

"Inferences in Contaminated Density Models with Applications to Microarray Data Analysis"

Richard Charnigo,
University of Kentucky

(Joint research with Hongying Dai, Columbus State University)

ABSTRACT

Mixture models are useful for describing populations with heterogeneity that cannot be readily linked to a known covariate. A special kind of mixture model arises when the parameters for one of the mixture components are known a priori. I refer to such a mixture as a contaminated density model. Recently a contaminated beta density model has been applied to describe the p-values obtained from an immense number of hypothesis tests in contexts such as microarray data analysis. In this talk I discuss two procedures for testing the null hypothesis that a contaminated density can be simplified to an uncontaminated density: a modified likelihood ratio test and a weighted D test. These procedures avoid the pathology of the conventional likelihood ratio test, which does not behave as anticipated in the contaminated density setting since the model parameters are not identifiable under the null hypothesis. The limiting distributions of the modified likelihood ratio and weighted D test statistics are chi-square and noncentral chi-square under the null hypothesis and under contiguous local alternatives, respectively. Thus the modified likelihood ratio and weighted D test statistics can be validly compared to chi-square quantiles in the contaminated density setting, while the conventional likelihood ratio test statistic cannot. In addition, simulation studies show that the modified likelihood ratio and weighted D tests are more powerful for detecting contamination of a beta density than the Kolmogorov-Smirnov test. An analysis of a real microarray data set illustrates the utility of the contaminated beta density model as well as of the two testing procedures.