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

Friday 4th October, 2019

Abstract

In Bayesian hypotheses testing framework, Bayes factor and its variants have been extensively studied and have been shown to have good performance in a wealth of complex testing problems. This motivates us to use Bayes factors to construct frequentist tests. In this paper, we investigate the asymptotic distribution of the Bayes factor and a class of its variants named generalized fractional Bayes factor. It shows that the classical Bayes factor has Wilks phenomenon only for a restricted class of priors while the generalized fractional Bayes factor has Wilks phenomenon for general priors. Frequentist tests based on Bayes factors are constructed using the Wilks phenomenon. We also extend the result to the general integrated likelihood ratio test. For regular models, the proposed tests have the same asymptotic local power as the likelihood ratio test. However, the proposed methodology has a wider application scope than the likelihood ratio test. In particular, our methodology can be applied even if the likelihood function is unbounded. We use three examples to illustrate the proposed methodology. In these examples, the likelihood ratio test may not be well defined or have undesirable behavior while the proposed tests have good performance.

Key words: Bayes consistency, power posterior, integrated likelihood, mixture model, posterior Bayes factor.

1 Introduction

The Likelihood ratio test plays a dominant role in parametric hypotheses testing. The fundamental lemma of Neyman and Pearson tells us that the likelihood ratio test (LRT) is the most powerful

*Corresponding author

Email address: xuxz@bit.edu.cn

test if the null and the alternative hypotheses are both simple. For composite hypotheses, the unknown parameters in the likelihood ratio test statistic (LRTS) are estimated by the maximum likelihood estimate (MLE). In a celebrated paper, (Wilks, 1938) proved that the asymptotic null distribution of the LRTS is free of nuisance parameters for regular models. With this important property, one can determine the critical value of the LRTS using its asymptotic null distribution. In this paper, we say a test statistic has Wilks phenomenon if its asymptotic null distribution does not depend on the nuisance parameters. Although the LRT is very successful in many specific problems, it also has some weakness. First, except for a rather restricted class of models, the explicit form of the LRTS is not available and numerical optimization method must be used to obtain the LRTS. Unfortunately, if the likelihood function is not concave and has multiple local maxima, the numerical optimization procedure may be highly nontrivial and there is no universally applicable optimization method. Second, even for some fairly regular models, the MLE does not exist with a positive probability; see, e.g, Fienberg and Rinaldo (2012) and Rinaldo et al. (2013). Hence the LRT is not always well defined. Worse still, in some problems the likelihood functions are unbounded with probability 1 and hence the LRT is not defined; see, e.g., Le Cam (1990). In fact, the unbounded likelihood occurs not only in artificial models, but also in some widely used models, such as the mixture models with unknown component location and scale parameters (Chen, 2017).

On the Bayesian side, the conventional tool for hypothesis testing is Bayes factor (Jeffreys, 1931). Bayes factor has been widely used by practitioners; see Kass and Raftery (1995) for a review. A delightful feature of the Bayes factor over the LRTS is that Bayes factor is well defined for any model provided the prior distributions are proper. Hence Bayes factor methodology is often used for complex models. The universal applicability of the Bayes factor motivates us to use the Bayes factor as a test statistic and construct a frequentist test. Note that the historical development of Bayes factor is almost orthogonal to frequentist test. A main reason is that the Bayes factor methodology does not intend to control the frequentist type I error rate and hence does not form a frequentist test. In fact, as shown by Clarke and Barron (1990), the asymptotic distribution of Bayes factor depends on the prior density at the true parameter. As a result, the Bayes factor does not have Wilks phenomenon in general. Nevertheless, we shall show that if the priors are carefully chosen, the Bayes factor indeed has Wilks phenomenon and can be used to construct a frequentist test. We prove that the test so constructed has the same asymptotic local power as the LRT. Our theoretical results does not require that the MLE exists or the likelihood is bounded. Hence theoretically, the test based on Bayes factor has a wider application scope than the LRT. In practice, however, the test based on Bayes factor may be difficult to implement. In fact, the priors which ensure the Wilks phenomenon of Bayes factor are often complicated. Also, it is known that the computation of Bayes factor is highly nontrivial. Fortunately, these problems can be solved by some variants of Bayes factor.

In Bayesian literature, several variants of Bayes factor have been proposed to reduce its sensitivity to priors. Two important variants of Bayes factor are the posterior Bayes factor (Aitkin,

1991) and the fractional Bayes factor (O'Hagan, 1995). We consider a class of statistics named generalized fractional Bayes factor which includes theses two variants of Bayes factor. We prove that the Wilks phenomenon of the generalized fractional Bayes factor holds for any reasonable prior. Also, the computation of the generalized fractional Bayes factor is straightforward provided sampling from the power posterior distribution is easy. Hence compared with Bayes factor, the generalized fractional Bayes factor is more suitable for constructing frequentist tests.

In another viewpoint, the generalized fractional Bayes factor is the ratio of the expectations of the power likelihood with respect to power posterior distributions. Hence the generalized fractional Bayes factor can be regarded as an integrated likelihood ratio test. For some complex models, the power posterior is intractable. In this case, a feasible strategy is to approximate the posterior distribution by simple form distributions by variational inference; see Blei et al. (2017) and the references therein. Motivated by variantional inference, we consider the general integrated likelihood ratio test statistic which uses the expectations of the power likelihood with respect to general weight functions. We prove that if the behavior of the weight functions is close to the power posterior, then the general integrated likelihood ratio test also has Wilks phenomenon. In particular, we show that the weight functions obtained by a variational method satisfies our requirements for the weight functions.

We use three examples to illustrate the good properties of the proposed methodology. The first example is the full-rank exponential family. We show that the conditions of our general theory is satisfied by the full-rank exponential family. Hence the proposed methodology can be used in common regular models. In the second and the third examples, we consider testing the homogeneity in two submodels of the normal mixture model. The LRT has bad behavior in these two examples. In fact, for the first submodel, the likelihood is unbounded and thus the LRT is not defined. For the second submodel, Hall and Stewart (2005) showed that the LRT has trivial power under $n^{-1/2}$ local alternative hypothesis. We prove that the proposed methodology has good asymptotic power behavior for both submodels.

The paper is organized as follow. In Section 2, we investigated the Wilks phenomenon of the Bayes factor and the generalized Bayes factor and use the Wilks phenomenon to construct frequentist tests. In Section 3, we use three examples to illustrate the behavior of the proposed methodology. Section 4 concludes the paper. All technical proofs are in Appendix.

2 Wilks phenomenon of Bayes factors

Let $\mathbf{X}^n = (X_1, \ldots, X_n)$ be independent identically distributed (iid) observations taking values in a standard measurable space $(\mathcal{X}; \mathscr{A})$. Suppose that the possible distribution P_θ of X_i has a density $p(x|\theta)$ with respect to μ, a σ-finite measure on $\mathcal{X}$. Denote by P_θ^n the joint distribution of $\mathbf{X}^n$. Let $p_n(\mathbf{x}^n|\theta) = \prod_{i=1}^n p(x_i|\theta)$ denote the density of P_θ^n with respect to the n-fold product measure μ^n. The parameter θ takes its values in Θ, an open subset of $\mathbb{R}^p$. Suppose $\theta = (\nu^\top, \xi^\top)^\top$, where ν is a

p_0 dimensional subvector and ξ is a $p - p_0$ dimensional subvector and "$\top$" means the transpose of a matrix. We would like to test the hypotheses

$$H : \theta \in \Theta_0 \quad \text{v.s.} \quad K : \theta \in \Theta \backslash \Theta_0,$$

where the null space Θ_0 is a p_0-dimensional subspace of Θ defined as

$$\Theta_0 = \{(\nu^\top, \xi^\top)^\top : (\nu^\top, \xi^\top)^\top \in \Theta,\, \xi = \xi_0\}.$$

If the null hypothesis is true, we denote by $\theta_0 = (\nu_0^\top, \xi_0^\top)^\top$ the true parameter which generates the data.

In Bayesian hypothesis testing framework, a fundamental tool is Bayes factor

$$\mathrm{BF}(\mathbf{X}^n) = \frac{\int_\Theta p_n(\mathbf{X}^n|\theta)\pi(\theta)\,\mathrm{d}\theta}{\int_{\tilde{\Theta}_0} p_n(\mathbf{X}^n|\nu, \xi_0)\pi_0(\nu)\,\mathrm{d}\nu},$$

where $\tilde{\Theta}_0 = \{\nu : (\nu^\top, \xi_0^\top)^\top \in \Theta_0\}$ and $\pi(\theta)$ and $\pi_0(\nu)$ are the prior densities of parameters under the alternative and the null hypotheses, respectively. The priors $\pi(\theta)$ and $\pi_0(\nu)$ may be improper, that is, $\int_\Theta \pi(\theta)\,\mathrm{d}\theta = +\infty$, $\int_{\tilde{\Theta}_0} \pi_0(\nu)\,\mathrm{d}\nu = +\infty$. Conventionally, the null hypothesis is rejected if $\mathrm{BF}(\mathbf{X}^n)$ is larger than certain threshold. The choice of threshold is mostly empirical in the literature. For example, Jeffreys (1961) suggested that the evidence against the null hypothesis is *decisive* if $\mathrm{BF}(\mathbf{X}^n) > 100$ while Kass and Raftery (1995) suggested that the evidence is *very strong* if $\mathrm{BF}(\mathbf{X}^n) > 150$. Unfortunately, these choices of threshold are not theoretically justified. Worse still, for improper priors, these choices of threshold suffer from the Lindley paradox, that is, the resulting test procedure largely depends on the arbitrary constants in priors densities; see, e.g., Shafer (1982).

In this paper, we treat the Bayes factor as a frequentist test statistic and the threshold is chosen to control the type I error rate. To achieve this, we need to investigate the asymptotic distribution of Bayes factor. We make the following assumption which is adapted from Kleijn and van der Vaart (2012).

Assumption 1. *The parameter spaces Θ and $\tilde{\Theta}_0$ are open subsets of $\mathbb{R}^p$ and $\mathbb{R}^{p_0}$, respectively. The parameters θ_0 and ν_0 are inner points of Θ and $\tilde{\Theta}_0$, respectively. The derivative*

$$\dot{\ell}_{\theta_0}(X) = \frac{\partial}{\partial \theta} \log p(X|\theta)\Big|_{\theta=\theta_0}$$

exists P_{θ_0}-a.s. and satisfies $P_{\theta_0}\dot{\ell}_{\theta_0} = \mathbf{0}_p$, where Pf means the expectation of $f(X)$ when X has distribution P. The Fisher information matrix $I(\theta_0) = P_{\theta_0}\dot{\ell}_{\theta_0}\dot{\ell}_{\theta_0}^\top$ is positive-definite. For every $M > 0$,

$$\sup_{\|h\| \le M} \left| R_n(\theta_0 \| \theta_0 + n^{-1/2}h) + h^\top I(\theta_0)\Delta_{n,\theta_0} - \frac{1}{2}h^\top I(\theta_0) h \right| \xrightarrow{P_{\theta_0}^n} 0,$$

where $R_n(\theta\|\theta') = \log\{p_n(\mathbf{X}^n|\theta)/p_n(\mathbf{X}^n|\theta')\}$ is the log-likelihood ratio between $p_n(\mathbf{X}^n|\theta)$ and $p_n(\mathbf{X}^n|\theta')$, and $\Delta_{n,\theta_0} = n^{-1/2}\sum_{i=1}^n I(\theta_0)^{-1}\dot{\ell}_{\theta_0}(X_i)$.

Assumption 1 assumes that the likelihood function $p_n(\mathbf{X}^n|\theta)$ has good local behavior when θ is close to θ_0. Although Assumption 1 is about the full likelihood function, it also implies the analogous assumption for the null likelihood function $p_n(\mathbf{X}^n|\nu,\xi_0)$. Let $\mathbf{I}_{p_0}$ denote the p_0 dimensional identity matrix, $\mathbf{J}=(\mathbf{I}_{p_0},\mathbf{0}_{p_0\times(p-p_0)})^\top$. Define

$$\dot\ell_{\theta_0}^{(0)}(X)=\mathbf{J}^\top\dot\ell_{\theta_0}(X),\quad I_\nu(\theta)=\mathbf{J}^\top I(\theta)\mathbf{J},\quad \Delta_{n,\theta_0}^{(0)}=\frac{1}{\sqrt{n}}\sum_{i=1}^n I_\nu(\theta_0)^{-1}\dot\ell_{\theta_0}^{(0)}(X_i).$$

Then Assumption 1 implies that

$$\sup_{\|h\|\leq M}\left|R_n(\theta_0\|\nu_0+n^{-1/2}h,\xi_0)+h^\top I_\nu(\theta_0)\Delta_{n,\theta_0}^{(0)}-\frac{1}{2}h^\top I_\nu(\theta_0)h\right|\xrightarrow{P_{\theta_0}^n}0.$$

Note that Assumption 1 imposes no condition on the likelihood function when θ is deviated from θ_0. In contrast, in the theory of the LRT, it is assumed that the MLE exists and is consistent; see, e.g., Wilks (1938) and van der Vaart (1998), Theorem 16.7. The existence and consistency of the MLE require that the likelihood value is negligible when θ is deviated from θ_0, which is not true for some important models. Hence we will not assume this strong condition. Instead, we shall utilize the $\sqrt{n}$-consistency of the posterior density $\pi(\theta|\mathbf{X}^n)=p_n(\mathbf{X}^n|\theta)\pi(\theta)/\int_\Theta p_n(\mathbf{X}^n|\theta)\pi(\theta)\,\mathrm{d}\theta$. We say $\pi(\theta|\mathbf{X}^n)$ is $\sqrt{n}$-consistent if for sufficiently large n, $\int_\Theta p_n(\mathbf{X}^n|\theta)\pi(\theta)\,\mathrm{d}\theta<\infty$, and for every $M_n\to\infty$,

$$\int_{\{\theta:\|\theta-\theta_0\|>M_n/\sqrt{n}\}}\pi(\theta|\mathbf{X}^n)\,\mathrm{d}\theta\xrightarrow{P_{\theta_0}^n}0.$$

The $\sqrt{n}$-consistency of $\pi_0(\nu|\mathbf{X}^n)$ is similarly defined.

We shall derive the asymptotic distribution of $\mathrm{BF}(\mathbf{X}^n)$ under the null hypothesis as well as the local alternative hypothesis. By local alternative hypothesis we means that the true parameter is θ_n and $\sqrt{n}(\theta_n-\theta_0)\to\eta$. Let

$$I_{\xi|\nu}(\theta)=\tilde{\mathbf{J}}^\top I(\theta)\tilde{\mathbf{J}}-\tilde{\mathbf{J}}^\top I(\theta)\mathbf{J}(\mathbf{J}^\top I(\theta)\mathbf{J})^{-1}\mathbf{J}^\top I(\theta)\tilde{\mathbf{J}},$$

where $\tilde{\mathbf{J}}=(\mathbf{0}_{(p-p_0)\times p_0},\mathbf{I}_{(p-p_0)})^\top$. It can be seen that $|I(\theta)|=|I_\nu(\theta)|\cdot|I_{\xi|\nu}(\theta)|$. Let $\chi^2(p-p_0,\eta^\top\tilde{\mathbf{J}}I_{\xi|\nu}(\theta_0)\tilde{\mathbf{J}}^\top\eta)$ denote a noncentral chi-squared random variable with $p-p_0$ degrees of freedom and noncentrality parameter $\eta^\top\tilde{\mathbf{J}}I_{\xi|\nu}(\theta_0)\tilde{\mathbf{J}}^\top\eta$. The following theorem gives the asymptotic distribution of Bayes factor.

Theorem 1. *Suppose that Assumption 1 holds, $\pi(\theta)$ is continuous at θ_0 with $\pi(\theta_0)>0$, $\pi_0(\nu)$ is continuous at ν_0 with $\pi_0(\nu_0)>0$, $\pi(\theta|\mathbf{X}^n)$, $\pi_0(\nu|\mathbf{X}^n)$ are $\sqrt{n}$-consistent. Suppose $\{\theta_n\}$ satisfies $\sqrt{n}(\theta_n-\theta_0)\to\eta$. Then*

$$2\log BF(\mathbf{X}^n)+(p-p_0)\log\left(\frac{n}{2\pi}\right)\overset{P_{\theta_n}^n}{\rightsquigarrow}\chi^2(p-p_0,\eta^\top\tilde{\mathbf{J}}I_{\xi|\nu}(\theta_0)\tilde{\mathbf{J}}^\top\eta)+2\log\frac{\left|I_{\xi|\nu}(\theta_0)\right|^{-\frac{1}{2}}\pi(\theta_0)}{\pi_0(\nu_0)},$$

where "$\overset{P_{\theta_n}^n}{\rightsquigarrow}$" means the weak convergence when $\mathbf{X}^n$ is from $P_{\theta_n}^n$.

Theorem 1 implies that the asymptotic null distribution of the Bayes factor depend on the nuisance parameters and the priors. Hence the Bayes factor does not have Wilks phenomenon for general priors. Nevertheless, the dependence of the asymptotic null distribution on the nuisance parameters can be cancelled by a class of priors. In fact, Theorem 1 implies that the asymptotic null distribution of $\mathrm{BF}_t(\mathbf{X}^n)$ is free of the nuisance parameter ν if and only if

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

for some constant c and for any $\nu \in \tilde{\Theta}$. Unfortunately, simple examples show that (1) does not hold for many popular objective priors for Bayes factor, including intrinsic priors (Berger and Pericchi, 1996), fractional intrinsic priors (De Santis and Spezzaferri, 1997), divergence-based priors (Bayarri and Garca-Donato, 2008), expected-posterior priors (Perez, 2002). Nevertheless, (1) holds for a large class of priors. For instance, the left hand side of (1) is equal to 1 if $\pi_0(\nu) = |I_\nu(\nu,\xi_0)|^{1/2}$ and $\pi(\theta) = |I(\theta)|^{1/2}$, that is, $\pi_0(\nu)$ and $\pi(\theta)$ are the Jeffreys priors under the null and the alternative hypotheses, respectively. In general, (1) is constant provided $\pi(\theta) = \pi(\xi|\nu)\pi(\nu)$ satisfies

$$\pi(\nu) = \pi_0(\nu), \quad \pi(\xi_0|\nu) \propto |I_{\xi|\nu}(\nu,\xi_0)|^{1/2}. \tag{2}$$

A class of priors satisfying (2) is the unit information priors proposed by Kass and Wasserman (1995), which is defined as

$$\pi(\nu) = \pi_0(\nu), \quad \pi(\xi|\nu) = |\Sigma_\xi(\nu)|^{-1/2} f\left((\xi-\xi_0)^\top \Sigma_\xi(\nu)^{-1}(\xi-\xi_0)\right),$$

where $\Sigma_\xi(\nu)$ satisfies $|\Sigma_\xi(\nu)| = |I_{\xi|\nu}(\nu,\xi_0)|^{-1}$.

