Response to Reviewers
" On the Wilks phenomenon of Bayes factors and the integrated likelihood ratio test "

Rui Wang[1] and Xingzhong Xu[1,2]

[1] School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081,China

[2] Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081,China

Saturday 24th October, 2020

We are grateful to the AE for informing us a number of useful relevant papers and helpful comments. He/she even provided us an example, namely binomial mixture model, which is very suitable for illustrating the advantages of the proposed method over the LRT. We thank both reviewers for their valuable and helpful comments and critiques. We have carefully revised our paper according to the comments and critiques and have made extensive modifications on the original manuscript. Now the paper is much improved. Below we respond to the AE and each reviewer in turn.

1   Response to the AE

First, the fractional posterior is a special case of a Gibbs posterior, and $\sqrt{n}$ consistency of Gibbs posteriors has been around in the literature for a while; see, for example, the paper "An MCMC approach to classical estimation" by Chernozhukov and Hong as a representative example. This is particularly not surprising for the fractional posterior since roughly speaking, it keeps the center of the posterior unaffected for large n and reduces the variance by a scale factor depending on the fractional power - hence the classical Bernstein–von Mises phenomenon should go through for the fractional posterior under the same conditions as the exact posterior. I found the integrability condition in Proposition 2 to be somewhat strong in this regard.

Answers: We thank the AE for informing the literatures on the Gibbs posterior. The fact that the fractional posterior is a special case of a Gibbs posterior is interesting, and we mention this

1

connection in the revised paper. I found that the results of Chernozhukov and Hong (2003) require certain conditions on the likelihood and only bounded likelihood can satisfy their conditions. Since we would like to include the unbounded likelihood case, we can not use the $\sqrt{n}$ consistency result in their paper. We note that the assumptions of Chernozhukov and Hong (2003) (that is, compact parameter space and continuous prior density) implies that the prior distribution is proper. In fact, we found that most existing results on the consistency of posterior require that the prior is proper. Since improper priors are often used for Bayes factor, existing results are not suitable for our purpose. The integrability condition in Proposition 2 is

$$\int_{\Theta} \exp\{-c^* D_{1-t}(\theta_0 \| \theta)\} \pi(\theta) \, \mathrm{d}\theta < \infty.$$

Note that if $\pi(\theta)$ is proper, then

$$\int_{\Theta} \exp\{-c^* D_{1-t}(\theta_0 \| \theta)\} \pi(\theta) \, \mathrm{d}\theta < \int_{\Theta} \pi(\theta) \, \mathrm{d}\theta < \infty.$$

Hence the integrability condition is satisfied for any proper priors. Furthermore, it also allows improper priors provided the tail probability of the prior density is not too thick. Hence our integrability condition is very weak. In the revised paper, we add a remark on this condition.

The paper doesn't discuss recent advances in the frequentist literature. I would point out the article "Generalized likelihood ratio statistics and Wilks phenomenon" by Fan et al. as an example; this has been cited many times and the authors need to check what is there beyond the usual LRT. Also take a look at the paper "Geometric Understanding of Likelihood Ratio Statistics" by Fan et al. which among other things provides a Bayesian argument for the derivation of the classical Wilk's phenomenon.

Answers: We thank the AE for informing us these references, which help us to improve the literature review of the paper.

2 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: Following the reviewer's suggestions, we improved the literature review in the revised paper. The idea of using the Bayes factor as a test statistic is definitely not a new one, and we point out this point explicitly in the revised paper. We add the following comment in the revised paper:

The idea is not new. In fact, this methodology dates back at least to Good (1967), and has been considered by many researchers since then. Good (1992) gave a review for some early literature using this idea. See Aerts et al. (2004), Zhou and Guan (2018) and Wang and Xu (2020) for some recent applications of the idea. However, to the best of our knowledge, the idea has been mostly used for specific models and has not been systematically studied with full mathematical rigour for general models. One cause of this fact may be the Lindley's paradox, that is, the distribution of the Bayes factor depends heavily on the prior density; see, e.g., Shafer (1982). ...

