The Annals of Statistics
2014, Vol. 42, No. 2, 483–492
DOI: 10.1214/13-AOS1175C
Main article DOI: 10.1214/13-AOS1175

DISCUSSION: "A SIGNIFICANCE TEST FOR THE LASSO"

BY JIANQING FAN[1] AND ZHENG TRACY KE[2]

Princeton University

We wholeheartedly congratulate Lockhart, Taylor, Tibshrani and Tibshrani on the stimulating paper, which provides insights into statistical inference based on the lasso solution path. The authors proposed novel covariance statistics for testing the significance of predictor variables as they enter the active set, which formalizes the data-adaptive test based on the lasso path. The observation that "shrinkage" balances "adaptivity" to yield to an asymptotic Exp(1) null distribution is inspiring, and the mathematical analysis is delicate and intriguing.

Adopting the notation from the paper under discussion, the main results are that the covariance statistics (Theorem 1)

$$(T_{k_0+1}, T_{k_0+2}, \dots, T_{k_0+d}) \xrightarrow{d} \bigl(\mathrm{Exp}(1), \mathrm{Exp}(1/2), \dots, \mathrm{Exp}(1/d)\bigr) \tag{1}$$

for orthogonal designs, and under the global null model (Theorem 2), $T_1 \xrightarrow{d} \mathrm{Exp}(1)$, and under the general model (Theorem 3), $P(T_{k_0+1} \geq t) \leq \exp(-t) + o(1)$. These remarkable results are derived under a number of critical assumptions such as the normality, the sure screening [borrowing the terminology of Fan and Lv (2008)] or model selection consistency of the lasso path. As pointed out in Fan and Li (2001), lasso introduces biases that are hard to account for. This together with the popularity of lasso give rise to the importance of this work, which results in informal statistical inference for the lasso. We welcome the opportunity to make a few comments.

1. Asymptotic null distributions. A natural question is how accurate the approximation (1) is and whether it holds for more general design matrices. We illustrate this using a small-scale numerical study. We take the same settings as in Section 5.2 (Table 2) by considering the global null true model with four types of design matrices: orthogonal, equal correlation, $AR(1)$ and block diagonal, where the parameter $\rho = 0.8$. We fix $n = 100$ and $p = 10$ and 50. When $p = 50$, the marginal distributions of $\{T_1, T_2, T_3\}$ are very close to the theoretical ones given by (1). However, when $p = 10$, the approximation is not accurate for the "equal correlation" and "$AR(1)$" designs. Figure 1 depicts the results for $p = 10$. The accuracies for the "orthogonal" and "block diagonal" designs are reasonable (omitted) and the accuracy for T_3 is in general worse than those for T_1 and T_2.

Received December 2013.
[1]Supported by NIH Grants R01-GM072611 and R01-GM100474 and NSF Grant DMS-12-06464.
[2]Supported by NIH Grant R01-GM072611 and NSF Grant DMS-12-06464.