We have seen that with carefully chosen priors, Bayes factor has Wilks phenomenon. In this case, Bayes factor can be treated as a frequentist statistic and the critical value is determined by the asymptotic distribution. However, this approach still has some weaknesses. First, the Wilks phenomenon of Bayes factor holds only for a restricted class of priors. As we have already pointed out, many popular objective priors do not satisfy (2). Also, for some models, Fisher information matrix has a complicated form. In this case, priors satisfying (2) may be undesirable. Second, the computation of Bayes factor is a difficult problem in general; see Friel and Wyse (2012) for a review on the computation of Bayes factor.

Now we turn to the variants of Bayes factor. We denote by $L_t(\mathbf{X}^n) = \int_\Theta [p_n(\mathbf{X}^n|\theta)]^t \pi(\theta)\,\mathrm{d}\theta$ the power marginal likelihood with power $t > 0$, and $\pi_t(\theta|\mathbf{X}^n) = [p_n(\mathbf{X}^n|\theta)]^t \pi(\theta)/L_t(\mathbf{X}^n)$ the power posterior density. We define $L_t^{(0)}(\mathbf{X}^n)$ and $\pi_{0,t}(\nu|\mathbf{X}^n)$ in a similar way. For $t > 0$, let $\mathrm{BF}_t(\mathbf{X}^n) = L_t(\mathbf{X}^n)/L_t^{(0)}(\mathbf{X}^n)$. Then $\mathrm{BF}_1(\mathbf{X}^n)$ is the usual Bayes factor. It can be expected that $\mathrm{BF}_t(\mathbf{X}^n)$ depends on the nuisance parameters and the priors in the same way for different values of $t > 0$. Hence the ratio of $\mathrm{BF}_t(\mathbf{X}^n)$ for two different t may cancel the dependence on the nuisance parameters and the priors. For $a > b > 0$, let $\Lambda_{a,b}(\mathbf{X}^n) = \mathrm{BF}_a(\mathbf{X}^n)/\mathrm{BF}_b(\mathbf{X}^n)$. This class of statistics includes two important variants of Bayes factor, namely, the posterior Bayes factor proposed by Aitkin (1991) and the fractional Bayes factor proposed by O'Hagan (1995). In fact, the posterior Bayes factor is equal to $\Lambda_{2,1}(\mathbf{X}^n)$ and the fractional Bayes factor is equal to $\Lambda_{1,b}(\mathbf{X}^n)$ for $b \in (0,1)$. For this reason we shall call $\Lambda_{a,b}(\mathbf{X}^n)$ the generalized fractional Bayes factor throughout the paper.

Since we treat $\Lambda_{a,b}(\mathbf{X}^n)$ as a frequentist statistic, the Wilks phenomenon of $\Lambda_{a,b}(\mathbf{X}^n)$ should be examined. We shall assume a is fixed as $n \to \infty$ and consider three settings for b: (a) b is fixed; (b) $b \to 0$ and $nb \to \infty$; (c) $nb \to b^* \in (0, +\infty)$. Note that the posterior Bayes factor uses fixed $a = 2$, $b = 1$ for all n, hence belongs to the setting (a). On the other hand, the fractional Bayes factor uses fixed $a = 1$ but a varying b as $n \to \infty$. In fact, O'Hagan (1995) proposed three ways to set b, the first one is $b = m_0/n$ for a fixed m_0, the second one is $b = n^{-1/2}$ and the third one is $b = \log(n)/n$. It can be seen that their first choice belongs to our setting (c) and their last two choices belong to our setting (b). In what follows, we shall derive the asymptotic distribution of $\Lambda_{a,b}(\mathbf{X}^n)$ in these three settings. Like Theorem 1, the $\sqrt{n}$-consistency of power posterior will play an important role. Suppose $t > 0$ is fixed as $n \to \infty$, we say $\pi_t(\theta|\mathbf{X}^n)$ is $\sqrt{n}$-consistent if for sufficiently large n, $\int_\Theta [p_n(\mathbf{X}^n|\theta)]^t \pi(\theta)\, \mathrm{d}\theta < \infty$, and for every $M_n \to \infty$,

$$\int_{\{\theta:\|\theta-\theta_0\|>M_n/\sqrt{n}\}} \pi_t(\theta|\mathbf{X}^n)\, \mathrm{d}\theta \xrightarrow{P^n_{\theta_0}} 0.$$

The $\sqrt{n}$-consistency of $\pi_{0,t}(\nu|\mathbf{X}^n)$ is similarly defined. In addition, further assumptions are required in the second and third settings.

For two parameters θ_1 and θ_2, the t order Rényi divergence ($0 < t < 1$) between P_{θ_1} and P_{θ_2} is defined as

$$D_t(\theta_1||\theta_2) = -\frac{1}{1-t} \log \rho_t(\theta_1, \theta_2),$$

where $\rho_t(\theta_1, \theta_2) = \int_{\mathcal{X}} p(X|\theta_1)^t p(X|\theta_2)^{1-t}\, \mathrm{d}\mu$ is the so-called Hellinger integral. Let $D_1(\theta_1\|\theta_2) = \int_{\mathcal{X}} \log(p(X|\theta_0)/p(X|\theta))\, p(X|\theta_0)\, \mathrm{d}\mu$ be the Kullback-Leibler divergence between P_{θ_1} and P_{θ_2}. It is known that $\lim_{t\uparrow 1} D_t(\theta_1\|\theta_2) = D_1(\theta_1\|\theta_2)$; see, e.g., van Erven and Harremoes (2014), Theorem 5. Let $V(\theta_0\|\theta) = P_{\theta_0}(\log(p(X|\theta_0)/p(X|\theta)) - D_1(\theta_0\|\theta))^2$. We shall make the following two assumptions in the second and third settings.

Assumption 2. *For any fixed $t \in (0,1]$, we assume $D_t(\theta_0\|\theta)$ satisfies the following conditions:*

- *$D_1(\theta_0\|\theta)$ is finite for all $\theta \in \Theta$;*
- *for each $\delta > 0$, there exists a $\epsilon > 0$ such that $D_t(\theta_0\|\theta) \geq \epsilon$ for $\|\theta - \theta_0\| \geq \delta$;*
- *as $\theta \to \theta_0$,*

$$D_t(\theta_0\|\theta) = (1 + o(1))\frac{t}{2}(\theta - \theta_0)^\top I(\theta_0)(\theta - \theta_0), \quad V(\theta_0\|\theta) = O\left(\|\theta - \theta_0\|^2\right).$$

See, e.g., van Erven and Harremoes (2014), Section III. H.

Assumption 3. *There exist $t^* \in (0,1)$, $c^*, c^\dagger > 0$, $t_0^* \in (0,1)$, $c_0^*, c_0^\dagger > 0$ such that*

$$\int_\Theta \exp\left\{-c^* D_{1-t^*}(\theta_0\|\theta)\right\}\pi(\theta)\,\mathrm{d}\theta < \infty, \quad \int_\Theta V(\theta_0\|\theta)\exp\left\{-c^\dagger D_1(\theta_0\|\theta)\right\}\pi(\theta)\,\mathrm{d}\theta < \infty,$$

$$\int_{\tilde{\Theta}_0} \exp\left\{-c_0^* D_{1-t_0^*}(\nu_0\|\nu)\right\}\pi_0(\nu)\,\mathrm{d}\nu < \infty, \quad \int_{\tilde{\Theta}_0} V(\nu_0\|\nu)\exp\left\{-c_0^\dagger D_1(\nu_0\|\nu)\right\}\pi_0(\nu)\,\mathrm{d}\nu < \infty.$$

Theorem 2. *Suppose that Assumption 1 holds. Suppose $\pi(\theta)$ is continuous at θ_0 with $\pi(\theta_0) > 0$, $\pi_0(\nu)$ is continuous at ν_0 with $\pi_0(\nu_0) > 0$. Suppose $a > b > 0$ and a is fixed as $n \to \infty$, Suppose $\pi_a(\theta|\mathbf{X}^n)$ and $\pi_{0,a}(\nu|\mathbf{X}^n)$ are $\sqrt{n}$-consistent. Suppose $\{\theta_n\}$ satisfies $\sqrt{n}(\theta_n - \theta_0) \to \eta$. Then the following assertions hold.*

(a) Suppose b is fixed as $n \to \infty$, $\pi_b(\theta|\mathbf{X}^n)$ and $\pi_{0,b}(\nu|\mathbf{X}^n)$ are $\sqrt{n}$-consistent. Then

$$\frac{2}{a-b}\log\Lambda_{a,b}(\mathbf{X}^n) + \frac{p-p_0}{a-b}\log\left(\frac{a}{b}\right) \overset{P^n_{\theta_n}}{\rightsquigarrow} \chi^2(p-p_0, \eta^\top\tilde{\mathbf{J}}I_{\xi|\nu}(\theta_0)\tilde{\mathbf{J}}^\top\eta).$$

(b) Suppose Assumptions 2, 3 hold and as $n \to \infty$, $b \to 0$, $bn \to \infty$. Then

$$\frac{2}{a-b}\log\Lambda_{a,b}(\mathbf{X}^n) + \frac{p-p_0}{a-b}\log\left(\frac{a}{b}\right) \overset{P^n_{\theta_n}}{\rightsquigarrow} \chi^2(p-p_0, \eta^\top\tilde{\mathbf{J}}I_{\xi|\nu}(\theta_0)\tilde{\mathbf{J}}^\top\eta).$$

(c) Suppose Assumptions 2, 3 hold and as $n \to \infty$, $nb \to b^* \in (c^*, +\infty)$. Then

$$2\log\Lambda_{a,b}(\mathbf{X}^n) + (p-p_0)\log\left(\frac{an}{2\pi}\right) \overset{P^n_{\theta_n}}{\rightsquigarrow} a\chi^2(p-p_0, \eta^\top\tilde{\mathbf{J}}I_{\xi|\nu}(\theta_0)\tilde{\mathbf{J}}^\top\eta) + 2\log\left(\frac{\left|I_{\xi|\nu}(\theta_0)\right|^{-\frac{1}{2}}\pi(\theta_0)}{\pi_0(\nu_0)}\right)$$
$$- 2\log\left(\frac{\int_\Theta \exp\left\{-b^* D_1(\theta_0\|\theta)\right\}\pi(\theta)\,\mathrm{d}\theta}{\log\int_{\tilde{\Theta}_0}\exp\left\{-b^* D_1(\theta_0\|\nu,\xi_0)\right\}\pi_0(\nu)\,\mathrm{d}\nu}\right).$$

Theorem 2 gives the asymptotic distribution of $\Lambda_{a,b}(\mathbf{X}^n)$ under the null hypothesis and the local alternative hypothesis. It can be seen that in the first two settings, $\Lambda_{a,b}(\mathbf{X}^n)$ has Wilks phenomenon for any continuous and positive prior density. However, when nb tends to a constant, $\Lambda_{a,b}(\mathbf{X}^n)$ does not have Wilks phenomenon in general, although it is invariant when multiplying the priors with constants. In the first two settings, we can construct a test with asymptotic type I error rate α. We reject the null hypothesis when $2\log\Lambda_{a,b}(\mathbf{X}^n) > (a-b)\chi^2_\alpha(p-p_0) - (p-p_0)\log(a/b)$ where $\chi^2_\alpha(p-p_0)$ is the upper α quantile of a chi-squared random variable with $p-p_0$ degrees of freedom. By Theorem 2, the resulting test has asymptotic local power

$$\Pr\left(\chi^2(p-p_0,\delta) > \chi^2_\alpha(p-p_0)\right). \tag{3}$$

It is known that, under certain regular conditions, (3) is also the asymptotic local power of the likelihood ratio test. In this view, $\Lambda_{a,b}(\mathbf{X}^n)$ enjoys good frequentist properties. O'Hagan (1995) argued that when robustness is no concern, it is natural to set b as small as possible since it makes maximal possible use of the data for model comparison. However, Theorem 2 implies that the frequentist test power of $\Lambda_{a,b}(\mathbf{X}^n)$ is in fact independent of the choice of b. Note that in Theorem 2, the second and the third settings require more assumptions than the first setting. Hence in frequentist perspective, it is preferred to use fixed b.

In both Theorem 1 and Theorem 2, a key assumption is the $\sqrt{n}$-consistency of the power posterior distribution, which enables us to not assume the existence and the consistency of the MLE. We would like to give sufficient conditions for the $\sqrt{n}$-consistency of $\pi_t(\theta|\mathbf{X}^n)$. In our definition of the $\sqrt{n}$-consistency of $\pi_t(\theta|\mathbf{X}^n)$, we assume the finiteness of $L_t(\mathbf{X}^n)$. It is known that this requirement is naturally satisfied if $t = 1$ and the prior is proper. In fact, the following proposition shows that when the prior is proper, $L_t(\mathbf{X}^n)$ is always finite for $t \leq 1$ and is not always finite for $t > 1$.

Proposition 1. *Suppose the prior density $\pi(\theta)$ is proper. If $t \leq 1$, for any model $\{P_\theta, \theta \in \Theta\}$ and any n, $L_t(\mathbf{X}^n) < +\infty$ $P_{\theta_0}^n$-a.s.. If $t > 1$, there exists a model such that for any n, $L_t(\mathbf{X}^n) = +\infty$ $P_{\theta_0}^n$-a.s..*

The above proposition implies that the behavior of $L_t(\mathbf{X}^n)$ for $t > 1$ may be undesirable. For this reason shall only consider the $\sqrt{n}$-consistency of $\pi_t(\theta|\mathbf{X}^n)$ for $t \leq 1$. For $t = 1$, the $\sqrt{n}$-consistency of $\pi_1(\theta|\mathbf{X}^n)$ is the $\sqrt{n}$-consistency of the posterior distribution. For the case where the prior distribution is proper, the consistency rate of the posterior has been well studied in the literature; see, e.g., Ghosal et al. (2000), Shen and Wasserman (2001), van der Vaart and Ghosal (2007). A popular and convenient way of establishing the consistency of posterior is through the condition that suitable test sequences exist. For example, Theorem 3.1 of Kleijn and van der Vaart (2012) assumes that for every $\epsilon > 0$, there exists a sequence of tests ϕ_n such that

$$P_{\theta_0}^n \phi_n \to 0, \qquad \sup_{\|\theta-\theta_0\|\geq\epsilon} P_\theta^n(1-\phi_n) \to 0. \tag{4}$$

This condition is satisfied when the parameter space is compact and the model is suitably continuous; see Theorem 3.2 of Kleijn and van der Vaart (2012). However, if the parameter space is not compact, one may have to manually construct a test sequence satisfying the condition (4). Also, if the prior distribution is not proper, existing results can not be directly applied.

The consistency of $\pi_t(\theta|\mathbf{X}^n)$ for $0 < t < 1$ is different from $t = 1$. Walker and Hjort (2001) considered the Hellinger consistency of $L_{1/2}(\theta|\mathbf{X}^n)$. They derived the consistency of $\pi_{1/2}(\theta|\mathbf{X}^n)$ under simple conditions. Recently, Bhattacharya et al. (2019) further developed the idea of Walker and Hjort (2001) and derived a general bounds for the consistency of $\pi_t(\theta|\mathbf{X}^n)$ for $0 < t < 1$. However, their result can not yield the $\sqrt{n}$-consistency for parametric models. We shall prove the $\sqrt{n}$-consistency of $\pi_t(\theta|\mathbf{X}^n)$ for $0 < t < 1$ under certain conditions on the Rényi divergence between distributions in the family $\{P_\theta : \theta \in \Theta\}$.

Assumption 4. *For some $t \in (0,1)$, there exist positive constants δ, ϵ and C such that, $D_t(\theta_0||\theta) \geq C\|\theta - \theta_0\|^2$ for $\|\theta - \theta_0\| \leq \delta$ and $D_t(\theta_0||\theta) \geq \epsilon$ for $\|\theta - \theta_0\| > \delta$.*

Remark 1. A remarkable property of Rényi divergence is the equivalence of all D_t, $t \in (0,1)$. If $0 < t_1 < t_2 < 1$, then

$$\frac{t_1}{1-t_1}\frac{1-t_2}{t_2} D_{t_2}(\theta_1||\theta_2) \leq D_{t_1}(\theta_1||\theta_2) \leq D_{t_2}(\theta_1||\theta_2).$$

See, e.g., van Erven and Harremoes (2014). As a result, if Assumption 4 holds for some $t \in (0,1)$, then it will hold for every $t \in (0,1)$.

Proposition 2. *Suppose Assumptions 1 and 4 hold and $t \in (0,1)$ is a fixed number. Suppose $\pi(\theta)$ is continuous at θ_0 with $\pi(\theta_0) > 0$. Suppose there exists a $c^* > 0$ such that*

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

Then $\pi_t(\theta|\mathbf{X}^n)$ is $\sqrt{n}$-consistent.

Note that Assumption 4 is a weaker version Assumption 2. In fact, Assumption 2 requires the exact form of the local expansion of $D_t(\theta_0\|\theta)$ while Assumption 4 only assumes $D_t(\theta_0\|\theta)$ has certain lower bound. For the normal mixture model in Section 4.2, it is relatively simple to verify Assumption 4 compared with Assumption 2. Also, it may be more convenient to verify Assumption 4 than to directly construct a test sequence satisfying the condition (4). Thus, it can be recommended to use $\Lambda_{a,b}(\mathbf{X}^n)$ with fixed $0 < b < a < 1$.

