Inference of mutual information and its generalizations from data

Charles Zheng and Yuval Benjamini

November 5, 2016

Abstract

1 Introduction

The concept of "information" plays a key role in areas as diverse as game theory, biology, neuroscience, and human engineering. A random quantity $\boldsymbol{Y}$ carries information about $\boldsymbol{X}$ if observing $\boldsymbol{Y}$ reduces our uncertainty about $\boldsymbol{X}$. In game theory, $\boldsymbol{Y}$ could be the opponent's bet, which reveals information about $\boldsymbol{X}$, the opponent's hand. In neuroscience, $\boldsymbol{Y}$ is brain activity which correlates to visual stimulus, $\boldsymbol{X}$. In communications, $\boldsymbol{X}$ is a plaintext message which is coded and transmitted as $\boldsymbol{Y}$, a series of recieved bits.

It is in the context of communication system that Shannon first proposed a quantification of the concept of "the amount of information that $\boldsymbol{Y}$ carries about $\boldsymbol{X}$", in the form of *mutual information*: if $\boldsymbol{X}$ and $\boldsymbol{Y}$ have joint density $p(\boldsymbol{x}, \boldsymbol{y})$, then the mutual information is defined as

$$I(\boldsymbol{X};\boldsymbol{Y}) = \int p(\boldsymbol{x},\boldsymbol{y}) \log \frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{x})p(\boldsymbol{y})} dxdy.$$

where $p(\boldsymbol{x})$ and $p(\boldsymbol{y})$ are the marginal densities. The word "mutual" is appropriate given the symmetry of $I(\boldsymbol{X};\boldsymbol{Y})$. However, at the intuitive level it does not seem *a priori* necessary that the measure of information be symmetric; indeed, generalizations of mutual information (Verdú 2015) are asymmetric.

Having a quantitative definition of information, such as Shannon's $I(\boldsymbol{X};\boldsymbol{Y})$, opens the possibility of computing, estimating or inferring the quantity for $(\boldsymbol{X},\boldsymbol{Y})$ pairs of interest in the natural world. Classical neuroscience experiments estimated the mutual information between the reaching angle of a monkey's arm, X, and the average firing rate of particular motor neurons, Y. Once estimates are available for individual pairs, cross-pair comparisons are possible. Scientists can check if motor neurons in one particular region of the brain are more "informative" (on an individual level) than motor neurons in another brain region. Even more interestingly, one can take an ensemble of neurons as Y. Without collecting any additional data, scientists can compare the information of individual neurons Y_1 and Y_2, and compare each to the information about X carried by the ensemble (Y_1, Y_2). Comparisons between ensembles motivate thinking about a "calculus of information." If

$$I(X;Y_1) + I(X;Y_2) > I(X;(Y_1,Y_2)),$$

then we say that Y_1 and Y_2 carry *redundant* information. If, on the other hand,

$$I(X;Y_1) + I(X;Y_2) < I(X;(Y_1,Y_2)),$$

one can say that Y_1 and Y_2 have *synergy*–the information carried by the whole is greater than the sum of the information of each part.