Integrated likelihood ratio test[☆]

author[1]

Radarweg 29, Amsterdam

Elsevier Inc[a,b], Global Customer Service[b,]*

[a]1600 John F Kennedy Boulevard, Philadelphia
[b]360 Park Avenue South, New York

Abstract

Likelihood ratio test is the most widely used test procedure. However, it has some weaknesses. Likelihood is unbounded for some important models. Even when the likelihood is bounded, the maximum may be not easy to obtain if it is not convex in parameters. Based on existing work on Bayesian hypothesis testing, we propose a new test procedure called integrated likelihood ratio test which shares the same asymptotic properties as that of likelihood ratio test. The proposed methodology is very flexible which takes posterior Bayes factor and fractional Bayes factor as special cases. It can also be used in the model where the posterior distribution is difficult to compute.

Keywords: Bayes consistency, Bayes factor, hypothesis testing

---

1. Introduction

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 paramters. 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 even in the large sample setting. In contrast, it is well known that the posterior distribution tends to become independent of the prior distribution as the sample size increases.

Several modifications of Bayes factor have been proposed. Aitkin (1991) proposed the posterior Bayes factor (PBF) which integrated the likelihood with respect to the posterior distribution. Another approach uses a portion of data as training sample. A posterior is computed using the training sample and then be used to calculate Bayes factor. Berger and Pericchi (1996) proposed the intrinsic Bayes factor by using all possible training samples of minimal size and averaging the resulting Bayes factor. Unfortunately, . O'Hagan (1995) found that the training sample approximates to the full likelihood raised to a fractional power. They called the resulting statistic an fractional Bayes factor (FBF). Compared with the Bayes factor, these testing methods are less sensitive to the prior distribution.

[☆]Fully documented templates are available in the elsarticle package on CTAN.

[*]Corresponding author
Email address: support@elsevier.com (Global Customer Service)
URL: www.elsevier.com (Elsevier Inc)
[1]Since 1880.

Existing research on the frequentist properties of Bayesian method is largely concerned with the consistency and asymptotic normality of the posetrior distribution. These results can be applied to Bayes point estimation and interval estimation. However, these testing methods lack a rigorous frequentist evaluation.

In this paper, we are concerned about the frequentist evaluations of Bayesian hypothesis testing.

Let $\mathbf{X}^{(n)} = (X_1, \dots, X_n)$ be independent identically distributed (i.i.d.) observations from a family parametrized by $\theta = (\nu^T, \xi^T)^T$, with $\dim(\theta) = p$ and $\dim(\nu) = p_0$. We would like to test the hypotheses

$$H_0 : \theta \in \Theta_0 \quad \text{v.s.} \quad H_1 : \theta \in \Theta,$$

where Θ is an open subset of $\mathbb{R}^p$ and Θ_0 is a p_0-dimensional subspace of Θ defined as

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

Bayesians put 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, several modifications of Bayes factor have been proposed. Aitkin (1991) proposed posterior Bayes factor (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 fractional Bayes factor (FBF) which is defiend to be

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

where for $t > 0$,

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

In this paper, we generalize the PBF and FBF and propose the integrated likelihood ratio test (ILRT) statistic, as follow

$$\frac{\int_{\Theta} \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^a \pi(\theta; \mathbf{X}^{(n)})\, d\theta}{\int_{\tilde{\Theta}_0} \left[p_n(\mathbf{X}^{(n)}|\nu, \xi_0)\right]^a \pi(\nu; \mathbf{X}^{(n)})\, d\nu}, \tag{1}$$

where $a > 0$ are hyperparameters, $\pi(\theta; \mathbf{X}^{(n)})$ and $\pi(\nu; \mathbf{X}^{(n)})$ are the weight 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

$$\pi(\theta; \mathbf{X}^{(n)}) = \frac{\left[p_n(\mathbf{X}^{(n)}|\theta)\right]^b \pi(\theta)}{\int_{\Theta} \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^b \pi(\theta)\, d\theta}, \tag{2}$$

then the ILRT statistic equals

$$\Lambda_{a,b}(\mathbf{X}^{(n)}) = \frac{L_{a+b}}{L_b} \cdot \frac{L_b^*}{L_{a+b}^*}.$$

We shall call $\Lambda_{a,b}(\mathbf{X}^{(n)})$ the generalized FBF throughout the paper. The FBF and PBF are both the special cases of the generalized FBF. In fact, the FBF is equal to $\Lambda_{1,b}(\mathbf{X}^{(n)})$, the PBF is equal to $\Lambda_{2,1}(\mathbf{X}^{(n)})$.

The ILRT methodology is very flexible. For some models, the quantity L_t is hard to compute. In this case, (2) may be complicated. Consequently, one may choose to use some simple form weight function, for example, the distribution obtained from variational inference.

The ILRT is computational feasible. Compared with LRT. Compared with Bayes factor.

The paper is organized as follow. In Section 2, we investegate the asymptotic properties of the generalzied FBF. Section 3 consider the ILRT with general weight function. Section 4 concludes the paper. All technical proves are in Apppendix.

2. Generalized FBF

2.1. setup

Let $\mathbf{X}^{(n)} = (X_1, \dots, X_n)$ be independent identically distributed (i.i.d.) observations with values in some space 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. 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 nested hypotheses

$$H_0 : \theta \in \Theta_0 \quad \text{v.s.} \quad \theta \in \Theta,$$

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.