3 Integrated likelihood ratio test

It is known that the computation of Bayes factor is not trivial even if it is easy to sample from the posterior distribution. Surprisingly, if b and $a-b$ are comparable, $\Lambda_{a,b}(\mathbf{X}^n)$ can be easily computed by sampling from the power posterior distribution. To see this, write

$$\Lambda_{a,b}(\mathbf{X}^n) = \frac{\int_\Theta \left[p_n(\mathbf{X}^n|\theta)\right]^{a-b} \pi_b(\theta|\mathbf{X}^n)\, \mathrm{d}\theta}{\int_{\tilde{\Theta}_0} \left[p_n(\mathbf{X}^n|\nu,\xi_0)\right]^{a-b} \pi_{0,b}(\nu|\mathbf{X}^n)\, \mathrm{d}\nu}.$$

We can independently generate $\theta_1, \ldots, \theta_m$ and $\nu_1, \ldots, \nu_m$ according to $\pi_b(\theta|\mathbf{X}^n)$ and $\pi_{0,b}(\nu|\mathbf{X}^n)$ for a large m and approximate $\Lambda_{a,b}(\mathbf{X}^n)$ as

$$\frac{\sum_{i=1}^m \left[p_n(\mathbf{X}^n|\theta_i)\right]^{a-b}}{\sum_{i=1}^m \left[p_n(\mathbf{X}^n|\nu_i,\xi_0)\right]^{a-b}}.$$

Usually, sampling from the power posterior can be implemented by a Markov chain Monte Carlo (MCMC) procedure. Alternatively, when models are complex or datasets are large, a popular strategy named variational inference is preferred; see, e.g., Blei et al. (2017). However, variational inference can not produce the exact posterior but only an approximation of it. This motivates us to consider

$$\Lambda^*_{a,b}(\mathbf{X}^n) = \frac{\int_\Theta [p_n(\mathbf{X}^n|\theta)]^{a-b}\,\pi_b(\theta;\mathbf{X}^n)\,\mathrm{d}\theta}{\int_{\tilde{\Theta}_0}[p_n(\mathbf{X}^n|\nu,\xi_0)]^{a-b}\,\pi_{0,b}(\nu;\mathbf{X}^n)\,\mathrm{d}\nu}, \tag{5}$$

where $\pi_b(\theta;\mathbf{X}^n)$ and $\pi_{0,b}(\nu;\mathbf{X}^n)$ are data dependent probability densities of parameters which are the approximations of the power posterior densities $\pi_b(\theta|\mathbf{X}^n)$ and $\pi_{0,b}(\nu|\mathbf{X}^n)$. We call $\pi_b(\theta;\mathbf{X}^n)$ and $\pi_{0,b}(\nu;\mathbf{X}^n)$ weight functions.

From frequentist perspective, the numerator and the denominator of (5) are both the integral of the likelihood with respect to certain weight functions. Note that in goodness of fit test, there are two common types of tests: extreme value type (Kolmogorov-Smirnov test, e.g.) and integral type (Cramér-von Mises test, e.g.). In classical parametric hypothesis testing, however, the focus is on the LRT which is an extreme value type statistic while little attention has been paid to the integrated likelihood functions. In this view, (5) fills in this gap of parametric hypothesis testing. Hence we would like to call $\Lambda^*_{a,b}(\mathbf{X}^n)$ the integrated likelihood ratio test statistic. Note that the LRT can also be regarded as an integrated likelihood ratio test statistic since the maximum likelihood can be regarded as the integral of the likelihood function with respect to the point mass on the MLE. However, the point mass measure is highly nonsmooth. For many models where the LRT fails, the likelihood function still has good properties for most θ but the MLE is unfortunately trapped in a fairly small area of θ where the likelihood has bad behavior. Intuitively, since the weight functions in (5) are fairly smooth, the defeat of the likelihood function in a small area will not introduce much effect on the integrated likelihood.

Now we investigate the asymptotic distribution of $\Lambda^*_{a,b}(\mathbf{X}^n)$. Let $h = \sqrt{n}(\theta-\theta_0)$ be the local parameter and $\pi_t(h;\mathbf{X}^n) = n^{-1/2}\pi_t(\theta_0 + n^{-1/2}h;\mathbf{X}^n)$ be the density in terms of h. If $\pi_1(\theta;\mathbf{X}^n)$ is exactly the posterior density of θ, then Bernstein-von Mises theorem asserts that under certain conditions, $\|\pi_1(h;\mathbf{X}^n) - \phi(h;\Delta_{n,\theta_0}, I_{\theta_0}^{-1})\|$ converges to 0 in $P^n_{\theta_0}$ probability, where for two densities $q_1(h)$ and $q_2(h)$, $\|q_1(h)-q_2(h)\| = \int |q_1(h)-q_2(h)|\,\mathrm{d}h$ is their total variation distance and $\phi(h;\mu,\Sigma)$ is the density function of a normal random variable with mean μ and variance matrix Σ evaluated at h. We shall assume that the approximate densities inherit such property.

Assumption 5. *Suppose b is fixed. Suppose that $\pi_b(h;\mathbf{X}^n)$ satisfies*

$$\left\|\pi_b(h;\mathbf{X}^n) - \phi(h;\Delta_{n,\theta_0}, b^{-1}I(\theta_0)^{-1})\right\| \xrightarrow{P^n_{\theta_0}} 0. \tag{6}$$

Similarly, let $h^{(0)} = \sqrt{n}(\nu-\nu_0)$. Define $\pi_{0,b}(h^{(0)};\mathbf{X}^n) = n^{-1/2}\pi_{0,b}(\nu;\mathbf{X}^n)$. Assume that

$$\|\pi_{0,b}(h^{(0)};\mathbf{X}^n) - \phi(h^{(0)};\Delta^{(0)}_{n,\theta_0}, b^{-1}I_\nu(\theta_0)^{-1})\| \overset{P^n_{\theta_0}}{\to} 0. \tag{7}$$

Furthermore, assume that for every $\epsilon > 0$, there exist Lebesgue integrable functions $T(h)$ and $T_0(h^{(0)})$ such that

$$\liminf_{n\to\infty} P^n_{\theta_0}\left\{\sup_{h\in\mathbb{R}^p}\left(\pi_b(h;\mathbf{X}^n) - T(h)\right) \le 0\right\} \ge 1-\epsilon, \tag{8}$$

$$\liminf_{n\to\infty} P_{\theta_0}^n \left\{ \sup_{h^{(0)}\in\mathbb{R}^{p_0}} \left(\pi_{0,b}(h^{(0)};\mathbf{X}^n) - T_0(h^{(0)})\right) \leq 0 \right\} \geq 1-\epsilon. \tag{9}$$

Remark 2. The conditions (6) and (7) assume that the weight functions satisfies the conclusion of the Berstein-von Mises theorem. These conditions are natural for power posteriors and their approximations. However, there are other weight functions also satisfy these conditions. For example, these conditions can also be satisfied by Generalized Fiducial distribution; see, e.g., Hannig et al. (2016). Hence the choice of the weight functions in the integrated likelihood ratio test statistic are not restricted to Bayesian methods.

Remark 3. The conditions (8) and (9) assume that there is a function controlling the tail of weight functions. We need to control the tail of the weight functions since the behavior of the likelihood may be undesirable when θ is far away from θ_0. If the weight functions $\pi_b(h;\mathbf{X}^n)$ and $\pi_{0,b}(h^{(0)};\mathbf{X}^n)$ are normal densities, then it can be shown that the conditions (6) and (7) imply (8) and (9).

The following theorem gives the asymptotic distribution of $\Lambda^*_{a,b}(\mathbf{X}^n)$.

Theorem 3. *Suppose a and b are fixed and $0 \leq a-b \leq 1$. Suppose Assumptions 1 and 5 hold. Then for $\{\theta_n\}$ such that $\sqrt{n}(\theta_n-\theta_0)\to\eta$, we have*

$$\frac{2}{a-b}\log\Lambda^*_{a,b}(\mathbf{X}^n) + \frac{p-p_0}{a-b}\log\left(\frac{a}{b}\right) \overset{P^n_{\theta_n}}{\rightsquigarrow} \chi^2(p-p_0, \eta^\top\tilde{\mathbf{J}} I_{\xi|\nu}(\theta_0)\tilde{\mathbf{J}}^\top\eta).$$

Theorem 3 shows that $\Lambda^*_{a,b}(\mathbf{X}^n)$ has the same asymptotic distribution as $\Lambda_{a,b}(\mathbf{X}^n)$. Now we consider a simple variational method which is guaranteed to yield a weight function satisfying Assumption 5. For comprehensive considerations of the statistical properties of variational methods; see Wang and Blei (2018), Pati et al. (2018) and Yang et al. (2017).

Let $\mathcal{Q}$ be the family of all p dimensional normal distribution. Let $\pi_b(\theta|\mathbf{X}^n)$ be the power posterior of order b and $\pi_b(h|\mathbf{X}^n) = n^{-1/2}\pi_b(\theta_0+n^{-1/2}h|\mathbf{X}^n)$ be the corresponding power posterior of h. Suppose that $\pi_b(h|\mathbf{X}^n)$ satisfies

$$\|\pi_b(h|\mathbf{X}^n) - \phi(h;\Delta_{n,\theta_0}, b^{-1}I(\theta_0)^{-1})\| \xrightarrow{P^n_{\theta_0}} 0. \tag{10}$$

Let the weight function $\pi_b^\dagger(\theta;\mathbf{X}^n)$ be the normal approximation of $\pi_b(\theta|\mathbf{X}^n)$ obtained from Rényi divergence variational inference (Li and Turner, 2016), that is,

$$\pi_b^\dagger(\theta;\mathbf{X}^n) = \operatorname*{arg\,min}_{q(\theta)\in\mathcal{Q}} -\frac{1}{1-\alpha}\log\int q(\theta)^\alpha \pi_b(\theta|\mathbf{X}^n)^{1-\alpha}\,\mathrm{d}\theta,$$

where $0<\alpha<1$ is an arbitrary constant. Let $\pi_b^\dagger(h;\mathbf{X}^n) = n^{-1/2}\pi_b^\dagger(\theta_0+n^{-1/2}h;\mathbf{X}^n)$ be the weight function of h. It can be seen that

$$\pi_b^\dagger(h;\mathbf{X}^n) = \operatorname*{arg\,min}_{q(h)\in\mathcal{Q}} -\frac{1}{1-\alpha}\log\int q(h)^\alpha \pi_b(h;\mathbf{X}^n)^{1-\alpha}\,\mathrm{d}h.$$

Hence we have

$$
\begin{aligned}
& -\frac{1}{1-\alpha}\log\int \pi_b^{\dagger}(h;\mathbf{X}^n)^{\alpha}\pi_b(h;\mathbf{X}^n)^{1-\alpha}\,\mathrm{d}h \\
\leq & -\frac{1}{1-\alpha}\log\int \phi(h;\Delta_{n,\theta_0},b^{-1}I(\theta_0)^{-1})^{\alpha}\pi_b(h|\mathbf{X}^n)^{1-\alpha}\,\mathrm{d}h.
\end{aligned}
\tag{11}
$$

Since Rényi divergence and total variation distance are topologically equivalent, (10) implies that the right hand side of (11) tends to 0 in $P_{\theta_0}^n$-probability. Again by the topological equivalence of Rényi divergence and total variation distance, we have

$$
\|\pi_b^{\dagger}(h;\mathbf{X}^n)-\phi(h;\Delta_{n,\theta_0},b^{-1}I(\theta_0)^{-1})\| \xrightarrow{P_{\theta_0}^n} 0.
$$

Note that $\pi_b^{\dagger}(h;\mathbf{X}^n)$ and $\phi(h;\Delta_{n,\theta_0},b^{-1}I(\theta_0)^{-1})$ are both normal density functions. For normal distributions, the convergence in total variation implies the convergence of parameters. Hence the mean and covariance parameters of $\pi_b^{\dagger}(h;\mathbf{X}^n)$ are bounded in probability. Then a dominating function $T(h)$ exists and thus (8) holds.

4 Examples

4.1 Full-rank exponential family

Exponential family possesses many desirable properties and includes many regular models. In this section, we apply the generalized fractional Bayes factor to the testing problem in the full-rank exponential family. Suppose

$$
p(X|\theta)=\exp\left\{\theta^{\top}T(X)-A(\theta)\right\}=\exp\left\{\nu^{\top}T_1(X)+\xi^{\top}T_2(X)-A(\theta)\right\}.
$$

We would like to test

$$
H:\xi=\xi_0 \quad v.s. \quad K:\xi\neq\xi_0.
$$

We assume Θ and $\tilde{\Theta}_0$ are open subsets of $\mathbb{R}^p$ and $\mathbb{R}^{p_0}$, respectively, and θ_0 and ν_0 are inner points of Θ and $\tilde{\Theta}_0$, respectively. We assume $I(\theta_0)$ is positive-definite.

We consider the test statistic $\Lambda_{a,b}(\mathbf{X}^n)$ with fixed $a>b>0$. To apply (a) of Theorem 2, we need to verify Assumption 1 and the $\sqrt{n}$-consistency of power posterior.

Proposition 3. *Suppose data are from the exponential family described above. Suppose the prior density $\pi(\theta)$ is continuous at θ_0 with $\pi(\theta_0)>0$, and there exists a $c^*>0$ such that*

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

Then Assumption 1 holds and $\pi_t(\theta;\mathbf{X}^n)$ is $\sqrt{n}$-consistent for any fixed $t>0$.

From Proposition 3, $\Lambda_{a,b}(\mathbf{X}^n)$ has the asymptotic distribution as stated in (a) of Theorem 2. Hence a test can be constructed which has the same asymptotic local power as the LRT. We note that for some models in exponential family, the MLE does not always exist; see, e.g., Rinaldo et al. (2013). Hence, the LRT is not always well defined. In contrast, $\Lambda_{a,b}(\mathbf{X}^n)$ is always well defined provided the priors are proper.

4.2 Normal mixture model

In this section, we apply the generalized fractional Bayes factor to testing the component number of normal mixture model. Normal mixture model is a highly irregular model. Due to partial loss of identifiability, the LRT has undesirable behavior. For example, if the component variances are totally unknown, the likelihood is unbounded and thus the LRT is not defined (Le Cam, 1990). See Chen (2017) for a review of the testing problems for mixture models. Since the integral of the likelihood can smooth the irregular behavior of the likelihood, it can be expected that Bayes methods may have better behavior than LRT. For example, for unknown variances case, the generalized fractional Bayes factor is at least well defined if the priors are proper.

Suppose $X_1, \dots, X_n$ are iid distributed as a mixture of normal distributions

$$p(X|\omega, \xi, \sigma^2) = \frac{1-\omega}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}X^2\right\} + \frac{\omega}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(X-\xi)^2\right\},$$

where $0 \le \omega \le 1$, $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}^+$. We shall consider two hypothesis testing problems for this model.

First, we assume $\omega = 1/2$ is known and consider testing the hypotheses

$$H : \xi = 0, \sigma = 1 \quad \text{vs.} \quad K : \xi \neq 0 \text{ or } \sigma \neq 1. \tag{12}$$

That is, we would like to test if the data are from a standard normal population or from a mixture of two normal populations with equal proportion. For this testing problem, the likelihood function is unbounded under the alternative hypothesis. In fact, if we take $\xi = X_1$ and let $\sigma^2 \to 0$, then the likelihood tends to infinity. See Figure 1. Thus, the LRT can not be defined. Apply (a) of Theorem 2 and Proposition 2, we can obtain the following proposition.

Proposition 4. *For hypotheses testing problem* (12), *if* $0 < b < a < 1$ *are fixed,* $\pi(\theta)$ *is continuous at* θ_0 *with* $\pi(\theta_0) > 0$, $\pi_0(\nu)$ *is continuous at* ν_0 *with* $\pi_0(\nu_0) > 0$, $\sqrt{n}((\xi, \sigma^2) - (0,1))^\top \to (\eta_1, \eta_2)^\top$, *then*

$$\frac{2}{a-b} \log \Lambda_{a,b}(\mathbf{X}^n) + \frac{2}{a-b} \log\left(\frac{a}{b}\right) \overset{P_{\theta_n}^n}{\rightsquigarrow} \chi^2(2, \eta_1^2/4 + \eta_2^2/8).$$

This example shows that even when the LRT fails, the generalized fractional Bayes factor can still be valid and has the expected asymptotic distribution. Thus, the proposed methodology has a wider application scope than the LRT.

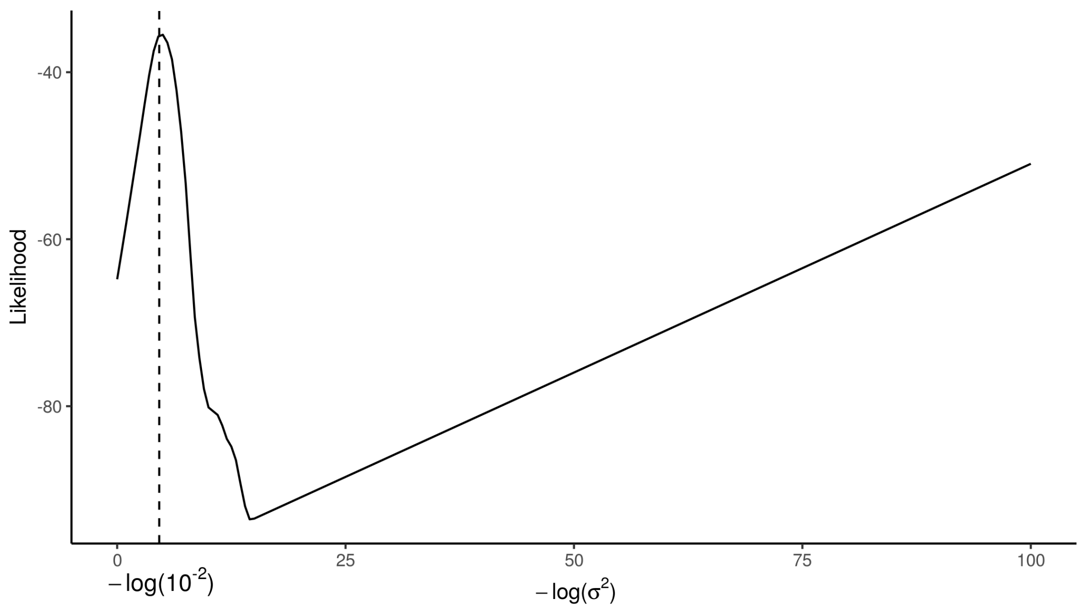

Figure 1: Data $X_1, \ldots, X_n$ are iid from the mixture model $(1-\omega)\mathcal{N}(0,1) + \omega\mathcal{N}(\xi, \sigma^2)$ with $(\omega, \xi, \sigma^2)^\top = (1/2, 1, 10^{-2})$ and $n = 50$. We plot the likelihood function in $-\log(\sigma^2)$ with $\omega = 1/2$ and $\xi = X_1$. The likelihood tends to infinity as $-\log(\sigma^2)$ tends to infinity, i.e., σ^2 tends to 0. In contrast, the likelihood has a local maximum around the true parameter $-\log(\sigma^2) = -\log(10^{-2})$.

