



Clustering of subsample means based on pairwise L_1 regularized empirical likelihood

Quynh Van Nong¹ · Chi Tim Ng²

Received: 18 April 2018 / Revised: 30 October 2019 / Published online: 12 December 2019
© The Institute of Statistical Mathematics, Tokyo 2019

Abstract

To classify a vast amount of strata or subsamples with unknown families of distributions according to their strata-means, a clustering approach is developed based on pairwise L_1 regularized empirical likelihood. Under such a clustering approach, all possible contradictory conclusions are ruled out automatically. On the contrary, the decision rules based on many existing pairwise comparison procedures can generate contradictory results. Moreover, under certain mild conditions, the proposed clustering method enjoys the consistency property that with probability going to one, all strata are classified correctly. An exterior point algorithm is presented for the clustering. The applications of the proposed methods are demonstrated using stock market data and microarray data of breast cancer patients.

Keywords Clustering · Empirical likelihood · Exterior point algorithm · Pairwise mean comparison · Pairwise L_1 regularization

✉ Chi Tim Ng
easterlyng@gmail.com

Quynh Van Nong
quynhvandhsp@gmail.com

¹ Department of Mathematics, Thai Nguyen University of Education, Luong Ngoc Quyen Street, Thai Nguyen City, Vietnam

² Department of Statistics, Chonnam National University, 77, Yongbong-ro, Buk-gu, Gwangju 61186, South Korea