

## Response to Reviewers

### “ On the Wilks phenomenon of Bayes factors and the integrated likelihood ratio test ”

Rui Wang<sup>1</sup> and Xingzhong Xu<sup>1,2</sup>

<sup>1</sup> School of Mathematics and Statistics, Beijing Institute of Technology, Beijing  
100081, China

<sup>2</sup> Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing  
100081, China

Wednesday 21<sup>st</sup> October, 2020

We thank the AE and both reviewers for their helpful comments and critiques. Over the past ten months, we have carefully revised our paper according to comments made by the AE and the reviewers. Below we respond to each reviewer in turn.

## 1 Response to reviewer 1

### 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.

### Answers:

We agree that literature review is lack. We thank.

a. Bayesian viewpoint of frequentist test. A lot of work. Many work on calibration of  $p$ -value to Bayes factor. For example, test based Bayes factor See ? for a review of this area.

b. significance test using Bayes factors. Not many. Bayesian-motivated. Existing work on the frequentist viewpoint of Bayes factor. Specific model. Linear model (yongtao guan) Posterior predictive  $p$ -value.

c. large sample properties of Bayes factors. Consistency of Bayes factor.

Using the Bayes factor as a frequentist test is of interest in history. Good 1967: My aim is to arrive at reasonably general-purpose initial distributions and significance tests, but the same strategy could be applied in any given concrete problem in which there is additional background evidence. Uniformly most powerful Bayesian tests.

The naive Bayes factor is not black box. The proposed method is a black box method. universal method.

Using Bayes factor as a frequentist test statistic is definitely not a new idea. But it is a good thing that it is not new. And we do not claim we propose it. In fact, this topic has been considered by many researchers. However, this idea has not been explored deeply. Most existing work focus on the consistency of the Bayes factor for specific models. To the best of our knowledge, the Wilks phenomenon of the Bayes factor has not been systematically explored. Also the asymptotic power, the variants of the Bayes factor. They have not been investigated in full rigor of mathematics.

**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.

Answers:

**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?

Answers:

**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?

Answers:

**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.

Answers:

Minor comments:

**1. P. 3. L. 19.** variantional  $\rightarrow$  variational

Answers:

The correction has been made.

**2. Assumption 1.** The term “inner point” is used but “interior point” is more standard term. Also here  $\Theta$  and  $\tilde{\Theta}$  is assumed open and “interior point” assumption

is not necessary.

**Answers:** The correction has been made. We drop the condition of “interior point”.

**3. P. 5. L. 36. means  $\rightarrow$  mean**

**Answers:**

The correction has been made.

**4. Equation (1). It is stated that null distribution of  $\text{BF}_t(X_t)$  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 formula in Theorem 1 is slightly different. It should be**

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

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

**Answers:**

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

**Answers:**

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

**Answers:**

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

**Answers:**

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

**Answers:**

## 2 Response to reviewer 2

1. Some related reference is missing such as Zhou and Guan (2018) JASA paper On the Null Distribution of Bayes Factors in Linear Regression. In this paper, they showed by considering the Bayes factor as a test statistics (like this submitted paper), they derived the null distribution of the Bayes factor for linear models, which is a weighted sum of chi-squared random variables. Even though this work is restricted to linear models under simple assumptions, it is worth to cite. Also, I believe that you can find more references that consider a Bayes factors as a test statistics in frequentist settings.

**Answers:**

2. On page 4, “The prior  $\pi(\theta)$  and  $\pi_0(\nu)$  may be improper...”. This is wrong in general. If they are improper, we cannot calculate the Bayes factor. Of course, I understand that even when the priors are improper, the fractional Bayes factor,

introduced later, circumvents this issue. But, I think that it would be better to briefly note this point at the place of the sentence “The prior  $\pi(\theta)$  and  $\pi_0(\nu)$  may be improper...”, to avoid a confusion.

Answers:

3. On line 31 on page 6, what is  $f$ ?

Answers:

4. On line 38 on page 6, you may want to note an extra weakness of the traditional approach. When the prior density is proportional to a function of the determinant of the Fisher information, its prior normalizing constant is difficult to evaluate, and the normalizing constant is critical for the Bayes factor. Although you noted “...Fisher information matrix has a complicated form...undesirable...”, it would be better to be more specific.

Answers:

5. This theoretical work is investigated under assumed that the posterior achieves  $\sqrt{n}$ -consistency (or the posterior contraction rate is  $n^{-1/2}$ ). In many practical statistical models, the optimal posterior contraction rate would be slower than  $n^{-1/2}$ . These examples include high-dimensional sparse problem like  $(\log p/n)^{1/2}$  or nonparametric function estimation  $n^{-\beta/(2\beta+p)}$ , where  $\beta$  is a smoothing factor for the true function. Under these interesting models, can you extend this result to more general models?

Answers:

6. You proposed theoretical results on asymptotic null distribution of Bayes factor which is a linear transformation of a chi-square distribution. But, you have not considered any simulation works to examine finite sample behavior of the approximated distribution. In practice, the accuracy of the asymptotic null distribution with finite samples is of interest. You may want to show that this asymptotic null distribution is useful in practice, especially for complicated and practical models. The examples you considered (without simulation studies) are too simple and far from a practical point of view.

Answers:

7. This paper is lack of tuning parameter selection which will be critical in the hypothesis testing result. Theoretically, it can satisfy some simple conditions, but in practice how to choose the tuning parameters  $a$  and  $b$  is very important.

Answers:

### 3 List of major changes

### References