In the above example, we assume $\omega = 1/2$ is known. If ω in unknown, then the mixture model suffers from loss of identifiability and the behavior of the likelihood is fairly complicated. For simplicity, we assume $\sigma^2 = 1$ is known and consider testing the hypotheses

$$H : \omega\xi = 0 \quad \text{vs.} \quad K : \omega\xi \neq 0. \tag{13}$$

It can be seen that this is equivalent to testing if the data are from a standard normal population or from a mixture of two normal populations. Although the LRT exists in this problem, its asymptotic behavior is complicated and its power behavior is not satisfactory. In fact, Hall and Stewart (2005) showed that in this problem, the LRT has trivial power under $n^{-1/2}$ local alternative hypothesis. For this irregular problem, Theorem 2 and Proposition 2 cannot be directly applied. This is because the second part is Assumption 4 is violated due to loss of identifiability. However, this does not mean that the proposed methodology is not applicable. In fact, the following proposition shows that $\Lambda_{a,b}(\mathbf{X}^n)$ has the desirable asymptotic properties.

Proposition 5. *Suppose* $\pi(\omega,\xi) = \pi_\omega(\omega)\pi_\xi(\xi)$, $\pi_\xi(\xi)$ *is positive and continuous at* $\xi = 0$, $\pi_\omega(\omega) \sim Beta(\alpha_1, \alpha_2)$ *with* $\alpha_1 > 1$. *Suppose* $0 < b < a < 1$. *Then,*

(i) under the null hypothesis,

$$\frac{2}{a-b}\log\Lambda_{a,b}(\mathbf{X}^n) + \frac{1}{a-b}\log\left(\frac{a}{b}\right) \overset{P^n_{\theta_0}}{\rightsquigarrow} \chi^2(1);$$

(ii) suppose for some $s < 1/4$, $\omega \geq n^{-s}$ *for large* n, $\sqrt{n}\omega\xi \to \eta$, *then*

$$\frac{2}{a-b}\log\Lambda_{a,b}(\mathbf{X}^n) + \frac{1}{a-b}\log\left(\frac{a}{b}\right) \overset{P^n_{\theta_n}}{\rightsquigarrow} \chi^2(1,\eta^2).$$

Proposition 5 shows that the test based on $\Lambda_{a,b}(\mathbf{X}^n)$ has nontrivial power if $\omega\xi$ is of order $n^{-1/2}$. In comparison, Hall and Stewart (2005) showed that the LRT has trivial power asymptotically if $\omega\xi = \gamma(n^{-1}\log\log n)^{1/2}$ with $|\gamma| < 1$.

5 Conclusion

In this paper, we investigated the Wilks phenomenon of the Bayes factor, the generalized fractional Bayes factor and more generally, the integrated likelihood ratio test. Using the Wilks phenomenon, these statistics can be used to construct the frequentist tests. We also apply the proposed methodology to three examples. These examples show that the proposed method can have good behavior even if the LRT is not defined or has poor properties. The integrated likelihood ratio test is easy to implement provided sampling from weight functions is convenient. If the weight functions are power posterior densities, then MCMC methods can be used to sample from weight functions. Furthermore, if MCMC is not efficient, one can use approximation methods, such as variational inference, and the resulting test procedure is still valid. Thus, the proposed methodology can be recommended when the classical LRT is not well defined or not easy to implement.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant No. 11471035.

Appendices

Appendix A Preliminary results

Lemma 1. *Suppose that Assumption 1 holds. Suppose* $\{\theta_n\}$ *satisfies* $\sqrt{n}(\theta_n - \theta_0) \to \eta$. *Then for any statistics* T_n, $T_n = o_{P^n_{\theta_0}}(1)$ *if and only if* $T_n = o_{P^n_{\theta_n}}(1)$.

Proof. Assumption 1 implies that $P^n_{\theta_0}$ and $P^n_{\theta_n}$ are mutually contiguous. Then the conclusion follows from Le Cam's first lemma (van der Vaart, 1998, Lemma 6.4). □

Lemma 2. *Suppose that Assumption 1 holds. Suppose* $\{\theta_n\}$ *satisfies* $\sqrt{n}(\theta_n - \theta_0) \to \eta$. *Then*

$$\Delta^\top_{n,\theta_0} I(\theta_0)\Delta_{n,\theta_0} - \Delta^{(0)\top}_{n,\theta_0} I_\nu(\theta_0)\Delta^{(0)}_{n,\theta_0} \overset{P^n_{\theta_n}}{\rightsquigarrow} \chi^2(p - p_0, \eta^\top \tilde{\mathbf{J}} I_{\xi|\nu}(\theta_0)\tilde{\mathbf{J}}^\top \eta).$$

Proof. It can be seen that $\Delta^{(0)}_{n,\theta_0} = (\mathbf{J}^\top I(\theta_0)\mathbf{J})^{-1}\mathbf{J}^\top I(\theta_0)\Delta_{n,\theta_0}$. Then

$$\begin{aligned}&\Delta^\top_{n,\theta_0} I(\theta_0)\Delta_{n,\theta_0} - \Delta^{(0)T}_{n,\theta_0} I_\nu(\theta_0)\Delta^{(0)}_{n,\theta_0}\\ =&\Delta^\top_{n,\theta_0} I(\theta_0)^{1/2}\big(\mathbf{I}_p - I(\theta_0)^{1/2}\mathbf{J}(\mathbf{J}^\top I(\theta_0)\mathbf{J})^{-1}\mathbf{J}^\top I(\theta_0)^{1/2}\big) I(\theta_0)^{1/2}\Delta_{n,\theta_0},\end{aligned}$$

where $\mathbf{I}_p - I(\theta_0)^{1/2}\mathbf{J}(\mathbf{J}^\top I(\theta_0)\mathbf{J})^{-1}\mathbf{J}^\top I(\theta_0)^{1/2}$ is a projection matrix with rank $p - p_0$. It remains to derive the asymptotic distribution of Δ_{n,θ_0}. Let $h_n = \sqrt{n}(\theta_n - \theta_0)$. From Assumption 1 and the central limit theorem, we have

$$\begin{pmatrix} \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^n \dot\ell_{\theta_0}(X_i) \\ \log\dfrac{p_n(\mathbf{X}^n|\theta_n)}{p_n(\mathbf{X}^n|\theta_0)} \end{pmatrix} \overset{P^n_0}{\rightsquigarrow} \mathcal{N}\left(\begin{pmatrix} 0 \\ -\frac{1}{2}\eta^\top I(\theta_0)\eta \end{pmatrix}, \begin{pmatrix} I(\theta_0) & I(\theta_0)\eta \\ \eta^\top I(\theta_0) & \eta^\top I(\theta_0)\eta \end{pmatrix}\right).$$

Then Le Cam's third lemma (van der Vaart, 1998, Example 6.7) implies that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \dot\ell_{\theta_0}(X_i) \overset{P^n_{\theta_n}}{\rightsquigarrow} \mathcal{N}(I(\theta_0)\eta, I(\theta_0)).$$

Consequently, Δ_{n,θ_0} weakly converges to $\mathcal{N}(\eta, I(\theta_0)^{-1})$ in $P^n_{\theta_n}$. It follows that

$$\Delta^\top_{n,\theta_0} I(\theta_0)\Delta_{n,\theta_0} - \Delta^{(0)T}_{n,\theta_0} I_\nu(\theta_0)\Delta^{(0)}_{n,\theta_0} \overset{P^n_{\eta_n}}{\rightsquigarrow} \chi^2(p - p_0, \delta),$$

where $\delta = \eta^\top \big(I(\theta_0) - I(\theta_0)\mathbf{J}(\mathbf{J}^\top I(\theta_0)\mathbf{J})^{-1}\mathbf{J}^\top I(\theta_0)\big)\eta$. Note that

$$\big(I(\theta_0) - I(\theta_0)\mathbf{J}(\mathbf{J}^\top I(\theta_0)\mathbf{J})^{-1}\mathbf{J}^\top I(\theta_0)\big)\mathbf{J}\mathbf{J}^\top = \mathbf{0}_{p\times p}.$$

Hence

$$\begin{aligned}
\delta =&\eta^\top \left(\mathbf{J}\mathbf{J}^\top + \tilde{\mathbf{J}}\tilde{\mathbf{J}}^\top\right)\big(I(\theta_0) - I(\theta_0)\mathbf{J}(\mathbf{J}^\top I(\theta_0)\mathbf{J})^{-1}\mathbf{J}^\top I(\theta_0)\big)\left(\mathbf{J}\mathbf{J}^\top + \tilde{\mathbf{J}}\tilde{\mathbf{J}}^\top\right)\eta \\
=&\eta^\top \tilde{\mathbf{J}}\tilde{\mathbf{J}}^\top \big(I(\theta_0) - I(\theta_0)\mathbf{J}(\mathbf{J}^\top I(\theta_0)\mathbf{J})^{-1}\mathbf{J}^\top I(\theta_0)\big)\tilde{\mathbf{J}}\tilde{\mathbf{J}}^\top\eta \\
=&\eta^\top \tilde{\mathbf{J}} I_{\xi|\nu}(\theta_0)\tilde{\mathbf{J}}^\top \eta.
\end{aligned}$$

This completes the proof. □

Lemma 3. *Suppose that Assumption 1 holds, $t \in (0, +\infty)$ is fixed, $\pi(\theta)$ is continuous at θ_0 with $\pi(\theta_0) > 0$. Then there is a sequence $M_n \to \infty$ such that*

$$\begin{aligned}
&\int_{\{\theta:\|\theta-\theta_0\|\le M_n/\sqrt{n}\}} \exp\{-tR_n(\theta_0\|\theta)\}\,\pi(\theta)\,\mathrm{d}\theta \\
=&(1+o_{P^n_{\theta_0}}(1))\pi(\theta_0)\left(\frac{2\pi}{tn}\right)^{p/2}|I(\theta_0)|^{-1/2}\exp\left\{\frac{t}{2}\Delta_{n,\theta_0}^\top I(\theta_0)\Delta_{n,\theta_0}\right\}.
\end{aligned}$$

Proof. For any fixed $M > 0$, we have

$$\begin{aligned}
&\int_{\{\theta:\|\theta-\theta_0\|\le M/\sqrt{n}\}} \exp\{-tR_n(\theta_0\|\theta)\}\,\pi(\theta)\,\mathrm{d}\theta \\
=&(1+o_{P^n_{\theta_0}}(1))n^{-p/2}\pi(\theta_0)\int_{\{h:\|h\|\le M\}}\exp\left\{-tR_n(\theta_0\|\theta_0+n^{-1/2}h)\right\}\,\mathrm{d}h \\
=&(1+o_{P^n_{\theta_0}}(1))n^{-p/2}\pi(\theta_0)\exp\left\{\frac{t}{2}\Delta_{n,\theta_0}^\top I(\theta_0)\Delta_{n,\theta_0}\right\} \\
&\cdot\int_{\{h:\|h\|\le M\}}\exp\left\{-\frac{t}{2}(h-\Delta_{n,\theta_0})^\top I(\theta_0)(h-\Delta_{n,\theta_0})\right\}\,\mathrm{d}h,
\end{aligned}$$

where the first equality follows from the continuity of $\pi(\theta)$ at θ_0 and the coordinate transformation $h = \sqrt{n}(\theta-\theta_0)$; and the second equality follows from the uniform expansion given by Assumption 1. This equality holds for every $M > 0$ and hence also for some $M_n \to \infty$. Since Δ_{n,θ_0} is bounded in probability, we have

$$\begin{aligned}
&\int_{\{h:\|h\|\le M_n\}}\exp\left\{-\frac{t}{2}(h-\Delta_{n,\theta_0})^\top I(\theta_0)(h-\Delta_{n,\theta_0})\right\}\,\mathrm{d}h \\
=&(1+o_{P^n_{\theta_0}}(1))\int_{\mathbb{R}^p}\exp\left\{-\frac{t}{2}(h-\Delta_{n,\theta_0})^\top I(\theta_0)(h-\Delta_{n,\theta_0})\right\}\,\mathrm{d}h \\
=&(1+o_{P^n_{\theta_0}}(1))\left(\frac{2\pi}{t}\right)^{p/2}|I(\theta_0)|^{-1/2}.
\end{aligned}$$

This completes the proof.

□

Lemma 4. *Suppose that Assumption 1 holds, $t \in (0, +\infty)$ is fixed, $\pi(\theta)$ is continuous at θ_0 with $\pi(\theta_0) > 0$, $\pi_t(\theta|\mathbf{X}^n)$ is $\sqrt{n}$-consistent. Then we have*

$$\int_\Theta \exp\left\{-tR_n(\theta_0\|\theta)\right\}\pi(\theta)\,\mathrm{d}\theta = (1 + o_{P^n_{\theta_0}}(1))\pi(\theta_0)\left(\frac{2\pi}{tn}\right)^{p/2}|I(\theta_0)|^{-1/2}\exp\left\{\frac{t}{2}\Delta_{n,\theta_0}^\top I(\theta_0)\Delta_{n,\theta_0}\right\}.$$

Proof. The $\sqrt{n}$-consistency of $\pi_t(\theta|\mathbf{X}^n)$ implies that for any $M_n \to \infty$,

$$\int_\Theta \exp\left\{-tR_n(\theta_0\|\theta)\right\}\pi(\theta)\,\mathrm{d}\theta =(1 + o_{P^n_{\theta_0}}(1))\int_{\left\{\theta:\|\theta-\theta_0\|\le M_n/\sqrt{n}\right\}}\exp\left\{-tR_n(\theta_0\|\theta)\right\}\pi(\theta)\,\mathrm{d}\theta.$$

Then the conclusion follows from Lemma 3. $\square$

Appendix B Proofs in Section 2

Proof of Theorem 1. From Lemma 4, we have

$$\int_\Theta \exp\left\{-R_n(\theta_0\|\theta)\right\}\pi(\theta)\,\mathrm{d}\theta = (1 + o_{P^n_{\theta_0}}(1))\pi(\theta_0)\left(\frac{2\pi}{n}\right)^{p/2}|I(\theta_0)|^{-1/2}\exp\left\{\frac{1}{2}\Delta_{n,\theta_0}^\top I(\theta_0)\Delta_{n,\theta_0}\right\},$$

and

$$\int_{\tilde\Theta} \exp\left\{-R_n\left(\theta_0\|\nu,\xi_0\right)\right\}\pi_0(\nu)\,\mathrm{d}\nu = (1+o_{P^n_{\theta_0}}(1))\pi_0(\nu_0)\left(\frac{2\pi}{n}\right)^{p_0/2}|I^{(0)}_{\theta_0}|^{-1/2}\exp\left\{\frac{1}{2}\Delta_{n,\theta_0}^{(0)\top} I_\nu(\theta_0)\Delta^{(0)}_{n,\theta_0}\right\}.$$

It follows that

$$\begin{aligned}
\log \mathrm{BF}_1(\mathbf{X}^n) =& \log\int_\Theta \exp\{-R_n(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta - \log\int_{\tilde\Theta}\exp\left\{-R_n\left(\theta_0\|\nu,\xi_0\right)\right\}\pi_0(\nu)\,\mathrm{d}\nu\\
=&\frac{p-p_0}{2}\log\left(\frac{2\pi}{n}\right) + \frac{1}{2}\left(\Delta_{n,\theta_0}^\top I(\theta_0)\Delta_{n,\theta_0} - \Delta_{n,\theta_0}^{(0)\top} I_\nu(\theta_0)\Delta^{(0)}_{n,\theta_0}\right)\\
&+\log\frac{|I(\theta_0)|^{-\frac{1}{2}}\pi(\theta_0)}{|I_\nu(\theta_0)|^{-\frac{1}{2}}\pi_0(\nu_0)} + o_{P^n_{\theta_0}}(1)\\
=&\frac{p-p_0}{2}\log\left(\frac{2\pi}{n}\right) + \frac{1}{2}\left(\Delta_{n,\theta_0}^\top I(\theta_0)\Delta_{n,\theta_0} - \Delta_{n,\theta_0}^{(0)\top} I_\nu(\theta_0)\Delta^{(0)}_{n,\theta_0}\right)\\
&+\log\frac{|I_{\xi|\nu}(\theta_0)|^{-\frac{1}{2}}\pi(\theta_0)}{\pi_0(\nu_0)} + o_{P^n_{\theta_0}}(1).
\end{aligned}$$

Combining the above equality and Lemma 1 and Lemma 2 leads to the conclusion. $\square$

Proposition 6. *Suppose that Assumptions 1, 2, 3 hold. Then the following assertions hold.*

(a) If $t \to 0$, $tn \to \infty$, then

$$\int_\Theta \exp\{-tR_n(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta = (1 + o_{P^n_{\theta_0}}(1))\pi(\theta_0)\left(\frac{2\pi}{tn}\right)^{p/2}|I(\theta_0)|^{-1/2}.$$

(b) If $tn \to c \in (c^*, +\infty)$, then

$$\int_\Theta \exp\{-tR_n(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta \xrightarrow{P^n_{\theta_0}} \int_\Theta \exp\{-cD_1(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta.$$

Proof. Assertion (a) follows from the following Lemma 6 and Lemma 5 with $t^\dagger = 0$. Assertion (b) follows directly from Lemma 6. □

