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

Sunday 13th May, 2018

Abstract

A general methodology called integrated likelihood ratio test is proposed which takes posterior Bayes factor and fractional Bayes factor as the main test statistics. Different from Bayesian hypothesis testing, we treat the resulting test procedure as frequentist test which controls the significance level. Under certain regular conditions, we prove the Wilks phenomenon of the integrated likelihood ratio test and derive its asymptotic local power. It is shown that the integrated likelihood ratio test shares similar frequency properties as the likelihood ratio test with fewer conditions. Our results hold even if the likelihood is unbounded where the likelihood ratio test can not be defined. We apply the proposed method to two testing problems in normal mixture model. The likelihood ratio test can not be defined for the first problem and has undesirable local power behavior for the second problem. In comparison, we prove that the integrated likelihood ratio test has good asymptotic power behavior in both problems.

KEYWORDS: *Bayes consistency; Bayes factor; Fractional posterior; Hypothesis testing; Mixture model; Rényi divergence; Variational inference.*

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 other hand, the fundamental lemma of Neyman and Pearson tells us that the likelihood ratio test

*Corresponding author

Email address: xuxz@bit.edu.cn

(LRT) is the most powerful test when the null and alternative hypotheses are both simple. For the general hypothesis testing problem

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

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

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

where $L(\theta)$ is the likelihood function, $\hat{\theta}_{\mathrm{MLE}}$ and $\hat{\theta}^{(0)}_{\mathrm{MLE}}$ 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, $2 \log \Lambda_{\mathrm{LRT}}$ converges to $\chi^2(p - p_0)$ in law under the null hypothesis. The LRT has been very successful in many specific problems. For some moderately complex problems, however, some difficulties may arise when using the LRT. For example, 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 is unbounded and hence the LRT is not defined. See, for example, Cam (1990). Note that the unbounded likelihood occurs not only in artifical models, but also in some widely used models, for example, the mixture models with unknown component location and scale (Chen, 2017).

In classical goodness of fit test, there are two common types of tests: one based on the maximum (Kolmogorov-Smirnov test, e.g.) and the other based on the integral (Cramér-von Mises test, e.g.). However, the integral type tests draw little attention to frequentist in parametric hypothesis testing problem. A natural integral type test statistic for hypothesis (1) is defined as

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

where Π and $\Pi^{(0)}$ are some probability measures on Θ and Θ_0, respectively. If $\Pi(d\theta)$ and $\Pi^{(0)}(d\theta)$ don't rely on data, then (2) has exactly the same form as the Bayes factor with Prior distributions Π and $\Pi^{(0)}$. The Bayes factor, proposed by Jeffreys (1931), is the conventional tool for Bayesian hypothesis testing and has been widely used by practitioners (See Kass and Raftery (1995) for a review). Compared with the methods in other Beyasian inference problem, such as point estimation and credible sets, Bayes factor is developed on its own ground and thus has its own nature. A notable feature of Bayes factor is that it can not be obtained solely from the posterior distribution of parameters. There are two consequence of this feature. First, the computation of Bayes factor is highly nontrivial. See Kass and Raftery (1995), Han and Carlin (2001), Raftery et al. (2006) and the references therein. Second, Bayes factor is sensitive to the choice of prior distribution. In fact, if the Bayes factor is treated as a frequentist test statistic, its asymptotic null distribution relies on the prior density evaluated at the true parameter value. See, for example, Clarke and Barron (1990). Thus, to formulate a frequentist test statistic, the measures $\Pi(d\theta)$ and $\Pi^{(0)}(d\theta)$ in (2) should rely on data.

