Robust Detection of Multiple Outliers in Grouped Multivariate Data

C. Caroni
National Technical University of Athens, Greece

N. Billor
Cuñorova University, Turkey

Abstract

Various methods have been developed for detecting multiple outliers in a multivariate sample (for example, Caroni & Prescott, 1992, Applied Statistics 41, 355-364). A robust method with a high breakdown point is provided by the BACON algorithm (Billor, Hadi and Velleman, 2000, Computational Statistics and Data Analysis 34, 279-298). The purpose of the present study is to extend this method for application when the sample has been drawn from a mixture distribution and is consequently composed of possibly overlapping subgroups; outliers in this situation are points that are distant from every subgroup. The extended algorithm is developed for any number of subgroups and the properties of the method are investigated in relation to sample size, number of variables, number of subgroups, separation between subgroups and location of the outliers. If the number of subgroups can not be specified a priori, then the method also has a useful application as a form of cluster analysis.

Key words: Outlier detection, robust methods, mixture distribution, cluster analysis