Stockholm University

Mathematical Statistics
Stockholm University
Master Thesis **2015:6**
http://www.math.su.se

Comparison of logistic and ridge regression in genetic association studies

Hanna Fues*

September 2015

Abstract

Genetic association studies are used to find regions of the genome that contribute to a specific disease by testing for an association between disease status and genetic variation(s). We use single nucleotide polymorphisms (SNPs) to investigate the genetic variability. SNPs are locations where single nucleotides differ on the DNA in at least 1% of the population. From a statistical point of view these investigations are commonly performed using logistic regression techniques, modeling disease risk as a function of markers, i.e. SNPs, but the need for a penalized regression approach arises when many markers are correlated. Furthermore, in genetic association studies one often has the situation where the number of markers studied exceed the number of observations, increasing the need for a penalized approach. In this thesis interest lies in the analysis of a genomic region, containing highly correlated markers, in relation to breast cancer risk. We do so by studying data on 89 050 individuals part of the Breast Cancer Associ- ation Consortium (BCAC) and compare logistic regression and ridge regression techniques, for localizing independent signals among the multiple markers in said genomic region. Furthermore, we use a data-driven method for estimation of the penalization parameter, proposed by Cule and De Iorio [2013], in our ridge regression analyses. We find that both regression approaches give similar results, deeming markers from the same genomic region as being significantly associated with breast cancer risk. Our analysis is the first step in a long chain of events leading to the identification of locations that are associated with breast cancer risk and this thesis gives an important indication as to what region future investigations should be focusing on.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: hannafues@gmail.com. Supervisor: Michael Höhle.