# A Perturbation Method for Inference on Regularized Regression Estimates

*Tianxi Cai*
*Department of Biostatistics*
*Harvard University*
*Boston, MA 02115*
tcai@hsph.harvard.edu

**Abstract**

Analysis of high dimensional data often seeks to identify a subset of important features and assess their effects on the outcome. Traditional statistical inference procedures based on standard regression methods often fail in the presence of high-dimensional features. In recent years, regularization methods have emerged as promising tools for analyzing high dimensional data. These methods simultaneously selects important features ad provide stable estimation of their effects. Adaptive LASSO and SCAD for instance, give consistent and asymptotically normal estimates with oracle properties. However, in finite samples, it remains difficult to obtain interval estimators for the regression parameters. In this paper, we propose perturbation resampling based procedures to approximate the distribution of a general class of penalized parameter estimates. Our proposal, justified by asymptotic theory, provides a simple way to estimate the covariance matrix and confidence regions. Through finite sample simulations, we verify the ability of this method to give accurate inference and compare it to other widely used standard deviation and confidence interval estimates. We also illustrate our proposals with a data set used to study the association of HIV drug resistance and a large number of genetic mutations.