Integrated Likelihood Ratio Test

Rui Wang[1], Na Li[2], and Xingzhong Xu[*1,3]

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

[2] School of Econometrics and Management, University of the Chinese Academy of Sciences, Beijing 100049, China.

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

Monday 18th June, 2018

Abstract

A general methodology named integrated likelihood ratio test is proposed which takes posterior Bayes factor and fractional Bayes factor as main test statistics. Under certain regular conditions, the Wilks phenomenon of the integrated likelihood ratio is proved, which can be used to determine the critical value of the test. We also give the asymptotic local power of the integrated likelihood ratio test. It is shown that the integrated likelihood ratio test shares similar frequency properties with the likelihood ratio test but requires fewer conditions. The integrated likelihood ratio test is available even if the likelihood functions are unbounded where the classical likelihood ratio test can not be defined. We apply the integrated likelihood ratio test to two submodels of the normal mixture model. The likelihood ratio test can not be defined for the first submodel and has undesirable local power behavior for the second submodel. In contrast, we show that the integrated likelihood ratio test has good asymptotic power behavior for both submodels.

KEYWORDS: *Bayes consistency; Fractional posterior; Integrated likelihood ratio; Mixture model; Posterior Bayes factor.*

1 Introduction

Likelihood inference plays a dominant role in parametric statistic inference. On the one hand, the maximum likelihood estimation is asymptotically optimal in a great variety of problems. On the

[*]Corresponding author

Email address: xuxz@bit.edu.cn

other hand, the fundamental lemma of Neyman and Pearson tells us that the likelihood ratio test (LRT) is the most powerful test if the null and alternative hypotheses are both simple. For testing complex hypotheses

$$H : \theta \in \Theta_0 \quad \text{vs.} \quad K : \theta \in \Theta_1, \tag{1}$$

where $\Theta_0 \cap \Theta_1 = \emptyset$, $\Theta_0 \cup \Theta_1 = \Theta$, Θ is an open subset of $\mathbb{R}^p$ and Θ_0 is a p_0-dimensional subset of $\mathbb{R}^p$, the LRT statistic is defined as

$$\Lambda_{\text{LRT}} = \frac{\max_{\theta \in \Theta} L(\theta)}{\max_{\theta \in \Theta_0} L(\theta)} = \frac{L(\hat{\theta}_{\text{MLE}})}{L(\hat{\theta}^{(0)}_{\text{MLE}})},$$

where $L(\theta)$ is the likelihood function, $\hat{\theta}_{\text{MLE}}$ and $\hat{\theta}^{(0)}_{\text{MLE}}$ are the MLE of θ in Θ and Θ_0, respectively. A key property of the LRT is Wilks phenomenon (Wilks, 1938) which asserts that for regular models, $2 \log \Lambda_{\text{LRT}}$ converges to $\chi^2(p - p_0)$ in law under the null hypothesis. The LRT has been very successful in many specific problems. However, for some moderately complex problems, some difficulties may arise when using the LRT. The maximization of $L(\theta)$ may be difficult if the likelihood function is not concave and has multiple local maxima. Worse still, in some problems the likelihood functions are unbounded and hence the LRT is not defined; see, for example, Le Cam (1990). Notice that 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).

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, no attention has been paid to the integrated likelihood functions. A natural integral type test statistic for hypothesis (1) is

$$\frac{\int_{\Theta} L(\theta) d\Pi(\theta)}{\int_{\Theta_0} L(\theta) d\Pi^{(0)}(\theta)}, \tag{2}$$

where Π and $\Pi^{(0)}$ are some probability measures on Θ and Θ_0, respectively. If Π and $\Pi^{(0)}$ are independent of data, then the statistic (2) is exactly the Bayes factor (Jeffreys, 1931) with the prior distributions Π and $\Pi^{(0)}$. The Bayes factor is the conventional tool for Bayesian hypothesis testing and has been widely used by practitioners; see Kass and Raftery (1995) for a review. However, Bayes factor is sensitive to the choice of prior distribution. In fact, the asymptotic distribution of Bayes factor depends on the prior density at the true parameter; see, for example, Clarke and Barron (1990). As a result, the Bayes factor cannot be treated as a frequentist test statistic. Thus, the measures Π and $\Pi^{(0)}$ considered in this paper will depend on data.

If $\Pi(\theta = \hat{\theta}_{\text{MLE}}) = 1$ and $\Pi^{(0)}(\theta = \hat{\theta}^{(0)}_{\text{MLE}}) = 1$, then the statistic (2) becomes the LRT statistic. In this case, the measure Π and $\Pi^{(0)}$ both concentrate on one point and are highly nonsmooth. For many models where the LRT fails, the likelihood function $L(\theta)$ still has good properties for most θ and the MLE is unfortunately trapped in a fairly small area of θ where $L(\theta)$ has bad

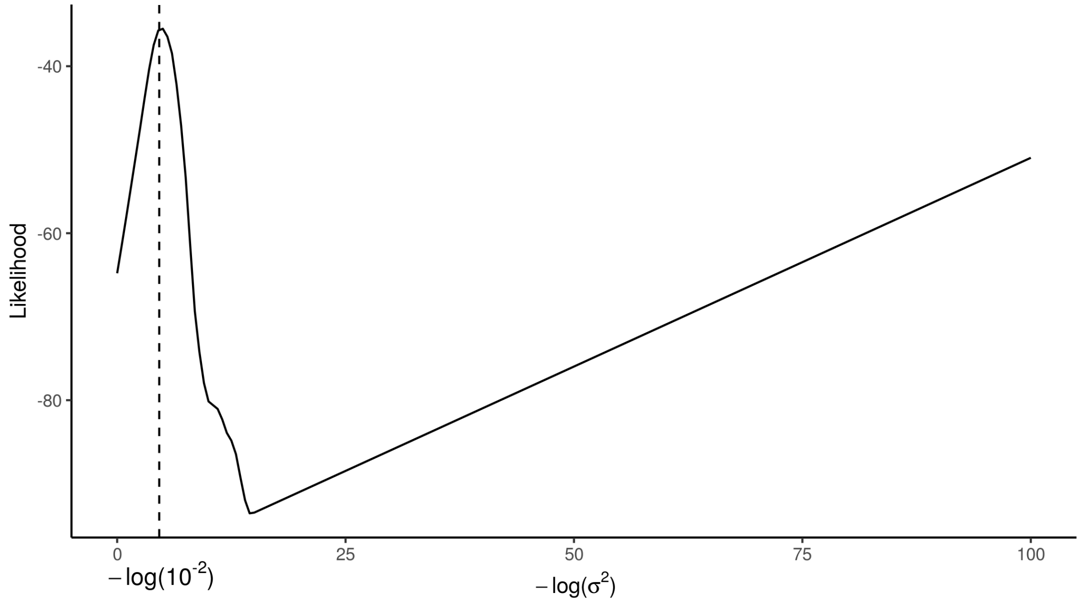

Figure 1: An example of unbounded likelihood. We data $X_1, \ldots, X_n$ which are iid from the mixture model $(1-\omega)\mathcal{N}(0,1) + \omega\mathcal{N}(\xi, \sigma^2)$ with $(\omega, \xi, \sigma^2)^T = (1/2, 1, 10^{-2})$ and $n = 50$. We plot the likelihood function in $-\log(\sigma^2)$ with $\omega = 1/2$ 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 $-\log(\sigma^2) = -\log(10^{-2})$.

behavior. Figure 1 exhibits this phenomenon. Intuitively, if Π and $\Pi^{(0)}$ are smooth and have small tail probability, then the defeat of the likelihood function in a small area will not introduce much effect on the integrated likelihood. Following this idea, a natural choice is to take Π and $\Pi^{(0)}$ as the posterior distribution (with certain prior distributions) of θ in Θ and Θ_0, respectively. In this case, the statistic (2) becomes the posterior Bayes factor proposed (PBF) by Aitkin (1991). Aitkin (1991) argued that if the likelihood is concentrated around the MLE, the PBF should approximately equal to $2^{(p-p_0)/2}\Lambda_{\mathrm{LRT}}$. This implies that PBF has the similar Wilks phenomenon as the LRT.

If we replace the likelihood function $L(\theta)$ by $L(\theta)^a$ for $a > 0$, then the LRT statistic becomes

$$\frac{\max_{\theta \in \Theta} L^a(\theta)}{\max_{\theta \in \Theta_0} L^a(\theta)} = \Lambda_{\mathrm{LRT}}^a,$$

which is equivalent to the LRT statistic. In contrast, the statistic

$$\frac{\int_{\Theta} L^a(\theta) d\Pi(\theta)}{\int_{\Theta_0} L^a(\theta) d\Pi^{(0)}(\theta)} \tag{3}$$

is not equivalent to the statistic (2). We will also consider the test statistic (3) with $0 < a < 1$. Correspondingly, the measure Π and $\Pi^{(0)}$ can also take the fractional posterior (Bhattacharya et al., 2016). Raising the likelihood to a fractional power has several advantages; see, e.g., Walker and Hjort (2001) and Bhattacharya et al. (2016). In particular, the consistency of the fractional posterior requires less conditions than the consistency of the usual posterior. A special case of the statistic (3) is the fractional Bayes factor (FBF) proposed by O'Hagan (1995). We call the statistic (3) the generalized FBF if Π and $\Pi^{(0)}$ are fractional posterior distributions.