Denote by P_θ^n the joint distribution of $\mathbf{X}^{(n)} = (X_1, \dots, X_n)$. 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.

Denote by P_θ^n the joint distribution of $\mathbf{X}^{(n)} = (X_1, \dots, X_n)$. 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.

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 this section, we study the asymptotic behavior of the generalized FBF.

Let $\Delta_{n,\theta_0} = \frac{1}{\sqrt{n}} \sum_{i=1}^n I_{\theta_0}^{-1} \dot{\ell}_{\theta_0}(X_i)$ be the 'locally sufficient' statistics. The corresponding quantities in the null space are

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

The following assumption is adapted from Kleijn and Vaart (2012).

Assumption 1. *The parameter space Θ is an open subset of $\mathbb{R}^p$. The null space $\tilde{\Theta}_0$ is an open subset of $\mathbb{R}^{p_0}$. The parameter θ_0 is an inner point of Θ, ν_0 is an inner point of $\tilde{\Theta}_0$. The function $\theta \mapsto \log p(X|\theta)$ is differentiable at θ_0 P_{θ_0}-a.s. with derivative*

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

There's an open neighborhood V of θ_0 such that for every $\theta_1, \theta_2 \in V$,

$$|\log p(X|\theta_1) - \log p(X|\theta_2)| \le m(X)\|\theta_1 - \theta_2\|,$$

where $m(X)$ is a measurable function satisfying $P_0 \exp[sm(X)] < \infty$ for some $s > 0$. The Fisher information matrix $I_{\theta_0} = P_{\theta_0} \dot{\ell}_{\theta_0} \dot{\ell}_{\theta_0}^T$ is positive-definite and as $\theta \to \theta_0$,

$$P_{\theta_0} \log \frac{p(X|\theta)}{\log(X|\theta_0)} = -\frac{1}{2}(\theta - \theta_0)^T I_{\theta_0} (\theta - \theta_0) + o(\|\theta - \theta_0\|^2).$$

Assumption 1 is satisfied by many common models, it ensures a local asymptotic normality expansion of likelihood. See Lemma 1 in Appendix.

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

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

where for a set $A \subset \Theta$,

$$L_t(A) = \int_A \Big[p_n(\mathbf{X}^{(n)}|\theta)\Big]^t \pi(\theta)\, d\theta.$$

The $\sqrt{n}$-consistency of L^*_t is defined similarly. 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} and L^*_b 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 $\sqrt{n}(\theta_n - \theta_0) \to \eta$,*

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

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 $2\log\Lambda_{a,b}(\mathbf{X}^{(n)})$ under the null hypothesis and the local alternative hypothesis. To obtain a test with asymptotic level α, the critical value of $2\log\Lambda_{a,b}(\mathbf{X}^{(n)})$ can be defined to be $-(p-p_0)\log(1+a/b) + a\chi^2_{p-p_0,1-\alpha}$, where $\chi^2_{p-p_0,1-\alpha}$ is the $1-\alpha$ quantile of a chi-squared random variable with $p - p_0$ degrees of freedom. The resulting test has local asymptotic power

$$\Pr\big(\chi^2_{p-p_0}(\delta) > \chi^2_{p-p_0,1-\alpha}\big). \tag{3}$$

It is known that, under certain regular conditions, (3) is also the local asymptotic power of the likelihood ratio test. In this view, $\Lambda_{a,b}(\mathbf{X}^{(n)})$ enjoys good frequentist properties.

The $\sqrt{n}$-consistency of L_t plays a key role in the proof of 1. We would like to give sufficient conditions for the $\sqrt{n}$-consistency of L_t. The following proposition shows that for full-rank exponential family, L_t is $\sqrt{n}$-consistent for all $t > 0$.

Proposition 1. *Suppose $p(X|\theta) = \exp\big[\theta^TT(X) - A(\theta)\big]$, 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 $t > 0$.

In general case, however, the $\sqrt{n}$-consistency of L_t needs further conditions. For $t = 1$, the $\sqrt{n}$-consistency of L_t is equivalent to the $\sqrt{n}$-consistency of posterior distribution. The consistency of posterior distribution have been considerable attention in the literature. See, for example, Ghosal et al. (2000), Shen and Wasserman (2001), van der Vaart and Ghosal (2007) and the references therein. 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*

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

Proposition 2 (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.*

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).

The consistency of L_t for $0 < t < 1$ is different from the consistency of posterior distribution. Walker and Hjort (2001) considered the Hellinger consistency of $L_{1/2}$. However, they only consider $t = 1/2$ and didn't prove the $\sqrt{n}$-convergence result. Next we shall prove the 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 will be assumed in 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

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

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 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 3, for fixed $t \in (0,1)$, L_t is consistent.*

The consistency of L_t for $t > 1$ can be proved under conditions similar to Assumption 2. However, while we require $\{\phi_n\}$ to be consistent tests, the requirement on the sequence $\{\phi_n\}$ for $t > 1$ lacks statistical interpretation. This implies that it may not be natural to use L_t for $t > 1$.

Note that L_1 is always well defined since it has finite integral. By holder inequality, L_t is also well defined for $0 < t < 1$. However, L_t is not always well defined. The following example is a counterexample.

