Review of "On the Wilks phenomenon of Bayes factors and the integrated likelihood ratio test"

The paper derives the asymptotic distributions of various Bayes factors, which are the usual Bayes factor, what they call generalized fractional Bayes factor, and integrated likelihood ratio test. From the derived asymptotic distributions, prior conditions for the Wilks phenomenon are obtained. If the null asymptotic distribution of the test statistic does not depend on the parameter, it is said that the Wilks phenomenon happens. I enjoyed reading the paper. The proofs are generally clearly written, although some notations are non-standard. The idea of using $\Delta_{a,b}$, ratio of Bayes factors, to get Wilks phenomenon was interesting. I find the derivation of the asymptotic distribution of the Bayes factors interesting. But I have some doubt that the proposed testing method has methodological innovation as the significant testing. I suggest that the author demonstrate that the proposed method has competitive strength over the existing methods.

Major comments

1. (Literature review) The paper proposes to use the Bayes factor as a test statistic for a significance test of the composite null hypothesis. I can hardly believe this idea is new. The paper does not mention any papers on this idea. I suggest to review the literature of significance test using Bayes factors or related issues. I also suggest to review the literature of large sample properties of Bayes factor, and to compare the results of this paper with existing results in the literature.

2. (numerical examples: simulation studies and real applications) The paper does not have any numerical study. I suggest to include extensive simulation studies comparing existing significance tests such as the likelihood ratio test (LRT). If the Bayes factors in the paper are used in hypothesis testing, is there anything that the user needs to be careful about? Please include in simulation study examples in which LRT fails and examples with wide applications.

3. (Choice of a and b) According to Theorem 2, there are many a and b such that $\Delta_{a,b}$ satisfies the Wilks phenomenon. In actual testing problem, what values of a and b would you suggest to use?

4. The paper discusses the fractional Bayes factor, the posterior Bayes factor and the integrated Bayes factor, but does not discuss the intrinsic Bayes factor. I suggest to include some discussion on the intrinsic Bayes factor. Can you get similar asymptotic results for the intrinsic Bayes factor?

5. The paper gives one example (mixtures of normals) for which the LRT fails. One example for the argument for the Bayes factor as the test statistic is weak. Please give more examples and numerical examples.

Minor comments

1. P.3. L. 19. variantional $\longrightarrow$ variational

2. Assumption 1. The term "inner point" is used but "interior point" is more standard term. Also here Θ and $\tilde{\Theta}$ is assumed open and "interior point" assumption is not necessary.

3. P. 5. L. 36. means $\longrightarrow$ mean

4. Equation (1). It is stated that "null distribution of $BF_t(X_n)$ is free of the nuisance parameter if and only if $\frac{|I_{\xi|\nu}(\nu,\xi_0)|^{-1/2}\pi(\theta_0)}{\pi_0(\nu)} \equiv c$ for some constant c. But the formular in Theorem 1 is slightly different. It should be

$$\frac{|I_{\xi|\nu}(\nu_0,\xi_0)|^{-1/2}\pi(\theta_0)}{\pi_0(\nu_0)} \equiv c$$

for some constant c. If this is the case, the subsequence discussion needs to be modified.

5. Please include some discussion on Assumptions 2 and 3. Please explain conditions in these assumptions.

6. P. 18. L. 44. It is said "This equality holds for every $M > 0$ and hence also for some $M_n \longrightarrow \infty$." Please prove this statement.

7. P. 24. L. 4. A typo. $L_t(\Theta; \mathbf{X}_n) \leq L_1^{1/t}(\Theta; \mathbf{X}_n) \longrightarrow L_t(\Theta; \mathbf{X}_n) \leq L_1^t(\Theta; \mathbf{X}_n)$.

8. P. 25. In the integral of the displayed formula. $\pi(\theta|\mathbf{X}_n) \longrightarrow \pi_t(\theta|\mathbf{X}_n)$