Under certain regular conditions, we rigorously prove the Wilks phenomenon of the generalized FBF. Based on the Wilks phenomenon, an asymptotically correct frequentist test procedure can be formulated. We also give the asymptotic local power of the resulting test procedure under contiguous alternative. It is shown that the generalized FBF has a similar asymptotic local power to the LRT. However, the generalized FBF can be applied to the cases where the likelihood is unbounded and thus has a wider application scope than the LRT.

The generalized FBF can be computed by sampling θ from the fractional posterior and calculate the sample mean of the fractional likelihood. For moderately complex model, however, sampling from the fractional posterior may be difficult and hence some approximation methods may be used in practice. Variational inference is a popular method for approximating intractable posterior distribution; see Blei et al. (2017) and the references therein. Such procedure produces a distribution other than the fractional posterior distribution. To accommodate such cases, we also give a theorem (Theorem 2 in Section 2) for the general measure Π and $\Pi^{(0)}$ in (3).

For some irregular problems, the behavior of likelihood is complicated. We apply the proposed method to testing the homogeneity in a two-component normal mixture model. This problem is fairly irregular and suffers from nonidentifiability and nonconcave likelihood. Hall and Stewart (2005) showed that the likelihood ratio test has trivial power under $n^{-1/2}$ local alternative hypothesis. In contrary, we show that the ILRT have nontrivial power under $n^{-1/2}$ local alternative hypothesis.

The paper is organized as follow. In Section 2, we prove the Wilks phenomenon of the ILRT statistic and gives the asymptotic local power of the corresponding test. In Section 3, we apply ILRT to testing the homogeneity in a two-component normal mixture model. Section 4 concludes the paper. All technical proofs are in Appendix.

2 Integrated likelihood ratio test

2.1 The test statistic

Let $\mathbf{X}^{(n)} = (X_1, \ldots, X_n)$ be independent identically distributed (iid) observations taking values in some space $(\mathcal{X}; \mathscr{A})$. Suppose that there is a σ-finite measure μ on $\mathcal{X}$ and that the possible distribution P_θ of X_i has a density $p(X|\theta)$ with respect to μ. 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 $\theta = (\nu^T, \xi^T)^T$, where ν is a p_0 dimensional subvector and ξ is a $p - p_0$ dimensional subvector. 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

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

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

In Bayesian hypothesis testing framework, one puts prior $\pi(\nu)$ and $\pi(\theta)$ on parameters under the null and alternative hypotheses, respectively. The PBF proposed by Aitkin (1991) is defined as

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

where $\tilde{\Theta}_0 = \{\nu : (\nu^T, \xi^T)^T \in \Theta_0\}$, $\pi(\nu|\mathbf{X}^{(n)})$ and $\pi(\theta|\mathbf{X}^{(n)})$ are the posterior densities under the null and alternative hypothesis, respectively. The FBF proposed by O'Hagan (1995) is defined as

$$\frac{L_1(\Theta; \mathbf{X}^{(n)})}{L_b(\Theta; \mathbf{X}^{(n)})} \cdot \frac{L_b^{(0)}(\tilde{\Theta}_0; \mathbf{X}^{(n)})}{L_1^{(0)}(\tilde{\Theta}_0; \mathbf{X}^{(n)})} \quad \text{for} \quad 0 < b < 1,$$

where for $t > 0$ and sets $A \subset \Theta$, $A_0 \subset \tilde{\Theta}_0$,

$$L_t(A; \mathbf{X}^{(n)}) = \int_A \big[p_n(\mathbf{X}^{(n)}|\theta)\big]^t \pi(\theta)\, d\theta, \quad L_t^{(0)}(A_0; \mathbf{X}^{(n)}) = \int_{A_0} \big[p_n(\mathbf{X}^{(n)}|\nu, \xi_0)\big]^t \pi(\nu)\, d\nu.$$

Being treated as frequentist test statistics, the PBF and FBF are both integrated type statistics. We generalize the PBF and FBF and propose the following integrated likelihood ratio test statistic

$$\frac{\int_\Theta \big[p_n(\mathbf{X}^{(n)}|\theta)\big]^a \pi(\theta; \mathbf{X}^{(n)})\, d\theta}{\int_{\tilde{\Theta}_0} \big[p_n(\mathbf{X}^{(n)}|\nu, \xi_0)\big]^a \pi(\nu; \mathbf{X}^{(n)})\, d\nu},$$

where $a > 0$ is a hyperparameter, the weight functions $\pi(\theta; \mathbf{X}^{(n)})$ and $\pi(\nu; \mathbf{X}^{(n)})$ are probability density functions in Θ and $\tilde{\Theta}_0$ given $\mathbf{X}^{(n)}$, respectively. Note that $\pi(\theta; \mathbf{X}^{(n)})$ and $\pi(\nu; \mathbf{X}^{(n)})$ are data dependent but does not need to be the posterior density.

If we take the weight function as the fractional posterior

$$\pi(\theta; \mathbf{X}^{(n)}) = \frac{\big[p_n(\mathbf{X}^{(n)}|\theta)\big]^b \pi(\theta)}{\int_\Theta \big[p_n(\mathbf{X}^{(n)}|\theta)\big]^b \pi(\theta)\, d\theta}, \quad \pi(\nu; \mathbf{X}^{(n)}) = \frac{\big[p_n(\mathbf{X}^{(n)}|\nu, \xi_0)\big]^b \pi(\nu)}{\int_{\tilde{\Theta}_0} \big[p_n(\mathbf{X}^{(n)}|\nu, \xi_0)\big]^b \pi(\nu)\, d\nu}, \tag{4}$$

then the ILRT statistic equals to

$$\Lambda_{a,b} = \frac{L_{a+b}(\Theta; \mathbf{X}^{(n)})}{L_b(\Theta; \mathbf{X}^{(n)})} \cdot \frac{L_b^{(0)}(\tilde{\Theta}_0; \mathbf{X}^{(n)})}{L_{a+b}^{(0)}(\tilde{\Theta}_0; \mathbf{X}^{(n)})}.$$

We call $\Lambda_{a,b}$ the generalized FBF throughout the paper. The FBF and PBF are both the special cases of the generalized FBF. In fact, the FBF equals to $\Lambda_{1-b,b}$, the PBF equals to $\Lambda_{1,1}$. For some moderately complex models, the fractional posterior (4) may be complicated. In this case, one may use some simple form weight function to approximate the fractional posterior (4). A popular method for approximating (4) is variational inference; see, e.g., Blei et al. (2017). In this case, the weight function in (11) is equals to the variational approximation of (4). The ILRT methodology also includes such approximate method.

The computation of the ILRT statistic is relatively simple. We can independently generate $\theta_1, \ldots, \theta_m$ and $\nu_1, \ldots, \nu_m$ according to $\pi(\theta; \mathbf{X}^{(n)})$ and $\pi(\nu; \mathbf{X}^{(n)})$ for a large m. Then the ILRT statistic can be approximated by

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

2.2 Generalized FBF

In this section, we investigate the asymptotic behavior of the generalized FBF. The following assumption is adapted from Kleijn and Vaart (2012) and is satisfied by many common models.

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 $P_{\theta_0}\dot{\ell}_{\theta_0} = 0_p$. The Fisher information matrix $I_{\theta_0} = P_{\theta_0}\dot{\ell}_{\theta_0}\dot{\ell}_{\theta_0}^T$ is positive-definite, where Pf means the expectation of $f(X)$ when X has distribution P. For every $M > 0$,

$$\sup_{\|h\|\le M} \Big| \log \frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)} - h^T I_{\theta_0}\Delta_{n,\theta_0} + \frac{1}{2}h^T I_{\theta_0} h\Big| \xrightarrow{P^n_{\theta_0}} 0,$$

where $\Delta_{n,\theta_0} = n^{-1/2}\sum_{i=1}^n I_{\theta_0}^{-1}\dot{\ell}_{\theta_0}(X_i)$.

For $t > 0$, we say $L_t(\cdot; \mathbf{X}^{(n)})$ is $\sqrt{n}$-consistent if for every $M_n \to \infty$,

$$\frac{L_t(\{\theta : \|\theta - \theta_0\| > M_n/\sqrt{n}\}; \mathbf{X}^{(n)})}{L_t(\Theta; \mathbf{X}^{(n)})} \xrightarrow{P^n_{\theta_0}} 0.$$

The $\sqrt{n}$-consistency of $L_t^{(0)}(\cdot; \mathbf{X}^{(n)})$ is similarly defined. Note that the consistency of $L_1(\cdot; \mathbf{X}^{(n)})$ is equivalent to the consistency of the posterior distribution. In Kleijn and Vaart (2012), the $\sqrt{n}$-consistency of posterior distribution is a key assumption to prove Bernstein-von Mises theorem. Likewise, the $\sqrt{n}$-consistency of $L_t(\cdot; \mathbf{X}^{(n)})$ is a key assumption of the following theorem.

