entropy

Entropy is defined as the uncertainty of a random variable. If $X$ is a random variable, with $p(x) = P(X = x)$, its probability mass function, the entropy can be mathematically defined as

$$H(X) = -\sum_{x \in X} p(x) log p(x) \tag{1}$$

The log is to the base 2 and entropy is expressed in bits.

### 2.2 Joint Entropy

The joint entropy H(X, Y) of a pair of discrete random variables (X, Y) with a joint distribution p(x, y) is defined as

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) log p(x, y) \tag{2}$$

### 2.3 Relative Entropy

The relative entropy is a measure of the distance between two distributions. $D(p||q)$ is a measure of the inefficiency of assuming that the distribution is q when the true distribution is p. Mathematically, it is defined as

$$D(p||q) = \sum_{x \in X} p(x) log \frac{p(x)}{q(x)} \tag{3}$$

Relative entropy is always non-negative and is zero if and only if p = q.

### 2.4 Mutual Information

The mutual information $I(X; Y)$ is the relative entropy between the joint distribution and the product distribution $p(x)p(y)$, i.e.,

$$I(X; Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) log \frac{p(x, y)}{p(x)p(y)} \tag{4}$$

### 2.5 Chain Rule

Let $X, Y$ and $Z$ be random variables in a topological ordering upto $n$ counts. We have the chain rule for entropy, mutual information and relative entropy defined as

$$H(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} H(X_i | X_{i-1}, ..X_1) \tag{5}$$

$$I(X_1, X_2, ..., X_n; Y) = \sum_{i=1}^{n} H(X_i; Y | X_{i-1}, ..X_1) \tag{6}$$

$$D(p(x,y)||q(x,y)) = D(p(x)||q(x)) + D(p(y|x)||q(y|x)) \tag{7}$$

## 3 Entropy Rates of a Stochastic Process

A stochastic process is an indexed sequence of random variables. It s said to be stationary if the joint distribution of any subset of the sequence of random variables is invariant with respect to shifts in the time index.

The entropy rate of a stochastic process is defined by

$$H(L) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, ...X_n) \tag{8}$$

if the limit exists

## References

[1] Sheri Edwards. Elements of information theory, thomas m. cover, joy a. thomas. john wiley & sons, inc.(2006), 2008.