Response to critiques for article EJS2102-039

Tun Lee Ng and Michael A. Newton

March 1, 2022

The critiques of EJS2102-039 point to clear limitations of that manuscript. We thank the editorial team for preparing a careful review of our manuscript, and we are grateful for the opportunity to present a further revision which fully addresses the issues raised. We hope you find the quality and significance of the revised work to meet the EJS standards.

In the following we provide responses to specific points, numbering comments as **ReviewerID.CommentNumber**.

1 Associate Editor

The paper is nice and is much improved.

(AE.1) I think the main problem is that purpose of the results has not been clearly explained and has got lost in search of a Bayesian interpretation. The referee is right about the complaint of the unnecessary complication of the randomly weighted lasso as an estimator when the ordinary lasso does the job fine. The focus should be about assessing uncertainty, when the approach produces an alternative to a Bayesian posterior, much like the bootstrap. The discussion at the beginning of Section 4 does go well and confuses the referee because the referred Bernstein-von Mises theorem for the high-dimensional sparse regression problem is unclear. (Even for fixed dimensional, if the error distribution is not parameterized, the parametric Bernstein-von Mises theorem in van der Vaart's book does not apply.) Thus I will recommend that the authors rewrite the paper to focus on uncertainty quantification and confidence region construction, rather than trying to give a far-fetched Bayesian interpretation, and also answer to the referee's questions.

Reply: The Associate Editor's comments are incisive. In the revision presented here, we have done exactly as suggested. We rewrote the paper with a focus on uncertainty quantification. We have completely downplayed the Bayesian angle. For example, we removed the difficult Section 4.1, we no longer emphasize Bayesian nonparametric derivations of random weighting in the introduction, and we drop the speculative comments about Bernstein-von Mises. This restructuring does not alter the mathematical/statistical contributions in any way; the sequence of technical results is retained as in the previous version. But the surrounding discussion is much better focused on the specific findings. In a final discussion section (new Section 5) we point to future work that may aim to establish connections to Bayesian methods, recognizing some recent literature on the topic, but we follow the AE's recommendation and drop our efforts to build such a connection with the present work.

2 Referee

The authors have presented a much improved article and their replies to my comments on the earlier version are quite satisfactory. This revised version is specially clear on the statement of results and also provides a wide simulation study and a real-data analysis example. The article, despite its improvements has failed to answer some issues which I point out below. It could be possibly because I may have missed out obvious connections to the Bayesian setup, due to my lack of expertise in that area.

(R1.1) If I have a data-set with (y, X), I am still not quite sure why I would run the weighted Lasso with penalized weights scheme. If I think of a parameter β_0, then I can use the standard Lasso toolbox based results to get model-selection, sign-consistency, limit laws, etc. Obviously, they would be done under suitable assumptions, which are the same assumptions that you make in your article. So, what is the need for bringing in the results you have presented? Do I need to approach the data-set with a different viewpoint about β_0 (a Bayesian viewpoint)? If so, then what is that alternative viewpoint, and how do your results help. Unfortunately, I have not got answer to this basic question.

And, I do not understand how to interpret the signs of Lasso estimates obtained from a randomly weighted criterion function. In Section 5, where you presented the real-data analysis, you talk about *marginal posterior distributions of β's obtained from these four methods*. So, indeed there is a prior on β_0? Then, where is that prior, when you develop asymptotic theory? If there is a connection then it will be better if you point it out clearly in your article.

Reply: In this revision, we have completely rewritten the opening and closing sections in order to better contextualize the new statistical results. To the referee's basic question, random weighting produces output aimed at uncertainty quantification in settings, such as LASSO regression, where a penalized objective function defines the estimation problem. As in bootstrapping, the random-weighting output expresses variation conditional upon data that we may use in various inference tasks. While we have been guided in some ways by a Bayesian perspective on random weighting, we have not established clear, definitive results that connect random weighting to Bayesian asymptotics. Therefore, we have changed the tenor of the paper to focus more clearly on the statistical findings we can prove. We no longer speculate on connections between random weighting and Bernstein-von Mises theory. We do emphasize that our results extend available theory for random-weighting methodology. This methodology is compelling in contemporary problems owing to advances in optimization (brief motivation, new Section 1); while tools for uncertainty quantification do exist for LASSO regression, we note that LASSO regression is a prototypical model for penalized estimation and a necessary case for the development of random-weighting theory.

Regarding the signs of Lasso estimates and the benchmark example, we no longer refer to *marginal posterior distributions of the β's from these four methods*. For numerical comparison purposes, we show the three random-weighting results against both residual bootstrap and the Bayes LASSO, both of which are available here for uncertainty quantification.

(R1.2) The theoretical results are developed under classical assumptions used in the widespread Lasso literature and that too with the added complexity of using random weights.

Reply: Thank you for taking note of the classical assumptions that we adopted in our manuscript. This concern is covered in our response to **(R1.1)**.

(R1.3) In Section 4, you write, the conditional distribution of the one-step random-weighting samples $\widehat{\boldsymbol{\beta}}_n^w$ converges to the same limit as in the Bernstein-von Mises Theorem, i.e., the conditional distribution of $\widehat{\boldsymbol{\beta}}_n^w$ is the same - at least up to the first order - as the posterior distribution of β under the regime of Bayesian inference.

Is there a clear proof of the part which I have underlined (either it is too obvious, so that you have not shown it), or it needs to be clearly presented. Where is the prior on β, under which you can establish this result? Are your asymptotic results based on this approach. Does your conditional consistency results automatically take care of any prior on β? You need to clear this part.

When you present Theorem 3.3, in what sense do you achieve the converges to the same limit part and how? Does $\arg\min_u V(u)$ have the same law as the posterior distribution of β?

Reply: The speculative connections to Bernstein-von Mises presented previously have been removed in the revised manuscript.

(R1.4) Can you connect the material in Sec. 3.2 to the post-Lasso OLS estimator (Belloni & Chernozukhov (Bernoulli/2011?)).

Reply: Thank you for suggesting this reference. We have included it into Section 3.2 of this second revision of manuscript. Belloni and Chernozhukov (2013) investigated the finite-sample and asymptotic properties of the post-LASSO OLS estimator but they have not considered the bootstrapped version of this estimator. Again, this issue is partially covered in our response to **(R1.1)**; Since our random-weighting (or weighted bootstrap) approach is **not** about coming up with another point estimator (and then study its asymptotic properties), perhaps a more related comparison to the results presented in Section 3.2 (especially our two-step random-weighting procedure) would be Liu and Yu (2013)'s residual bootstrap to their LASSO+LS estimator, which we have already discussed in our 'Sampling Theory Interpretation' subsection (previously Section 4.2, now Section 3.3).

(R1.5) The article is promising, but it lacks clarity in some aspects. I am sorry for my delayed report, but I hope my comments will be helpful in improving this article.

Reply: Again, we thank you for your helpful comments and suggestions that have guided our re-writing. We hope that our responses for this revision sufficiently clarify your questions and concerns.

References

Alexandre Belloni and Victor Chernozhukov. Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2):521–547, 2013.

Hanzhong Liu and Bin Yu. Asymptotic properties of lasso+mls and lasso+ridge in sparse high-dimensional linear regression. *Electronic Journal of Statistics*, 7:3124–3169, 2013.