Theorem 1. *Suppose that Assumption 1 holds, $L_{a+b}(\cdot; \mathbf{X}^{(n)})$, $L_b(\cdot; \mathbf{X}^{(n)})$, $L_{a+b}^{(0)}(\cdot; \mathbf{X}^{(n)})$ and $L_b^{(0)}(\cdot; \mathbf{X}^{(n)})$ are $\sqrt{n}$-consistent, $\pi(\theta)$ is continuous at θ_0 with $\pi(\theta_0) > 0$, $\pi(\nu)$ is continuous at ν_0 with $\pi(\nu_0) > 0$. Then for $\{\theta_n\}$ such that $\sqrt{n}(\theta_n - \theta_0) \to \eta$,*

$$2\log\Lambda_{a,b} \overset{P^n_{\theta_n}}{\rightsquigarrow} -(p - p_0)\log(1 + \frac{a}{b}) + a\chi^2(p - p_0, \delta),$$

where $\chi^2(p-p_0,\delta)$ is a noncentral chi-squared random variable with $p-p_0$ degrees of freedom and noncentrality parameter $\delta = \eta^T\big(I_{\theta_0} - I_{\theta_0}J(J^TI_{\theta_0}J)^{-1}J^TI_{\theta_0}\big)\eta$ and $J = (I_{p_0}, 0_{p_0\times(p-p_0)})^T$, "$\rightsquigarrow$" means weak convergence.

Theorem 1 gives the asymptotic distribution of $2\log\Lambda_{a,b}$ under the null hypothesis and the local alternative hypothesis. To formulate a test with asymptotic type I error rate α, the critical value of $2\log\Lambda_{a,b}$ can be defined to be $-(p-p_0)\log(1+a/b)+a\chi^2_{1-\alpha}(p-p_0)$ where $\chi^2_{1-\alpha}(p-p_0)$ is the $1-\alpha$ quantile of a chi-squared random variable with $p-p_0$ degrees of freedom. By Theorem 1, the resulting test has local asymptotic power

$$\Pr\left(\chi^2(p-p_0,\delta) > \chi^2_{1-\alpha}(p-p_0)\right). \tag{5}$$

It is known that, under certain regular conditions, (5) is also the local asymptotic power of the likelihood ratio test. In this view, $\Lambda_{a,b}$ enjoys good frequentist properties.

The $\sqrt{n}$-consistency of $L_t(\cdot;\mathbf{X}^{(n)})$ is a key assumption of Theorem 1. Hence we would like to give sufficient conditions for the $\sqrt{n}$-consistency of $L_t(\cdot;\mathbf{X}^{(n)})$. First we consider the exponential family of distributions.

Proposition 1. *Suppose $p(X|\theta) = \exp\left[\theta^T T(X) - A(\theta)\right]$, Θ is an open subset of $\mathbb{R}^p$, θ_0 is an interior point of Θ,*

$$I_{\theta_0} = \frac{\partial^2}{\partial\theta\partial\theta^T}A(\theta_0) > 0.$$

Then $L_t(\cdot;\mathbf{X}^{(n)})$ is consistent for $t>0$.

Proposition 1 establishes the $\sqrt{n}$-consistent of $L_t(\cdot;\mathbf{X}^{(n)})$ for all $t>0$ under full-rank exponential family models. If the full model and the null model both belong to the full-rank exponential family, Assumption 1 is also satisfied. Hence, Theorem 1 implies that the generalized FBF has the similar asymptotic properties with the classical LRT statistic. However, for any test methodology, the success in the full-rank exponential family models is just a minimal requirement since the LRT is also easy to implement and enjoys good asymptotic properties. We would like to consider more general models.

For general models, the likelihood function may not be concave. This often makes it hard to implement the LRT. For some models, a more serious problem may occur, that is, the likelihood may be unbounded and hence the LRT can not be defined. This problem may occur even if the likelihood function has good local analytical properties, such as location-scale mixture models. See Le Cam (1990) for more examples. A natural question is that if the fractional integrated likelihood $L_t(\Theta;\mathbf{X}^{(n)})$ is always well defined. The following theorem shows that $L_t(\Theta;\mathbf{X}^{(n)})$ is always well defined for $t\le 1$ and is not well defined for some model for $t>1$.

Proposition 2. *If $t\le 1$, $L_t(\Theta;\mathbf{X}^{(n)}) < +\infty$ $P^n_{\theta_0}$-a.s. for any models. If $t>1$, $L_t(\Theta;\mathbf{X}^{(n)}) = +\infty$ for some models.*

Because of the bad behavior of $L_t(\Theta;\mathbf{X}^{(n)})$ for $t>1$, next we only consider $L_t(\Theta;\mathbf{X}^{(n)})$ for $t\leq 1$. For $t=1$, the $\sqrt{n}$-consistency of $L_t(\cdot;\mathbf{X}^{(n)})$ is equivalent to the $\sqrt{n}$-consistency of the posterior distribution which is a well studied problem; 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. This approach is adopted by Ghosal et al. (2000), van der Vaart and Ghosal (2007) and Kleijn and Vaart (2012). For example, Theorem 3.1 of Kleijn and Vaart (2012) assumes that for every $\epsilon>0$, there exists a sequence of tests ϕ_n such that

$$P_{\theta_0}^n\phi_n\to 0,\qquad \sup_{\|\theta-\theta_0\|\geq\epsilon} P_\theta^n(1-\phi_n)\to 0. \tag{6}$$

This condition is satisfied when the parameter space is compact and the model is suitably continuous; see Theorem 3.2 of Kleijn and Vaart (2012). However, if the parameter space is not compact, one may have to manually construct a test sequence satisfying the condition (6).

The consistency of $L_t(\cdot;\mathbf{X}^{(n)})$ for $0<t<1$ is different from $t=1$. Walker and Hjort (2001) considered the Hellinger consistency of $L_{1/2}(\cdot;\mathbf{X}^{(n)})$. They derived the consistency of $L_{1/2}(\cdot;\mathbf{X}^{(n)})$ under simple conditions. Recently, Bhattacharya et al. (2016) further developed the idea of Walker and Hjort (2001) and derived a general bounds for the consistency of $L_t(\cdot;\mathbf{X}^{(n)})$ 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 $L_t(\cdot;\mathbf{X}^{(n)})$ for $0<t<1$ under certain conditions on the Rényi divergence between distributions in the family $\{P_\theta:\theta\in\Theta\}$.

For two parameters θ_1 and θ_2, the α order Rényi divergence ($0<\alpha<1$) of P_{θ_1} from P_{θ_2} is defined to be

$$D_\alpha(\theta_1||\theta_2)=-\frac{1}{1-\alpha}\log\rho_\alpha(\theta_1,\theta_2),$$

where $\rho_\alpha(\theta_1,\theta_2)=\int_{\mathcal{X}} p(X|\theta_1)^\alpha p(X|\theta_2)^{1-\alpha}\,d\mu$ is the so-called Hellinger integral. The following assumption is needed for our $\sqrt{n}$-consistency result.

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

Remark 1. A remarkable property of Rényi divergence is the equivalence of all D_α: If $0<\alpha<\beta<1$, then

$$\frac{\alpha}{1-\alpha}\frac{1-\beta}{\beta}D_\beta(\theta_1||\theta_2)\leq D_\alpha(\theta_1||\theta_2)\leq D_\beta(\theta_1||\theta_2).$$

See, e.g., Bobkov et al. (2016). As a result, if Assumption 2 holds for some $\alpha\in(0,1)$, then it will hold for every $\alpha\in(0,1)$.

To appreciate Assumption 2, suppose, for example, that $D_\alpha(\theta||\theta_0)$ is twice continuously differentiable in θ. Since $\theta=\theta_0$ is a minimum point of $D_\alpha(\theta||\theta_0)$, the first order derivative of $D_\alpha(\theta||\theta)$

at $\theta = \theta_0$ is zero and the second order derivative at $\theta = \theta_0$ is positive semidefinite. By Taylor theorem, in a small neighbourhood of θ_0,

$$D_\alpha(\theta||\theta_0) = \frac{1}{2}(\theta - \theta_0)^T \frac{\partial^2}{\partial\theta\partial\theta^T} D_\alpha(\theta||\theta_0)\Big|_{\theta=\theta^*}(\theta - \theta_0),$$

where θ^* is between θ_0 and θ. If we further assume the second order derivative is positive definite at $\theta = \theta_0$, then in a small neighbourhood of θ_0, there is a positive constant C such that $D_\alpha(\theta||\theta_0) \geq C\|\theta - \theta_0\|^2$. Thus, Assumption 2 is a fairly weak condition.

