Chapter 2

**Interaction Screening for Ultra-High Dimensional Data (Title of iForm paper)**

2.1 Introduction

Variable selection is a heavily researched area in statistical literature. Aspects that need to be considered when selecting predictors for a final model are incorporating meaningful predictors, maintaining accuracy of estimation and the computational cost of conducting the procedure. Fan and Lv (2008) proposed the property of sure independent screening as a way of identifying meaningful variables. Their property ensures that important variables survive after applying variable screening procedures with probability tending to 1.

Recently, Candes and Tao have proposed the Dantzig selector using *L*1-regularization and showed that it achieves the ideal risk up to a logarithmic factor log*.p/*. Their innovative procedure and remarkable result are challenged when the dimensionality is ultrahigh

Sure independence screening : a property that all the important variables survive after applying a

variable screening procedure with probability tending to 1.

Interaction models provide a better approximation to the response surface, improve prediction accuracy, and bring new insight on the interplay between predictors. (Interaction Screening for Ultra-High Dimensional Data, Hoa and Zhang 2014)

2.2 Model

* iForm Results
  + Adjusted R-Sqaures
  + Markers selected in final model
    - Identify cis and trans-QTLs
  + Look for the “hotspots” identified in 2010 c. elegans paper

iForm Selection Algorithm

D is chose as a reasonable upperbound of to terminate the procedure. BIC can also be used for the optimal stopping position. Since the main effects and interaction effects are chosen dynamically we only need to use the selection criteria 1 time.

2.3 Results

2.4 Discussion