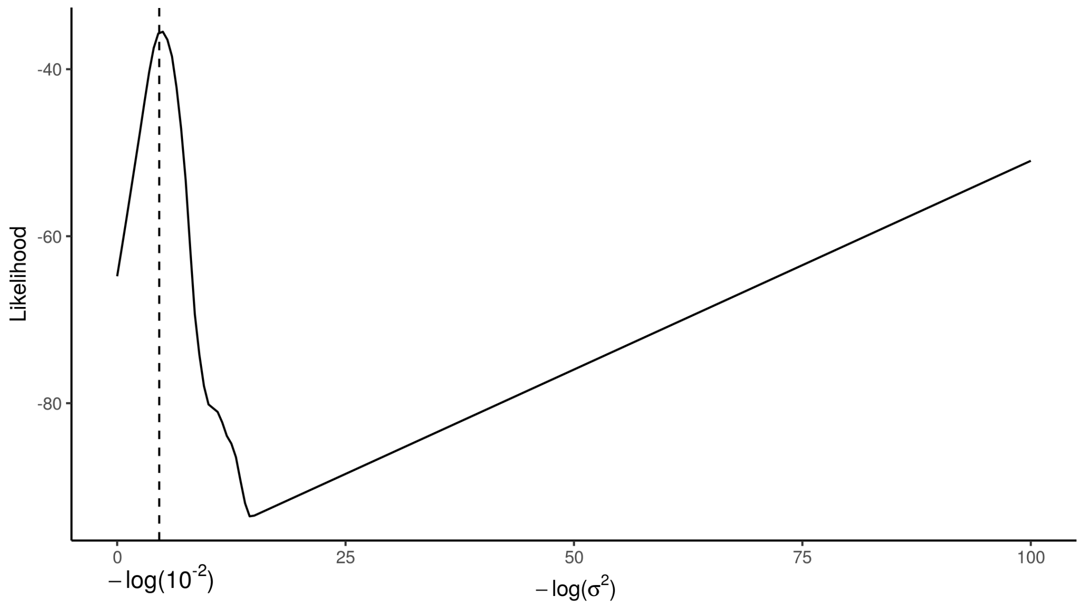

Figure 1: An example of unbounded likelihood. We generate data $X_1, \ldots, X_n$ are iid from the mixture model $(1-\omega)\mathcal{N}(0,1) + \omega\mathcal{N}(\xi, \sigma^2)$ with $(\omega, \xi, \sigma^2)^T = (1/2, 1, 10^{-2})$ and $n = 50$. We plot the likelihood function versus $-\log(\sigma^2)$ with $\omega = 1/2$ and $\xi = X_1$. The likelihood tends to infinity as $-\log(\sigma^2)$ tends to infinity. However, the behavior of the likelihood around the true parameter is fairly well.

If $\Pi(\theta = \hat{\theta}_{\mathrm{MLE}}) = 1$ and $\Pi^{(0)}(\theta = \hat{\theta}^{(0)}_{\mathrm{MLE}}) = 1$, then (2) equals to the LRT statistic. In this case, the measure Π and $\Pi^{(0)}$ both concentrate on one point which are highly nonsmooth. For many models, although 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 behavior. See Figure 1. In such cases, some smoother Π and $\Pi^{(0)}$ may perform better intuitively. Following this idea, a natural choice is to take Π and $\Pi^{(0)}$ as the posterior distribution (with respect to certain predefined prior distribution) 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 $2^{(p-p_0)}\Lambda_{\mathrm{LRT}}$. This implies that PBF has a similar Wilks phenomenon as the LRT.

Note that for any $a > 0$, the LRT is equivalent to

$$\frac{\max_{\theta \in \Theta} L^a(\theta)}{\max_{\theta \in \Theta_0} L^a(\theta)}.$$

In the contrary, 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 (2). We shall 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, for example, 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 (2) is the fractional Bayes factor (FBF) proposed by O'Hagan (1995). We call (2) 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 power of the resulting test procedure under contiguous alternative. It is shown that the generalized FBF has similar asymptotic behavior as 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 produce a distribution other than the fractional posterior distribution. To accommodate such cases, we also give a theorem 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 nonconvex 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 asympototic local power. In Section 3, we apply ILRT to testing the homogeneity in a two-component normal mixture model. Section 4 concludes the paper. All technical proves 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 with 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, a 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 conventional Bayes factor is defined as