Proposition 3. *Suppose θ_0 is an interior of Θ, $\pi(\theta)$ is continuous at θ_0 and $\pi(\theta_0) > 0$. Under Assumptions 1 and 2, for fixed $t \in (0, 1)$, $L_t(\cdot; \mathbf{X}^{(n)})$ is consistent.*

Note that if the conditions of Theorem 1 are satisfied, the asymptotic power of $\Lambda_{a,b}$ is independent of a, b. Hence specific choices of a, b are not crucial provided $L_{a+b}(\cdot; \mathbf{X}^{(n)})$, $L_b(\cdot; \mathbf{X}^{(n)})$, $L^{(0)}_{a+b}(\cdot; \mathbf{X}^{(n)})$ and $L^{(0)}_b(\cdot; \mathbf{X}^{(n)})$ are $\sqrt{n}$-consistent. For some models, it is more convenient to verify Assumption 2 than to directly construct a test sequence satisfying the condition (6). In such cases, it can be recommended to use the generalized FBF with $a + b < 1$.

2.3 General weight function

For some moderately complex models, the fractional posterior (4) are not easy to calculate. In such cases, one may want to use simpler weight functions to approximate (4). In this section, we consider the asymptotic properties of the ILRT statistic with general weight functions.

Let $h = \sqrt{n}(\theta - \theta_0)$ be the local parameter and $\pi_n(h; \mathbf{X}^{(n)}) = \pi(\theta_0 + n^{-1/2}h; \mathbf{X}^{(n)})$ be the weight function in terms of h. If $\pi(\theta; \mathbf{X}^{(n)})$ is the posterior density of θ, then Bernstein-von Mises theorem asserts that under certain conditions, $\|\pi_n(h; \mathbf{X}^{(n)}) - \phi(h; \Delta_{n,\theta_0}, I^{-1}_{\theta_0})\|$ converges to 0 in $P^n_{\theta_0}$ probability, where for two densities $q_1(h)$ and $q_2(h)$, $\|q_1(h) - q_2(h)\| = \int |q_1(h) - q_2(h)|\, dh$ is their total variation distance. Similarly, if $\pi(\theta; \mathbf{X}^{(n)})$ is the fractional posterior density of θ with fractional power b, it can be proved that under certain conditions, $\|\pi_n(h; \mathbf{X}^{(n)}) - \phi(h; \Delta_{n,\theta_0}, b^{-1}I^{-1}_{\theta_0})\|$ converges to 0 in $P^n_{\theta_0}$ probability. We shall assume that the weight function inherits such properties.

Assumption 3. *Let $b \in (0, 1)$ be a fixed number. Assume that $\pi_n(h; \mathbf{X}^{(n)})$ satisfies*

$$\|\pi_n(h; \mathbf{X}^{(n)}) - \phi(h; \Delta_{n,\theta_0}, b^{-1}I^{-1}_{\theta_0})\| \xrightarrow{P^n_{\theta_0}} 0. \tag{7}$$

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

$$\|\pi_n(h^{(0)}; \mathbf{X}^{(n)}) - \phi(h^{(0)}; \Delta^{(0)}_{n,\theta_0}, b^{-1}I^{(0)-1}_{\theta_0})\| \xrightarrow{P^n_{\theta_0}} 0, \tag{8}$$

where

$$\Delta^{(0)}_{n,\theta_0} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I^{(0)-1}_{\theta_0} \dot{\ell}^{(0)}_{\theta_0}(X_i), \quad I^{(0)}_{\theta_0} = P_{\theta_0}\dot{\ell}^{(0)}_{\theta_0}\dot{\ell}^{(0)T}_{\theta_0}, \quad \dot{\ell}^{(0)}(X) = \frac{\partial}{\partial\nu} \log p(X|\nu, \xi_0)\Big|_{\nu=\nu_0}.$$

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

$$\lim_{n\to\infty} P_{\theta_0}^n \left\{ \sup_{h\in\mathbb{R}^p} (\pi_n(h;\mathbf{X}^{(n)}) - T(h)) \le 0 \right\} \ge 1-\epsilon. \tag{9}$$

$$\lim_{n\to\infty} P_{\theta_0}^n \left\{ \sup_{h^{(0)}\in\mathbb{R}^{p_0}} (\pi_n(h^{(0)};\mathbf{X}^{(n)}) - T^{(0)}(h^{(0)})) \le 0 \right\} \ge 1-\epsilon. \tag{10}$$

The conditions (9) and (10) in Assumption 3 assume that there is a function controlling the tail of the weight functions. We need to control the tail of the weight function since the behavior of the likelihood may be undesirable when θ is far away from θ_0. In fact, even for some fairly regular models, the likelihood may tends to infinity, which invalidates LRT; see, e.g., Le Cam (1990). So we control the tail of the weight function to avoid too much weights on the tail of likelihood. If the weight function $\pi_n(h;\mathbf{X}^{(n)})$ is normal density, then it can be shown that the conditions (7) and (8) implies (9) and (10).

Under Assumption 3, we consider the ILRT statistic

$$\Lambda_{a,b}^* = \frac{\int_\Theta \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^a \pi(\theta;\mathbf{X}^{(n)})\,d\theta}{\int_{\tilde{\Theta}_0} \left[p_n(\mathbf{X}^{(n)}|\nu,\xi_0)\right]^a \pi(\nu;\mathbf{X}^{(n)})\,d\nu}. \tag{11}$$

The following theorem gives the asymptotic distribution of ILRT statistic.

Theorem 2. *Suppose that Assumptions 1 and 3 hold with $a+b\le 1$. Then for $\{\theta_n\}$ such that $\sqrt{n}(\theta_n-\theta_0)\to\eta$, we have*

$$2\log\Lambda_{a,b}^* \overset{P_{\theta_n}^n}{\rightsquigarrow} -(p-p_0)\log(1+\frac{a}{b}) + a\chi^2(p-p_0,\delta),$$

where δ is defined as in Theorem 1.

Theorem 2 shows that even with approximate weight function, the ILRT statistic can still produce an asymptotic optimal test. A practical method to obtain simple form weight function $\pi_n(h;\mathbf{X}^{(n)})$ is the variational inference; see, e.g., Blei et al. (2017). Next we shall consider a simple variational method which is guaranteed to yield a weight function satisfying Assumption 3. For comprehensive considerations of the statistical properties of variational methods; see the recent works of Wang and Blei (2017), Pati et al. (2017) and Yang et al. (2017).

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

$$\|\pi_n(h;\mathbf{X}^{(n)}) - \phi(h;\Delta_{n,\theta_0}, b^{-1}I_{\theta_0}^{-1})\| \xrightarrow{P_{\theta_0}^n} 0. \tag{12}$$

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

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

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

$$\pi_n^\dagger(h;\mathbf{X}^{(n)}) = \mathop{\arg\min}_{q(h)\in\mathcal{Q}} -\frac{1}{1-\alpha}\log\int q(h)^\alpha \pi_n(h;\mathbf{X}^{(n)})^{1-\alpha}\,dh.$$

Hence we have

$$-\frac{1}{1-\alpha}\log\int \pi_n^\dagger(h;\mathbf{X}^{(n)})^\alpha \pi_n(h;\mathbf{X}^{(n)})^{1-\alpha}\,dh \le -\frac{1}{1-\alpha}\log\int \phi(h;\Delta_{n,\theta_0}, I_{\theta_0}^{-1})^\alpha \pi_n(h|\mathbf{X}^{(n)})^{1-\alpha}\,dh. \tag{13}$$

Since Rényi divergence and total variation distance are equivalent, (12) implies that the right hand side of (13) tends to 0 in $P_{\theta_0}^n$-probability. Again by the equivalence of Rényi divergence and total variation distance, we have

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

Note that $\pi_n^\dagger(h;\mathbf{X}^{(n)})$ 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 $\pi_n^\dagger(h;\mathbf{X}^{(n)})$ are bounded in probability. Then a dominating function $T(h)$ exists and thus (9) holds.

3 Normal mixture model

In this section, we apply the ILRT methodology to the testing the component number of normal mixture model. Normal mixture model is a highly irregular model. Due to partial loss of identifiability, the likelihood ratio test has undesirable behavior. For example, if the component variances are totally unknown, the likelihood is unbounded and thus likelihood ratio test 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 ILRT may have better behavior than likelihood ratio test. For example, for unknown variances case, ILRT is at least well defined.

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

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

where $0\le\omega\le1$, $\mu\in\mathbb{R}$ and $\sigma^2\in\mathbb{R}^+$. First, we assume $\omega=1/2$ is known and consider testing the hypotheses

$$H:\xi=0,\sigma=1 \quad \text{vs.} \quad K:\xi\ne0 \text{ or } \sigma\ne1. \tag{14}$$

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. Thus, the LRT can not be defined. Using Theorem 1 and Proposition 3, we can obtain the following proposition.

Proposition 4. *For hypotheses testing problem* (14)*, if* $\sqrt{n}((\xi,\sigma^2)-(0,1))^T \to (\eta_1,\eta_2)^T$*, then the generalized FBF with* $a>0$*,* $b>0$ *and* $a+b<1$ *satisfies*

$$2\log\Lambda_{a,b} \overset{P_{\theta_n}^n}{\rightsquigarrow} -2\log(1+\frac{a}{b}) + a\chi^2(2,\eta_1^2/4+\eta_2^2/8).$$

This example shows that even when the LRT fails, the ILRT may still be valid and has the expected asymptotic distribution. Thus, the ILRT methodology has a wider application scope than the LRT.

