Integrated likelihood ratio test

The authors studied Wilks phenomenon for integrated likelihood ratio test statistics (ILRT) using powered likelihood and fractional posterior as weight functions. The main motivation (as claimed by the authors) is the advantage of ILRT over usual LRT where the likelihood can be unbounded. The authors specifically show that in the example of mixture models, the ILRT is defined and does have a Willk's phenomenon whereas similar results do not hold for LRT.

My main concern after reading the paper is that the general idea is too diluted and not properly fleshed out since there is only one example (Gaussian mixture model) to go along with it. More so, the hypothesis (3.1) seems not properly motivated; may be the authors are interested in testing whether two components are necessary to fit the data? A more interesting problem (perhaps a bit more ambitious) is to consider testing for the number of components in a mixture model. Does the ILRT offer any advantage in that regard? Currently the results seem to be too specific to make any practical impact. Nevertheless, I should at least encourage the authors to make an effort to showcase the benefit in more than one example. Here are more specific issues that need to be resolved.

1. Theorem 2.1 is nice in the sense if $a + b < 1$, then the conditions required to satisfy $\sqrt{n}$ consistency of L_t and $L_t^{(0)}$ is rather mild. However, Theorem 2.1 is not useful in practice, i.e., to derive asymptotic cut-offs unless the right hand side can be made free of unknown parameters, i.e., replacing θ_0 by an empirical estimate, e.g. the fractional posterior mean.

2. Assumption 2.2 does not necessarily hold for moderate to high dimensional settings where the smallest eigen value of the Hessian can be actually very small. How generalizable is this to high-dimensional θ?

3. For the variational inference result, the authors stated the result for the Renyi divergence variational inference, which is not what is commonly used in practice. The main technical difficulty in the theoretical analysis in variational inference is due to the fact that mismatch between the divergence with respect to which the objective function is minimized and the divergence with respect to which we seek the estimation accuracy. Is there any way such a result can be generalized for the usual KL distance?

4. For singular models such as the Gaussian mixtures, there are well-known expansions of marginal likelihood e.g. Watanabe [2009] which takes care of the fact that the Fisher information matrix is singular. In the proof of Theorem 3.1 authors somehow bypassed this by directly working with the marginal likelihood which is commendable, but obscures the main message at the same time. Do the authors envision a more general theory in this direction using Integrated LRT for singular models?

References

Sumio Watanabe. *Algebraic geometry and statistical learning theory*, volume 25. Cambridge University Press, 2009.