The Annals of Statistics
2014, Vol. 42, No. 2, 501–508
DOI: 10.1214/13-AOS1175E
Main article DOI: 10.1214/13-AOS1175

DISCUSSION: "A SIGNIFICANCE TEST FOR THE LASSO"

BY LARRY WASSERMAN

Carnegie Mellon University

The paper by Lockhart, Taylor, Tibshirani and Tibshirani (LTTT) is an important advancement in our understanding of inference for high-dimensional regression. The paper is a tour de force, bringing together an impressive array of results, culminating in a set of very satisfying convergence results. The fact that the test statistic automatically balances the effect of shrinkage and the effect of adaptive variable selection is remarkable.

The authors make very strong assumptions. This is quite reasonable: to make significant theoretical advances in our understanding of complex procedures, one has to begin with strong assumptions. The following question then arises: what can we do without these assumptions?

1. The assumptions. The assumptions in this paper—and in most theoretical papers on high-dimensional regression—have several components. These include:

(1) The linear model is correct.
(2) The variance is constant.
(3) The errors have a Normal distribution.
(4) The parameter vector is sparse.
(5) The design matrix has very weak collinearity. This is usually stated in the form of incoherence, eigenvalue restrictions or incompatibility assumptions.

To the best of my knowledge, these assumptions are not testable when $p > n$. They are certainly a good starting place for theoretical investigations but they are indeed very strong. The regression function $m(x) = \mathbb{E}(Y|X = x)$ can be any function. There is no reason to think it will be close to linear. Design assumptions are also highly suspect. High collinearity is the rule rather than the exception especially in high-dimensional problems. An exception is signal processing, in particular compressed sensing, where the user gets to construct the design matrix. In this case, if the design matrix is filled with independent random Normals, the design matrix will be incoherent with high probability. But this is a rather special situation.

None of this is meant as a criticism of the paper. Rather, I am trying to motivate interest in the question I asked earlier, namely: what can we do without these assumptions?

Received December 2013.