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

$$\omega\xi = 0 \quad \text{vs.} \quad \omega\xi \neq 0. \tag{15}$$

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 it has trivial power under $n^{-1/2}$ local alternative hypothesis. For this irregular problem, Theorem 1 and Proposition 3 cannot be directly applied. This is because the second part is Assumption (3) is violated due to loss of identifiability. However, this does not means that the ILRT is not applicable. In fact, the following theorem shows that the generalized FBF with $a+b<1$ has the desirable asymptotic properties.

Theorem 3. *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* $a+b<1$*. Then,*

(i) under the null hypothesis,

$$2\log\Lambda_{a,b} \overset{P_{\theta_0}^n}{\rightsquigarrow} \log(1+\frac{a}{b}) + a\chi^2(1);$$

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

$$2\log\Lambda_{a,b} \overset{P_{\theta_n}^n}{\rightsquigarrow} \log(1+\frac{a}{b}) + a\chi^2(1,\eta^2).$$

Theorem 3 shows that the ILRT 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 $|\gamma| < 1$.

4 Conclusion

In this paper, we proposed a flexible methodology ILRT which includes some existing method as special cases. We give the asymptotic distribution of the generalized FBF, which is a special case of ILRT. We also investigate the asymptotic behavior of the ILRT for general weight functions. It is shown that the generalized FBF has the Wilks phenomenon similar to the LRT. The asymptotic local power is also given. We also apply the ILRT methodology to two submodels of the normal mixture model. These examples show that the ILRT can have good behavior even if the LRT is not defined or has poor properties.

The ILRT statistic is easy to implement provided sampling from weight functions is simple. If the weight functions are fractional posterior densities, then Markov chain Monte Carlo (MCMC) methods can be used to sample from weight functions. If MCMC is not efficient, one can use approximation methods such as variational inference and the resulting test procedure is still valid. Thus, the ILRT methodology can also be recommended when the classical LRT is not easy to implement.

It is interesting to apply the ILRT methodology to specific complex testing problems. We leave it for future research.

Acknowledgements

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

Appendices

Appendix A Proofs in Section 2

Proof of Theorem 1. For fixed $t > 0$ and $M > 0$, we have

$$
\begin{aligned}
&\log \int_{\{\theta:\|\theta-\theta_0\|\leq M/\sqrt{n}\}} \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^t \pi(\theta)\, d\theta \\
=&\log \int_{\{\theta:\|\theta-\theta_0\|\leq M/\sqrt{n}\}} \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^t d\theta + \log \pi(\theta_0) + o_{P_{\theta_0}^n}(1) \\
=&\log \int_{\{h:\|h\|\leq M\}} \exp\left[t \log p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)\right] dh - \frac{p}{2}\log n + \log \pi(\theta_0) + o_{P_{\theta_0}^n}(1),
\end{aligned}
$$

where the first equality holds since $\pi(\theta)$ is continuous at θ_0 and the second equality follows from the coordinate transformation $h = \sqrt{n}(\theta - \theta_0)$. By the uniform expansion given by Assumption 1 and a little algebra, we have

$$\begin{aligned}
&\log \int_{\{h:\|h\|\leq M\}} \exp\left[t \log p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)\right] dh \\
=&\log \int_{\{h:\|h\|\leq M\}} \exp\left[-\frac{t}{2}(h-\Delta_{n,\theta_0})^T I_{\theta_0}(h-\Delta_{n,\theta_0})\right] dh + \frac{t}{2}\Delta_{n,\theta_0}^T I_{\theta_0}\Delta_{n,\theta_0} + t\log p_n(\mathbf{X}^{(n)}|\theta_0) + o_{P_{\theta_0}^n}(1).
\end{aligned}$$

Thus

$$\begin{aligned}
&\log \int_{\{\theta:\|\theta-\theta_0\|\leq M/\sqrt{n}\}} \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^t \pi(\theta)\, d\theta \\
=&\log \int_{\{h:\|h\|\leq M\}} \exp\left[-\frac{t}{2}(h-\Delta_{n,\theta_0})^T I_{\theta_0}(h-\Delta_{n,\theta_0})\right] dh \\
&+ \frac{t}{2}\Delta_{n,\theta_0}^T I_{\theta_0}\Delta_{n,\theta_0} + t\log p_n(\mathbf{X}^{(n)}|\theta_0) - \frac{p}{2}\log n + \log \pi(\theta_0) + o_{P_{\theta_0}^n}(1).
\end{aligned}$$

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

$$\begin{aligned}
&\log \int_{\{h:\|h\|\leq M_n\}} \exp\left[-\frac{t}{2}(h-\Delta_{n,\theta_0})^T I_{\theta_0}(h-\Delta_{n,\theta_0})\right] dh \\
=&\log \int_{\mathbb{R}^p} \exp\left[-\frac{t}{2}(h-\Delta_{n,\theta_0})^T I_{\theta_0}(h-\Delta_{n,\theta_0})\right] dh + o_{P_{\theta_0}^n}(1) \\
=&\frac{p}{2}\log(2\pi) - \frac{p}{2}\log t - \frac{1}{2}\log|I_{\theta_0}| + o_{P_{\theta_0}^n}(1).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\log \int_{\{\theta:\|\theta-\theta_0\|\leq M_n/\sqrt{n}\}} \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^t \pi(\theta)\, d\theta \\
=&\frac{p}{2}\log\left(\frac{2\pi}{n}\right) - \frac{p}{2}\log t - \frac{1}{2}\log|I_{\theta_0}| + \log\pi(\theta_0) + \frac{t}{2}\Delta_{n,\theta_0}^T I_{\theta_0}\Delta_{n,\theta_0} + t\log p_n(\mathbf{X}^{(n)}|\theta_0) + o_{P_{\theta_0}^n}(1).
\end{aligned}$$

If $L_t(\cdot;\mathbf{X}^{(n)})$ is $\sqrt{n}$-consistent, then

$$\begin{aligned}
&\log L_t(\Theta;\mathbf{X}^{(n)}) = \log \int_{\Theta} \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^t \pi(\theta)\, d\theta \\
=&\frac{p}{2}\log\left(\frac{2\pi}{n}\right) - \frac{p}{2}\log t - \frac{1}{2}\log|I_{\theta_0}| + \log\pi(\theta_0) + \frac{t}{2}\Delta_{n,\theta_0}^T I_{\theta_0}\Delta_{n,\theta_0} + t\log p_n(\mathbf{X}^{(n)}|\theta_0) + o_{P_{\theta_0}^n}(1).
\end{aligned}$$

Similarly, if $L_t^{(0)}(\cdot;\mathbf{X}^{(n)})$ is $\sqrt{n}$-consistent,

$$\begin{aligned}
&\log L_t^{(0)}(\tilde{\Theta};\mathbf{X}^{(n)}) = \log \int_{\tilde{\Theta}_0} \left[p_n(\mathbf{X}^{(n)}|\nu,\xi_0)\right]^t \pi(\nu)\, d\nu \\
=&\frac{p_0}{2}\log\left(\frac{2\pi}{n}\right) - \frac{p_0}{2}\log t - \frac{1}{2}\log|I_{\theta_0}^{(0)}| + \log\pi(\nu_0) + \frac{t}{2}\Delta_{n,\theta_0}^{(0)T} I_{\theta_0}^{(0)}\Delta_{n,\theta_0}^{(0)} + t\log p_n(\mathbf{X}^{(n)}|\theta_0) + o_{P_{\theta_0}^n}(1).
\end{aligned}$$

By the mutual contiguity of $P_{\theta_0}^n$ and $P_{\theta_n}^n$, the term $o_{P_{\theta_0}^n}(1)$ is also $o_{P_{\theta_n}^n}(1)$. Hence

$$\begin{aligned}
\log\Lambda_{a,b} =& \log L_{a+b}(\Theta;\mathbf{X}^{(n)}) - \log L_b(\Theta;\mathbf{X}^{(n)}) - \log L_{a+b}^{(0)}(\tilde{\Theta}_0;\mathbf{X}^{(n)}) + \log L_b^{(0)}(\tilde{\Theta}_0;\mathbf{X}^{(n)}) \\
=& -\frac{p-p_0}{2}\log\left(1+\frac{a}{b}\right) + \frac{a}{2}\left(\Delta_{n,\theta_0}^T I_{\theta_0}\Delta_{n,\theta_0} - \Delta_{n,\theta_0}^{(0)T} I_{\theta_0}^{(0)}\Delta_{n,\theta_0}^{(0)}\right) + o_{P_{\theta_n}^n}(1).
\end{aligned}$$

