Discussion to 'Uncertainty Quantification for the Horseshoe' by Stéphanie van der Pas, Botond Szabó, and Aad van der Vaart*

Juho Piironen† Michael Betancourt‡ Daniel Simpson§
Aki Vehtari†

11 October 2018

The authors present a detailed analysis of the asymptotic frequentist properties of credible sets derived from posteriors with normal-linear measurement models and horseshoe priors. Although we disagree with the claim that *"In Bayesian practice credible balls are nevertheless used as if they were confidence sets"*, the results in the paper are important for identifying where the horseshoe priors are fragile asymptotically, and hence particularly dangerous in the non-asymptotic regimes more typical of the applied problems where sparse models are needed.

One clarification we believe is warranted is that the horseshoe family of prior distributions does not encode sparsity as is typically interpreted. Instead of partitioning parameters into those that are zero and non-zero, the horseshoe priors actually separate parameters into those that are resolvable by measurements and those that are not. In particular, as with any model the horseshoe priors cannot be interpreted outside of the context of a particular likelihood (Gelman et al., 2017). Consequently the statement that *"τ can be interpreted as the proportion of nonzero parameters, up to a logarithmic factor"* is not quite true.

Piironen and Vehtari (2017b; 2017c) demonstrate that the effects of τ in horseshoe priors are intimately related to the measurement variability σ, even for the simple normal-linear measurement model. Figure 6 of Piironen and Vehtari (2017c), for example, clearly illustrates that rescaling the data changes the impact of the horseshoe prior unless τ is scaled by σ, even with an oracle prior information about the true number of significant parameters, $p_0 = p_n$. In particular, the resolution threshold $\sqrt{2\log(n/p_n)}$ arising in the paper implicitly assumes that the measurement variability σ is equal to 1,

*To appear in *Bayesian Analysis*

†Helsinki Institute for Information Technology HIIT, Department of Computer Science, Aalto University, Finland.

‡Department of Statistics, Columbia University, New York.

§Department of Statistical Sciences, University of Toronto, Canada.

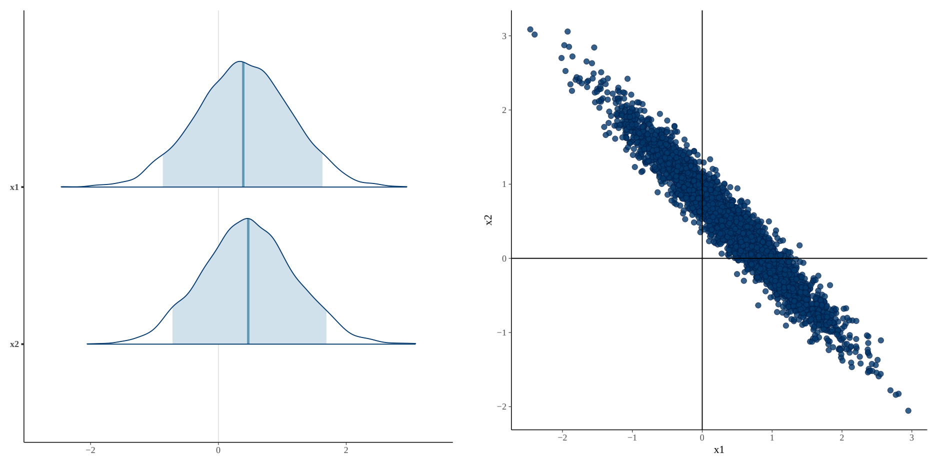

Figure 1: The left plot shows marginal posteriors of effects which overlap zero. The right plot shows the corresponding joint distribution which reveals strong posterior dependency and the fact that zero is not included in the joint credible region.

but a more realistic threshold has to take into account the value of σ, which is typically unknown a priori. We are very curious as to how robust the results presented in the paper are to these circumstances where also σ must be inferred.

Additionally, we find that the focus on marginal credible intervals is a significant limitation. One of the defining features of the family of horseshoe priors, and indeed a strong reason for their utility, is that they do not regularize each parameter independently but rather induce a joint regularization over the entire parameter space. In particular, joint credible intervals can behave much differently from marginal intervals. Figure 1 illustrates that with a linear model that employs a uniform prior over the slopes of two correlating predictors x_1 and x_2 it may happen that the joint posterior concentrates away from the origin without either of the marginals clearly distinguished from zero.

The situation becomes even more difficult with a large number of correlating predictors when utilizing the horseshoe prior. In this case even for the most relevant variables most of the posterior mass can concentrate around zero, see for example Figure 9 in Piironen and Vehtari (2017c), which makes a reliable variable selection based on the posterior intervals challenging.

Moreover, the method by Carvalho et al. (2010) of including all variables with $\kappa_j > 1/2$ would fail in this case because none of the predictors have $\kappa_j > 1/2$. Consequently, we believe the only reliable variable selection strategy in these situations is based on the estimated effect on the predictive distribution, for example using the projection predictive variable selection (Piironen and Vehtari, 2017a). This framework has the added benefit that it provides guidance on how to select out significant parameters jointly, instead of one by one as discussed in the paper.

Finally, we advise caution with regard to the recommendation of the maximum marginal likelihood estimator for τ in practical problems. The large p, small n applications where horseshoe priors are most needed lie far away from the asymptotic regime that stabilizes the MMLE. Any complexity of the measurement model beyond the normal-linear model only makes the matter worse.

References

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

Gelman, A., Simpson, D., and Betancourt, M. (2017). The prior can generally only be understood in the context of the likelihood. *arXiv preprint arXiv:1708.07487*.

Piironen, J. and Vehtari, A. (2017a). Comparison of Bayesian predictive methods for model selection. *Statistics and Computing*, 27(3):711–735.

Piironen, J. and Vehtari, A. (2017b). On the hyperprior choice for the global shrinkage parameter in the horseshoe prior. In Singh, A. and Zhu, J., editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 905–913.

Piironen, J. and Vehtari, A. (2017c). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*. Accepted for publication. arXiv preprint arXiv:1707.01694.