
KJELL DOKSUM, Univ. of Wisconsin, Dept. of Statistics, 1300 Univ. Ave, Madison, WI
Genomics, Matrices, and Statistical Inference

Genomics research often involves a 0-1 response such as “control” and “case” (nondisease and disease) and a number of covariates such as genetic markers (up to 1.2 million at this date), population membership, conventional covariates such as age, sex, body mass index (BMI), and treatment levels. Population geneticists have developed clever procedures for analysing data from experiments involving such genomics data. These procedures involve using what could be called inverse principal component analysis (PCA) and rely on matrix equations that connect regular and inverse PCA, where inverse PCA refers to PCA with variables exchanged with sample values. That is, inverse PCA is PCA with the design matrix transposed. We examine the properties of the procedures proposed based on this inverse PCA. The population geneticists propose intuitive models for their data. However, in their analysis they ignore their own models and use simple Stat 101 nonparametric chi-square tests. We compare these tests with likelihood ratio tests and Bayesian procedures. More generally, we examine statistical inference appropriate for the models proposed by population geneticists.

This is joint work with Kam Tsui and Fan Yang.