Report on the article titled:

Random-weighting in LASSO regression - R1

by T. Ng and M. A. Newton.

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.

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.

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.

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

   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 $\operatorname{argmin}_{\mathbf{u}} V(\mathbf{u})$ have the same law as the as the posterior distribution of β?

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

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.