Since $I^{(0)}_{\theta_0} = J^T I_{\theta_0} J$ and $\Delta^{(0)}_{n,\theta_0} = (J^T I_{\theta_0} J)^{-1} J^T I_{\theta_0} \Delta_{n,\theta_0}$, we have

$$\Delta^T_{n,\theta_0} I_{\theta_0} \Delta_{n,\theta_0} - \Delta^{(0)T}_{n,\theta_0} I^{(0)}_{\theta_0} \Delta^{(0)}_{n,\theta_0} = \Delta^T_{n,\theta_0} I^{1/2}_{\theta_0} \big(I_p - I^{1/2}_{\theta_0} J (J^T I_{\theta_0} J)^{-1} J^T I^{1/2}_{\theta_0}\big) I^{1/2}_{\theta_0} \Delta_{n,\theta_0},$$

where $I_p - I^{1/2}_{\theta_0} J (J^T I_{\theta_0} J)^{-1} J^T I^{1/2}_{\theta_0}$ is a projection matrix with rank $p - p_0$. It remains to derive the asymptotic distribution of Δ_{n,θ_0}. Let $h_n = \sqrt{n}(\theta_n - \theta_0)$. By Assumption 1 and CLT,

$$\begin{pmatrix} \dfrac{1}{\sqrt{n}} \displaystyle\sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \\ \log \dfrac{p_n(\mathbf{X}^{(n)}|\theta_n)}{p_n(\mathbf{X}^{(n)}|\theta_0)} \end{pmatrix} \overset{P_0^n}{\rightsquigarrow} \mathcal{N}\left(\begin{pmatrix} 0 \\ -\frac{1}{2}\eta^T I_{\theta_0}\eta \end{pmatrix}, \begin{pmatrix} I_{\theta_0} & I_{\theta_0}\eta \\ \eta^T I_{\theta_0} & \eta^T I_{\theta_0}\eta \end{pmatrix} \right).$$

Hence by Le Cam's third lemma,

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

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

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

which completes the proof.

□

Proof of Proposition 1. For exponential family, we have

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

and

$$\log \frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)} = h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0} h - g_n(h),$$

where

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

Without loss of generality, we assume $M_n \to \infty$ and $M_n^3/\sqrt{n} \to 0$. By Taylor's theorem and the continuity of the third derivative of $A(\theta)$,

$$\max_{\{h:\|h\|\le M_n\}} |g_n(h)| = O\left(\frac{M_n^3}{\sqrt{n}}\right) \to 0.$$

This allows us to derive the following lower bound for $L_t(\Theta; \mathbf{X}^{(n)})$.

$$\begin{aligned}
L_t(\Theta; \mathbf{X}^{(n)}) \ge& \int_{\{\theta:\|\theta-\theta_0\|\le M_n/\sqrt{n}\}} \big[p_n(\mathbf{X}^{(n)}|\theta)\big]^t \pi(\theta)\, d\theta \\
=&(1 + o_{P^n_{\theta_0}}(1)) n^{-p/2}\pi(\theta_0)\big[p_n(\mathbf{X}^{(n)}|\theta_0)\big]^t \int_{\{h:h\le M_n\}} \exp\big[th^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{t}{2}h^T I_{\theta_0} h\big]\, dh \\
=&(1 + o_{P^n_{\theta_0}}(1)) n^{-p/2}\pi(\theta_0)\big[p_n(\mathbf{X}^{(n)}|\theta_0)\big]^t \int_{\mathbb{R}^p} \exp\big[th^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{t}{2}h^T I_{\theta_0} h\big]\, dh \\
=&(1 + o_{P^n_{\theta_0}}(1)) n^{-p/2}\pi(\theta_0)\big[p_n(\mathbf{X}^{(n)}|\theta_0)\big]^t \exp\big[-\frac{t}{2}\Delta^T_{n,\theta_0} I_{\theta_0}\Delta_{n,\theta_0}\big](2\pi)^{p/2} t^{-p/2} |I_{\theta_0}|^{-1/2}.
\end{aligned}$$

Next we upper bound $\log(p_n(\mathbf{X}^{(n)}|\theta)/p_n(\mathbf{X}^{(n)}|\theta_0))$ for $\|\theta-\theta_0\| \geq M_n/\sqrt{n}$. We have

$$\begin{aligned}
&\max_{\{\theta:\|\theta-\theta_0\|=M_n/\sqrt{n}\}} \log \frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)} = \max_{\{h:\|h\|=M_n\}} \log \frac{p_n(\mathbf{X}^{(n)}|\theta_0+n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)} \\
&\leq \|I_{\theta_0}\Delta_{n,\theta_0}\|M_n - \frac{\lambda_{\min}(I_{\theta_0})}{2}M_n^2 + \max_{\{h:\|h\|=M_n\}} |g_n(h)|,
\end{aligned}$$

where $\lambda_{\min}(I_{\theta_0}) > 0$ is the minimum eigenvalue of I_{θ_0}. Also note that $I_{\theta_0}\Delta_{n,\theta_0}$ is bounded in probability. Hence with probability tending to 1,

$$\max_{\{\theta:\|\theta-\theta_0\|=M_n/\sqrt{n}\}} \log \frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)} \leq -\frac{\lambda_{\min}(I_{\theta_0})}{4}M_n^2.$$

By the concavity of $\log p_n(\mathbf{X}^{(n)}|\theta)$, for $\|\theta-\theta_0\| \geq M_n/\sqrt{n}$,

$$\frac{M_n/\sqrt{n}}{\|\theta-\theta_0\|}\Big(\log p_n(\mathbf{X}^{(n)}|\theta) - \log p_n(\mathbf{X}^{(n)}|\theta_0)\Big) \leq \log p_n\Big(\mathbf{X}^{(n)}\Big|\theta_0 + \frac{M_n/\sqrt{n}}{\|\theta-\theta_0\|}(\theta-\theta_0)\Big) - \log p_n(\mathbf{X}^{(n)}|\theta_0).$$

Thus,

$$\begin{aligned}
\log \frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)} &\leq \frac{\sqrt{n}\|\theta-\theta_0\|}{M_n} \log \frac{p_n\Big(\mathbf{X}^{(n)}\Big|\theta_0 + \frac{M_n/\sqrt{n}}{\|\theta-\theta_0\|}(\theta-\theta_0)\Big)}{p_n(\mathbf{X}^{(n)}|\theta_0)} \\
&\leq \frac{\sqrt{n}\|\theta-\theta_0\|}{M_n}\Big(-\frac{\lambda_{\min}(I_{\theta_0})}{4}M_n^2\Big) \\
&= -\frac{\lambda_{\min}(I_{\theta_0})}{4}\sqrt{n}\|\theta-\theta_0\|M_n.
\end{aligned}$$

Fix an $\epsilon > 0$ such that $\sup_{\|\theta-\theta_0\|<\epsilon} \pi(\theta) < +\infty$. We have

$$\begin{aligned}
&\int_{\{\theta:\|\theta-\theta_0\|>M_n/\sqrt{n}\}} \big[p_n(\mathbf{X}^{(n)}|\theta)\big]^t \pi(\theta)\,d\theta \\
\leq& \big[p_n(\mathbf{X}^{(n)}|\theta_0)\big]^t \int_{\{\theta:\|\theta-\theta_0\|>M_n/\sqrt{n}\}} \exp\Big[-\frac{t\lambda_{\min}(I_{\theta_0})}{4}\sqrt{n}\|\theta-\theta_0\|M_n\Big]\pi(\theta)\,d\theta \\
=& \big[p_n(\mathbf{X}^{(n)}|\theta_0)\big]^t \Big(\int_{\{\theta:M_n/\sqrt{n}\leq\|\theta-\theta_0\|\leq\epsilon\}} \exp\Big[-\frac{t\lambda_{\min}(I_{\theta_0})}{4}\sqrt{n}\|\theta-\theta_0\|M_n\Big]\pi(\theta)\,d\theta \\
&+ \int_{\{\theta:\|\theta-\theta_0\|>\epsilon\}} \exp\Big[-\frac{t\lambda_{\min}(I_{\theta_0})}{4}\sqrt{n}\|\theta-\theta_0\|M_n\Big]\pi(\theta)\,d\theta\Big) \\
\leq& \big[p_n(\mathbf{X}^{(n)}|\theta_0)\big]^t \Big(\big(\sup_{\|\theta-\theta_0\|<\epsilon} \pi(\theta)\big) \int_{\{\theta:\|\theta-\theta_0\|\geq M_n/\sqrt{n}\}} \exp\Big[-\frac{t\lambda_{\min}(I_{\theta_0})}{4}\sqrt{n}\|\theta-\theta_0\|M_n\Big]\,d\theta \\
&+ \exp\Big[-\frac{t\lambda_{\min}(I_{\theta_0})}{4}\epsilon\sqrt{n}M_n\Big]\Big) \\
=& \big[p_n(\mathbf{X}^{(n)}|\theta_0)\big]^t \Big(\big(\sup_{\|\theta-\theta_0\|<\epsilon} \pi(\theta)\big) n^{-p/2} \int_{\{h:\|h\|\geq M_n\}} \exp\Big[-\frac{t\lambda_{\min}(I_{\theta_0})}{4}\|h\|M_n\Big]\,dh \\
&+ \exp\Big[-\frac{t\lambda_{\min}(I_{\theta_0})}{4}\epsilon\sqrt{n}M_n\Big]\Big).
\end{aligned}$$