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

where $\tilde{\Theta}_0 = \{\nu : (\nu^T, \xi^T)^T \in \Theta_0\}$. However, Bayes factor is sensitive to the specification of prior, which may cause difficulties in the absense of a well-formulated subjective prior. See, for example, Shafer (1982). To deal with this problem, Aitkin (1991) proposed PBF which is defined to be

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

where $\pi(\nu|\mathbf{X}^{(n)})$ and $\pi(\theta|\mathbf{X}^{(n)})$ are the posterior densities under the null and alternative hypothesis, respectively. O'Hagan (1995) proposed FBF which is defined as

$$\frac{L_1}{L_b} \cdot \frac{L_b^{(0)}}{L_1^{(0)}} \quad \text{for} \quad 0 < b < 1,$$

where for $t > 0$,

$$L_t = \int_\Theta \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^t \pi(\theta)\, d\theta, \quad L_t^{(0)} = \int_{\Theta_0} \left[p_n(\mathbf{X}^{(n)}|\nu, \xi_0)\right]^t \pi(\nu)\, d\nu.$$

We generalize the PBF and FBF and propose the ILRT statistic as

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

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 $\tilde{\Theta}_0$ respectively. Note that $\pi(\theta; \mathbf{X}^{(n)})$ and $\pi(\nu; \mathbf{X}^{(n)})$ may be data dependent but does not need to be the posterior density. If we take the weight function as

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

then the ILRT statistic equals to

$$\Lambda_{a,b} = \frac{L_{a+b}}{L_b} \cdot \frac{L_b^{(0)}}{L_{a+b}^{(0)}}.$$

We shall 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 $\Lambda_{1,b}$, the PBF equals to $\Lambda_{2,1}$.

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

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

For some moderately complex models, (4) may be complicated. Consequently, sampling from (4) may be intractable. In this case, one may use some simple form weight function to approximate (4). A popular method for approximating (4) is variational inference. See. See, for example, Blei et al. (2017). In this case, the weight function in (10) is equals to the variational apprixomation of (4). The ILRT methodology also includes such approximate method.

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 space Θ and $\tilde{\Theta}_0$ are open subsets of $\mathbb{R}^p$ and $\mathbb{R}^{p_0}$, respectively. The parameter 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. For every $M > 0$,

$$\sup_{\|h\|\leq M} \left| \log \frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)} - h^T I_{\theta_0}\Delta_{n,\theta_0} + \frac{1}{2}h^T I_{\theta_0} h \right| \xrightarrow{P_{\theta_0}^n} 0,$$

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

For $t > 0$, we say L_t is $\sqrt{n}$-consistent if for every $M_n \to \infty$,

$$\frac{L_t(\{\theta : \|\theta - \theta_0\| > M_n/\sqrt{n}\})}{L_t} \xrightarrow{P_{\theta_0}^n} 0,$$

where for a set $A \subset \Theta$, $L_t(A) = \int_A [p_n(\mathbf{X}^{(n)}|\theta)]^t \pi(\theta)\, d\theta$. The $\sqrt{n}$-consistency of $L_t^{(0)}$ is similarly defined. Note that the consistency of L_1 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 is a key assumption of the following theorem.

Theorem 1. *Suppose that Assumption 1 holds, L_{a+b}, L_b, $L_{a+b}^{(0)}$ and $L_b^{(0)}$ 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 $\sqrt{n}(\theta_n - \theta_0) \to \eta$,*

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

where $\chi^2(p-p_0,\delta)$ is a noncentral chi-squared random variable with $p-p_0$ degrees of freedom and noncentrality parameter $\delta = \eta^T\left(I_{\theta_0} - I_{\theta_0}J(J^T I_{\theta_0}J)^{-1}J^T I_{\theta_0}\right)\eta$ 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 obtain 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 is a key assumption of Theorem 1. Hence we would like to give sufficient conditions for the $\sqrt{n}$-consistency of L_t. First we consider the exponential family.