Lemma 5. *Suppose that Assumptions 1, 2, 3 hold. Suppose as $n \to \infty$, $t \to 0$, $tn \to \infty$. Then for any fixed $t^\dagger \in [0, t^*)$,*

$$\int_\Theta \exp\left\{-tnD_{1-t^\dagger}(\theta_0\|\theta)\right\}\pi(\theta)\,\mathrm{d}\theta = (1+o(1))\pi(\theta_0)\left(\frac{2\pi}{(1-t^\dagger)tn}\right)^{p/2}|I(\theta_0)|^{-1/2}.$$

Proof. Assumption 2 implies that

$$\begin{aligned}
&\int_\Theta \exp\left\{-tnD_{1-t^\dagger}(\theta_0\|\theta)\right\}\pi(\theta)\,\mathrm{d}\theta \\
\geq &\int_{\left\{\theta:\|\theta-\theta_0\|\leq (tn)^{-1/4}\right\}} \exp\left\{-tnD_{1-t^\dagger}(\theta_0\|\theta)\right\}\pi(\theta)\,\mathrm{d}\theta \\
=&(1+o(1))\pi(\theta_0)\int_{\left\{\theta:\|\theta-\theta_0\|\leq (tn)^{-1/4}\right\}} \exp\left\{-\frac{tn(1-t^\dagger)}{2}(\theta-\theta_0)^\top I(\theta_0)(\theta-\theta_0)\right\}\mathrm{d}\theta \\
=&(1+o(1))\pi(\theta_0)(tn)^{-p/2}\int_{\left\{\vartheta:\|\vartheta\|\leq (tn)^{1/4}\right\}} \exp\left\{-\frac{1-t^\dagger}{2}\vartheta^\top I(\theta_0)\vartheta\right\}\mathrm{d}\vartheta \\
=&(1+o(1))\pi(\theta_0)\left(\frac{2\pi}{(1-t^\dagger)tn}\right)^{p/2}|I(\theta_0)|^{-1/2}.
\end{aligned}$$

Now we prove the other direction of the inequality. Assumption 2 allows us to choose $\epsilon \in (0,1)$ and $\delta > 0$ such that for $\|\theta - \theta_0\| \leq \delta$,

$$D_{1-t^\dagger}(\theta_0||\theta) \geq (1-\epsilon)\frac{1-t^\dagger}{2}(\theta-\theta_0)^\top I(\theta_0)(\theta-\theta_0), \quad \pi(\theta) \leq (1+\epsilon)\pi(\theta_0).$$

Also by Assumption 2, there exists a $\epsilon^* > 0$ such that $D_{t^\dagger}(\theta_0\|\theta) \geq \epsilon^*$ for $\|\theta-\theta_0\| \geq \delta$. Hence for sufficiently large n such that $tn > c^*$, we have

$$
\begin{aligned}
&\int_\Theta \exp\{-tnD_{1-t^\dagger}(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta \\
\leq &(1+\epsilon)\pi(\theta_0)\int_{\{\theta:\|\theta-\theta_0\|\leq\delta\}} \exp\left\{-tn(1-\epsilon)\frac{1-t^\dagger}{2}(\theta-\theta_0)^\top I(\theta_0)(\theta-\theta_0)\right\}\mathrm{d}\theta \\
&+\exp\{-tn\epsilon^*\}\int_{\{\theta:\|\theta-\theta_0\|>\delta\}} \exp\{-tn\left(D_{1-t^\dagger}(\theta_0\|\theta)-\epsilon^*\right)\}\pi(\theta)\,\mathrm{d}\theta \\
\leq &(1+\epsilon)\pi(\theta_0)\int_\Theta \exp\left\{-tn(1-\epsilon)\frac{1-t^\dagger}{2}(\theta-\theta_0)^\top I(\theta_0)(\theta-\theta_0)\right\}\mathrm{d}\theta \\
&+\exp\{-tn\epsilon^*\}\int_\Theta \exp\{-c^*\left(D_{1-t^*}(\theta_0\|\theta)-\epsilon^*\right)\}\pi(\theta)\,\mathrm{d}\theta \\
= &(1+\epsilon)(1-\epsilon)^{-p/2}\pi(\theta_0)\left(\frac{2\pi}{(1-t^\dagger)tn}\right)^{p/2}|I(\theta_0)|^{-1/2} \\
&+\exp\{-(tn-c^*)\epsilon^*\}\int_\Theta \exp\{-c^*D_{1-t^*}(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta \\
= &(1+o(1))(1+\epsilon)(1-\epsilon)^{-p/2}\pi(\theta_0)\left(\frac{2\pi}{(1-t^\dagger)tn}\right)^{p/2}|I(\theta_0)|^{-1/2}.
\end{aligned}
$$

Note that ϵ can be arbitrarily small. This completes the proof.

□

Lemma 6. *Suppose that Assumptions 1, 2, 3 hold. Suppose as $n\to\infty$, $t\to 0$, $tn\to c\in(c^*,+\infty]$. Then*

$$
\int_\Theta \exp\{-tR_n(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta = (1+o_{P^n_{\theta_0}}(1))\int_\Theta \exp\{-tnD_1(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta.
$$

Proof. Without loss of generality, suppose $tn > c^*$. Define

$$
w_n(\theta) = \frac{\exp\{-tnD_1(\theta_0\|\theta)\}\pi(\theta)}{\int_\Theta \exp\{-tnD_1(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta}.
$$

Note that

$$
\int_\Theta \exp\{-tnD_1(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta \leq \int_\Theta \exp\{-c^*D_{1-t^*}(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta < \infty.
$$

Hence $w_n(\theta)$ is well defined. It is easy to verify the following equality which will play an important role in our proof.

$$
\begin{aligned}
D_1\left(w_n(\theta)\,\mathrm{d}\theta\middle\|\pi_t(\theta|\mathbf{X}^n)\,\mathrm{d}\theta\right) = &\int_\Theta \left[tR_n(\theta_0\|\theta) - tnD_1(\theta_0\|\theta)\right]w_n(\theta)\,\mathrm{d}\theta \\
&+\log\frac{\int_\Theta \exp\{-tR_n(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta}{\int_\Theta \exp\{-tnD_1(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta}.
\end{aligned}
\tag{14}
$$

In view of (14), we only need to prove

$$P_{\theta_0}^n D_1\left(w_n(\theta)\,\mathrm{d}\theta \| \pi_t(\theta|\mathbf{X}^n)\,\mathrm{d}\theta\right) \to 0 \tag{15}$$

and

$$P_{\theta_0}^n \left(\int_\Theta \left[tR_n(\theta_0\|\theta) - tnD_1(\theta_0\|\theta) \right] w_n(\theta)\,\mathrm{d}\theta \right)^2 \to 0. \tag{16}$$

Proof of (15)*:* From Fubini's theorem and the fact $P_{\theta_0}^n R_n(\theta_0\|\theta) = nD_1(\theta_0\|\theta)$, we have

$$P_{\theta_0}^n \int_\Theta \left[tR_n(\theta_0\|\theta) - tnD_1(\theta_0\|\theta) \right] w_n(\theta)\,\mathrm{d}\theta = 0.$$

Jensen's inequality implies that

$$\begin{aligned} P_{\theta_0}^n \log \int_\Theta \exp\{-tR_n(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta &\le \log P_{\theta_0}^n \int_\Theta \exp\{-tR_n(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta \\ &= \log \int_\Theta \exp\{-tnD_{1-t}(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta. \end{aligned}$$

Then from (14), we have the upper bound

$$P_{\theta_0}^n D_1\left(w_n(\theta)\,\mathrm{d}\theta \middle\| \pi_t(\theta|\mathbf{X}^n)\,\mathrm{d}\theta\right) \le \log \frac{\int_\Theta \exp\{-tnD_{1-t}(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta}{\int_\Theta \exp\{-tnD_1(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta}. \tag{17}$$

If $tn \to c \in (0, +\infty)$, then the dominated convergence theorem implies that

$$\begin{aligned} \lim_{n\to\infty} \int_\Theta \exp\{-tnD_{1-t}(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta &= \int_\Theta \exp\{-cD_1(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta, \\ \lim_{n\to\infty} \int_\Theta \exp\{-tnD_1(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta &= \int_\Theta \exp\{-cD_1(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta. \end{aligned}$$

Hence the right hand side of (17) converges to 0.

We turn to the case $tn \to \infty$. For any $t^\dagger \in (0, t^*)$, since $t < t^\dagger$ for sufficiently large n, we have

$$\begin{aligned} \limsup_{n\to\infty} \log \frac{\int_\Theta \exp\{-tnD_{1-t}(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta}{\int_\Theta \exp\{-tnD_1(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta} &\le \limsup_{n\to\infty} \log \frac{\int_\Theta \exp\{-tnD_{1-t^\dagger}(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta}{\int_\Theta \exp\{-tnD_1(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta} \\ &= -\frac{p}{2}\log(1-t^\dagger), \end{aligned}$$

where the last equality follows from Lemma 5. Let $t^\dagger \to 0$, then the right hand side of (17) converges to 0. This completes the proof of (15).

Proof of (16)*:* It can be seen that

$$\begin{aligned} &P_{\theta_0}^n \left(\int_\Theta \left[tR_n(\theta_0\|\theta) - tnD_1(\theta_0\|\theta) \right] w_n(\theta)\,\mathrm{d}\theta \right)^2 \\ \le &\int_\Theta P_{\theta_0}^n \left[tR_n(\theta_0\|\theta) - tnD_1(\theta_0\|\theta) \right]^2 w_n(\theta)\,\mathrm{d}\theta \\ = &t^2 n \frac{\int_\Theta V(\theta_0\|\theta) \exp\{-tnD_1\left(\theta_0\|\theta\right)\}\pi(\theta)\,\mathrm{d}\theta}{\int_\Theta \exp\{-tnD_1\left(\theta_0\|\theta\right)\}\pi(\theta)\,\mathrm{d}\theta}. \end{aligned}$$

If $tn \to c \in (0, +\infty)$, the above expression obviously tends to 0. Now we assume $tn \to \infty$. Assumption 2 allows us to choose $\epsilon \in (0,1)$ and $\delta > 0$ such that for $\|\theta - \theta_0\| \le \delta$,

$$D_1(\theta_0\|\theta) \ge \frac{1-\epsilon}{2}(\theta-\theta_0)^\top I(\theta_0)(\theta-\theta_0), \quad V(\theta_0\|\theta) \le C\|\theta-\theta_0\|^2, \quad \pi(\theta) \le (1+\epsilon)\pi(\theta_0).$$

Also by Assumption 2, there exists a $\epsilon^* > 0$ such that $D_1(\theta_0\|\theta) \ge \epsilon^*$ for $\|\theta - \theta_0\| \ge \delta$. Then for sufficiently large n such that $tn > c^\dagger$, we have

$$\begin{aligned}
&\int_\Theta V(\theta_0\|\theta) \exp\{-tnD_1(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta \\
\le&(1+\epsilon)\pi(\theta_0)\int_{\{\theta:\|\theta-\theta_0\|\le\delta\}} C\|\theta-\theta_0\|^2 \exp\left\{-tn(1-\epsilon)\frac{1}{2}(\theta-\theta_0)^\top I(\theta_0)(\theta-\theta_0)\right\}\mathrm{d}\theta \\
&+ \exp\{-tn\epsilon^*\}\int_{\{\theta:\|\theta-\theta_0\|>\delta\}} V(\theta_0\|\theta)\exp\{-tn(D_1(\theta_0\|\theta)-\epsilon^*)\}\pi(\theta)\,\mathrm{d}\theta. \\
\le&(1+\epsilon)\pi(\theta_0)(tn)^{-p/2-1}\int_{\mathbb{R}^p} C\|\vartheta\|^2\exp\left\{-(1-\epsilon)\frac{1}{2}\vartheta^\top I(\theta_0)\vartheta\right\}\mathrm{d}\vartheta \\
&+ \exp\left\{-(tn-c^\dagger)\epsilon^*\right\}\int_\Theta V(\theta_0\|\theta)\exp\left\{-c^\dagger D_1(\theta_0\|\theta)\right\}\pi(\theta)\,\mathrm{d}\theta \\
=&O\left((tn)^{-p/2-1}\right).
\end{aligned}$$

The last inequality, combined with Lemma 5, leads to

$$P_{\theta_0}^n\left(\int_\Theta [tR_n(\theta_0\|\theta) - tnD_1(\theta_0\|\theta)]\, w_n(\theta)\,\mathrm{d}\theta\right)^2 = t^2 n \frac{O\left((tn)^{-p/2-1}\right)}{\pi(\theta_0)\,(2\pi)^{p/2}\,(tn)^{-p/2}\,|I(\theta_0)|^{-1/2}.} \to 0.$$

Hence (16) holds. This completes the proof.

□

Proof of Theorem 2.

$$\begin{aligned}
\log\Lambda_{a,b}(\mathbf{X}^n) =& \log\int_\Theta \exp\{-aR_n(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta - \log\int_\Theta\exp\{-bR_n(\theta_0\|\theta)\}\pi(\theta)\,\mathrm{d}\theta \\
&- \log\int_{\tilde\Theta_0}\exp\{-aR_n(\theta_0\|\nu,\xi_0)\}\pi_0(\nu)\,\mathrm{d}\nu + \log\int_{\tilde\Theta_0}\exp\{-bR_n(\theta_0\|\nu,\xi_0)\}\pi_0(\nu)\,\mathrm{d}\nu.
\end{aligned}$$

If a and b are fixed, we apply Lemma 4 to these four terms respectively. Then

$$\log\Lambda_{a,b} = -\frac{p-p_0}{2}\log\left(\frac{a}{b}\right) + \frac{a-b}{2}\left(\Delta_{n,\theta_0}^\top I(\theta_0)\Delta_{n,\theta_0} - \Delta_{n,\theta_0}^{(0)T} I_\nu(\theta_0)\Delta_{n,\theta_0}^{(0)}\right) + o_{P_{\theta_0}^n}(1).$$

Then the assertion (a) follows from Lemma 1 and Lemma 2.

Similarly, assertion (b) and (c) follows from Proposition 6, Lemma 1 and Lemma 2.

□

Proof of Proposition 1. Note that $L_1(\Theta;\mathbf{X}^n)$ is well defined $P_{\theta_0}^n$-a.s. since it has finite integral

$$\int_{\mathcal{X}^n} L_1(\Theta;\mathbf{X}^n)\,\mathrm{d}\mu^n = \int_\Theta\left(\int_{\mathcal{X}^n} p_n(\mathbf{X}^n|\theta)\,\mathrm{d}\mu^n\right)\pi(\theta)\,\mathrm{d}\theta = 1.$$

For $0 < t < 1$, by Hölder's inequality, we have $L_t(\Theta; \mathbf{X}^n) \le L_1^{1/t}(\Theta; \mathbf{X}^n)$. This proves the first part of the proposition.

We turn to the second part of the proposition. The following observation is critical in our proof. For $\epsilon \in (0, +\infty]$, $\gamma \in (0, +\infty)$,

$$\begin{aligned}\int_{\{x \in \mathbb{R}^p : \|x\| < \epsilon\}} \|x\|^{\gamma - p} \exp\{-\|x\|\} \,\mathrm{d}x =& \frac{\pi^{p/2}}{\Gamma(p/2)} \int_0^{\epsilon^2} y^{\gamma/2 - 1} \exp\{-\sqrt{y}\} \,\mathrm{d}y \\ =& \frac{2\pi^{p/2}}{\Gamma(p/2)} \int_0^{\epsilon} y^{\gamma - 1} \exp\{-y\} \,\mathrm{d}y,\end{aligned} \tag{18}$$

where the first equality follows from Fang et al. (1990), Lemma 1.4. If $\gamma \le 0$, the above integral is infinity.