Following the reviewer's suggestion, we also comment on the literature of large sample properties of Bayes factor, as follow:

In this work, we treat the Bayes factor and its variants as frequentist test statistics. Hence a closely related area of research is the frequentist properties of the Bayes factor. Most existing results in this area focus on the consistency of the Bayes factor; see Berger et al. (2003), Moreno et al. (2010), Wang and Maruyama (2016), Chatterjee et al. (2018) and the references therein. There are still relatively few researches on the asymptotic distribution of the Bayes factor. Clarke and Barron (1990) derived the asymptotic distribution of the ratio of the marginal likelihood to the likelihood at the true parameter, which can be regarded as the Bayes factor for a point null hypothesis. Gelfand and Dey (1994) derived formal approximations to the Bayas factor and some of its variants. In this work, in order to use the Bayes factor to construct frequentist tests, we give a thorough study of the asymptotic distribution of the Bayes factor and its variants for general models. Thess results fill a theoretical gap in the literature and are interesting in their own right.

We add a remark below Theorem 1 which gives a comparison of Theorem 1 and the result in Clarke and Barron (1990):

Clarke and Barron (1990) derived the asymptotic null distribution of

$$\frac{\int_{\Theta} p_n(\mathbf{X}^n \theta)\pi(\theta)\,\mathrm{d}\theta}{p_n(\mathbf{X}^n \theta_0)},$$

which can be regarded as the Bayes factor for the point null hypothesis, i.e., $p_0 = 0$. Theorem 1 extends their result and gives the asymptotic distribution of the general Bayes factor under both the null hypothesis and the local alternative hypothesis. Clarke and Barron (1990) imposed some conditions on the likelihood around the true parameter, which may not be easy to verify for some moderately complex models. In comparison, Theorem 1 only assumes the likelihood can be expanded in a $n^{-1/2}$ neighborhood of the true parameter, which is satisfied by most regular models.

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: We follow the reviewer's suggestion. In the revised paper, we add simulations to examine the performance of the proposed method and compare it with the LRT. The simulation examples include two hard problem, that is, normal mixture model and binomial mixture model. For these models, researchers have proved that the LRT has irregular behavior. The testing problem for binomial mixture model has applications in genetics. Numerical results show that the proposed method outperforms the LRT in terms of both test level and test power.

In the revised paper, we also discussed that the users needs to be careful about the degree of freedom when they deal with irregular models, as follows:

The proposed method is quite universal. However, when applied to irregular models, one needs to be careful about the freedom of the asymptotic chi-squared distribution. For example, in Proposition 5 and Proposition 6, the freedom of the asymptotic chi-squared distribution is 1 instead of $p - p_0 = 2$. This phenomenon is essentially caused by the loss of identifiability of the mixture models. In general, when the model has a loss of identifiability, the degree of freedom of the asymptotic chi-squred distribution may be less than $p - p_0$. In this case, if the true degree of freedom is not easy to obtain, the user can simply use $p - p_0$ as the degree of freedom, and the resulting test can still preserve the test level. Of course, this simple strategy may lead to decreased power.

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:

Our results imply that the choice of a and b does not affect the Wilks phenomenon and asymptotic power provided a and b are fixed and $0 < b < a < 1$. In the simulation studies of the revised paper, the default value is $a = 2/3$ and $b = 1/3$. we also add a simulation to examine the effect of a and b on the finite sample performance of the test procedure. It turns out that the proposed methdod is not sensitive to the choice a and b. In practice, if there is no practical evidence for the choice of a and b, one can simply choose $a = 2/3$ and $b = 1/3$.

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: The asymptotic behavior of the intrinsic Bayes factor is relatively simple. In the asymptotic sense, the intrinsic Bayes factor is equivalent to the Bayes factor with intrinsic priors; see, e.g., Berger and Pericchi (1996) and Moreno et al. (1998). Simple examples show that the Bayes factor with intrinsic priors does not have Wilks phenomenon. Hence intrinsic Bayes factor can not be directly used as a frequentist test statistic. This point is made clear in our paper.

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: We follow the reviewer's suggestion. In the revised paper, we add an example of binomial mixture model. This model is useful in genetics. It is known that the LRT for this model has complicated behavior. We proved that the proposed method has Wilks phenomenon in this problem. We also conduct simulations to examine the performance of the proposed method. It turns out that the proposed method outperforms the LRT in terms of both test level and test power.

Minor comments:

1. P. 3. L. 19. variantional → 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 Θ 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 → mean

Answers: The correction has been made.