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 is consistent for $t > 0$.

Proposition 1 establishes the $\sqrt{n}$-consistent of L_t 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. In this case, Theorem 1 implies that the generalized FBF can be used as frequentist test. However, for any test methodology, the success in the full-rank exponential family models is just a minimal requirement since for these models, 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 convex. 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 Cam (1990) for more examples. A natural question is that if the fractional integrated likelihood L_t is always well defined. The following theorem shows that L_t is always well defined for $t \le 1$ and is not well defined for some model for $t > 1$.

Proposition 2. *If $t \leq 1$, $L_t < +\infty$ $P^n_{\theta_0}$-a.s. for any models. If $t > 1$, $L_t = +\infty$ for some models.*

Because of the bad behavior of L_t for $t > 1$, next we only consider L_t for $t \leq 1$. For $t = 1$, the $\sqrt{n}$-consistency of L_t is equivalent to the $\sqrt{n}$-consistency of the posterior distribution. The consistency of posterior distribution has drawn much attention in the literature. See, for example, 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).

Assumption 2. *For every $\epsilon > 0$, there exists a sequence of tests ϕ_n such that*

$$P^n_{\theta_0}\phi_n \to 0, \qquad \sup_{\|\theta-\theta_0\|\geq\epsilon} P^n_\theta(1-\phi_n) \to 0.$$

Assumption 2 is satisfied when the parameter space is compact and the model is suitably continuous. See Theorem 3.2 of Kleijn and Vaart (2012).

Proposition 3 (Kleijn and Vaart (2012), Theorem 3.1)**.** *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, L_1 is consistent.*

The consistency of L_t for $0 < t < 1$ is different from $t = 1$. Walker and Hjort (2001) considered the Hellinger consistency of $L_{1/2}$. It is shown that the consistency of $L_{1/2}$ does not need Assumption 2. However, they only consider $t = 1/2$ and didn't consider the $\sqrt{n}$-convergence result. 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 for $0 < t < 1$. However, their result can not yield the $\sqrt{n}$-consistency for parametric models. We shall prove the $\sqrt{n}$-consisency of L_t 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 3. *For some $\alpha \in (0,1)$, there exist positive constancts 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, for example, Bobkov et al. (2016). As a result, if Assumption 3 holds for some $\alpha \in (0,1)$, then it will hold for every $\alpha \in (0,1)$.

To appreciate Assumption 3, suppose, for example, that $D_\alpha(\theta||\theta_0)$ is twice continuously differentiable in θ. Since $\theta = \theta_0$ is a minumum 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 3 is a fairly weak condition.

Proposition 4. *Suppose θ_0 is an interior of Θ, $\pi(\theta)$ is continuous at θ_0 and $\pi(\theta_0) > 0$. Under Assumptions 1 and 3, for fixed $t \in (0, 1)$, L_t is consistent.*

Compared with Assumption 2, it appears that Assumption 3 is easier to verify. Note that the asymptotic power of $\Lambda_{a,b}$ is independent of a, b. Hence 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 this case, we can use simpler weight functions to approximate (4).

Let $h = \sqrt{n}(\theta - \theta_0)$ be the local parameter. Then the posterior density of h is $\pi_n(h|\mathbf{X}^{(n)}) = n^{-1/2}\pi_n(\theta|\mathbf{X}^{(n)})$. Theorem 2.1 of Kleijn and Vaart (2012) states that under Assumptions 1 and 2,

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

where $\Delta_{n,\theta_0} = n^{-1/2}\sum_{i=1}^n I_{\theta_0}^{-1}\dot{\ell}_{\theta_0}(X_i)$ and 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. We shall assume that the weight function inherits this property.