Thus,

$$
\begin{aligned}
&\frac{\int_{\{\theta:\|\theta-\theta_0\|>M_n/\sqrt{n}\}}\big[p_n(\mathbf{X}^{(n)}|\theta)\big]^t\pi(\theta)\,d\theta}{\int_\Theta\big[p_n(\mathbf{X}^{(n)}|\theta)\big]^t\pi(\theta)\,d\theta}\\
=&O_{P^n_{\theta_0}}(1)\Big(\int_{\{h:\|h\|\geq M_n\}}\exp\Big[-\frac{t\lambda_{\min}(I_{\theta_0})}{4}\|h\|M_n\Big]\,dh+n^{p/2}\exp\Big[-\frac{t\lambda_{\min}(I_{\theta_0})}{4}\epsilon\sqrt{n}M_n\Big]\Big)\\
=&o_{P^n_{\theta_0}}(1).
\end{aligned}
$$

□

Proof of Proposition 2. Note that $L_1(\Theta;\mathbf{X}^{(n)})$ is well defined $P^n_{\theta_0}$-a.s. since it has finite integral

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

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

To prove the second part of the proposition, consider the following example. Suppose $X_1,\ldots,X_n$ are iid from the density

$$
p(X|\theta)=C|X-\theta|^{-\gamma}\exp\big[-(X-\theta)^2\big],
$$

where C is the normalizing constant and $\gamma\in(0,1)$ is a known hyperparameter. The parameter space Θ is equal to $\mathbb{R}$. Then

$$
L_t(\Theta;\mathbf{X}^{(n)})=C^n\int_{-\infty}^{+\infty}\Big[\prod_{i=1}^n|X_i-\theta|\Big]^{-t\gamma}\exp\big[-t\sum_{i=1}^n(X_i-\theta)^2\big]\pi(\theta)\,d\theta.
$$

Note that almost surely, there is no tie among $X_1,\ldots,X_n$. Consequently, $L_t(\Theta;\mathbf{X}^{(n)})=+\infty$ almost surely if and only if $t\geq\gamma^{-1}$. Since $\gamma^{-1}\in(1,+\infty)$, this example shows that $L_t(\Theta;\mathbf{X}^{(n)})$ is not always well defined for $t>1$.

□

Proof of Proposition 3. Note that

$$
\frac{L_t(\{\theta:\|\theta-\theta_0\|\geq\frac{M_n}{\sqrt{n}}\};\mathbf{X}^{(n)})}{L_t(\Theta;\mathbf{X}^{(n)})}=\frac{\displaystyle\int_{\{\theta:\|\theta-\theta_0\|\geq\frac{M_n}{\sqrt{n}}\}}\Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^t\pi(\theta)\,d\theta}{\displaystyle\int_\Theta\Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^t\pi(\theta)\,d\theta}. \tag{16}
$$

Without loss of generality, we assume $M_n/\sqrt{n}\to 0$.

Consider the expactation of the numerator of 16. It follows from Fubini's theorem that

$$
\begin{aligned}
&P_{\theta_0}^n \int_{\{\theta:\|\theta-\theta_0\|\geq \frac{M_n}{\sqrt{n}}\}} \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^t \pi(\theta)\, d\theta \\
=&\int_{\{\theta:\|\theta-\theta_0\|\geq \frac{M_n}{\sqrt{n}}\}} \left\{\int_{\mathcal{X}^n} \big[p_n(\mathbf{X}^{(n)}|\theta)\big]^t \big[p_n(\mathbf{X}^{(n)}|\theta_0)\big]^{1-t}\, d\mu^n\right\} \pi(\theta)\, d\theta \\
=&\int_{\{\theta:\|\theta-\theta_0\|\geq \frac{M_n}{\sqrt{n}}\}} \big[\rho_t(\theta,\theta_0)\big]^n \pi(\theta)\, d\theta \\
=&\int_{\{\theta:\|\theta-\theta_0\|\geq \frac{M_n}{\sqrt{n}}\}} \exp\big[-(1-t)nD_t(\theta||\theta_0)\big]\pi(\theta)\, d\theta.
\end{aligned}
$$

Decompose the integral region into two parts $\{\theta : \frac{M_n}{\sqrt{n}} \leq \|\theta-\theta_0\| \leq \delta\}$ and $\{\theta : \|\theta-\theta_0\| > \delta\}$. Then Assumption 2 implies that

$$
\begin{aligned}
&\int_{\{\theta:\|\theta-\theta_0\|\geq \frac{M_n}{\sqrt{n}}\}} \exp\big[-(1-t)nD_t(\theta||\theta_0)\big]\pi(\theta)\, d\theta \\
=&\left(\int_{\{\theta:\frac{M_n}{\sqrt{n}}\leq\|\theta-\theta_0\|\leq\delta\}} + \int_{\{\theta:\|\theta-\theta_0\|>\delta\}}\right) \exp\big[-(1-t)nD_t(\theta||\theta_0)\big]\pi(\theta)\, d\theta \\
\leq& \max_{\|\theta-\theta_0\|\leq\delta} \pi(\theta) \int_{\left\{\theta:\|\theta-\theta_0\|\geq \frac{M_n}{\sqrt{n}}\right\}} \exp\big[-(1-t)Cn\|\theta-\theta_0\|^2\big]\, d\theta + \exp\big[-(1-t)\epsilon n\big] \\
=&\big(\max_{\|\theta-\theta_0\|\leq\delta} \pi(\theta)\big) n^{-p/2} \int_{\left\{h:\|h\|\geq M_n\right\}} \exp\big[-(1-t)C\|h\|^2\big]\, dh + \exp\big[-(1-t)\epsilon n\big].
\end{aligned}
$$

Hence

$$
n^{p/2} \int_{\{\theta:\|\theta-\theta_0\|\geq \frac{M_n}{\sqrt{n}}\}} \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^t \pi(\theta)\, d\theta = o_{P_{\theta_0}^n}(1). \tag{17}
$$

Now we consider the denominator of (16).

$$
\begin{aligned}
&\int_{\Theta} \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^t \pi(\theta)\, d\theta \geq \int_{\{\theta:\|\theta-\theta_0\|\leq n^{-1/2}\}} \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^t \pi(\theta)\, d\theta \\
\geq&\Big(\min_{\|\theta-\theta_0\|\leq n^{-1/2}} \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^t \pi(\theta)\Big) \int_{\{\theta:\|\theta-\theta_0\|\leq n^{-1/2}\}} 1\, d\theta \\
\geq&\Big(\exp\Big[t \min_{\|\theta-\theta_0\|\leq n^{-1/2}} \log \frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]\Big)\Big(\min_{\|\theta-\theta_0\|\leq n^{-1/2}} \pi(\theta)\Big) n^{-p/2} \frac{2\pi^{p/2}}{\Gamma(p/2)}.
\end{aligned}
$$

By Assumption 1,

$$
\min_{\|\theta-\theta_0\|\leq n^{-1/2}} \log \frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)} \geq -\|I_{\theta_0}\Delta_{n,\theta_0}\| - \frac{1}{2}\|I_{\theta_0}\| + o_{P_{\theta_0}^n}(1),
$$

where $I_{\theta_0}\Delta_{n,\theta_0}$ is bounded in probability. Also note that

$$
\min_{\|\theta-\theta_0\|\leq n^{-1/2}} \pi(\theta) \to \pi(\theta_0) > 0.
$$

Then for every $\epsilon' > 0$, there is a constant $c > 0$ such that with probability at least $1 - \epsilon'$,

$$\int_\Theta \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^t \pi(\theta)\, d\theta \geq c n^{-p/2}.$$

This inequality, together with (17), proves the consistency of $L_t(\cdot;\mathbf{X}^{(n)})$. □

Proof of Theorem 2. By contiguity, we only need to prove the convergence in $P^n_{\theta_0}$. Let $M > 0$ be any fixed number. Assumption 1 implies that

$$\begin{aligned}&\int_{\|h\|\leq M} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^a \pi_n(h;\mathbf{X}^{(n)})\, dh \\ &= \exp[o_{p_0^n}(1)] \int_{\|h\|\leq M} \exp\left[a h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{a}{2} h^T I_{\theta_0} h\right] \pi_n(h;\mathbf{X}^{(n)})\, dh.\end{aligned} \tag{18}$$

Since Δ_{n,θ_0} is bounded in probability, we have

$$\begin{aligned}&\int_{\|h\|\leq M} \exp\left[a h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{a}{2} h^T I_{\theta_0} h\right] \left|\pi_n(h;\mathbf{X}^{(n)}) - \phi(h;\Delta_{n,\theta_0}, b^{-1} I_{\theta_0}^{-1})\right| dh \\ &\leq \sup_{\|h\|\leq M} \exp\left[a h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{a}{2} h^T I_{\theta_0} h\right] \int_{\|h\|\leq M} \left|\pi_n(h;\mathbf{X}^{(n)}) - \phi(h;\Delta_{n,\theta_0}, b^{-1} I_{\theta_0}^{-1})\right| dh \xrightarrow{P^n_{\theta_0}} 0.\end{aligned}$$

This, combined with (18), yields

$$\begin{aligned}&\int_{\|h\|\leq M} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^a \pi_n(h;\mathbf{X}^{(n)})\, dh \\ &= \int_{\|h\|\leq M} \exp\left[a h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{a}{2} h^T I_{\theta_0} h\right] \phi(h;\Delta_{n,\theta_0}, b^{-1} I_{\theta_0}^{-1})\, dh + o_{P^n_{\theta_0}}(1).\end{aligned} \tag{19}$$

This is true for every $M > 0$ and hence also for some $M_n \to \infty$.