Example 1. *Suppose $X_1, \ldots, X_n$ are i.i.d. from the density*

$$p(x|\theta) = C|x-\theta|^{-1/2}\exp\big[-(x-\theta)^2\big],$$

where C is the normalizing constant. The parameter space Θ is equal to $\mathbb{R}$. We would like to test the hypotheses $H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$. The likelihood is

$$p_n(\mathbf{X}^{(n)}|\theta) = C^n\Big[\prod_{i=1}^n |X_i - \theta|\Big]^{-1/2}\exp\big[-\sum_{i=1}^n (X_i-\theta)^2\big].$$

Under the alternative hypothesis, the likelihood tends to infinity if θ tends to X_i, $i = 1, \dots, n$. Consequently, LRT fails in this model. We impose a prior $\pi(\theta)$. Suppose that $\pi(\theta)$ is positive for all θ. Then

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

The likelihood will almost surely have no ties and consequently $L_t(\mathbf{X}^{(n)}) = +\infty$ if and only if $t \geq 2$.

Based on our theoretical result and this example, we suggest to use FBF with $a + b \leq 1$.

3. General weight function

In some cases, the posterior density or the general FBF is not easy to calculate or have unsatisfactory properties. Thanks to the flexibility of ILRT, we can consider general weight function in such cases.

Let $h = \sqrt{n}(\theta - \theta_0)$. Kleijn and Vaart (2012), Theorem 2.1 states that under Assumption 1, 2,

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

where for two density p and q, $\|p - q\| = \int |p - q|$ is the total variation distance between p and q. We shall assume that the weight function inherits this property.

Assumption 4. *Let $\pi_n(h; \mathbf{X}^{(n)})$ be a weight function satisfying*

$$\|\pi_n(h; \mathbf{X}^{(n)}) - \phi(h; \Delta_{n,\theta_0}, I_{\theta_0}^{-1})\| \overset{P_{\theta_0}^n}{\to} 0 \tag{5}$$

Furthermore, assume that for every $\epsilon > 0$, there's a Lebesgue integrable function $T(h)$, a $K > 0$ and an $A > 0$ such that

$$\lim_{n\to\infty} P_{\theta_0}^n\big(\sup_{\|h\|\geq K\sqrt{n}} (\pi_n(h; \mathbf{X}^{(n)}) - T(h)) \leq 0\big) \geq 1 - \epsilon \tag{6}$$

$$\lim_{n\to\infty} P_{\theta_0}^n\big(\sup_{\|h\|\leq K\sqrt{n}} \pi_n(h; \mathbf{X}^{(n)}) \leq A\big) \geq 1 - \epsilon \tag{7}$$

The condition 6 assumes there is a function controlling the tail of weight function. For a statistical model, the likelihood value makes no sense when θ is far away from θ_0, or $\sqrt{n}h$ is large. The bad behavior of the tail of likelihood function may affect the behavior of posterior distribution. To avoid the bad behavior of the likelihood function when $\sqrt{n}h$ is large, most existing literatures impose conditions on the model. Here we impose 6 on weight function instead. The condition 7 is satisfied in most usual case.

Theorem 2. *Suppose the true parameter θ_0 is an interior point of Θ, ν is a relative interior point of $\tilde{\Theta}_0$. Under Assumptions 1, 2 and4, for bounded real numbers η_n, we have*

$$2\log(\Lambda(X)) \overset{P_{\eta_n}^n}{\rightsquigarrow} -(p_2 - p_1)\log 2 + \chi^2_{p_2-p_1}(\delta)$$

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). The following example shows that the weight function obtained from Rényi divergence variational inference satisfies Assumption 4.

Example 2. *Suppose $\pi_n(h; \mathbf{X}^{(n)})$ is obtain from Rényi divergence variational inference (Li and Turner, 2016):*

$$\pi_n(h; \mathbf{X}^{(n)}) = \min_{q\in\mathcal{Q}} -\frac{1}{1-\alpha}\log\int_{\mathcal{X}} q(h)^{\alpha}\pi(h|\mathbf{X}^{(n)})^{1-\alpha}\, d\mu,$$

where $\mathcal{Q}$ is the family of all p dimensional normal distribution. Since

$$-\frac{1}{1-\alpha}\log\int_{\mathcal{X}}\pi(h;\mathbf{X}^{(n)})^{\alpha}\pi(h|\mathbf{X}^{(n)})^{1-\alpha}\,d\mu \leq -\frac{1}{1-\alpha}\log\int_{\mathcal{X}}\phi(h;\Delta_{n,\theta_0},I_{\theta_0}^{-1})^{\alpha}\pi(h|\mathbf{X}^{(n)})^{1-\alpha}\,d\mu. \quad (8)$$

By the equivalence of Rényi divergence and total variation distance and Bernstein-von Mise theorem, the right hand side of (8) *tends to* 0*. Again by the equivalence of Rényi divegence and total variation distance,* (5) *holds. Since $\pi_n(h;\mathbf{X}^{(n)})$ is a normal density,* (5) *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* (6) *and* (7) *hold.*

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 FPF, which is a special case of ILRT. We also investigates the asymptotic behavior of ILRT for general weight functions. This allows one to use a simple form approximation of the posterior distribution as weight function. In particular, we show that the weight function can be obtained from Rényi divergence variational inference.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant No. 11471035, 11471030.

