Report on manuscript "Bayesian Restricted Likelihood Methods" by John R. Lewis et al.

Comments on the content

The manuscript develops a new methodology for using a restricted likelihood in a Bayesian context. It derives an ingenious way of generating complete data samples that match the observed data as far as the conditioning statistics is concerned. The resulting algorithm is presented in sufficient detail to be useful to the practitioner. In view of these qualities, the manuscript deserves to be published.

However, in the example chosen by the authors the gain in performance as compared to classical methods is marginal at best, at the price of a presumably much larger computational complexity. In the simple regression model the restricted likelihood is outperformed by the heavy-tailed model for the small sample size, whereas for larger sample sizes its performance is virtually identical to the classical robust estimators. In the hierarchical model, it is indeed slightly better than the classical estimators, but I wonder whether this slight gain justifies the additional effort of finding reasonable priors for the hyperparameters and the much more involved computation of the estimates. If the priors are not chosen carefully, the results could even be worse than the ones from the classical estimators. The authors should therefore provide an additional example where the superiority of their approach shows more clearly. It would also be useful to tell the reader whether the code used in the example is already publicly available, and how it compares in terms of computational load with the classical M-estimators and also with other robust estimators such as LMS (least median of squares) and LTS (least trimmed squares).

In addition to the usual concerns about the sensitivity of the results on the prior in the Bayesian context, the method as presented here raises the problem of choosing the type of the M-estimator, to which the example data seem to be quite sensitive. It is not difficult to understand why Tukey's redescending M-estimator performs better than the Huber-type M-estimator, but maybe not so easy why the difference is much more pronounced in the hierarchical model as compared to the simple model. If the authors have an explanation for this effect it should be given in the paper. For other data sets, however, a good choice of the M-estimator could well be different, so some guidelines would be welcome to the practitioner.

The authors should also discuss in more detail whether it is possible to go beyond M-estimators and to use high breakdown-point estimators such as LMS and LTS as the conditioning statistic, and which scale estimators could or should be used as companions. As the LMS regression has no tunable parameters, and the LTS regression has only a single one, the user would be relieved from studying the sensitivity to the choice of the M-estimator from an infinite number of possibilities.

Finally, I have to remark that the first sentence of the Discussion is dangerously close to an insult to the numerous researchers and practitioners who have applied Bayesian methods in data analysis for years, even decades. The approach chosen by the authors is certainly innovative and in my opinion potentially fruitful, but it can hardly claim to *begin* to reconcile the two fields. I think it is fair to ask that the authors find a slightly more modest expression of their enthusiasm.

In summary, I recommend publication of the manuscript under the provision that the authors submit a revised version which addresses the points raised above. It should contain an additional example that shows a significant edge of their method over the classical ones. In this new example robust estimators with high breakdown point (LMS and LTS) should be studied along with suitable M-estimators. As is well known, the LMS-estimator has worse asymptotic properties than M-estimators and LTS, and it will be interesting to see whether this is visible in the results.

Minor comment

Figure 4 is very hard to read, the panels should have the same size as the ones in Figure 5.