Entropy and information (divergence) measures

Mohtashami Borzadaran, G. R.

Department of Statistics, Faculty of Mathematical Sciences
Ferdowsi University of Mashhad, Mashhad, Iran
Email: gmb1334@yahoo.com

The extension of notion for the measure of information with application in communication theory back to the experiences of the C. E. Shannon during the second World War II. In 1948, he introduced that the entropy is a real number associated with a random variable which is equal to the expected value of the surprise that we would receive upon getting a realization of the random variable. Let, $S_a(p) = -\log_a p$ (base-2 is often used) be the measure of surprise we feel when an event with the probability p is occurring actually occurs. Then, entropy for a random variable is calculated using the probability mass (or pdf) function of the random variable via $H_a(X) = E[S_a(p(X))]$. Some of the properties and characterizations of the Shannon entropy and its extension versions are mentioned here.

Also, finding expressions for the multivariate distributions (discrete or continuous) and information measure such as mutual information with some of their properties and discussing in view of copula are reviewed.

The principle of maximum entropy provides a method to select the unknown pdf (or pmf) compatible to entropy under a specified constraint. This idea was introduced by Jaynes 1957 and obtained via a theorem by Kagan et al. (1973). Applying to maximum Renyi or Tsallis entropy and also $\phi-$entropy, as a general format subsume many special cases. Similar arguments are applicable to a multivariate set-up.

In probability theory and information theory, the Kullback Leibler divergence (also information divergence, information gain, relative entropy) is a non-symmetric measure of the difference between two probability distributions. Specifically, the Kullback Leibler divergence is typically represents the "true" distribution of data and a theoretical model for approximation of the true distribution. Although it is often intuited as a metric or distance, the KL divergence is not a true metric. Various applications in statistics and properties of it is one of our aim in here. The link between maximum likelihood and maximum entropy and Kullback Leibler information is important for a discussion which is coming in this note. There are several types of information divergence measure that are studied in literature as extensions of the Shannon entropy and Kullback Leibler information. Some of them can be collected in Csiszar $\phi-$divergence as special cases. So, minimization of them is important and finding these optimal measures is the other direction that is discussed in this paper with the related special states such as Kullback Leibler information, χ^2-divergence, total variation, squared perimeter distance, Renyi divergence, Hellinger distance, directed divergence and so on.

Keywords: Entropy, Maximum entropy, Kullback Leibler information, Information measures, Minimization of Kullback Leibler information.