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

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

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

Answers: This problem is a little bit brain twisting. But I think our original statement is correct. In the paper, under the null hypothesis, we assume the true parameter is $\theta_0 = (\nu_0^\top, \xi_0^\top)^\top$. Here ξ_0 is specified by the null hypothesis ($H : \xi = \xi_0$). But ν_0 is an unknown nuisance parameter. Hence according to Theorem 1, to make the null distribution of $\mathrm{BF}(\mathbf{X}^n)$ free of the nuisance parameter ν_0, one should require that

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

for any nuisance parameter ν_0. But the above statement is equivalent to say that

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

for any nuisance parameter ν.

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

Answers: We follow the reviewer's suggestion. In the revised paper, we add the following comment bellow Assumption 2:

The first condition in Assumption 2 avoids infinite Kullback-Leibler divergence. If the second condition in Assumption 2 does not hold, then there is a $\delta > 0$ and a sequence of parameters $\{\theta_n\}$ such that $\|\theta_n - \theta_0\| \geq \delta$ and $D_t(\theta_0\|\theta_n) \to 0$. In this case, the model will suffer from loss of identifiability. Hence the second condition assumes that the model is identifiable in a reasonable sense. The third condition in Assumption 2 assumes that $D_t(\theta_0\|\theta)$ has a reasonable Taylor approximation around $\theta = \theta_0$; see, e.g., van Erven and Harremoes (2014), Section III. H.

We add the following comment bellow Assumption 3:

Assumption 3 assumes that the tail of the prior density is not too thick. To appreciate the conditions, suppose P_θ is the normal distribution $\mathcal{N}(\theta, 1)$ and $\theta_0 = 0$. Then the first two conditions of Assumption 3 becomes

$$\int_\Theta \exp\left\{-c^\dagger \theta^2/2\right\} \pi(\theta)\, \mathrm{d}\theta < \infty, \quad \int_\Theta \theta^2 \exp\left\{-c^\dagger \theta^2/2\right\} \pi(\theta)\, \mathrm{d}\theta < \infty.$$

The above condition is satisfied for $\pi(\theta) \equiv 1$. This implies that Assumption 3 is weak and it allows improper priors.

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

Answers:

We have obtained that for any fixed $M > 0$,

$$\int_{\{\theta:\|\theta-\theta_0\|\leq M/\sqrt{n}\}} \exp\left\{-tR_n(\theta_0\|\theta)\right\} \pi(\theta)\, \mathrm{d}\theta$$
$$= (1 + o_{P_{\theta_0}^n}(1))n^{-p/2}\pi(\theta_0)\exp\left\{\frac{t}{2}\Delta_{n,\theta_0}^\top I(\theta_0)\Delta_{n,\theta_0}\right\}$$
$$\cdot \int_{\{h:\|h\|\leq M\}} \exp\left\{-\frac{t}{2}(h - \Delta_{n,\theta_0})^\top I(\theta_0)(h - \Delta_{n,\theta_0})\right\} \mathrm{d}h.$$

We claim that this equality holds for every $M > 0$ and hence also for some $M_n \to \infty$.

In essence, this claim follows from a result in mathematical analysis and the fact that convergence in probability is metrizable. To make the proof more readable, we add a lemma in the revised paper (Lemma 2 in the revised paper).

Lemma. *Let $T_{m,n}$, $m = 1, 2, \ldots$, $n = 1, 2, \ldots$, be random variables such that for fixed m, $T_{m,n}$ converges in probability to 0 as $n \to \infty$. Then there exists a sequence $\{h(n)\}$ such that $h(n) \to \infty$ and $T_{h(n),n}$ converges in probability to 0 as $n \to \infty$.*

Using this lemma, we can prove our claim as follow:

Note that for every fixed $M > 0$, the term $o_{P_{\theta_0}^n}(1)$ in the above equality converges in probability to 0 as $n \to \infty$. Hence by Lemma 2, this term also converges in probability to 0 for some $M_n \to \infty$. Therefore the above equality still holds if we replace M by M_n.

In the revised paper, we also simplify some other proofs to improve the readability of the paper.

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

Answers: The correction has been made.

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

Answers: The correction has been made. During the revision of the paper, we also corrected many other typos and mistakes. We are sorry for such typos and mistakes.

