***Joint Colloquium***

UNIVERSITY OF GEORGIA
DEPARTMENT OF STATISTICS

DEPARTMENT OF EPIDEMIOLOGY AND BIOSTATISTICS

**Yongdai Kim**

**Seoul National University, South Korea**

**“Sparse Regression with Incentive”**

 Spare regularization methods for high dimensional regression have received much attention recently as an alternative of subset selection methods. Examples are lasso (Tibshirani 1996), bridge regression (1993), scad (2001), to name just few. An advantage of sparse regularization methods is that it gives a stable estimator with automatic variable selection and hence the resulting estimator performs well in prediction. Also, sparse regularization methods have many desirable properties when the true model is sparse.

 However, there are several disadvantages of sparse regularization methods as discussed by Zou and Hastie (2005). First, when p > n, the solution has at most n many nonzero coeﬃcients. Second, if there is a group of covariates whose correlations are very high, the solution usually takes one covariate from the group and does not care which one is selected. Third, empirically, where there high correlations between covariates, the prediction performance of sparse solutions are dominated by non-sparse solutions such as ridge regression (Tibshirani 1996) where all of the highly correlated covariates are used.

 Hence, there is a need of less sparse solutions than the aforementioned sparse regularization methods. To make a less sparse solution, Zou and Hastie (2005) proposed the elastic net penalty, which is a linear combination of the lasso and ridge penalties. Friedman and Popescu (2004) proposed the gradient directed regularization which is a modiﬁed gradient descent method where more than one predictor variables are updated at each iteration.

 There are limitations in the elastic net and modiﬁed gradient directed regularization. The elastic net requires the rescaling of the solution to avoid overshrinkage. Even though Zou and Hastie (2005) proposed the rescaling factor heuristically for linear regression, it is not clear how to rescale the solution for other problems such as logistic regression. The estimator of the gradient directed regularization is not deﬁned by a minimizer of a penalized empirical risk and hence it is diﬃcult to study properties of the estimator.

 In this talk, we propose a new regularization method called the sparse regression with incentive (SRI) to overcome deﬁciencies of the elastic net and gradient directed regularization. Advantages of the SRI over the elastic net and gradient directed regularization is that the estimator is deﬁned by a minimizer of a penalized empirical risk and it does not require the post-hoc rescaling as the elastic net does. Also, we can incorporate prior information of the group structure of covariates.

**Thursday August 26, 2010**

ROOM 306

Statistics Building

University of Georgia

Athens, GA 30602

3:30 P.M. – Room 306, Statistics Building

**Refreshments following talk at 4:30 P.M. in room 230 (The Cohen Room)**