Appendices

For two measure sequence P_n and Q_n on measurable spaces $(\Omega_n,\mathcal{A}_n)$, denote by $P_n \lhd\rhd Q_n$ that P_n and Q_n are mutually contiguous. That is, for any statistics $T_n\colon \Omega_n \mapsto \mathbb{R}^k$, we have $T_n \overset{P_n}{\rightsquigarrow} 0 \Leftrightarrow T_n \overset{Q_n}{\rightsquigarrow} 0$.

Lemma 1 (Kleijn and Vaart (2012), Lemma 2.1.)**.** *Under Assumption 1, we have $\|\dot{\ell}_{\theta_0}(X)\| \leq m(X)$ P_0-a.s., $P_0\dot{\ell}_{\theta_0}(X)=0$ and for every $M>0$*

$$\sup_{\|h\|\leq M}\left|\log\frac{p_n(\mathbf{X}^{(n)}|\theta_0+n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)} - h^T I_{\theta_0}\Delta_{n,\theta_0} + \frac{1}{2}h^T I_{\theta_0}h\right| \xrightarrow{P_0^n} 0.$$

Lemma 2. *Under Assumptions 1 and 2, there exists for every $M_n\to\infty$ a sequence of tests ϕ_n and a constant $\delta>0$ such that, for every sufficiently large n and every $\|\theta-\theta_0\|\geq M_n/\sqrt{n}$,*

$$P_0^n\phi_n \to 0, \quad P_\theta^n(1-\phi_n) \leq \exp[-\delta n(\|\theta-\theta\|^2\wedge 1)].$$

(See der Vaart (2000) Lemma 10.3., Kleijn and Vaart (2012))

Appendix A Proofs in Section 2

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

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

By Proposition 1,

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

Thus

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

This equality holds for every $M > 0$ and hence also for some $M_n \to \infty$. Note that Δ_{n,θ_0} is bounded in probability. Hence

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

Thus,

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

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

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

Similarly, if $L^*_t(\mathbf{X}^{(n)})$ is consistent,

$$
\begin{aligned}
&\log L^*_t(\mathbf{X}^{(n)}) = \log \int_{\tilde{\Theta}_0} \big[p_n(\mathbf{X}^{(n)}|\nu,\xi_0)\big]^t \pi(\nu)\, d\nu \\
=&\frac{p_1}{2}\log\big(\frac{2\pi}{n}\big) - \frac{p_1}{2}\log t - \frac{1}{2}\log|I^*_{\theta_0}| + \log\pi(\nu_0) + \frac{t}{2}\Delta^{*T}_{n,\theta_0} I^*_{\theta_0}\Delta^*_{n,\theta_0} + t\log p_n(\mathbf{X}^{(n)}|\theta_0) + o_{P^n_{\theta_0}}(1).
\end{aligned}
$$

These expansions, combined with the mutually contiguity of $P^n_{\theta_0}$ and $P^n_{\theta_n}$, yield

$$
\begin{aligned}
\log \Lambda_{a,b}(\mathbf{X}^{(n)}) =& \log L_a(\mathbf{X}^{(n)}) - \log L_b(\mathbf{X}^{(n)}) - \log L^*_a(\mathbf{X}^{(n)}) + \log L^*_b(\mathbf{X}^{(n)}) \\
=& -\frac{p-p_1}{2}\log\frac{a}{b} + \frac{a-b}{2}\Big(\Delta^T_{n,\theta_0} I_{\theta_0}\Delta_{n,\theta_0} - \Delta^{*T}_{n,\theta_0} I^*_{\theta_0}\Delta^*_{n,\theta_0}\Big) + o_{P^n_{\theta_n}}(1).
\end{aligned}
$$

Note that

$$I_{\theta_0}^* = J^T I_{\theta_0} J, \quad \Delta_{n,\theta_0}^* = (J^T I_{\theta_0} J)^{-1} J^T I_{\theta_0} \Delta_{n,\theta_0}.$$

Then

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

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 $p - p_1$.

Now we need to derive the asymptotic distribution of Δ_{n,θ_0}. Let $h_n = \sqrt{n}(\theta_n - \theta_0)$. By Proposition 1 and CLT,

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

Hence by Le Cam's third lemma,

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

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

$$\Delta_{n,\theta_0}^T I_{\theta_0} \Delta_{n,\theta_0} - \Delta_{n,\theta_0}^{*T} I_{\theta_0}^* \Delta_{n,\theta_0}^* \overset{P_{\eta_n}^n}{\rightsquigarrow} \chi_{p-p_1}^2(\delta).$$

This completes the proof.

