## A two-stage penalized logistic regression approach to case-control genome-wide association studies

 $Zehua \ Chen$ 

National University of Singapore Department of Statistics and Applied Probability 3 Science Drive 2, Singapore 117546 stachenz@nus.edu.sg

## Abstract

In this talk, we discuss a two-stage penalized logistic regression approach to case-control genome-wide association studies. This approach consists of a screening stage and a selection stage. In the screening stage, main-effect and interaction-effect features are screened by using  $L_1$  penalized logistic likelihood in a tournament procedure. In the selection stage, the retained features are ranked by the logistic likelihood with SCAD penalty and Jeferry's Prior penalty, a sequence of nested candidate models are formed, and the models are assessed by a family of extended Bayesian information criteria. The approach is justified by its asymptotic property: the false discovery rate and the positive detection rate converge to 0 and 1 respectively as sample size goes to infinity. Its performance in finite sample case is studied by simulation studies. It is compared with a pair-wise multiple testing approach and the LASSO-patternsearch algorithm. The approach is also applied to the analysis of CGEMS prostate cancer data.