

A Robust Estimation of Location and Scatter

Maman A. Djauhari

*Faculty of Mathematics and Natural Sciences,
Institut Teknologi Bandung
Jalan Ganesa 10, Bandung 40132, Indonesia
E-mail: maman@dns.math.itb.ac.id*

ABSTRACT

Statisticians face increasingly the task of analyzing large and high dimension multivariate data sets. This is due to the advances in computer technology which have facilitated greatly the collection of large data sets and, on the other hand, to the fact that most statistical experiments are multivariate in nature. One of the primary problems encountered in this task is robust estimation of location and scatter. In the literature the most popular and widely used robust parametric method for such parameter estimation is the so-called Fast MCD. However, although it is affine-equivariant and has high breakdown point, it is not apt when the data sets are of high dimension because its computational efficiency becomes lower. This is a direct consequence from the use of Mahalanobis distance or, equivalently, Mahalanobis depth in data ordering process which needs the inversion of covariance matrix and the use of MCD as the objective function. In this paper we propose a method which is as effective as Fast MCD but computationally more efficient. For this purpose, in multivariate ordering step, we use a new depth function which is equivalent to Mahalanobis depth and has lower computational complexity. Furthermore, in data concentration step, we use vector variance as the measure of multivariate scatter instead of covariance determinant and we replace the objective function MCD with minimum vector variance to reduce the complexity of this step. At the end of the paper we illustrate the effectiveness of this method using a simulation experiment.

Keywords: affine-equivariant, breakdown point, center-outward ordering, data depth, multivariate scatter, robust estimation of location and scatter.