

Comments on: High-dimensional simultaneous inference with the bootstrap

Arindam Chatterjee¹

Published online: 9 October 2017
© Sociedad de Estadística e Investigación Operativa 2017

Professors Dezeure, Bühlmann and Zhang are to be congratulated for presenting this nicely written article. They provide theoretical justification for bootstrap-based statistical inference in high-dimensional linear models, for both homoscedastic and heteroscedastic errors which can possibly be non-Gaussian. They have shown that, for each j , the studentized statistic T_j , and its suitably defined bootstrap version T_j^* , are pivotal quantities and they converge marginally to standard normal distribution. They also prove bootstrap distributional consistency of $\max_{j \in G} T_j^*$, when the size of G can be as large as p . This holds *irrespective* of the true value of the underlying regression coefficient. This is an extremely important feature of their proposed methodology, specially in the context of high-dimensional inference.

In particular, they have shown consistency of the residual bootstrap, which is a commonly used bootstrap technique for linear models with fixed design and also the wild bootstrap, which can deal with possible heteroscedasticity. As a consequence of their results, it is possible to construct simultaneous confidence intervals (CIs) and tests for an increasing number of regression coefficients, without imposing any ‘beta-min’-type condition on the underlying regression parameter. The key idea is the use of a *regular* estimator, which avoids the superefficiency phenomenon associated with bootstrap for Lasso-type sparse estimators.

It is interesting to note that the construction of de-sparsified Lasso estimator uses the naive Lasso estimator $(p + 1)$ times. Firstly, the Lasso is used for computing the residuals Z_j and then for obtaining the bias correction term. It is reasonable to ask whether

This comment refers to the invited paper available at: doi:[10.1007/s11749-017-0554-2](https://doi.org/10.1007/s11749-017-0554-2).

✉ Arindam Chatterjee
cha@isid.ac.in

¹ Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Delhi, New Delhi 110016, India

the Lasso estimator is an *optimal* choice in some sense for this purpose. Optimality could be defined in terms of coverage errors for confidence intervals for individual coefficients, or an increasing number of coefficients, or in terms of controlling the FWER in multiple testing, or some other suitable manner. Obviously, the Lasso seems to be the most natural choice and computationally least demanding, in comparison with other sparsity inducing alternatives, viz., SCAD, MCP or any other sparse estimator. Since the SCAD and MCP are nearly unbiased, these estimators might provide a better choice for the bias correction part, although it may require some assumptions on the underlying regression coefficient. In order to find an answer, a finer analysis of the asymptotic properties of the de-sparsified Lasso may be required, which takes into account the effects of using the Lasso in the initial stage.

Also, as seen from the simulation results, the coverage accuracy of the bootstrapped de-sparsified Lasso-based CIs is superior to that obtained by the naive de-sparsified Lasso and that obtained by the bootstrapping technique proposed by [Zhang and Cheng \(2016\)](#). This implies that the bootstrapped de-sparsified Lasso reduces coverage errors as expected. The proofs provided in the article indicate that the asymptotic normality for individual components of the bootstrapped de-sparsified Lasso estimator uses the linearity of the leading term in the expansion for $\sqrt{n}(\hat{b}_j - \beta_{0,j})$. As stated by the authors, there is little theoretical difference (in terms of distributional convergence) between bootstrapping the entire de-sparsified Lasso estimator and bootstrapping its leading linear term, as done by [Zhang and Cheng \(2016\)](#). However, the numerical results point out the effects of including the remainder term in the expansion for the de-sparsified Lasso estimator. It is likely that including this term can have higher-order effects on the coverage accuracy of the resulting CIs. Developing a higher-order asymptotic theory becomes an important issue in this context. This will enable us to clearly understand the reasons for reduced coverage errors obtained by using the bootstrapped de-sparsified Lasso estimator. A more challenging problem is to develop a similar higher-order asymptotic theory for the statistic $\max_{j \in G} T_j$ and its bootstrap version, where the size of the set G is increasing with the sample size n .

At this point, it is important to note that from the practitioners point of view, who is facing a statistical modeling problem with an enormously large number of variables, the problem of inference after model selection remains [cf. [Lee et al. \(2016\)](#)].

This article is definitely an important theoretical contribution in the context of inference in high-dimensional linear models, and it widely extends the applicability of the bootstrap in this setup. The de-sparsification approach can possibly be extended to other sparse estimators used in the context of high-dimensional linear models. There are some challenging theoretical issues which should be investigated to uncover the finer aspects of the de-sparsified Lasso estimator and its bootstrap version.

References

- Lee JD, Sun DL, Sun Y, Taylor JE (2016) Exact post-selection inference, with application to the Lasso. *Ann Stat* 44(3):907–927
- Zhang X, Cheng G (2016) Simultaneous inference for high-dimensional linear models. *J Am Stat Assoc*. doi:[10.1080/01621459.2016.1166114](https://doi.org/10.1080/01621459.2016.1166114)