3   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: We thank the reviewer for informing us this useful paper. In the revised paper, we reviewed the literature that consider a Bayes factors as a frequentist test statistic, and add the following comment in the revised paper:

. . . In fact, this methodology dates back at least to Good (1967), and has been considered by many researchers since then. Good (1992) gave a review for some early literature using this idea. See Aerts et al. (2004), Zhou and Guan (2018) and Wang and Xu (2020) for some recent applications of the idea. However, to the best of our knowledge, the idea has been mostly used for specific models and has not been systematically studied with full mathematical rigour for general models. One cause of this fact may be the Lindley's paradox, that is, the distribution of the Bayes factor depends heavily on the prior density; see, e.g., Shafer (1982). . . .

Following another reviewer's suggestion, we also comment on the literature of large sample properties of Bayes factor, as follow:

In this work, we treat the Bayes factor and its variants as frequentist test statistics. Hence a closely related area of research is the frequentist properties of the Bayes factor. Most existing results in this area focus on the consistency of the Bayes factor; see Berger et al. (2003), Moreno et al. (2010), Wang and Maruyama (2016), Chatterjee et al. (2018) and

the references therein. There are still relatively few researches on the asymptotic distribution of the Bayes factor. Clarke and Barron (1990) derived the asymptotic distribution of the ratio of the marginal likelihood to the likelihood at the true parameter, which can be regarded as the Bayes factor for a point null hypothesis. Gelfand and Dey (1994) derived formal approximations to the Bayas factor and some of its variants. In this work, in order to use the Bayes factor to construct frequentist tests, we give a thorough study of the asymptotic distribution of the Bayes factor and its variants for general models. Thess results fill a theoretical gap in the literature and are interesting in their own right.

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: A direct method to compute the Bayes factor is to sample parameters from the prior and then use the sampled parameters to calculate the average of the likelihood. If the priors are improper, then one can not sample from the prior and the direct method can not be used. In this case, the Bayes factor may be computed by some other indirect methods (although it may be a difficult problem). See, Friel and Wyse (2012) for a review of the computation of the Bayes factor. In fact, it is a common practice to use improper prior for certain parameters. For example, for Bayes factor of linear model, popular choice of the prior for the variance parameter is improper; see, e.g., Liang et al. (2008). Therefore, although the computation of the Bayes factor is difficult, we think that it is not exact to say "we cannot calculate the Bayes factor".

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

Answers: Here f is a probability density function. We are sorry that we did not explain the meaning of f in the original paper.

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: We agree that the original statement is not specific enough. Following the reviewer's suggestion, we the following comment to the revised paper:

Second, in order to construct priors satisfying (2), the determinant of the Fisher information matrix should be evaluated, which is a difficult task for many models. Hence it may not be easy to construct priors satisfying (2), especially for complex models.

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 β is a smoothing factor for the true function. Under these interesting models, can you extend this result to more general models?

Answers: In this paper, we systematically investigate the low-dimensional parametric models. It is possible to investigate the case for certain high-diensional problem, as we have done in Wang and Xu (2020). It is also possible to investigate specific irregular models as we have done in the mixture examples. We believe the methodology can also be applied to nonparametric problem with a suitable modification. However, it is hard to provide a general theory for these problems. The application of the proposed method to challenging models is an interesting problem, but it is definitely beyond the scope of the current paper. We would like to report further applications in the future.

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: We follow the reviewer's suggestion. In the revised paper, we conduct simulations to examine the finite sample performance of the proposed method. The simulation results show that the empirical size of the proposed method is very close to the nominal test level. This indicates that our asymptotic results are useful.

In the revised paper, we consider four examples, including exponential family models, two normal mixture models and a newly added example of the binomial mixture model which arises in genetics. It is known that the LRT for these problems has nonstandard behavior for these models. For these models, we prove that the generalized FBF has Wilks phenomenon. We also conduct simulation studies which show that the proposed method outperforms the LRT in terms of both test level and test power.

It is relatively simple to apply the proposed method to the examples. However, we would like to point out that the normal mixture models and the binomial mixture model are irregular models and the LRT for these models are not very simple to use. Take the second normal mixture model in the paper as an example. For this model, the asymptotic distribution of the LRT is a difficult problem. Bickel and Chernoff (1993) studied this problem, but they did not fully solve it. Liu and Shao (2004) gave the asymptotic null distribution of the LRT. Hall and Stewart (2005) derived the asymptotic power of the LRT. According to the result of Hall and Stewart (2005), the

LRT has trivial power under $n^{-1/2}$ local alternative hypothesis. For this irregualar model, the generalized FBF still has Wilks phenomenon and has better performance than the LRT. We think these examples illustrate the wide application scope of the proposed method.

