ESTIMATING MUTUAL INFORMATION FROM DATA WITH ERROR BARS

Anthony Gotera[1,2], Charles T, Curry[3,4], Kevin R. Wheeler[5],
Karen A. Huyser[1], William B. Rossow[6], and Kevin H. Knuth[5]

(1) Ed. Assoc., NASA Ames Research Center, Moffett Field CA, USA
(2) California State University East Bay, Hayward CA, USA
(3) QSS, NASA Ames Research Center, Moffett Field CA, USA
(4) University of California Santa Cruz, CA, USA
(5) Intelligent Sys. Div., NASA Ames, Moffett Field CA, USA
(6) NASA Goddard Institute for Space Studies, New York NY, USA

Abstract

The information-theoretic quantity known as mutual information can be used as an indication of a relationship between two variables. While computation of the mutual information from data is straightforward, it is generally inaccurate and lacks a measure of the uncertainty in the result. We present a Bayesian method to estimate mutual information with error bars from data. First an optimal binning approach is utilized to infer the underlying probability density function from the data. This leads to a posterior probability, which is Dirichlet distribution. By directly sampling from the Dirichlet distribution, we obtain a chain of probability density functions from which we compute a chain of entropies with associated error bars. Because the mutual information is a sum and difference of entropies, we can compute an estimate of the mutual information along with its associated error bars. This technique is being applied to Earth science data, which is helping to identify relationships among climate variables.

Key Words: Mutual Information, Entropy, Optimal Binning, Density Functions, Dirichlet Distribution, Earth Science