Assumption 4. *Let $b \in (0, 1)$ be a fixed number. Assume that $\pi_n(h; \mathbf{X}^{(n)})$ satisfies*

$$\|\pi_n(h; \mathbf{X}^{(n)}) - \phi(h; \Delta_{n,\theta_0}, b^{-1}I_{\theta_0}^{-1})\| \xrightarrow{P_{\theta_0}^n} 0. \tag{6}$$

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

$$\|\pi_n(h^{(0)}; \mathbf{X}^{(n)}) - \phi(h^{(0)}; \Delta_{n,\theta_0}^{(0)}, b^{-1}I_{\theta_0}^{(0)-1})\| \xrightarrow{P_{\theta_0}^n} 0, \tag{7}$$

where

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

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

$$\lim_{n\to\infty} P_{\theta_0}^n \left\{ \sup_{h\in\mathbb{R}^p} (\pi_n(h;\mathbf{X}^{(n)}) - T(h)) \le 0 \right\} \ge 1-\epsilon. \tag{8}$$

$$\lim_{n\to\infty} P_{\theta_0}^n \left\{ \sup_{h^{(0)}\in\mathbb{R}^{p_0}} (\pi_n(h^{(0)};\mathbf{X}^{(n)}) - T^{(0)}(h^{(0)})) \le 0 \right\} \ge 1-\epsilon. \tag{9}$$

The conditions (8) and (9) in Assumption 4 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 undersirable 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, for example, 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 (6) and (7) implies (8) and (9).

Under Assumption 4, we consider the ILRT statistic

$$\Lambda_{a,b}^* = \frac{\int_\Theta \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^a \pi(\theta;\mathbf{X}^{(n)})\, d\theta}{\int_{\tilde{\Theta}_0} \left[p_n(\mathbf{X}^{(n)}|\nu,\xi_0)\right]^a \pi(\nu;\mathbf{X}^{(n)})\, d\nu}. \tag{10}$$

The following theorem gives the asymptotic distribution of ILRT statistic.

Theorem 2. *Suppose that Assumptions 1 and 4 hold, the true parameter θ_0 is an interior point of Θ, ν is a relative interior point of $\tilde{\Theta}_0$. Suppose $a+b\le 1$. 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, for example, Blei et al. (2017). Next we shall consider a simple variational method which is guaranteed to yield a weight function satisfying Assumption 4. 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 $\pi_n(\theta;\mathbf{X}^{(n)})$ be the fractional posterior of order b and $\pi_n(h;\mathbf{X}^{(n)}) = n^{-1/2}\pi_n(\theta_0+n^{-1/2}h;\mathbf{X}^{(n)})$ be the corresponding fractional posterior of h. Suppose that $\pi_n(h;\mathbf{X}^{(n)})$ satisfies Bernstein-von Mises theorem,

$$\|\pi_n(h;\mathbf{X}^{(n)}) - \phi(h;\Delta_{n,\theta_0}, b^{-1}I_{\theta_0}^{-1})\| \xrightarrow{P_{\theta_0}^n} 0. \tag{11}$$

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

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

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

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

Hence we have

$$-\frac{1}{1-\alpha} \log \int \pi^{\dagger}(h; \mathbf{X}^{(n)})^{\alpha} \pi_n(h; \mathbf{X}^{(n)})^{1-\alpha}\, dh \le -\frac{1}{1-\alpha} \log \int \phi(h; \Delta_{n,\theta_0}, I_{\theta_0}^{-1})^{\alpha} \pi_n(h|\mathbf{X}^{(n)})^{1-\alpha}\, dh. \tag{12}$$

Since Rényi divergence and total variation distance are equivalent, (11) implies that the right hand side of (12) tends to 0 in $P_{\theta_0}^n$-probability. Again by the equivalence of Rényi divegence and total variation distance, we have

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

Since $\pi_n(h; \mathbf{X}^{(n)})$ is a normal density, (6) implies the mean and covariance parameter of $\pi_n(h; \mathbf{X}^{(n)})$ converges to Δ_{n,θ_0} and $I_{\theta_0}^{-1}$ respectively. Then a dominating function $T(h)$ exists and thus (8) 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 bounded and thus likelihood ratio test is not defined (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 $0 \le \omega \le 1$, $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}^+$. First, we assume $\omega = 1/2$ and consider testing the hypotheses

$$H: \xi = 0, \sigma = 1 \quad \text{vs.} \quad K: \xi \ne 0 \text{ or } \sigma \ne 1. \tag{13}$$

For this testing problem, the likelihhood 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 4, we can obtain the following proposition.