□

Proof of Proposition 1. By some algebra, we have

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

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 $M_n^3/\sqrt{n} \to 0$. Then by Taylor's theorem and the continuity of the third derivative of $A(\theta)$,

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

Then

$$\begin{aligned}
&\int_\Theta \big[p_n(\mathbf{X}^{(n)}|\theta)\big]^t \pi(\theta)\, d\theta \ge \int_{\{\theta: \|\theta - \theta_0\| \le M_n/\sqrt{n}\}} \big[p_n(\mathbf{X}^{(n)}|\theta)\big]^t \pi(\theta)\, d\theta \\
&= (1 + o_{P_0^n}(1)) n^{-p/2} \pi(\theta_0) \big[p_n(\mathbf{X}^{(n)}|\theta_0)\big]^t \int_{\{h: h \le M_n\}} \exp\big[t h^T I_{\theta_0} \Delta_{n,\theta_0} - \frac{t}{2} h^T I_{\theta_0} h\big]\, dh \\
&= (1 + o_{P_0^n}(1)) n^{-p/2} \pi(\theta_0) \big[p_n(\mathbf{X}^{(n)}|\theta_0)\big]^t \int_{\mathbb{R}^p} \exp\big[t h^T I_{\theta_0} \Delta_{n,\theta_0} - \frac{t}{2} h^T I_{\theta_0} h\big]\, dh \\
&= (1 + o_{P_0^n}(1)) n^{-p/2} \pi(\theta_0) \big[p_n(\mathbf{X}^{(n)}|\theta_0)\big]^t \exp\big[-\frac{t}{2} \Delta_{n,\theta_0}^T I_{\theta_0} \Delta_{n,\theta_0}\big] (2\pi)^{p/2} t^{-p/2} |I_{\theta_0}|^{-1/2}.
\end{aligned}$$

We have

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

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,

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

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

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

Thus,

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

For $\epsilon>0$ such that $\sup_{\|\theta-\theta_0\|<\epsilon}\pi(\theta)\le+\infty$, we have

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

Thus,

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

$\square$

Proof of Proposition 3. Note that

$$\frac{L_t(\{\theta : \|\theta - \theta_0\| \geq \frac{M_n}{\sqrt{n}}\})}{L_t} = \frac{\int_{\{\theta:\|\theta-\theta_0\|\geq \frac{M_n}{\sqrt{n}}\}} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^t \pi(\theta)\, d\theta}{\int_\Theta \left[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^t \pi(\theta)\, d\theta}. \tag{9}$$

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

Consider the expactation of the numerator of 9. It follows from Fubini's theorem that

$$\begin{aligned}
&P_0^n \int_{\{\theta:\|\theta-\theta_0\|\geq \frac{M_n}{\sqrt{n}}\}} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^t \pi(\theta)\, d\theta \\
=& \int_{\{\theta:\|\theta-\theta_0\|\geq \frac{M_n}{\sqrt{n}}\}} \left\{\int_{\mathcal{X}^n} \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^t \left[p_n(\mathbf{X}^{(n)}|\theta_0)\right]^{1-t} d\mu^n\right\} \pi(\theta)\, d\theta \\
=& \int_{\{\theta:\|\theta-\theta_0\|\geq \frac{M_n}{\sqrt{n}}\}} \left[\rho_t(\theta,\theta_0)\right]^n \pi(\theta)\, d\theta \\
=& \int_{\{\theta:\|\theta-\theta_0\|\geq \frac{M_n}{\sqrt{n}}\}} \exp\left[-(1-t)nD_t(\theta||\theta_0)\right]\pi(\theta)\, d\theta.
\end{aligned}$$

Decompose the integral region into two parts $\{\theta : \frac{M_n}{\sqrt{n}} \leq \|\theta - \theta_0\| \leq \delta\}$ and $\{\theta : \|\theta - \theta_0\| > \delta\}$,

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

Now we consider the denominator of (9).

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

By Proposition 1,

$$\min_{\|\theta-\theta_0\|\leq n^{-1/2}} \log\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)} \geq -\|I_{\theta_0}\Delta_{n,\theta_0}\| - \frac{1}{2}\|I_{\theta_0}\| + o_{P_0^n}(1).$$

Since $I_{\theta_0}\Delta_{n,\theta_0}$ is bounded in probability,

$$\min_{\|\theta-\theta_0\|\leq n^{-1/2}} \log\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}$$

is lower bounded in probability. Note that

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

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

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

Combining the upper bound and the lower bound yields that with probability at least $1-\epsilon'$,

$$\begin{aligned}&\frac{L_t(\{\theta : \|\theta-\theta_0\| \ge \frac{M_n}{\sqrt{n}}\})}{L_t}\\ \le& c^{-1}\big(\max_{\|\theta-\theta_0\|\le\delta} \pi(\theta)\big)\int_{\{h:\|h\|\ge M_n\}} \exp\big[-(1-t)C\|h\|^2\big]\, d\theta + c^{-1}n^{p/2}\exp\big[-(1-t)\epsilon n\big] \to 0.\end{aligned}$$

Since ϵ is arbitrary, the theorem follows. $\square$

Appendix B Proofs in Section 3

Proof of Theorem 2. By contiguity, we only need to prove the convergence in P_0^n.

