Jensen-Shannon Divergence: Estimation and Hypothesis Testing

Ann Marie Stewart

Abstract

Jensen-Shannon divergence is one reasonable solution to the problem of measuring the level of difference or "distance" between two probability distributions on a multinomial population. If one of the distributions is assumed to be known a priori, estimation is a one-sample problem; if the two probability distributions are both assumed to be unknown, estimation becomes a two-sample problem. In both cases, the simple plug-in estimator has a bias that is $O(1/N)$, and hence bias reduction is explored in this dissertation. Using the well-known the jackknife method for both the one-sample and two-sample cases, an estimator with a bias of $O(1/N^2)$ is achieved. The asymptotic distributions of the estimators are determined to be chi-squared when the two distributions are equal, and normal when the two distributions are different. Then, hypothesis tests for the equality of the two multinomial distributions in both cases are established using test statistics based upon the jackknifed estimators. Finally, simulation studies are shown to verify the results numerically, and then the results are applied to real-world datasets.

Department of Mathematics and Statistics, UNC Charlotte, Charlotte, NC 28223