In view of (18), we can define a family of density functions indexed by $\theta \in \mathbb{R}^p$ as

$$p(X|\theta) = \frac{\Gamma(p/2)}{2\pi^{p/2}\Gamma(\gamma)} \|X - \theta\|^{\gamma - p} \exp\{-\|X - \theta\|\},$$

where $\gamma > 0$ is a known hyperparameter. Suppose $\Theta = \mathbb{R}^p$, $X_1, \ldots, X_n \in \mathbb{R}^p$ are iid random vectors with density function $p(X|\theta_0)$. Let $\pi(\theta)$ be any proper prior density which is continuous and positive. Then

$$L_t(\Theta; \mathbf{X}^n) = \left(\frac{\Gamma(p/2)}{2\pi^{p/2}\Gamma(\gamma)}\right)^{tn} \int_{-\infty}^{+\infty} \left(\prod_{i=1}^n \|X_i - \theta\|\right)^{t(\gamma - p)} \exp\left\{-t\sum_{i=1}^n \|X_i - \theta\|\right\} \pi(\theta) \,\mathrm{d}\theta.$$

Note that with probability 1, there is no tie among $X_1, \ldots, X_n$. Consequently, there exists a sufficiently small $\epsilon > 0$ such that the sets $A_i := \{\theta : \|X_i - \theta\| < \epsilon\}$, $i = 1, \ldots, n$, are disjoint. It can be seen that for $i = 1, \ldots, n$, $p(X_i|\theta)$ is continuous and bounded for $\theta \notin A_i$. Hence

$$\int_{\mathbb{R}^p \setminus \cup_{i=1}^n A_i} [p_n(\mathbf{X}^n|\theta)]^t \pi(\theta) \,\mathrm{d}\theta \le \prod_{i=1}^n \sup_{\theta \notin A_i} [p(X_i|\theta)]^t < \infty.$$

On the other hand, for $i = 1, \ldots, n$,

$$\int_{A_i} [p_n(\mathbf{X}^n|\theta)]^t \pi(\theta) \,\mathrm{d}\theta \le \left(\prod_{j \ne i}^n \sup_{\theta \in A_i} [p(X_j|\theta)]^t\right) \left(\sup_{\theta \in A_i} \pi(\theta)\right) \int_{A_i} [p(X_i|\theta)]^t \,\mathrm{d}\theta,$$

$$\int_{A_i} [p_n(\mathbf{X}^n|\theta)]^t \pi(\theta) \,\mathrm{d}\theta \ge \left(\prod_{j \ne i}^n \inf_{\theta \in A_i} [p(X_j|\theta)]^t\right) \left(\inf_{\theta \in A_i} \pi(\theta)\right) \int_{A_i} [p(X_i|\theta)]^t \,\mathrm{d}\theta.$$

Hence $\int_{A_i} [p_n(\mathbf{X}^n|\theta)]^t \pi(\theta) \,\mathrm{d}\theta$ is finite if and only if $\int_{A_i} [p(X_i|\theta)]^t \,\mathrm{d}\theta$ is finite. Note that

$$\int_{A_i} [p(X_i|\theta)]^t \,\mathrm{d}\theta = \left(\frac{\Gamma(p/2)}{2\pi^{p/2}\Gamma(\gamma)}\right)^t \int_{\{\|\theta\| < \epsilon\}} \|\theta\|^{t(\gamma - p) + p - p} \exp\{-t\|\theta\|\} \,\mathrm{d}\theta.$$

It follows that $\int_{A_i} [p(X_i|\theta)]^t \,\mathrm{d}\theta$ is finite if and only if $t(\gamma - p) + p > 0$, or equivalently, $\gamma > p(t-1)/t$. Thus, $L_t(\theta; \mathbf{X}^n)$ is finite if and only if $\gamma > p(t-1)/t$. If $t > 1$, $p(t-1)/t > 0$, then $L_t(\theta; \mathbf{X}^n) = +\infty$ provided $\gamma \in (0, p(t-1)/t]$. This completes the proof. $\square$

Proof of Proposition 2. Without loss of generality, we assume $M_n/\sqrt{n} \to 0$. Note that

$$\int_{\{\theta:\|\theta-\theta_0\|\geq M_n/\sqrt{n}\}} \pi(\theta|\mathbf{X}^n)\,\mathrm{d}\theta = \frac{\int_{\{\theta:\|\theta-\theta_0\|\geq M_n/\sqrt{n}\}} \exp\{-tR_n(\theta_0\|\theta)\}\,\pi(\theta)\,\mathrm{d}\theta}{\int_\Theta \exp\{-tR_n(\theta_0\|\theta)\}\,\pi(\theta)\,\mathrm{d}\theta}.$$

From Lemma 3,

$$\begin{aligned}
&\int_\Theta \exp\{-tR_n(\theta_0\|\theta)\}\,\pi(\theta)\,\mathrm{d}\theta \\
\geq &\int_{\{\theta:\|\theta-\theta_0\|\leq M_n/\sqrt{n}\}} \exp\{-tR_n(\theta_0\|\theta)\}\,\pi(\theta)\,\mathrm{d}\theta \\
=&(1+o_{P^n_{\theta_0}}(1))\pi(\theta_0)\left(\frac{2\pi}{tn}\right)^{p/2}|I(\theta_0)|^{-1/2}\exp\left\{\frac{t}{2}\Delta^\top_{n,\theta_0}I(\theta_0)\Delta_{n,\theta_0}\right\}.
\end{aligned}\tag{19}$$

On the other hand, it follows from Fubini's theorem that

$$\begin{aligned}
&P^n_{\theta_0}\int_{\{\theta:\|\theta-\theta_0\|\geq M_n/\sqrt{n}\}} \exp\{-tR_n(\theta_0\|\theta)\}\,\pi(\theta)\,\mathrm{d}\theta \\
=&\int_{\{\theta:\|\theta-\theta_0\|\geq M_n/\sqrt{n}\}} \exp\{-tnD_{1-t}(\theta_0||\theta)\}\,\pi(\theta)\,\mathrm{d}\theta.
\end{aligned}$$

Assumption 4 implies that there exist $\delta > 0$, $C > 0$ and $\epsilon > 0$ such that $D_{1-t}(\theta_0\|\theta) \geq C\|\theta-\theta_0\|^2$ for $\|\theta-\theta_0\| \leq \delta$ and $D_{1-t}(\theta_0\|\theta) \geq \epsilon$ for $\|\theta-\theta_0\| > \delta$. Decompose the integral region into two parts $\{\theta : M_n/\sqrt{n} \leq \|\theta-\theta_0\| \leq \delta\}$ and $\{\theta : \|\theta-\theta_0\| > \delta\}$. Then for sufficiently large n,

$$\begin{aligned}
&\int_{\{\theta:\|\theta-\theta_0\|\geq M_n/\sqrt{n}\}} \exp\{-tnD_{1-t}(\theta_0||\theta)\}\,\pi(\theta)\,\mathrm{d}\theta \\
\leq &\max_{\|\theta-\theta_0\|\leq\delta}\pi(\theta)\int_{\left\{\theta:M_n/\sqrt{n}\leq\|\theta-\theta_0\|<\delta\right\}} \exp\left\{-tCn\|\theta-\theta_0\|^2\right\}\,\mathrm{d}\theta \\
&+\exp\{-t\epsilon n\}\int_{\{\theta:\|\theta-\theta_0\|\geq\delta\}} \exp\{-tn(D_{1-t}(\theta_0||\theta)-\epsilon)\}\,\pi(\theta)\,\mathrm{d}\theta \\
\leq&\big(\max_{\|\theta-\theta_0\|\leq\delta}\pi(\theta)\big)n^{-p/2}\int_{\{h:\|h\|\geq M_n\}} \exp\left\{-tC\|h\|^2\right\}\,\mathrm{d}h \\
&+\exp\{-t\epsilon n\}\int_{\{\theta:\|\theta-\theta_0\|\geq\delta\}} \exp\{-c^*(D_{1-t}(\theta_0||\theta)-\epsilon)\}\,\pi(\theta)\,\mathrm{d}\theta.
\end{aligned}$$

Hence

$$\int_{\{\theta:\|\theta-\theta_0\|\geq M_n/\sqrt{n}\}} \exp\{-tR_n(\theta_0\|\theta)\}\,\pi(\theta)\,\mathrm{d}\theta = o_{P^n_{\theta_0}}\left(n^{-p/2}\right).\tag{20}$$

Then the $\sqrt{n}$-consistency of $L_t(\cdot;\mathbf{X}^n)$ follows from (19) and (20).

□

Proof of Theorem 3. It can be seen that

$$\Lambda^*_{a,b}(\mathbf{X}^n) = \frac{\int \exp\left\{-(a-b)R_n\left(\theta_0\|\theta_0+n^{-1/2}h\right)\right\}\pi_b(h;\mathbf{X}^n)\,\mathrm{d}h}{\int \exp\left\{-(a-b)R_n\left(\theta_0\|\nu_0+n^{-1/2}h^{(0)},\xi_0\right)\right\}\pi_b(h^{(0)};\mathbf{X}^n)\,\mathrm{d}h^{(0)}}.$$

Let $M > 0$ be any fixed number. We have

$$
\begin{aligned}
&\int_{\|h\|\leq M} \exp\left\{-(a-b)R_n(\theta_0\|\theta_0+n^{-1/2}h)\right\}\pi_b(h;\mathbf{X}^n)\,\mathrm{d}h \\
=&(1+o_{P^n_{\theta_0}}(1))\int_{\|h\|\leq M} \exp\left\{(a-b)h^\top I(\theta_0)\Delta_{n,\theta_0}-\frac{a-b}{2}h^\top I(\theta_0)h\right\}\pi_b(h;\mathbf{X}^n)\,\mathrm{d}h \qquad (21)\\
=&\int_{\|h\|\leq M} \exp\left\{(a-b)h^\top I(\theta_0)\Delta_{n,\theta_0}-\frac{a-b}{2}h^\top I(\theta_0)h\right\}\phi(h;\Delta_{n,\theta_0},b^{-1}I(\theta_0)^{-1})\,\mathrm{d}h+o_{P^n_{\theta_0}}(1),
\end{aligned}
$$

where the first equality follows from Assumption 1 and the second equality follows from (6). This is true for every $M > 0$ and hence also for some $M_n \to \infty$.

Now we prove that for any $M_n \to +\infty$,

$$
\int_{\|h\|>M_n} \exp\left\{-(a-b)R_n(\theta_0\|\theta_0+n^{-1/2}h)\right\}\pi_b(h;\mathbf{X}^n)\,\mathrm{d}h \xrightarrow{P^n_{\theta_0}} 0. \qquad (22)
$$

By Assumption 5, for any $\epsilon > 0$, with probability at least $1-\epsilon$,

$$
\begin{aligned}
&\int_{\|h\|>M_n} \exp\left\{-(a-b)R_n(\theta_0\|\theta_0+n^{-1/2}h)\right\}\pi_b(h;\mathbf{X}^n)\,\mathrm{d}h \\
\leq&\int_{\|h\|>M_n} \exp\left\{-(a-b)R_n(\theta_0\|\theta_0+n^{-1/2}h)\right\}T(h)\,\mathrm{d}h.
\end{aligned} \qquad (23)
$$

Since $a-b\leq 1$, Hölder's inequality implies that

$$
P^n_{\theta_0}\exp\left\{-(a-b)R_n(\theta_0\|\theta_0+n^{-1/2}h)\right\}\leq\left(P^n_{\theta_0}\exp\left\{-R_n(\theta_0\|\theta_0+n^{-1/2}h)\right\}\right)^{a-b}=1.
$$

Hence the expectation of the right hand side of (23) satisfies

$$
P^n_{\theta_0}\int_{\|h\|>M_n} \exp\left\{-(a-b)R_n(\theta_0\|\theta_0+n^{-1/2}h)\right\}T(h)\,\mathrm{d}h\leq\int_{\|h\|>M_n}T(h)\,\mathrm{d}h\to 0.
$$

This verifies (22).

Combining (21) and (22) yields

$$
\begin{aligned}
&\int \exp\left\{-(a-b)R_n\left(\theta_0\|\theta_0+n^{-1/2}h\right)\right\}\pi_b(h;\mathbf{X}^n)\,\mathrm{d}h \\
=&\left(\frac{a}{b}\right)^{-p/2}\exp\left\{\frac{a-b}{2}\Delta^\top_{n,\theta_0}I(\theta_0)\Delta_{n,\theta_0}\right\}+o_{P^n_{\theta_0}}(1).
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
&\int \exp\left\{-(a-b)R_n\left(\theta_0\|\nu_0+n^{-1/2}h^{(0)},\xi_0\right)\right\}\pi_b(h^{(0)};\mathbf{X}^n)\,\mathrm{d}h^{(0)} \\
=&\left(\frac{a}{b}\right)^{-p_0/2}\exp\left\{\frac{a-b}{2}\Delta^{(0)\top}_{n,\theta_0}I_\nu(\theta_0)\Delta^{(0)}_{n,\theta_0}\right\}+o_{P^n_{\theta_0}}(1).
\end{aligned}
$$

Hence

$$
2\log\Lambda^*_{a,b}(\mathbf{X}^n)=-(p-p_0)\log\left(\frac{a}{b}\right)+(a-b)\left(\Delta^\top_{n,\theta_0}I(\theta_0)\Delta_{n,\theta_0}-\Delta^{(0)\top}_{n,\theta_0}I_\nu(\theta_0)\Delta^{(0)}_{n,\theta_0}\right)+o_{P^n_{\theta_0}}(1).
$$

Then the conclusion follows from Lemma 1 and Lemma 2. $\square$

Appendix C Proofs in Section 3

Proof of Proposition 3. For exponential family, we have $\dot{\ell}_{\theta_0}(X) = T(X) - (\partial/\partial\theta)A(\theta_0)$ and $I(\theta_0) = (\partial^2/\partial\theta\partial\theta^\top)A(\theta_0)$. Thus,

$$I(\theta_0)\Delta_{n,\theta_0} = n^{-1/2}\sum_{i=1}^{n} T(X_i) - \sqrt{n}\frac{\partial}{\partial\theta}A(\theta_0)$$

and

$$R_n(\theta_0\|\theta_0 + n^{-1/2}h) = -h^\top I(\theta_0)\Delta_{n,\theta_0} + \frac{1}{2}h^\top I(\theta_0)h + g_n(h),$$

where

$$g_n(h) = n\Big(A(\theta_0 + n^{-1/2}h) - A(\theta_0) - n^{-1/2}h\frac{\partial}{\partial\theta}A(\theta_0) - \frac{1}{2n}h^\top I(\theta_0)h\Big).$$

From Taylor's theorem and the continuity of the third derivative of $A(\theta)$, for any fixed $M > 0$,

$$\max_{\{h:\|h\|\le M\}} |g_n(h)| = O\left(\frac{1}{\sqrt{n}}\right) \to 0.$$

Hence Assumption 1 holds.

We turn to the $\sqrt{n}$-consistency of $\pi_t(\theta;\mathbf{X}^n)$. From Lemma 3, there exists a sequence $M_n \to \infty$ such that

$$\begin{aligned}
&\int_\Theta \exp\{-tR_n(\theta_0\|\theta)\}\,\pi(\theta)\,\mathrm{d}\theta \\
\ge&\int_{\{\theta:\|\theta-\theta_0\|\le M_n/\sqrt{n}\}} \exp\{-tR_n(\theta_0\|\theta)\}\,\pi(\theta)\,\mathrm{d}\theta \\
=&(1+o_{P^n_{\theta_0}}(1))\pi(\theta_0)\left(\frac{2\pi}{tn}\right)^{p/2}|I(\theta_0)|^{-1/2}\exp\left\{\frac{t}{2}\Delta_{n,\theta_0}^\top I(\theta_0)\Delta_{n,\theta_0}\right\}.
\end{aligned} \tag{24}$$

Next we lower bound $R_n(\theta_0\|\theta)$ for $\|\theta-\theta_0\| \ge M_n/\sqrt{n}$. We have

$$\begin{aligned}
\min_{\{\theta:\|\theta-\theta_0\|=M_n/\sqrt{n}\}} R_n(\theta_0\|\theta) &= \min_{\{h:\|h\|=M_n\}} R_n(\theta_0\|\theta_0+n^{-1/2}h) \\
&\ge -\|I(\theta_0)\Delta_{n,\theta_0}\|M_n + \frac{\lambda_{\min}(I(\theta_0))}{2}M_n^2 - \max_{\{h:\|h\|=M_n\}}|g_n(h)|,
\end{aligned}$$

where $\lambda_{\min}(I(\theta_0)) > 0$ is the minimum eigenvalue of $I(\theta_0)$. Also note that $I(\theta_0)\Delta_{n,\theta_0}$ is bounded in probability. Hence with probability tending to 1,

$$\min_{\{\theta:\|\theta-\theta_0\|=M_n/\sqrt{n}\}} R_n(\theta_0\|\theta) \ge \frac{\lambda_{\min}(I(\theta_0))}{4}M_n^2.$$

Note that $R_n(\theta_0\|\theta)$ is convex in θ and $R_n(\theta_0\|\theta_0) = 0$. Then Jensen's inequality implies that for $\|\theta-\theta_0\| \ge M_n/\sqrt{n}$,

$$\frac{M_n/\sqrt{n}}{\|\theta-\theta_0\|}R_n(\theta_0\|\theta) \ge R_n\left(\theta_0\middle\|\theta_0 + \frac{M_n/\sqrt{n}}{\|\theta-\theta_0\|}(\theta-\theta_0)\right).$$

The last two inequalities imply that with probability tending to 1, for all θ such that $\|\theta-\theta_0\| \geq M_n/\sqrt{n}$,

$$
\begin{aligned}
R_n(\theta_0\|\theta) &\geq \frac{\sqrt{n}\|\theta-\theta_0\|}{M_n} R_n\left(\theta_0 \Big\| \theta_0 + \frac{M_n/\sqrt{n}}{\|\theta-\theta_0\|}(\theta-\theta_0)\right) \\
&\geq \frac{\sqrt{n}\|\theta-\theta_0\|}{M_n} \min_{\{\theta:\|\theta-\theta_0\|=M_n/\sqrt{n}\}} R_n(\theta_0\|\theta) \\
&\geq \frac{\lambda_{\min}(I(\theta_0))}{4}\sqrt{n}\|\theta-\theta_0\|M_n.
\end{aligned}
$$

Fix an $\epsilon > 0$ such that $\sup_{\|\theta-\theta_0\|<\epsilon}\pi(\theta) < +\infty$. For sufficiently large n, with probability tending to 1, we have

$$
\begin{aligned}
&\int_{\{\theta:\|\theta-\theta_0\|>M_n/\sqrt{n}\}} \exp\left\{-tR_n(\theta_0\|\theta)\right\}\pi(\theta)\,\mathrm{d}\theta \\
\leq &\int_{\{\theta:\|\theta-\theta_0\|>M_n/\sqrt{n}\}} \exp\left\{-\frac{t\lambda_{\min}(I(\theta_0))}{4}\sqrt{n}\|\theta-\theta_0\|M_n\right\}\pi(\theta)\,\mathrm{d}\theta \\
= &\int_{\{\theta:M_n/\sqrt{n}<\|\theta-\theta_0\|\leq\epsilon\}} \exp\left\{-\frac{t\lambda_{\min}(I(\theta_0))}{4}\sqrt{n}\|\theta-\theta_0\|M_n\right\}\pi(\theta)\,\mathrm{d}\theta \\
&+\exp\left\{-\frac{t\lambda_{\min}(I(\theta_0))}{4}\epsilon\sqrt{n}M_n\right\}\int_{\{\theta:\|\theta-\theta_0\|>\epsilon\}} \exp\left\{-\frac{t\lambda_{\min}(I(\theta_0))}{4}\sqrt{n}(\|\theta-\theta_0\|-\epsilon)M_n\right\}\pi(\theta)\,\mathrm{d}\theta \\
\leq &\Big(\sup_{\|\theta-\theta_0\|<\epsilon}\pi(\theta)\Big)\int_{\{\theta:\|\theta-\theta_0\|\geq M_n/\sqrt{n}\}} \exp\left\{-\frac{t\lambda_{\min}(I(\theta_0))}{4}\sqrt{n}\|\theta-\theta_0\|M_n\right\}\,\mathrm{d}\theta \\
&+\exp\left\{-\frac{t\lambda_{\min}(I(\theta_0))}{4}\epsilon\sqrt{n}M_n\right\}\int_{\{\theta:\|\theta-\theta_0\|>\epsilon\}} \exp\left\{-c^*(\|\theta-\theta_0\|-\epsilon)\right\}\pi(\theta)\,\mathrm{d}\theta \\
= &\Big(\sup_{\|\theta-\theta_0\|<\epsilon}\pi(\theta)\Big)(\sqrt{n}M_n)^{-p}\int_{\{h:\|h\|\geq M_n^2\}} \exp\left\{-\frac{t\lambda_{\min}(I(\theta_0))}{4}\|h\|\right\}\,\mathrm{d}h \\
&+\exp\left\{-\frac{t\lambda_{\min}(I(\theta_0))}{4}\epsilon\sqrt{n}M_n\right\}\int_{\{\theta:\|\theta-\theta_0\|>\epsilon\}} \exp\left\{-c^*(\|\theta-\theta_0\|-\epsilon)\right\}\pi(\theta)\,\mathrm{d}\theta.
\end{aligned}
$$

It follows that

$$
\int_{\{\theta:\|\theta-\theta_0\|>M_n/\sqrt{n}\}} \exp\left\{-tR_n(\theta_0\|\theta)\right\}\pi(\theta)\,\mathrm{d}\theta = O_{P_{\theta_0}^n}((\sqrt{n}M_n)^{-p}).
$$

Combining the last display and (24) leads to

$$
\frac{\int_{\{\theta:\|\theta-\theta_0\|>M_n/\sqrt{n}\}} \exp\left\{-tR_n(\theta_0\|\theta)\right\}\pi(\theta)\,\mathrm{d}\theta}{\int_\Theta \exp\left\{-tR_n(\theta_0\|\theta)\right\}\pi(\theta)\,\mathrm{d}\theta} = O_{P_{\theta_0}^n}\left(M_n^{-p}\right) \xrightarrow{P_{\theta_0}^n} 0.
$$

This completes the proof.