It is interesing to use the proposed method to solve more complex problems. But it is difficult to seriously investigate more problems within limited space. Hence we would like to leave it for future work.

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: This question is asked by both reviewers. Our Answer is as follow:

Our results imply that the choice of a and b does not affect the Wilks phenomenon and asymptotic power provided a and b are fixed and $0 < b < a < 1$. In the simulation studies of the revised paper, the default value is $a = 2/3$ and $b = 1/3$. we also add a simulation to examine the effect of a and b on the finite sample performance of the test procedure. It turns out that the proposed methdod is not sensitive to the choice a and b. In practice, if there is no practical evidence for the choice of a and b, one can simply choose $a = 2/3$ and $b = 1/3$.

4 List of major changes

- We add an example of binomial mixture model. It is known that the LRT for this model has complicated behavior. We proved that the proposed method has Wilks phenomenon in this problem.

- We add simulations studies for the normal mixture model and the binomial mixture model.

- The literature review is much improved.

- We add many remarks to explain our results and assumptions.

- We modify and simplify some of the theoretical proofs.

References

Aerts, M., Claeskens, G., and Hart, J. D. (2004). Bayesian-motivated tests of function fit and their asymptotic frequentist properties. *The Annals of Statistics*, 32(6):2580–2615.

Berger, J. O., Ghosh, J. K., and Mukhopadhyay, N. (2003). Approximations and consistency of bayes factors as model dimension grows. *Journal of Statistical Planning and Inference*, 112:241–258.

Berger, J. O. and Pericchi, L. R. (1996). The intrinsic bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91(433):109–122.

Bickel, P. J. and Chernoff, . (1993). Asymptotic distribution of the likelihood ratio statistic in a prototypical non regular problem. In Ghosh, e. a., editor, *Statistics and Probability: A Raghu Raj Bahadur Festschcrift*, pages 83–96. Eastern Limited.

Chatterjee, D., Maitra, T., and Bhattacharya, S. (2018). A short note on almost sure convergence of bayes factors in the general set-up. *The American Statistician*. Online.

Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation. *Journal of Econometrics*, 115(2):293–346.

Clarke, B. S. and Barron, A. R. (1990). Information-theoretic asymptotics of bayes methods. *IEEE Transactions on Information Theory*, 36(3):453–471.

Friel, N. and Wyse, J. (2012). Estimating the evidence - a review. *Statistica Neerlandica*, 66(3):288–308.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(3):501–514.

Good, I. J. (1967). A Bayesian significance test for multinomial distributions. (With discussion). *Journal of the Royal Statistical Society. Series B. Methodological*, 29:399–431.

Good, I. J. (1992). The Bayes/non-Bayes compromise: a brief review. *Journal of the American Statistical Association*, 87(419):597–606.

Hall, P. and Stewart, M. (2005). Theoretical analysis of power in a two-component normal mixture model. *Journal of Statistical Planning and Inference*, 134(1):158 – 179.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association*, 103(481):410–423.

Liu, X. and Shao, Y. (2004). Asymptotics for the likelihood ratio test in a two-component normal mixture model. *Journal of Statistical Planning and Inference*, 123(1):61 – 81.

Moreno, E., Bertolino, F., and Racugno, W. (1998). An intrinsic limiting procedure for model selection and hypotheses testing. *Journal of the American Statistical Association*, 93(444):1451–1460.

Moreno, E., Girn, F. J., and Casella, G. (2010). Consistency of objective bayes factors as the model dimension grows. *The Annals of Statistics*, 38(4):1937–1952.

Shafer, G. (1982). Lindley's paradox. *Journal of the American Statistical Association*, 77(378):325–351. With discussion and with a reply by the author.

van Erven, T. and Harremoes, P. (2014). Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820.

Wang, M. and Maruyama, Y. (2016). Consistency of bayes factor for nonnested model selection when the model dimension grows. *Bernoulli*, 22(4):2080–2100.

Wang, R. and Xu, X. (2020). A bayesian-motivated test for high-dimensional linear regression models with fixed design matrix. *Statistical Papers*.

Zhou, Q. and Guan, Y. (2018). On the null distribution of bayes factors in linear regression. *Journal of the American Statistical Association*, 113(523):1362–1371.