Proposition 5. *For hypotheses testing problem* (13)*, 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*

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

This example shows that every 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 comlicated. For simplicity, we assume $\sigma^2 = 1$ is known and consider testing the hypotheses

$$\omega\xi = 0 \quad \text{vs.} \quad \omega\xi \neq 0. \tag{14}$$

Even for this simple case, the likelihood has undesirable behavior due to loss of identifiability. Although the LRT exists in this problem, 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, theorems in Section 2 can not be directly applied. This is because the second part is Assumption (4) 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*(α_1, α_2) *with* $\alpha_1 > 1$*. Suppose* $a + b < 1$*. Then,*

(i) under the null hypothesis,

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

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

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

Theorem 3 shows that the ILRT has nontrival 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 $|\gamma| < 1$. To have a nontrivial power,

4 Conclusion

In this paper, we proposed a flexible methodology ILRT which includes some existing method as special cases. We gave the asymptotic distribution of the generalized FBF, which is a special case of ILRT. We also investigates the asymptotic behavior of the ILRT for general weight functions. It is shown that the generalized FBF has a Wilks phenomenon similar to the LRT. The asymptotic local power is also given. We also apply the ILRT methodology to two testing problems for 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 integral can smooth the likelihood. hence it can be expected that the ILRT method can have better properties than the LRT when the likilihood has complicated behavior. The success of the ILRT methodology in our examples verifies this point. It is interesting to apply the ILRT methodology to more problems where classical methods don't have good properties. 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

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

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

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

Thus

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

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

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

Thus,

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

If L_t is consistent, then

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

Similarly, if $L_t^{(0)}$ is consistent,

$$
\begin{aligned}
&\log L_t^{(0)} = \log\int_{\tilde\Theta_0} \big[p_n(\mathbf{X}^{(n)}|\nu,\xi_0)\big]^t \pi(\nu)\,d\nu \\
=&\frac{p_0}{2}\log\big(\frac{2\pi}{n}\big) - \frac{p_0}{2}\log t - \frac{1}{2}\log|I^{(0)}_{\theta_0}| + \log\pi(\nu_0) + \frac{t}{2}\Delta_{n,\theta_0}^{(0)T} I^{(0)}_{\theta_0}\Delta^{(0)}_{n,\theta_0} + t\log p_n(\mathbf{X}^{(n)}|\theta_0) + o_{P^n_{\theta_0}}(1).
\end{aligned}
$$

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

$$
\begin{aligned}
\log\Lambda_{a,b} =& \log L_{a+b} - \log L_b - \log L^{(0)}_{a+b} + \log L^{(0)}_b \\
=& -\frac{p-p_0}{2}\log(1+\frac{a}{b}) + \frac{a}{2}\Big(\Delta_{n,\theta_0}^T I_{\theta_0}\Delta_{n,\theta_0} - \Delta_{n,\theta_0}^{(0)T} I^{(0)}_{\theta_0}\Delta^{(0)}_{n,\theta_0}\Big) + o_{P^n_{\theta_n}}(1).
\end{aligned}
$$

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

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

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

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

Hence by Le Cam's third lemma,

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

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

$$\Delta_{n,\theta_0}^T I_{\theta_0} \Delta_{n,\theta_0} - \Delta_{n,\theta_0}^{(0)T} I_{\theta_0}^{(0)} \Delta_{n,\theta_0}^{(0)} \overset{P_{\eta_n}^n}{\rightsquigarrow} \chi^2(p - p_0, \delta),$$

which completes the proof.

$\square$

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.