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

$$\int_{\|h\|> M_n} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^a \pi_n(h;\mathbf{X}^{(n)})\, dh \xrightarrow{P^n_{\theta_0}} 0. \tag{20}$$

By Assumption 3, for any $\epsilon > 0$, with probability at least $1 - \epsilon$,

$$\int_{\|h\|> M_n} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^a \pi_n(h;\mathbf{X}^{(n)})\, dh \leq \int_{\|h\|> M_n} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^a T(h)\, dh. \tag{21}$$

By Hölder's inequality,

$$P^n_{\theta_0} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^a = \int p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)^a p_n(\mathbf{X}^{(n)}|\theta_0)^{1-a} d\mu^n \leq 1.$$

Hence the expectation of the right hand side of (21) satisfies

$$P^n_{\theta_0} \int_{\|h\|> M_n} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^a T(h)\, dh \leq \int_{\|h\|> M_n} T(h)\, dh \to 0.$$

This verifies (20).

Combining (19) and (20) yields

$$\int_{h\in\mathbb{R}^p}\left[\frac{p_n(\mathbf{X}^{(n)}|\theta_0+n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^a \pi_n(h;\mathbf{X}^{(n)})\,dh = \left(1+\frac{a}{b}\right)^{-p/2}\exp\left(\frac{a}{2}\Delta_{n,\theta_0}^T I_{\theta_0}\Delta_{n,\theta_0}\right)+o_{P_{\theta_0}^n}(1).$$

Similarly, we have

$$\int_{h^{(0)}\in\mathbb{R}^{p_0}}\left[\frac{p_n(\mathbf{X}^{(n)}|\nu_0+n^{-1/2}h^{(0)},\xi_0)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^a \pi_n(h^{(0)};\mathbf{X}^{(n)})\,dh^{(0)} = \left(1+\frac{a}{b}\right)^{-p_0/2}\exp\left(\frac{a}{2}\Delta_{n,\theta_0}^{(0)T} I_{\theta_0}^{(0)}\Delta_{n,\theta_0}^{(0)}\right)+o_{P_{\theta_0}^n}(1).$$

Hence

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

Then the conclusion follows by the same argument as the last part of Theorem 1. □

Appendix B Proofs in Section 3

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

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

By direct calculation, we have

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

Hence $P_{\theta_0}^n\dot{\ell}_{\theta_0}=0_2$ and

$$I_{\theta_0}=\begin{pmatrix}\frac{1}{4} & 0\\ 0 & \frac{1}{8}\end{pmatrix}.$$

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

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

It is known that $\max_{1\le i\le n}|X_i|=O_{P_{\theta_0}^n}(\sqrt{\log n})$. Then by 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

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

On the other hand,

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

Combining these two expansion yields

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

Use Taylor expansion $\log(1+x)=x-x^2/2+O(x^3)$ for $x\in(-1,1)$, we have

$$\begin{aligned}
&\log\frac{p(X_i|\theta_0+n^{-1/2}h)}{p(X_i|\theta_0)}\\
=&\frac{h_1}{2\sqrt{n}}X_i+\left(\frac{h_2}{4\sqrt{n}}-\frac{h_1^2}{4n}\right)(1-X_i^2)+\frac{h_2^2}{16n}X_i^4-\frac{h_2^2}{8n}X_i^2-\frac{h_2^2}{16n}+\frac{3h_1h_2}{4n}X_i-\frac{h_1h_2}{4n}X_i^3\\
&-\frac{h_1^2}{8n}X_i^2-\frac{h_2^2}{32n}(1-X_i^2)^2-\frac{h_1h_2}{8n}X_i(1-X_i^2)+O_{P_{\theta_0}^n}\left(\frac{\log^6 n}{n^{3/2}}\right).
\end{aligned}$$

This expansion is uniform for $\|h\|\le M$ and $i=1,\ldots,n$. Thus,

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

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

Now we verify Assumption 2. We have

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

We have

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

On the one hand,

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

Hence if $(\xi, \tau)^T$ is close enough to $(0,1)^T$, we have

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

On the other hand, for $(\xi, \tau)^T$ close enough to $(0,1)^T$,

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

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

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

For $\sqrt{\xi^2 + (\tau-1)^2} \geq \delta$, we have

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

The Hellinger distance between normal distribution can be explicitly derived. And it can be easily seen that it has a lower bound $\epsilon > 0$ for $\sqrt{\xi^2 + (\sigma^2-1)^2} \geq \delta$. Thus Assumption 2 holds.

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

To prove Theorem 3, the following result is useful.

Proposition 5. *Suppose the conditions of Theorem 3 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* $M_n \geq \sqrt{\log n/(2(t \wedge (1-t)))}$, *we have*

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

Proof.

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

Note that

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

This implies that

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

If $M_n \geq \sqrt{\log n/2(t\wedge(1-t))}$,

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

This completes the proof.

$\square$

Proof of Theorem 3. 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 $i=1,\ldots,n$, 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 $A((\log n)^r)\cap\{\omega\geq n^{-s}\}$, we have $\max_{1\leq i\leq n}|Y_i| = O_{P^n_{\theta_0}}(\sqrt{\log n})$. On $A((\log n)^r)\cap\{\omega\geq n^{-s}\}$, 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 $A((\log n)^r)\cap\{\omega\geq n^{-s}\}$,

$$\sum_{i=1}^n \log\frac{p(X_i|\omega,\xi)}{p(X_i|0,0)} = \omega\xi\sum_{i=1}^n X_i - \frac{1}{2}n\omega^2\xi^2 + o_{P^n_{\theta_0}}(1). \tag{22}$$

As a result,

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

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

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

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

$$\int_{-2\Phi^{-1}\big((\log n)^r/(2\omega\sqrt{n})+1/2\big)}^{2\Phi^{-1}\big((\log n)^r/(2\omega\sqrt{n})+1/2\big)}\exp\big\{t\omega\xi\sum_{i=1}^n X_i-\frac{1}{2}nt\omega^2\xi^2\big\}d\xi=\frac{1}{\omega}\sqrt{\frac{2\pi}{tn}}\exp\Big\{\frac{t}{2n}(\sum_{i=1}^n X_i)^2\Big\}(1+o_{P_{\theta_0}^n}(1)),$$

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:

$$\begin{aligned}&\Pi\big(\omega(2\Phi(|\xi|/2)-1)\leq(\log n)^r n^{-1/2},\omega\leq n^{-s}\big)\\&\leq\Pi\big(\omega(2\Phi(|\xi|/2)-1)\leq(\log n)^r n^{-1/2},\omega\leq 2(\log n)^r n^{-1/2}\big)\\&\quad+\Pi\big(\omega(2\Phi(|\xi|/2)-1)\leq(\log n)^r n^{-1/2},2(\log n)^r n^{-1/2}\leq\omega\leq n^{-s}\big).\end{aligned}$$

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 $(2\Phi(|\xi|/2)-1)\leq\omega^{-1}(\log n)^r n^{-1/2}\leq 1/2$, which implies the boundedness of ξ. It follows that $|\xi|\leq C\omega^{-1}(\log n)^r n^{-1/2}$ for some constant $C>0$ on this event. Thus,

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

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

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

It follows that

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

Combine these arguments and Proposition 5, we have

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

This implies that

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

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 $s < 1/4$, $\omega \ge n^{-s}$ for large n and $\sqrt{n}\omega\xi \to \eta$. Then it follows from (22) and Le Cam's first lemma (Vaart, 1998, Theorem 6.4) that $P^n_{\theta_n}$ and $P^n_{\theta_0}$ are mutually contiguous. As a result,

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

Note that (22) implies that

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

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

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

This proves the conclusion of (ii).

$\square$

References

Aitkin, M. (1991). Posterior bayes factors. journal of the royal statistical society series b-methodological.

Bhattacharya, A., Pati, D., and Yang, Y. (2016). Bayesian fractional posteriors. *ArXiv e-prints*.

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

Bobkov, S. G., Chistyakov, G. P., and Götze, F. (2016). Rényi divergence and the central limit theorem. *ArXiv e-prints*.

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.

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.

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

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

Kleijn, B. and Vaart, A. (2012). The bernstein-von-mises theorem under misspecification. *Electron. J. Stat.*, 6(1):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. 57:99–138.

Pati, D., Bhattacharya, A., and Yang, Y. (2017). On Statistical Optimality of Variational Bayes. *ArXiv e-prints*.

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

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

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

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. (2017). Frequentist Consistency of Variational Bayes. *ArXiv e-prints*.

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*.