The proof consists of two steps. In the first part of the proof, let M be a fixed positive number. We prove

$$\left|\int_{\|h\|\le M} \frac{p_n(\mathbf{X}^{(n)}|\theta_0+n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\pi_n(h;\mathbf{X}^{(n)})\, dh - \int_{\|h\|\le M} \exp\big[h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0} h\big]\phi(h;\Delta_{n,\theta_0}, I_{\theta_0}^{-1})\, dh\right| \xrightarrow{P_0^n} 0 \tag{10}$$

Propostion 1 implies that

$$\int_{\|h\|\le M} \frac{p_n(\mathbf{X}^{(n)}|\theta_0+n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\pi_n(h;\mathbf{X}^{(n)})\, dh = \exp[o_{p_0^n}(1)]\int_{\|h\|\le M} \exp\big[h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0} h\big]\pi_n(h;\mathbf{X}^{(n)})\, dh \tag{11}$$

So we only need to consider $\int_{\|h\|\le M} \exp\big[h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0} h\big]\pi_n(h;\mathbf{X}^{(n)})\, dh$. By central limit theorem, Δ_{n,θ_0} weakly converges to a normal distribution. As a result, $\sup_{\|h\|\le M} \exp\big[h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0} h\big]$ is bounded in probability. It follows that

$$\begin{aligned}&\int_{\|h\|\le M} \exp\big[h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0} h\big]\big|\pi_n(h;\mathbf{X}^{(n)}) - \phi(h;\Delta_{n,\theta_0}, I_{\theta_0}^{-1})\big|\, dh\\ \le& \sup_{\|h\|\le M} \exp\big[h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0} h\big]\int_{\|h\|\le M} \big|\pi_n(h;\mathbf{X}^{(n)}) - \phi(h;\Delta_{n,\theta_0}, I_{\theta_0}^{-1})\big|\, dh \xrightarrow{P_0^n} 0.\end{aligned}$$

This, combined with (11), proves (10). This is true for every M and hence also for some $M_n \to \infty$.

In the second part, we prove

$$\psi(M) \overset{def}{=} \frac{\int_{\|h\|>M} p_n(\mathbf{X}^{(n)}|\theta_0+n^{-1/2}h)\pi_n(h;\mathbf{X}^{(n)})\, dh}{\int p_n(\mathbf{X}^{(n)}|\theta_0+n^{-1/2}h)\pi_n(h;\mathbf{X}^{(n)})\, dh} \xrightarrow{P_0^n} 0. \tag{12}$$

Let ϕ_n be a test function satisfying the conclusion of Lemma 2. We have

$$\psi(M) = \psi(M)\phi_n + \psi(M)(1-\phi_n).$$

Since $\psi(M) \le 1$, $\psi(M)\phi_n \le \phi_n \xrightarrow{P_0^n} 0$. So it's enough to prove

$$\psi(M)(1-\phi_n) \xrightarrow{P_0^n} 0$$

Fix a positive number U. Then

$$\psi(M)(1-\phi_n) \le \frac{\int_{\|h\|>M_n} p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)\pi_n(h;\mathbf{X}^{(n)})\,dh}{\int_{\|h\|\le U} p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)\pi_n(h;\mathbf{X}^{(n)})\,dh}(1-\phi_n). \tag{13}$$

First we give a lower bound of $\int_{\|h\|\le U} p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)\pi_n(h;\mathbf{X}^{(n)})\,dh$. Note that

$$\begin{aligned}
&\int_{\|h\|\le U} p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)\pi_n(h;\mathbf{X}^{(n)})\,dh \\
=&\exp[o_{P_0^n}(1)]p_n(\mathbf{X}^{(n)}|\theta_0)\int_{\|h\|\le U} \exp\big[h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0}h\big]\pi_n(h;\mathbf{X}^{(n)})\,dh \\
\ge&\exp[o_{P_0^n}(1)]p_n(\mathbf{X}^{(n)}|\theta_0)\Big\{\int_{\|h\|\le U} \exp\big[h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0}h\big]\phi(h;\Delta_{n,\theta_0}, I_{\theta_0}^{-1})\,dh \\
&- \sup_{\|h\|\le U}\exp\big[h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0}h\big]\int_{\|h\|\le U}\big|\pi_n(h;\mathbf{X}^{(n)}) - \phi(h;\Delta_{n,\theta_0}, I_{\theta_0}^{-1})\big|\,dh\Big\} \\
=&\exp[o_{P_0^n}(1)]p_n(\mathbf{X}^{(n)}|\theta_0)\Big\{\int_{\|h\|\le U} \exp\big[h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0}h\big]\phi(h;\Delta_{n,\theta_0}, I_{\theta_0}^{-1})\,dh - O_P(1)o_P(1)\Big\}.
\end{aligned}$$

Fix an $\epsilon > 0$. Since Δ_{n,θ_0} is uniformly tight, with probability at least $1-\epsilon/2$, $|\Delta_{n,\theta_0}| \le C$ for a constant C. If this event happens, we have

$$\int_{\|h\|\le U} \exp\big[h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0}h\big]\phi(h;\Delta_{n,\theta_0}, I_{\theta_0}^{-1})\,dh > 2c$$

for some $c > 0$. Thus, there is a $c > 0$ and an event $D_{1,n}$ with probability at least $1-\epsilon$ on which

$$\int_{\|h\|\le U} p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)\pi_n(h;\mathbf{X}^{(n)})\,dh \ge c p_n(\mathbf{X}^{(n)}|\theta_0)$$

for sufficiently large n.

On the other hand, by Assumption 4, there is a $K > 0$, a $A > 0$ and an event $D_{2,n}$ with probability at least $1-\epsilon$ on which

$$\sup_{\|h\|>K\sqrt{n}}(\pi_n(h;\mathbf{X}^{(n)}) - T(h)) \le 0, \qquad \sup_{\|h\|\le K\sqrt{n}}\pi_n(h;\mathbf{X}^{(n)}) \le A$$

for sufficiently large n.