$\square$

Proof of Proposition 4. We shall verify Assumption 1 and Assumption 4. We use the parameterization $\theta = (\xi,\tau)^\top = (\xi,\sigma^{-2})^\top$. Then

$$
p(X|\theta) = \frac{1}{2}\phi(X) + \frac{1}{2}\sqrt{\tau}\phi(\sqrt{\tau}(X-\xi)).
$$

By direct calculation, we have

$$\dot{\ell}_{\theta_0}(X) = \left(\frac{1}{2}X, \frac{1}{4}(1-X^2)\right)^\top.$$

Hence $P_{\theta_0}\dot{\ell}_{\theta_0} = \mathbf{0}_2$ and $I(\theta_0) = \mathrm{diag}(1/4, 1/8)$.

Let $M > 0$ be a fixed constant. For $h = (h_1, h_2)^\top \in \mathbb{R}^2$ such that $\|h\| \le M$ and $i = 1, \ldots, n$, we have

$$\begin{aligned}\frac{p(X_i|\theta_0 + n^{-1/2}h)}{p(X_i|\theta_0)} =& \frac{1}{2} + \frac{1}{2}\sqrt{1+\frac{h_2}{\sqrt{n}}}\exp\left\{-\frac{h_2}{2\sqrt{n}}X_i^2\right.\\ &\left. + \left(1+\frac{h_2}{\sqrt{n}}\right)\frac{h_1}{\sqrt{n}}X_i - \frac{1}{2}\left(1+\frac{h_2}{\sqrt{n}}\right)\frac{h_1^2}{n}\right\}.\end{aligned}$$

It is well known that $\max_{1\le i\le n}|X_i| = O_{P_{\theta_0}^n}(\sqrt{\log n})$. Then from Taylor expansion $\exp(x) = 1 + x + x^2/2 + O(x^3)$, we have, uniformly for $\|h\| \le M$ and $i = 1, \ldots, n$, that

$$\begin{aligned}&\exp\left\{-\frac{h_2}{2\sqrt{n}}X_i^2 + \left(1+\frac{h_2}{\sqrt{n}}\right)\frac{h_1}{\sqrt{n}}X_i - \frac{1}{2}\left(1+\frac{h_2}{\sqrt{n}}\right)\frac{h_1^2}{n}\right\}\\ =&1 - \frac{h_1^2}{2n} + \left(\frac{h_1}{\sqrt{n}} + \frac{h_1h_2}{n}\right)X_i + \left(-\frac{h_2}{2\sqrt{n}} + \frac{h_1^2}{2n}\right)X_i^2\\ &- \frac{h_1h_2}{2n}X_i^3 + \frac{h_2^2}{8n}X_i^4 + O_{P_{\theta_0}^n}\left(\frac{\log^3 n}{n^{3/2}}\right).\end{aligned}$$

On the other hand, for sufficiently large n such that $M/\sqrt{n} \le 1/2$, we have, uniformly for $\|h\| \le M$, that

$$\sqrt{1+\frac{h_2}{\sqrt{n}}} = 1 + \frac{h_2}{2\sqrt{n}} - \frac{h_2^2}{8n} + O\left(\frac{1}{n^3}\right).$$

Multiplying the above two expansions yields

$$\begin{aligned}\frac{p(X_i|\theta_0 + n^{-1/2}h)}{p(X_i|\theta_0)} =&1 + \frac{h_1}{2\sqrt{n}}X_i + \left(\frac{h_2}{4\sqrt{n}} - \frac{h_1^2}{4n}\right)(1-X_i^2) + \frac{h_2^2}{16n}X_i^4\\ &- \frac{h_2^2}{8n}X_i^2 - \frac{h_2^2}{16n} + \frac{3h_1h_2}{4n}X_i - \frac{h_1h_2}{4n}X_i^3 + O_{P_{\theta_0}^n}\left(\frac{\log^3 n}{n^{3/2}}\right).\end{aligned}$$

From Taylor expansion $\log(1+x) = x - x^2/2 + O(x^3)$ for $x \in (-1, 1)$, we have, uniformly for $\|h\| \le M$ and $i = 1, \ldots, n$, that

$$\begin{aligned}&\log\frac{p(X_i|\theta_0 + n^{-1/2}h)}{p(X_i|\theta_0)}\\ =&\frac{h_1}{2\sqrt{n}}X_i + \left(\frac{h_2}{4\sqrt{n}} - \frac{h_1^2}{4n}\right)(1-X_i^2) + \frac{h_2^2}{16n}X_i^4 - \frac{h_2^2}{8n}X_i^2 - \frac{h_2^2}{16n} + \frac{3h_1h_2}{4n}X_i\\ &- \frac{h_1h_2}{4n}X_i^3 - \frac{h_1^2}{8n}X_i^2 - \frac{h_2^2}{32n}(1-X_i^2)^2 - \frac{h_1h_2}{8n}X_i(1-X_i^2) + O_{P_{\theta_0}^n}\left(\frac{\log^3 n}{n^{3/2}}\right).\end{aligned}$$

Thus,

$$\begin{aligned}
\log \frac{p_n(\mathbf{X}^n|\theta_0+n^{-1/2}h)}{p_n(\mathbf{X}^n|\theta_0)} =& \sum_{i=1}^n \log \frac{p(X_i|\theta_0+n^{-1/2}h)}{p(X_i|\theta_0)} \\
=& \frac{h_1}{2\sqrt{n}}\sum_{i=1}^n X_i + \frac{h_2}{4\sqrt{n}}\sum_{i=1}^n (1-X_i^2) - \frac{h_1^2}{8} - \frac{h_2^2}{16} + o_{P_{\theta_0}^n}(1),
\end{aligned}$$

where the $o_{P_{\theta_0}^n}(1)$ term is uniform for $\|h\|\leq M$. This verifies Assumption 1.

Now we verify Assumption 4. We have

$$\begin{aligned}
D_{1/2}(\theta_0||\theta) =& -2\log\int\sqrt{p(X|\theta)p(X|\theta_0)}\,\mathrm{d}\mu \\
\geq& 2\left(1-\int\sqrt{p(X|\theta)p(X|\theta_0)}\,\mathrm{d}\mu\right) \\
=& \int\left(\sqrt{p(X|\theta)}-\sqrt{p(X|\theta_0)}\right)^2\mathrm{d}\mu \\
\geq& \frac{1}{4}\left(\int|p(X|\theta)-p(X|\theta_0)|\,\mathrm{d}\mu\right)^2 \\
=& \frac{1}{16}\left(\int\left|\sqrt{\tau}\phi(\sqrt{\tau}(X-\xi))-\phi(X)\right|\,\mathrm{d}\mu\right)^2.
\end{aligned}$$

Note that

$$\begin{aligned}
\int\left|\sqrt{\tau}\phi(\sqrt{\tau}(X-\xi))-\phi(X)\right|\,\mathrm{d}\mu \geq& \left|\int \exp(iX)\sqrt{\tau}\phi(\sqrt{\tau}(X-\xi))-\exp(itX)\phi(X)\,\mathrm{d}\mu\right| \\
=& \left|\exp(i\xi-1/(2\tau))-\exp(-1/2)\right|.
\end{aligned}$$

The last display has two consequences. On the one hand,

$$\int\left|\sqrt{\tau}\phi(\sqrt{\tau}(X-\xi))-\phi(X)\right|d\mu \geq |\sin\xi|\exp(-1/(2\tau)).$$

Hence if $(\xi,\tau)^\top$ is close enough to $(0,1)^\top$, then

$$\int\left|\sqrt{\tau}\phi(\sqrt{\tau}(X-\xi))-\phi(X)\right|d\mu \gtrsim |\xi|.$$

On the other hand, it is not hard to see that

$$\left|\exp(i\xi-1/(2\tau))-\exp(-1/2)\right| \geq \left|\exp(-1/(2\tau))-\exp(-1/2)\right|.$$

Hence if $(\xi,\tau)^\top$ close enough to $(0,1)^\top$, then

$$\int\left|\sqrt{\tau}\phi(\sqrt{\tau}(X-\xi))-\phi(X)\right|d\mu \gtrsim |\tau-1|.$$

The above equalities imply that there exist $\delta>0$ and $C>0$ such that for $\sqrt{\xi^2+(\tau-1)^2}<\delta$,

$$D_{1/2}(\theta_0||\theta)\geq C(\xi^2+(\tau-1)^2).$$

We turn to the case $\sqrt{\xi^2+(\tau-1)^2}\geq\delta$. We have

$$
\begin{aligned}
D_{1/2}(\theta_0||\theta) &\geq \frac{1}{16}\left(\int \left|\sqrt{\tau}\phi(\sqrt{\tau}(X-\xi))-\phi(X)\right| \,\mathrm{d}\mu\right)^2 \\
&\geq \frac{1}{16}\left(\int \left(\sqrt{\sqrt{\tau}\phi(\sqrt{\tau}(X-\xi))}-\sqrt{\phi(X)}\right)^2 \mathrm{d}\mu\right)^2 \\
&= \frac{1}{4}\left(1-\sqrt{\frac{2\sqrt{\tau}}{1+\tau}}\exp\left\{-\frac{1}{4}\frac{\tau\xi^2}{1+\tau}\right\}\right)^2.
\end{aligned}
$$

Note that if $\sqrt{\xi^2+(\tau-1)^2}\geq\delta$, $(\tau-1)^2\geq\delta^2/2$ or else $(\tau-1)^2<\delta^2/2$ and $\xi^2\geq\delta^2/2$. If $(\tau-1)^2\geq\delta^2/2$, then

$$
1-\sqrt{\frac{2\sqrt{\tau}}{1+\tau}}\exp\left\{-\frac{1}{4}\frac{\tau\xi^2}{1+\tau}\right\}\geq 1-\sqrt{\frac{2\sqrt{\tau}}{1+\tau}},
$$

which obviously has a positive lower bound. On the other hand, suppose $(\tau-1)^2<\delta^2/2$ and $\xi^2\geq\delta^2/2$, then

$$
1-\sqrt{\frac{2\sqrt{\tau}}{1+\tau}}\exp\left\{-\frac{1}{4}\frac{\tau\xi^2}{1+\tau}\right\}\geq 1-\exp\left\{-\frac{1}{4}\frac{\tau\xi^2}{1+\tau}\right\}\geq 1-\exp\left\{-\frac{\delta^2}{8}\frac{1-\delta/\sqrt{2}}{2-\delta/\sqrt{2}}\right\}.
$$

Thus, $D_{1/2}(\theta_0\|\theta)$ has a positive lower bound for $\sqrt{\xi^2+(\tau-1)^2}\geq\delta$. This verifies Assumption 4.

If $\sqrt{n}((\xi,\sigma^2)-(0,1))^\top\to(\eta_1,\eta_2)^\top$, then $\sqrt{n}((\xi,\tau)-(0,1))^\top\to(\eta_1,-\eta_2)^\top$ and the conclusion follows from (a) of Theorem 2. □

To prove Proposition 5, the following result is useful.

Proposition 7. *Suppose the conditions of Proposition 5 holds. Let* $A(M_n)=\{(\omega,\xi):\omega\big(2\Phi(|\xi|/2)-1\big)\leq M_n n^{-1/2}\}$. *Let* $0<t<1$ *be a constant. If* $M_n\geq\sqrt{\log n/(2(t\wedge(1-t)))}$, *then*

$$
P_{\theta_0}^n\int_{A(M_n)^c}\Big[\prod_{i=1}^n\frac{p(X_i|\omega,\xi)}{p(X_i|0,0)}\Big]^t\pi(\omega,\xi)\,\mathrm{d}\omega d\mu=o(n^{-1/2}).
$$

Proof. It can be seen that

$$
\begin{aligned}
&P_{\theta_0}^n\int_{A(M_n)^c}\Big[\prod_{i=1}^n\frac{p(X_i|\omega,\xi)}{p(X_i|0,0)}\Big]^t\pi(\omega,\xi)\,\mathrm{d}\omega d\xi \\
=&\int_{A(M_n)^c}\Big(\int p(X_1|\omega,\xi)^t p(X_1|0,0)^{1-t}\,d\mu\Big)^n\pi(\omega,\xi)\,d\omega d\xi.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\int p(X_i|\omega,\xi)^t p(X_i|0,0)^{1-t}\,d\mu \\
\leq&\Big(\int \sqrt{p(X_i|\omega,\xi)p(X_i|0,0)}\,d\mu\Big)^{2(t\wedge(1-t))} \\
=&\Big(1-\frac{1}{2}\int \big(\sqrt{p(X_i|\omega,\xi)}-\sqrt{p(X_i|0,0)}\big)^2\,d\mu\Big)^{2(t\wedge(1-t))} \\
\leq&\exp\Big(-(t\wedge(1-t))\int \big(\sqrt{p(X_i|\omega,\xi)}-\sqrt{p(X_i|0,0)}\big)^2\,d\mu\Big) \\
\leq&\exp\Big(-\frac{1}{4}(t\wedge(1-t))\big(\int \big|p(X_i|\omega,\xi)-p(X_i|0,0)\big|\,d\mu\big)^2\Big) \\
=&\exp\Big(-\frac{1}{4}(t\wedge(1-t))\omega^2\big(\int \big|\phi(X_i-\xi)-\phi(X_i)\big|\,d\mu\big)^2\Big) \\
=&\exp\Big(-(t\wedge(1-t))\omega^2\big(2\Phi(|\xi|/2)-1\big)^2\Big).
\end{aligned}
$$

The last display implies that

$$
\begin{aligned}
&P^n_{\theta_0}\int_{A(M_n)^c}\Big[\prod_{i=1}^n \frac{p(X_i|\omega,\xi)}{p(X_i|0,0)}\Big]^t \pi_\omega(\omega)\pi_\xi(\xi)\,d\omega d\xi \\
\leq&\int_{A(M_n)^c}\exp\Big[-(t\wedge(1-t))n\omega^2\big(2\Phi(|\xi|/2)-1\big)^2\Big]\pi_\omega(\omega)\pi_\xi(\xi)\,d\omega d\xi \\
\leq&\int_{A(M_n)^c}\exp\big[-(t\wedge(1-t))M_n^2\big]\pi_\omega(\omega)\pi_\xi(\xi)\,d\omega d\xi \\
\leq&n^{-1/2}\int_{A(M_n)^c}\pi_\omega(\omega)\pi_\xi(\xi)\,d\omega d\xi \\
=&o(n^{-1/2}).
\end{aligned}
$$

This completes the proof.

□

Proof of Proposition 5. We have

$$
\sum_{i=1}^n \log\frac{p(X_i|\omega,\xi)}{p(X_i|0,0)}=\sum_{i=1}^n\log\Big(1+\omega\big(\exp(\xi X_i-\xi^2/2)-1\big)\Big)=\sum_{i=1}^n\log(1+\omega\xi Y_i),
$$

where $Y_i=\big(\exp(\xi X_i-\xi^2/2)-1\big)/\xi$ if $\xi\neq 0$ and $Y_i=X_i$ if $\xi=0$.