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

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 is well defined $P^n_{\theta_0}$-a.s. since it has finite integral

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

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

To prove the second part of the proposition, consider the following example. Suppose $X_1,\dots,X_n$ 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

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

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

□

Proof of Proposition 4. Note that

$$
\frac{L_t(\{\theta:\|\theta-\theta_0\|\geq\frac{M_n}{\sqrt{n}}\})}{L_t}=\frac{\displaystyle\int_{\{\theta:\|\theta-\theta_0\|\geq\frac{M_n}{\sqrt{n}}\}}\Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^t\pi(\theta)\,d\theta}{\displaystyle\int_\Theta\Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^t\pi(\theta)\,d\theta}. \tag{15}
$$

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

Consider the expactation of the numerator of 15. It follows from Fubini's theorem that

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

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 3 implies that

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

Hence

$$
n^{p/2} \int_{\{\theta:\|\theta-\theta_0\|\geq \frac{M_n}{\sqrt{n}}\}} \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^t \pi(\theta)\, d\theta = o_{P_{\theta_0}^n}(1). \tag{16}
$$

Now we consider the denominator of (15).

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

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 $1 - \epsilon'$,

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

This inequality, together with (16), proves the consistency of L_t. □

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

$$\begin{aligned}
&\int_{\|h\|\leq M} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^a \pi_n(h; \mathbf{X}^{(n)})\, dh \\
&= \exp[o_{p_0^n}(1)] \int_{\|h\|\leq M} \exp\big[ah^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{a}{2}h^T I_{\theta_0} h\big]\pi_n(h; \mathbf{X}^{(n)})\, dh.
\end{aligned} \tag{17}$$

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

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

This, combined with (17), yields

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

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

$$\int_{\|h\|> M_n} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^a \pi_n(h; \mathbf{X}^{(n)})\, dh \xrightarrow{P_{\theta_0}^n} 0. \tag{19}$$

By Assumption 4, for any $\epsilon > 0$, with probability at least $1 - \epsilon$,

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

By Hölder's inequality,

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

Hence the expectation of the right hand side of (20) satisfies

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

This verifies (19).

Combining (18) and (19) 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 5. We shall verify Assumption 1 and Assumption 3. 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

$$I_{\theta_0}=\begin{pmatrix}\frac14 & 0\\ 0 & \frac18\end{pmatrix}.$$

Let $M>0$ is a fixed constant. For $h=(h_1,h_2)^T\in\mathbb{R}^2$ such that $\|h\|\le M$ and $i=1,\dots,n$, 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 $i=1,\dots,n$, 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 $i=1,\dots,n$. 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 3. 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

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

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

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

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

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

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 3 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 6. *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

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

This implies that

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

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

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

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 $\xi\ne0$ and $Y_i=X_i$ if $\xi=0$.

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

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

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 $A((\log n)^r)\cap\{\omega\geq n^{-s}\}$,

$$\sum_{i=1}^n \log\frac{p(X_i|\omega,\xi)}{p(X_i|0,0)} = \omega\xi\sum_{i=1}^n X_i - \frac{1}{2}n\omega^2\xi^2 + o_{P^n_{\theta_0}}(1). \tag{21}$$

As a result,

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

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

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

By direct calculation, we have

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

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 $A((\log n)^r)\cap\{\omega\leq n^{-s}\}$. 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 $(2\Phi(|\xi|/2)-1)\leq\omega^{-1}(\log n)^r n^{-1/2}\leq 1/2$, 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

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

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 6, 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

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

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 (21) 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 (21) 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

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

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). R\'enyi divergence and the central limit theorem. *ArXiv e-prints*.

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

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.

Han, C. and Carlin, B. P. (2001). Markov chain monte carlo methods for computing bayes factors. *Journal of the American Statistical Association*, 96(455):1122–1132.

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.

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*.

Raftery, A. E., Newton, M. A., Satagopan, J. M., and Krivitsky, P. N. (2006). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity.

Shafer, G. (1982). Lindley's paradox. *Journal of the American Statistical Association*, 77(378):325–334.

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*.