

Lasso confidence intervals: Another look at the bootstrap

Logan Harris
Department of Biostatistics
University of Iowa

Patrick Breheny
Department of Biostatistics
University of Iowa

August 11, 2024

Abstract

Abstract goes here. Test.

1 Introduction

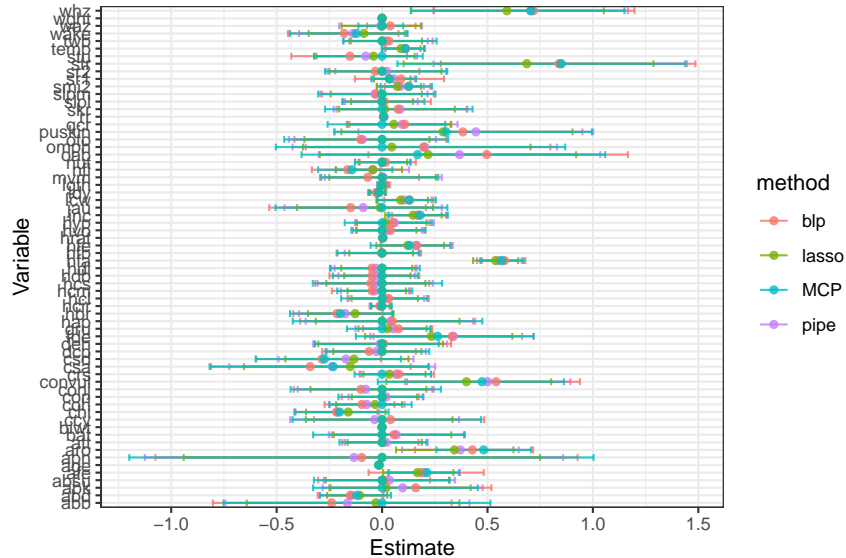


Figure 1: Confidence intervals produced by three different methods for all 66 variables in the `whoari` dataset ...

2 Random thoughts / stuff from other papers

As covered in [cite previous paper](#), a simple debiasing procedure only accounts for the first order bias.

That is, for a given variable, it directly accounts for the penalty induced bias but it does not account for the effect the of the bias introduced by biased estimates that a given variable is correlated with. This, what we call second order, bias can still partially be chalked up to the penalty but occurs in a somewhat back door way.

The effect of this bias suggest that the method is not useful, however, the conditions considered in this previous manuscript are not necessarily all that fair. In fact, the main simulation set up considered is primed to highlight this bias. Furthermore, this manuscript also indicated that λ_{CV} is not necessarily the best λ for the Debiased method.

Additionally as mentioned in that previous manuscript, if one is concerned with covering larger magnitude values of β , then we suggested that the lasso penalty may not be the best option and that the practitioner may want to consider alternative penalties, such as the Minimax Concave Penalty (MCP).

As an extension of these two ideas, we then bring these two streams of thought together and present the debiased method for MCP.

Note that when minimizing CVE, we are not necessarily minimizing the bias of the estimates... but rather the error in the holdout set. Estimating the bias is a much more difficult task

- **Debiased bootstrap:** Uses the correlation between \mathbf{x}_j^b and the partial residuals obtained from the model for each bootstrap sample as the bootstrap draw for variable j .

Regardless, although Debiased does reduce the rate of intervals missing toward zero for larger magnitude β s, it is unable to correct all the bias, still missing below at a rate higher than ideal.

The primary point to note for Debiased is that the value of λ that produces closest to nominal coverage is outside of the range of λ s selected by CV. Furthermore, this value occurs at a place that appears to produce nominal coverage across the range of β values. As will be covered in the Discussion, the bias of z_j is directly related to the bias of $\hat{\beta}_{-j}$. The smaller value of λ providing the best performance in this scenario suggests that the performance of Debiased is optimized when λ is as small as possible to avoid bias explicitly introduced by λ but large enough to produce a stable estimate for β . For Debiased, the far left hand side (λ_{\max}) represents marginal regression estimates whereas if $\lambda \rightarrow 0$ (towards the right) would be increasingly similar to fully joint estimates for a multiple linear regression model.

This leads to a limitation of the methods proposed here. There is additional bias in the estimates beyond the bias introduced by the penalty. This is most easily observed when considering the debias method but applies to all four methods. Recall that the estimate used for the debiased method is for variable j is $\mathbf{z}_j = \frac{1}{n} \mathbf{x}_j^T \mathbf{r}_j$. As shown in Breheny (2019), in the non-orthogonal case,

$$\frac{1}{n} \mathbf{x}_j^T \mathbf{r}_j = \beta_j^* + \frac{1}{n} \mathbf{x}_j^T \epsilon + \frac{1}{n} \mathbf{x}_j^T \mathbf{X}_{-j} (\beta_{-j}^* - \hat{\beta}_{-j}).$$

Focusing on the last term, this shows that there is bias in z_j from the bias in estimating all other β s that variable j is correlated with. Of course, z_j is used in the construction of the full conditional posterior and the traditional estimate is essentially z_j with the penalty applied. This indicated that such bias is pervasive in each method. Since this bias depends on the sample correlation, this bias can be present even when features are truly independent. Furthermore, during bootstrap resamples, chance correlation will occur even if the original dataset was generated with orthogonality. That said, this appears not to be considerably detrimental in practice as the methods fare well even with high levels of correlation. Future work could be done to correct this bias, however, consideration should be given estimating this bias for each $\hat{\beta}_j$ depends on accurately estimating the bias of all other coefficients.

I actually think the main example (laplace) here is a very severe case in terms of the bias that would be observed.

References

BREHENY, P. J. (2019). Marginal false discovery rates for penalized regression models. *Biostatistics*, **20** 299–314.

Supplemental materials

Nothing