Let $r>1/2$ and $s<1/4$, on $A((\log n)^r)\cap\{\omega\geq n^{-s}\}$, we have $|\xi|=O\big((\log n)^r/n^{1/2-s}\big)$. It is known that $\max_{1\leq i\leq n}|X_i|=O_P(\sqrt{\log n})$. On $A((\log n)^r)\cap\{\omega\geq n^{-s}\}$, we have $\max_{1\leq i\leq n}|\xi X_i-\xi^2/2|\leq|\xi|\max_{1\leq i\leq n}|X_i|+\xi^2/2=O_P(|\xi|(\log n)^{1/2})$. Then on $A((\log n)^r)\cap\{\omega\geq n^{-s}\}$, uniformly for $i=1,\dots,n$, we have

$$
\begin{aligned}
Y_i&=\xi^{-1}\Big(\xi X_i-\xi^2/2+\frac{1}{2}(\xi X_i-\xi^2/2)^2+O_{P^n_{\theta_0}}\big(|\xi|^3(\log n)^{3/2}\big)\Big) \\
&=X_i-\frac{1}{2}\xi+\frac{1}{2}\xi X_i^2-\frac{1}{2}\xi^2X_i+\frac{1}{8}\xi^3+O_{P^n_{\theta_0}}\big(|\xi|^2(\log n)^{3/2}\big) \\
&=X_i+\frac{1}{2}\xi(X_i^2-1)+O_{P^n_{\theta_0}}\big(|\xi|^2(\log n)^{3/2}\big).
\end{aligned}
$$

In particular, on $A((\log n)^r)\cap\{\omega\ge n^{-s}\}$, we have $\max_{1\le i\le n}|Y_i| = O_{P^n_{\theta_0}}(\sqrt{\log n})$. On $A((\log n)^r)\cap\{\omega\ge n^{-s}\}$, we have $\omega\xi = O((\log n)^r/\sqrt{n})$, then by Taylor expansion,

$$\begin{aligned}\sum_{i=1}^n \log(1+\omega\xi Y_i) =&\omega\xi\sum_{i=1}^n Y_i - \frac{1}{2}\omega^2\xi^2\sum_{i=1}^n Y_i^2 + O_{P^n_{\theta_0}}(n\omega^3\xi^3(\log n)^{3/2})\\ =&\omega\xi\sum_{i=1}^n Y_i - \frac{1}{2}\omega^2\xi^2\sum_{i=1}^n Y_i^2 + o_{P^n_{\theta_0}}(1).\end{aligned}$$

Note that

$$\begin{aligned}\omega\xi\sum_{i=1}^n Y_i =&\omega\xi\sum_{i=1}^n X_i + \frac{1}{2}\omega\xi^2\sum_{i=1}^n (X_i^2-1) + O_{P^n_{\theta_0}}\big(n\omega|\xi|^3(\log n)^{3/2}\big)\\ =&\omega\xi\sum_{i=1}^n X_i + O_{P^n_{\theta_0}}\Big(\frac{(\log n)^{3r+3/2}}{n^{1/2-2s}}\Big)\\ =&\omega\xi\sum_{i=1}^n X_i + o_{P^n_{\theta_0}}(1).\end{aligned}$$

On the other hand, $\omega^2\xi^2\sum_{i=1}^n Y_i^2 = n\omega^2\xi^2 + o_{P^n_{\theta_0}}(1)$. Then uniformly on $A((\log n)^r)\cap\{\omega\ge n^{-s}\}$,

$$\sum_{i=1}^n \log\frac{p(X_i|\omega,\xi)}{p(X_i|0,0)} = \omega\xi\sum_{i=1}^n X_i - \frac{1}{2}n\omega^2\xi^2 + o_{P^n_{\theta_0}}(1). \tag{25}$$

As a result,

$$\begin{aligned}&\int_{A((\log n)^r)\cap\{\omega\ge n^{-s}\}}\Big[\prod_{i=1}^n\frac{p(X_i|\omega,\xi)}{p(X_i|0,0)}\Big]^t\pi(\omega,\xi)\,d\omega d\xi\\ =&(1+o_{P^n_{\theta_0}}(1))\int_{A((\log n)^r)\cap\{\omega\ge n^{-s}\}}\exp\left\{t\omega\xi\sum_{i=1}^n X_i - \frac{1}{2}nt\omega^2\xi^2\right\}\pi(\omega,\xi)\,d\omega d\xi.\end{aligned}$$

Note that on $A((\log n)^r)\cap\{\omega\ge n^{-s}\}$, $\pi_\xi(\xi) = (1+o(1))\pi_\xi(0)$. Then

$$\begin{aligned}&\int_{A((\log n)^r)\cap\{\omega\ge n^{-s}\}}\exp\big\{t\omega\xi\sum_{i=1}^n X_i - \frac{1}{2}nt\omega^2\xi^2\big\}\pi(\omega,\xi)\,d\omega d\xi\\ =&(1+o_{P^n_{\theta_0}}(1))\pi_\xi(0)\int_{A((\log n)^r)\cap\{\omega\ge n^{-s}\}}\exp\big\{t\omega\xi\sum_{i=1}^n X_i - \frac{1}{2}nt\omega^2\xi^2\big\}\pi_\omega(\omega)\,d\omega d\xi\\ =&(1+o_{P^n_{\theta_0}}(1))\pi_\xi(0)\int_{n^{-s}}^1\pi_\omega(\omega)\,d\omega\int_{-2\Phi^{-1}\big((\log n)^r/(2\omega\sqrt{n})+1/2\big)}^{2\Phi^{-1}\big((\log n)^r/(2\omega\sqrt{n})+1/2\big)}\exp\big\{t\omega\xi\sum_{i=1}^n X_i - \frac{1}{2}nt\omega^2\xi^2\big\}d\xi.\end{aligned}$$

By direct calculation, we have

$$\begin{aligned}&\int_{-2\Phi^{-1}\big((\log n)^r/(2\omega\sqrt{n})+1/2\big)}^{2\Phi^{-1}\big((\log n)^r/(2\omega\sqrt{n})+1/2\big)}\exp\big\{t\omega\xi\sum_{i=1}^n X_i - \frac{1}{2}nt\omega^2\xi^2\big\}d\xi\\ =&\frac{1}{\omega}\sqrt{\frac{2\pi}{tn}}\exp\Big\{\frac{t}{2n}(\sum_{i=1}^n X_i)^2\Big\}\Big[\Phi\Big(2\sqrt{tn}\omega\Phi^{-1}\Big(\frac{(\log n)^r}{2\omega\sqrt{n}}+\frac{1}{2}\Big)-\sqrt{\frac{t}{n}}\sum_{i=1}^n X_i\Big)\\ &-\Phi\Big(-2\sqrt{tn}\omega\Phi^{-1}\Big(\frac{(\log n)^r}{2\omega\sqrt{n}}+\frac{1}{2}\Big)-\sqrt{\frac{t}{n}}\sum_{i=1}^n X_i\Big)\Big].\end{aligned}$$

Since
$$2\sqrt{tn}\omega\Phi^{-1}\Big(\frac{(\log n)^r}{2\omega\sqrt{n}}+\frac{1}{2}\Big)\geq\sqrt{2\pi t}(\log n)^r,$$
we have
$$\begin{aligned}&\int_{-2\Phi^{-1}\left((\log n)^r/(2\omega\sqrt{n})+1/2\right)}^{2\Phi^{-1}\left((\log n)^r/(2\omega\sqrt{n})+1/2\right)}\exp\big\{t\omega\xi\sum_{i=1}^n X_i-\frac{1}{2}nt\omega^2\xi^2\big\}d\xi\\ =&\frac{1}{\omega}\sqrt{\frac{2\pi}{tn}}\exp\Big\{\frac{t}{2n}(\sum_{i=1}^n X_i)^2\Big\}(1+o_{P_{\theta_0}^n}(1)),\end{aligned}$$
where the $o_{P_{\theta_0}^n}(1)$ term is uniform for ω. Thus,
$$\begin{aligned}&\int_{A((\log n)^r)\cap\{\omega\geq n^{-s}\}}\exp\big\{t\omega\xi\sum_{i=1}^n X_i-\frac{1}{2}nt\omega^2\xi^2\big\}\pi(\omega,\xi)\,d\omega d\xi\\ =&(1+o_{P_{\theta_0}^n}(1))\pi_\xi(0)\sqrt{\frac{2\pi}{tn}}\exp\Big\{\frac{t}{2n}(\sum_{i=1}^n X_i)^2\Big\}\int_0^1\frac{1}{\omega}\pi_\omega(\omega)\,d\omega.\end{aligned}$$

Now we consider the event $A((\log n)^r)\cap\{\omega\leq n^{-s}\}$. By Theorem 2 of Liu and Shao (2004), we have
$$\sup_{\omega\in[0,1],t\in\mathbb{R}}\sum_{i=1}^n\big(\log p(X_i|\omega,\xi)-\log p(X_i|0,0)\big)=O_{P_{\theta_0}^n}(\log\log n).$$
Thus,
$$\begin{aligned}&\int_{A((\log n)^r)\cap\{\omega\leq n^{-s}\}}\Big[\prod_{i=1}^n\frac{p(X_i|\omega,\xi)}{p(X_i|0,0)}\Big]^t\pi(\omega,\xi)\,d\omega d\xi\\ =&\exp\big\{O_{P_{\theta_0}^n}(\log(\log n))\big\}\Pi\big(\omega(2\Phi(|\xi|/2)-1)\leq(\log n)^r n^{-1/2},\omega\leq n^{-s}\big).\end{aligned}$$
We break the probability into two parts:
$$\begin{aligned}&\Pi\big(\omega(2\Phi(|\xi|/2)-1)\leq(\log n)^r n^{-1/2},\omega\leq n^{-s}\big)\\ \leq&\Pi\big(\omega(2\Phi(|\xi|/2)-1)\leq(\log n)^r n^{-1/2},\omega\leq 2(\log n)^r n^{-1/2}\big)\\ &+\Pi\big(\omega(2\Phi(|\xi|/2)-1)\leq(\log n)^r n^{-1/2},2(\log n)^r n^{-1/2}\leq\omega\leq n^{-s}\big).\end{aligned}$$
The first probability satisfies
$$\begin{aligned}\Pi\big(\omega(2\Phi(|\xi|/2)-1)\leq&(\log n)^r n^{-1/2},\omega\leq 2(\log n)^r n^{-1/2}\big)\\ \leq&\Pi\big(\omega\leq 2(\log n)^r n^{-1/2}\big)\\ \lesssim&\int_0^{2(\log n)^r n^{-1/2}}w^{\alpha_1-1}\,d\omega\\ \lesssim&\Big(\frac{(\log n)^r}{\sqrt{n}}\Big)^{\alpha_1}.\end{aligned}$$

Next we deal with the second probability. On the event of the second probability, we have $(2\Phi(|\xi|/2)-1) \le \omega^{-1}(\log n)^r n^{-1/2} \le 1/2$, which implies the boundedness of ξ. It follows that $|\xi| \le C\omega^{-1}(\log n)^r n^{-1/2}$ for some constant $C>0$ on this event. Thus,

$$\begin{aligned}
&\Pi\big(\omega(2\Phi(|\xi|/2)-1) \le (\log n)^r n^{-1/2}, 2(\log n)^r n^{-1/2} \le \omega \le n^{-s}\big) \\
\le &\Pi\big(|\xi| \le C\omega^{-1}(\log n)^r n^{-1/2}, 2(\log n)^r n^{-1/2} \le \omega \le n^{-s}\big) \\
\le &\Pi\big(|\xi| \le C\omega^{-1}(\log n)^r n^{-1/2}, \omega \le n^{-s}\big) \\
\lesssim &\int_0^{n^{-s}} \omega^{\alpha_1 - 1}\, d\omega \int_{-C\omega^{-1}(\log n)^r n^{-1/2}}^{C\omega^{-1}(\log n)^r n^{-1/2}} \pi_\xi(\xi)\, d\xi.
\end{aligned}$$

There exits $\epsilon > 0$ and $M > 0$ such that $\pi_\xi(\xi) \le M$ for $\xi \in [-\epsilon, \epsilon]$. Then

$$\begin{aligned}
&\int_0^{n^{-s}} \omega^{\alpha_1 - 1}\, d\omega \int_{-C\omega^{-1}(\log n)^r n^{-1/2}}^{C\omega^{-1}(\log n)^r n^{-1/2}} \pi_\xi(\xi)\, d\xi \\
\le &\int_0^{C(\log n)^r/(\epsilon\sqrt{n})} \omega^{\alpha_1 - 1}\, d\omega + \int_{C(\log n)^r/(\epsilon\sqrt{n})}^{n^{-s}} 2MC\omega^{\alpha_1-2}(\log n)^r n^{-1/2}\, d\omega \\
\lesssim &\Big(\frac{(\log n)^r}{\sqrt{n}}\Big)^{\alpha_1} + \frac{(\log n)^r}{\sqrt{n}}\left(\Big(\frac{(\log n)^r}{\sqrt{n}}\Big)^{\alpha_1-1} \vee \Big(\frac{1}{n^s}\Big)^{\alpha_1-1}\right) \\
= &\Big(\frac{(\log n)^r}{\sqrt{n}}\Big)^{\alpha_1} \vee \frac{(\log n)^r}{n^{1/2+s(\alpha_1-1)}}.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\int_{A((\log n)^r)\cap\{\omega \le n^{-s}\}} \Big[\prod_{i=1}^n \frac{p(X_i|\omega,\xi)}{p(X_i|0,0)}\Big]^t \pi(\omega,\xi)\, d\omega d\xi \\
= &\exp\big\{O_{P^2_{\theta_0}}(\log(\log n))\big\}\left(\Big(\frac{(\log n)^r}{\sqrt{n}}\Big)^{\alpha_1} \vee \frac{(\log n)^r}{n^{1/2+s(\alpha_1-1)}}\right) = o_{P^n_{\theta_0}}(n^{-1/2}).
\end{aligned}$$

Combine these arguments and Proposition 7, we have

$$\begin{aligned}
&\int \Big[\prod_{i=1}^n \frac{p(X_i|\omega,\xi)}{p(X_i|0,0)}\Big]^t \pi(\omega,\xi)\, d\omega d\xi \\
= &\left(\int_{A((\log n)^r)^c} + \int_{A((\log n)^r)\cap\{\omega < n^{-s}\}} + \int_{A((\log n)^r)\cap\{\omega \ge n^{-s}\}}\right)\Big[\prod_{i=1}^n \frac{p(X_i|\omega,\xi)}{p(X_i|0,0)}\Big]^t \pi(\omega,\xi)\, d\omega d\xi \\
= &(1 + o_{P^n_{\theta_0}}(1))\pi_\xi(0)\sqrt{\frac{2\pi}{tn}} \exp\Big\{\frac{t}{2n}(\sum_{i=1}^n X_i)^2\Big\}\int_0^1 \frac{1}{\omega}\pi_\omega(\omega)\, d\omega.
\end{aligned}$$

This implies that

$$2\log\Lambda_{a,b}(\mathbf{X}^n) = -\log(a/b) + \frac{a-b}{n}(\sum_{i=1}^n X_i)^2 + o_{P^n_{\theta_0}}(1).$$

Then the conclusion of (i) holds since $(\sum_{i=1}^n X_i)^2/n$ weakly converges to $\chi^2(1)$ under $P^n_{\theta_0}$.

Now we prove (ii). Suppose that $\theta_n = (\omega, \xi)$ satisfies that for some $s < 1/4$, $\omega \geq n^{-s}$ for large n and $\sqrt{n}\omega\xi \to \eta$. Then it follows from (25) and Le Cam's first lemma (van der Vaart, 1998, Theorem 6.4) that $P^n_{\theta_n}$ and $P^n_{\theta_0}$ are mutually contiguous. As a result,

$$2 \log \Lambda_{a,b}(\mathbf{X}^n) = -\log(a/b) + \frac{a-b}{n}(\sum_{i=1}^n X_i)^2 + o_{P^n_{\theta_n}}(1).$$

Note that (25) implies that

$$\left(n^{-1/2}\sum_{i=1}^n X_i, \log\frac{p_n(\mathbf{X}^n|\theta)}{p_n(\mathbf{X}^n|\theta_0)}\right)^\top \overset{P^n_{\theta_0}}{\rightsquigarrow} \mathcal{N}_2\left(\begin{pmatrix} 0 \\ -\eta^2/2 \end{pmatrix}, \begin{pmatrix} 1 & \eta \\ \eta & \eta^2 \end{pmatrix}\right).$$

By Le Cam's third lemma (van der Vaart, 1998, Example 6.7), we have

$$\sum_{i=1}^n X_i \overset{P^n_{\theta_n}}{\rightsquigarrow} \mathcal{N}(\eta, 1).$$

This proves the conclusion of (ii).

□

References

Aitkin, M. (1991). Posterior bayes factors. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(1):111–128.

Bayarri, M. J. and Garca-Donato, G. (2008). Generalization of jeffreys divergence-based priors for bayesian hypothesis testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5):981–1003.

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.

Bhattacharya, A., Pati, D., and Yang, Y. (2019). Bayesian fractional posteriors. *The Annals of Statistics*, 47(1):39–66.

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *arxiv*.

Chen, J. (2017). On finite mixture models. *Statistical Theory and Related Fields*, 1(1):15–27.

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

De Santis, F. and Spezzaferri, F. (1997). Alternative bayes factors for model selection. *Canadian Journal of Statistics*, 25(4):503–515.

Fang, K. T., Kotz, S., and Ng, K. W. (1990). *Symmetric multivariate and related distributions*, volume 36 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd., London.

Fienberg, S. E. and Rinaldo, A. (2012). Maximum likelihood estimation in log-linear models. *The Annals of Statistics*, 40(2):996–1023.

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

Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.*, 28(2):500–531.

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.

Hannig, J., Iyer, H., Lai, R. C. S., and Lee, T. C. M. (2016). Generalized fiducial inference: a review and new results. *Journal of the American Statistical Association*, 111(515):1346–1361.

Jeffreys, H. (1931). *Scientific Inference*. Cambridge University Press, Cambridge, 1 edition.

Jeffreys, H. (1961). *Theory of probability*. Third edition. Clarendon Press, Oxford.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kass, R. E. and Wasserman, L. (1995). A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the American Statistical Association*, 90(431):928–934.

Kleijn, B. J. K. and van der Vaart, A. W. (2012). The Bernstein-Von-Mises theorem under misspecification. *Electronic Journal of Statistics*, 6:354–381.

Le Cam, L. (1990). Maximum likelihood: An introduction. *International Statistical Review*, 58(2):153–171.

Li, Y. and Turner, R. E. (2016). Rényi divergence variational inference. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29*, pages 1073–1081. Curran Associates, Inc.

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.

O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society. Series B. Methodological*, 57(1):99–138. With discussion and a reply by the author.

Pati, D., Bhattacharya, A., and Yang, Y. (2018). On statistical optimality of variational bayes. In Storkey, A. and Perez-Cruz, F., editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 1579–1588, Playa Blanca, Lanzarote, Canary Islands. PMLR.

Perez, J. M. (2002). Expected-posterior prior distributions for model selection. *Biometrika*, 89(3):491–512.

Rinaldo, A., Petrović, S., and Fienberg, S. E. (2013). Maximum likelihood estimation in the β-model. *The Annals of Statistics*, 41(3):1085–1110.

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.

Shen, X. and Wasserman, L. (2001). Rates of convergence of posterior distributions. *Annals of Statistics*, 29(3):687–714.

van der Vaart, A. and Ghosal, S. (2007). Convergence rates of posterior distributions for noniid observations. *Annals of Statistics*, 35(1):192–223.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

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

Walker, S. G. and Hjort, N. L. (2001). On bayesian consistency. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):811–821.

Wang, Y. and Blei, D. M. (2018). Frequentist consistency of variational bayes. *Journal of the American Statistical Association*, pages 1–15.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9(1):60–62.

Yang, Y., Pati, D., and Bhattacharya, A. (2017). α-Variational Inference with Statistical Guarantees. *ArXiv e-prints*.