Combining these bounds yields

$$\psi(M)(1-\phi_n) \le \frac{\int_{\|h\|>M_n} p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)\big(A\mathbf{1}_{M_n\le\|h\|\le K\sqrt{n}} + T(h)\mathbf{1}_{\|h\|>K\sqrt{n}}\big)\,dh}{c p_n(\mathbf{X}^{(n)}|\theta_0)}(1-\phi_n) + \mathbf{1}\{D_{1,n}^C \cup D_{2,n}^C\}.$$

Hence for sufficiently large n,

$$\begin{aligned}
&P_0^n\psi(M)(1-\phi_n) \\
\le& c^{-1}\int_{\mathcal{X}^n}\int_{\|h\|>M_n}(1-\phi_n)p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)\big(A\mathbf{1}_{M_n\le\|h\|\le K\sqrt{n}} + T(h)\mathbf{1}_{\|h\|>K\sqrt{n}}\big)\,dh\,d\mu^n + 2\epsilon \\
=& c^{-1}\int_{\|h\|>M_n}\Big(\int_{\mathcal{X}^n}(1-\phi_n)p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)\,d\mu^n\Big)\big(A\mathbf{1}_{M_n\le\|h\|\le K\sqrt{n}} + T(h)\mathbf{1}_{\|h\|>K\sqrt{n}}\big)\,dh + 2\epsilon \\
\le& c^{-1}\int_{\|h\|>M_n}\exp\big[-\delta(\|h\|^2 \wedge n)\big]\big(A\mathbf{1}_{M_n\le\|h\|\le K\sqrt{n}} + T(h)\mathbf{1}_{\|h\|>K\sqrt{n}}\big)\,dh + 2\epsilon.
\end{aligned}$$

Note that $\delta(\|h\|^2 \cap n) \geq \delta^*(\|h\|^2 \wedge K^2 n)$, where $\delta^* = \delta \min(1, K^{-2})$. Hence

$$\begin{aligned}
&\int_{\|h\|>M_n} \exp\big[-\delta(\|h\|^2 \wedge n)\big]\big(A\mathbf{1}_{M_n \leq \|h\| \leq K\sqrt{n}} + T(h)\mathbf{1}_{\|h\|>K\sqrt{n}}\big)\,dh \\
\leq &\int_{\|h\|>M_n} \exp\big[-\delta^*(\|h\|^2 \wedge K^2 n)\big]\big(A\mathbf{1}_{M_n \leq \|h\| \leq K\sqrt{n}} + T(h)\mathbf{1}_{\|h\|>K\sqrt{n}}\big)\,dh \\
\leq &A\int_{\|h\|\geq M_n} e^{-\delta^*\|h\|^2}\,dh + e^{-\delta^* K^2 n}\int_{\|h\|>K\sqrt{n}} T(h)\,dh \to 0.
\end{aligned}$$

Therefore $\psi(M) \xrightarrow{P_0^n} 0$.

Finally we have

$$\begin{aligned}
&\left|\int \frac{p(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p(\mathbf{X}^{(n)}|\theta_0)}\pi_n(h;\mathbf{X}^{(n)})\,dh - 2^{-\frac{p}{2}}\exp\big[\frac{1}{2}\Delta_{n,\theta_0}^T I_{\theta_0}\Delta_{n,\theta_0}\big]\right| \\
&= \left|\int \frac{p(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p(\mathbf{X}^{(n)}|\theta_0)}\pi_n(h;\mathbf{X}^{(n)})\,dh - \int_{\|h\|\leq M_n} \frac{p(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p(\mathbf{X}^{(n)}|\theta_0)}\pi_n(h;\mathbf{X}^{(n)})\,dh\right| \\
&+ \left|\int_{\|h\|\leq M_n} \frac{p(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p(\mathbf{X}^{(n)}|\theta_0)}\pi_n(h;\mathbf{X}^{(n)})\,dh - \int_{\|h\|\leq M_n} \exp\big[h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0}h\big]\phi(h;\Delta_{n,\theta_0}, I_{\theta_0}^{-1})\,dh\right| \\
&+ \left|\int_{\|h\|\leq M_n} \exp\big[h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0}h\big]\phi(h;\Delta_{n,\theta_0}, I_{\theta_0}^{-1})\,dh - 2^{-\frac{p}{2}}\exp\big[\frac{1}{2}\Delta_{n,\theta_0}^T I_{\theta_0}\Delta_{n,\theta_0}\big]\right| \\
&= J_1 + J_2 + J_3
\end{aligned}$$

By the first step of the proof, we have $J_2 \xrightarrow{P_0^n} 0$. Hence

$$\int_{\|h\|\leq M_n} \frac{p(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p(\mathbf{X}^{(n)}|\theta_0)}\pi_n(h;\mathbf{X}^{(n)})\,dh$$

is bounded in probability. Therefore

$$\begin{aligned}
J_1 &= \int_{\|h\|\leq M_n} \frac{p(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p(\mathbf{X}^{(n)}|\theta_0)}\pi_n(h;\mathbf{X}^{(n)})\,dh \left|\frac{\int p(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)\pi_n(h;\mathbf{X}^{(n)})\,dh}{\int_{\|h\|\leq M_n} p(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)\pi_n(h;\mathbf{X}^{(n)})\,dh} - 1\right| \\
&= O_{P_0^n}(1)o_{P_0^n}(1)
\end{aligned}$$

And J_3 convenges to 0 for trivial reason.

