

DISCUSSION: “A SIGNIFICANCE TEST FOR THE LASSO”¹

BY JINCHI LV AND ZEMIN ZHENG

University of Southern California

Professors Lockhart, Taylor, Tibshirani and Tibshirani are to be congratulated for their innovative and valuable contribution to the important and timely problem of testing the significance of covariates for the Lasso. Since the invention of the Lasso in Tibshirani (1996) for variable selection, there has been a huge growing literature devoted to its theory and implementation, its extensions to various model settings and different variants and developing more general regularization methods. Most of existing studies have focused on the prediction, estimation and variable selection properties ranging from consistency in prediction and estimation to consistency in model selection in terms of recovery of the true underlying sparse model. The problem of deriving the asymptotic distributions for regularized estimators, as the global or computable solutions, in high dimensions is relatively less well studied.

How to develop efficient significance testing procedures for the regularization methods is particularly important since in real applications one would like to assess the significance of selected covariates with their p -values. Such p -values are also crucial for multiple comparisons in testing the significance of a large number of covariates simultaneously. In contrast to the use of some resampling or data splitting techniques for evaluating the significance, in the present paper Lockhart, Taylor, Tibshirani and Tibshirani propose a novel powerful yet simple covariance test statistic T_k for testing the significance of the covariate X_j that enters the model at the k th step of the piecewise linear Lasso solution path in the linear regression model setting. Such a test statistic is shown to have an exact $\text{Exp}(1)$ asymptotic null distribution in the case of orthonormal design matrix and the case of $k = 1$ (i.e., the global null with zero true regression coefficient vector) for general design matrix. In the general case, the $\text{Exp}(1)$ distribution provides a conservative asymptotic null distribution. The significance test for the Lasso proposed in the paper is elegant thanks to its simplicity and theoretical guarantees in high dimensions.

We appreciate the opportunity to comment on several aspects of this paper. In particular, our discussion will focus on four issues: (1) alternative test statistics, (2) the event B and generalized irrerepresentable conditions, (3) model misspecification, and (4) more general regularization methods.

Received December 2013.

¹Supported by NSF CAREER Award DMS-09-55316.