

Hypothesis testing in finite mixture models: from the likelihood ratio test to EM-test

Abstract

In the presence of heterogeneity, a mixture model is most natural to characterize the random behavior of the samples taken from such populations. Such strategy has been widely employed in applications ranging from genetics, information technology, marketing, to finance. Studying the mixing structure behind a random sample from the population allows us to infer the degree of heterogeneity with important implications in applications such as the presence of disease subgroups in genetics. The statistical problem is to test the hypotheses on the order of the finite mixture models. There has been continued interest in the limiting behavior of the likelihood ratio tests. The non-regularity of the finite mixture models has provided statisticians ample examples of unusual limiting distributions. Yet many of such results are not convenient for conducting hypothesis tests. Motivated at overcoming such difficulties, we have developed a number of strategies to obtain tests with high efficiency yet easy to use limiting distributions. The latest development is a class of EM-tests which are advantageous in many respects. Their limiting distributions are easier to derive mathematically, simple for implementation in data analysis and valid for more general class of mixture models without restrictions on the space of the mixing distribution. The simulation indicates the limiting distributions have good precision at approximating the finite sample distributions in the examples investigated.