Response to reviewer 1

Marnix Van Soom Bart de Boer

As you already stated during your talk it seems that you can come over some problems with your first prior by simply employing nested sampling not suffering from the label switching problem. Maybe you could just add a further paragraph (as an outlook???) summing up your insights from talking to John and how this will set up an agenda for ongoing work.

While nested sampling of course has no theoretical problem with calculating Z for a posterior exhibiting exchange symmetry, the latter does pose some difficult practical issues, which π_3 tries to address in a "plug and play" approach; i.e., compatible with off-the-shelve exploration algorithms which can be found in libraries such as dynesty.

The main practical issue is that for most exploration algorithms, the posterior modes due to exchange symmetry may (will) prove to be too *numerous* and too *sticky* to enable convergence of the nested sampling run.

Numerous because many exploration algorithms work with topology and retain some kind of bounds to restrict the prior subject to the current lowest likelihood value, such as elliptical bounds based on clustering the particles by their mean and covariances. If the live particles are spread along many modes, this will cause problems to estimate accurate bounds and will prevent convergence, as practical experience bears out.

Sticky because the "gates" between primary and induced modes must be explored in order to arrive at the correct value of Z, and practical exploration algorithms might be too biased towards the sticky modes such that they miss these important gates. This is a subtle point but important to understand how nested sampling infers the value of Z for posterior landscapes with exchange symmetry. The gates are crucial because they create regions of enclosed prior mass X in which likelihood values are larger than they would be if the model would not have the exchange symmetry, because the latter interpolates smoothly between primary and induced modes. Such early discovery (while X is still closer to 1) of larger likelihood values timely corrects the value of Z upwards to account for the extra mass contained by the induced modes. Indeed, once the current likelihood lower bound L^* is such that all enclosed regions of the prior for which $L(\boldsymbol{x}) > L^*$ have disconnected into islands, and there is at least one particle in each primary mode or one of its induced copies, nested sampling does not make any distinction between the problem with or without exchange symmetry[1]. The upward revision of Z must happen entirely before L reaches that L^*, and it happens by the aforementioned early discovery of higher likelihood values residing in those gates. If K is moderately large, the number of possible gates grows quickly, as can be seen in the pairwise projections in Figure 1 in the revised paper. In addition, multiplets (for example modes close to the bissectrice in Figure 1) create ample opportunity for such gates. A proliferation of gates nearby sticky modes poses practical problems for exploration.

It took me quite a while to understand this, and I did not have an answer ready at the time of the presentation. I should also say that I have not yet talked to John about this, but I'm planning to do so soon. He did make the comment that priors imposing an ordering (such as π_3) tend to hinder exploration, which intuitively makes sense. He also mentioned that the issue with exchange-symmetric problems would indeed lie in practical exploration.

I have decided not to add all this detail to the paper in order to keep its conciseness, and the fact that the discussion of the prior should not depend too much on the actual computational method used (the above discussion applies to nested sampling, while many other methods exist to estimate Z). In addition, the original formulation in the paper still stands (lines 36-41 in the revised paper):

[1] Assuming adequate exploration and no mode die-off.

> The trouble is that correctly taking into account these induced modes during the evaluation of $Z(K)$ requires a surprising amount of extra work besides tuning the MCMC method of choice, and that is the label switching problem in our setting. In fact there is currently no widely accepted solution for the label switching problem in the context of mixture models either [**?**, **?**].

I have, however, added the following to the Discussion:

- Lines 158-160: "The prior is compatible with off-the-shelf exploration algorithms and solves the label switching problem without any special tuning or post processing. It would be interesting to compare it to other approaches [**?**, e.g.,]]Buscicchio2019, especially in terms of exploration efficiency."

However, don't you have to shorten the paper to 8 pages?

Yes, we had misinterpreted the page limit to not include references.