Doctoral Dissertation

Clustering of strata means based on pairwise L1 regularized empirical likelihood

Department of Statistics

Graduate School of Chonnam National University

Nong Quynh Van

August 2019

Doctoral Dissertation

Clustering of strata means based on pairwise L1 regularized empirical likelihood

Department of Statistics

Graduate School of Chonnam National University

Nong Quynh Van

Directed by Professor Chi Tim Ng

Department of Statistics, Chonnam National University

August 2019

Clustering of strata means based on pairwise L1 regularized empirical likelihood

Department of Statistics

Graduate School of Chonnam National University

Nong Quynh Van

The dissertation entitled above, by the graduate student Nong Quynh Van, in partial fulfillment of the requirements for the **Doctor of Philosophy in Statistics** has been deemed by the Professors below.

Jang Sun Baek, Ph.D .....

Jae Sik Jeong, Ph.D .....

Myung Hwan Na, Ph.D .....

Chi Tim Ng, Ph.D .....

Woo Joo Lee, Ph.D .....

August 2019

Dedicated to

My Parents, My Younger Brother and My Family

Acknowledgements

First and foremost, I am cordially thankful for my supervisor, Dr. Chi Tim Ng who offered the opportunity to study PhD degree for me. Without his support, guidance, knowledge and patience, this research would not have been possible. I dare to say that he is a good PhD advisor.

I wish to express my gratitude to the Vietnamese Ministry of Education and Training.

I would like to thank all students and staff of the Department of Statistics of Chonnam National University for their friendly and good social environment. I also would like to thank members of the Mathematical Statistics Research Laboratory, especially Nguyen Van Cuong, Zhang Lili, Zhang Kaimeng, 박진경, 김태경 for their great help and friendship.

I would like to dedicate this thesis to my parents and my brother who have given me the eternal love and have encouraged me to pursue the long academic journey. Without their sacrifices, I would not have finished my studies and my thesis.

Finally, I would like to take this opportunity to thank my husband for his love and unconditional support. The last word goes to my lovely children, Ngo Bao Chau and Ngo Duc Thanh who have brought me the true meaning of love and have given me the strength to get all things done.

Contents

List of Tables

List of Figures

Abstract

To determine the pairwise equalities-in-mean of a vast amount of subsamples with unknown distributions, a clustering approach is developed based on L_1 regularized empirical likelihood. Under the 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, the equalities-in-mean of all pairs of subsamples can be determined 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.