Then we can apply the argument to both the numerator and denominator of integrated likelihood ratio statistics. By CLT,

$$I_{\theta_0}\Delta_{n,\theta_0} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \overset{P_0^n}{\rightsquigarrow} \xi, \tag{14}$$

where $\xi \sim N(0, I_{\theta_0})$.

$$I_{\theta_0}^*\Delta_{n,\theta_0}^* = \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_{\theta_0}^*(X_i) \overset{P_0^n}{\rightsquigarrow} \xi^*, \tag{15}$$

where ξ^* is the first p_1 coordinates of ξ. Hence

$$\begin{aligned}
\Lambda(X) &= \frac{2^{-\frac{p_2}{2}}\exp\{\frac{1}{2}\Delta_{n,\theta_0}^T I_{\theta_0}\Delta_{n,\theta_0}\} + o_{P_0^n}(1)}{2^{-\frac{p_1}{2}}\exp\{\frac{1}{2}\Delta_{n,\theta_0}^{*T} I_{\theta_0}^*\Delta_{n,\theta_0}^*\} + o_{P_0^n}(1)} \\
&\overset{P_0^n}{\rightsquigarrow} \frac{2^{-\frac{p_2}{2}}\exp\{\frac{1}{2}\xi^T I_{\theta_0}^{-1}\xi\}}{2^{-\frac{p_1}{2}}\exp\{\frac{1}{2}\xi^{*T} I_{\theta_0}^{*-1}\xi^*\}}.
\end{aligned} \tag{16}$$

But

$$\xi^T I_{\theta_0}^{-1}\xi - \xi^{*T} I_{\theta_0}^{*-1}\xi^* = (I_{\theta_0}^{-\frac{1}{2}}\xi)^T \left(I_{p_2\times p_2} - I_{\theta_0}^{\frac{1}{2}} \begin{pmatrix} I_{\theta_0}^{*-1} & 0 \\ 0 & 0 \end{pmatrix} I_{\theta_0}^{\frac{1}{2}}\right)(I_{\theta_0}^{-\frac{1}{2}}\xi). \tag{17}$$

$I_{\theta_0}^{-\frac{1}{2}}\xi$ is a p_2-dimensional standard normal distribution, The middle term is a projection matrix with rank $p_2 - p_1$. Hence we have

$$2\log(\Lambda(X)) \overset{P_0^n}{\rightsquigarrow} \chi^2_{p_2-p_1} - (p_2 - p_1)\log(2). \tag{18}$$

□

References

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

Berger JO, Pericchi LR. The intrinsic bayes factor for model selection and prediction. Journal of the American Statistical Association 1996;91(433):109–22. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1996.10476668. doi:10.1080/01621459.1996.10476668. arXiv:http://www.tandfonline.com/doi/pdf/10.1080/01621459.1996.10476668.

Blei DM, Kucukelbir A, McAuliffe JD. Variational inference: A review for statisticians. arxiv 2017;.

Bobkov SG, Chistyakov GP, Götze F. R\'enyi divergence and the central limit theorem. ArXiv e-prints 2016;arXiv:1608.01805.

Ghosal S, Ghosh JK, van der Vaart AW. Convergence rates of posterior distributions. Ann Statist 2000;28(2):500–31. URL: https://doi.org/10.1214/aos/1016218228. doi:10.1214/aos/1016218228.

Han C, Carlin BP. Markov chain monte carlo methods for computing bayes factors. Journal of the American Statistical Association 2001;96(455):1122–32. URL: https://doi.org/10.1198/016214501753208780. doi:10.1198/016214501753208780. arXiv:https://doi.org/10.1198/016214501753208780.

Jeffreys H. Scientific Inference. 1st ed. Cambridge: Cambridge University Press, 1931.

Kass RE, Raftery AE. Bayes factors. Journal of the American Statistical Association 1995;90(430):773–95.

Kleijn B, Vaart A. The bernstein-von-mises theorem under misspecification. Electron J Stat 2012;6(1):354–81.

Li Y, Turner RE. Rényi divergence variational inference. In: Lee DD, Sugiyama M, Luxburg UV, Guyon I, Garnett R, editors. Advances in Neural Information Processing Systems 29. Curran Associates, Inc.; 2016. p. 1073–81. URL: http://papers.nips.cc/paper/6208-renyi-divergence-variational-inference.pdf.

O'Hagan A. Fractional bayes factors for model comparison 1995;57:99–138.

Raftery AE, Newton MA, Satagopan JM, Krivitsky PN. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity 2006;.

Shafer G. Lindley's paradox. Journal of the American Statistical Association 1982;77(378):325–34. doi:10.1080/01621459.1982.10477809. arXiv:http://amstat.tandfonline.com/doi/pdf/10.1080/01621459.1982.10477809.

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

der Vaart A. Asymptotic Statistics. ???: Cambridge university press, 2000.

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

Walker SG, Hjort NL. On bayesian consistency. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2001;63(4):811–21. URL: http://kar.kent.ac.uk/10563/.