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 (LRT) 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. We propose a new test procedure called integrated likelihood ratio test (ILRT) which can overcome the above difficulties. Posterior Bayes factor is a special case of ILRT. We proof the Wilks phenomenon of ILRT and give the asymptotic local power.

Keywords:

1. Introduction

Suppose that we have n observations $\mathbf{X}^{(n)} = (X_1, \ldots, X_n)$ which are independent identically distributed (i.i.d.) random variables with values in some space space $(\mathcal{X}; \mathscr{A})$. Assume 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 some set Θ.

Suppose we are interested in testing the hypotheses $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta$ for a subset Θ_0 of Θ. The well known likelihood ratio test (LRT) is defined as

$$\frac{\sup_{\Theta} p_n(\mathbf{X}^{(n)}|\theta)}{\sup_{\Theta_0} p_n(\mathbf{X}^{(n)}|\theta)}, \tag{1}$$

where $p_n(\mathbf{X}^{(n)}|\theta) = \prod_{i=1}^{n} p(X_i|\theta)$ is the density of $\mathbf{X}^{(n)}$ with respect to μ^n, the n-fold product measure of μ. LRT is the most widely used statistical method which enjoys many optimal properties. For example, by Neyman-Pearson lemma, it's the most powerful test (MPT) in simple null and simple alternative case (Lehmann J. P. R, 2005). In multi-dimensional parameter case, MPT does not exist. Nevertheless, the LRT is asymptotic optimal in the sense of Bahadur efficiency (Bahadur, 1971). However, even in some widely used models, likelihood may be unbounded. See Cam (1990) for some examples. In this case, LRT does not exist. Another weakness of LRT occurs when the likelihood is not convex in parameters. In this case, numerical algorithms for maximizing likelihood may trap in local maxima.

[☆]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.

In Bayesian framework, Bayes factor is the most popular methodology. However, the frequency property of Bayes factor is not satisfactory. Several modifications of Bayes factor have been proposed. See, for example, xxxxxx. Among them, Aitkin (1991) proposed posterior Bayes factor (PBF)

$$\Lambda_P(\mathbf{X}^{(n)}) = \frac{\int_\Theta p(\mathbf{X}^{(n)}|\theta)\pi(\theta|\mathbf{X}^{(n)})\,d\theta}{\int_{\Theta_0} p(\mathbf{X}^{(n)}|\theta)\pi^*(\theta|\mathbf{X}^{(n)})\,d\theta},$$

where $\pi^*(\theta|\mathbf{X}^{(n)})$ and $\pi(\theta|\mathbf{X}^{(n)})$ are the posterior densities under null hypotheses and alternative hypothesis. For $t > 0$, define $z_t(\mathbf{X}^{(n)}) = \int_\Theta \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^t \pi(\theta)\,d\theta$, $z_t^*(\mathbf{X}^{(n)}) = \int_{\Theta_0} \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^t \pi^*(\theta)\,d\theta$. Then PBF can be written as

$$\Lambda_P(\mathbf{X}^{(n)}) = \frac{z_2(\mathbf{X}^{(n)})}{z_1(\mathbf{X}^{(n)})} \cdot \frac{z_1^*(\mathbf{X}^{(n)})}{z_2^*(\mathbf{X}^{(n)})}.$$

where Gelfand D. K. D (1993) derived the null distribution of PBF. However, they didn't explicitly give the conditions needed. In fact, their proof relies on Laplace approximation, which assumes the existence of maximum likelihood estimator (MLE). Note that the existence of MLE implies the existence of LRT. Hence the scope of their method doesn't exceed that of classical LRT.

O'Hagan (1995) proposed the fractional Bayes factor (FBF)

$$\Lambda_F(\mathbf{X}^{(n)}) = \frac{z_1(\mathbf{X}^{(n)})}{z_{1/2}(\mathbf{X}^{(n)})} \cdot \frac{z_{1/2}^*(\mathbf{X}^{(n)})}{z_1^*(\mathbf{X}^{(n)})}.$$

The idea of fractional likelihood is also adopted by Walker and Hjort (2001). We will see that FBF has a wider applicable scope than PBF.

Both PBF and FBF is a special case of the general ILRT.

Based on the proof of Bernstein-von Mises theorem (See der Vaart (2000) and Kleijn and Vaart (2012)), we give the proof of the Wilks phenomenon and local power of ILRT under fairly weak assumptions.

2. Integrated likelihood ratio test

The posterior Bayes factor can be generalized to the integrated likelihood ratio test (ILRT) statistic, as follow

$$\Lambda(\mathbf{X}^{(n)}) = \frac{\int_\Theta \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^a \pi(\theta;\mathbf{X}^{(n)})\,d\theta}{\int_{\Theta_0} \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^a \pi_0(\theta;\mathbf{X}^{(n)})\,d\theta}, \tag{2}$$

where $a > 0$ is a hyperparameter, $\pi(\theta; X)$ and $\pi^*(\theta; X)$ are weight functions which may be data dependent but does not need to be the posterior density of θ.

Let $z_t(\mathbf{X}^{(n)}) = \int_\Theta \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^t \pi(\theta)\,d\theta$. Similarly define z_t^*. If

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

then

$$\Lambda(\mathbf{X}^{(n)}) = \frac{z_{a+b}(\mathbf{X}^{(n)})}{z_b(\mathbf{X}^{(n)})} \cdot \frac{z_b^*(\mathbf{X}^{(n)})}{z_{a+b}^*(\mathbf{X}^{(n)})}.$$

The case $a = b = 1/2$ corresponds to the fractional Bayes factor (FBF) (O'Hagan, 1995). The case $a = b = 1$ corresponds to the posterior Bayes factor (PBF).

The parameter space Θ is an open subset of $\mathbb{R}^{p_2}$. The null space Θ_0 is a p_1-dimensional subspace of Θ

$$\Theta_0 = \{\theta \in \Theta : \theta_{p_1+1} = \theta_{0,p_1+1}, \ldots, \theta_{p_2} = \theta_{0,p_2}\}, \tag{3}$$

where the last $p_2 - p_1$ parameters $\theta_{0,p_1+1}, \dots, \theta_{0,p_2}$ are fixed. We want to test the hypothesis

$$H_0 : \theta \in \Theta_0 \quad vs. \quad H_1 : \theta \in \Theta. \tag{4}$$

The first p_1 parameters are nuisance parameters.

Θ_0 can be regarded as a open subset of $\mathbb{R}^{p_1}$. To simplify notations, we denote $\tilde{\Theta}_0 = \{(\theta_1, \dots, \theta_{p_1})^T : (\theta_1, \dots, \theta_{p_1}, \theta_{0,p_1+1}, \theta_{0,p_2}) \in \Theta_0\}$. We use p_1-dimensional vector $\tilde{\theta} \in \tilde{\Theta}_0$ to represent $\theta \in \Theta_0$ and regard $\tilde{\Theta}_0$ as the null space. Let $\pi(\theta; \mathbf{X})$ and $\tilde{\pi}(\tilde{\theta}; \mathbf{X})$ be the weight functions in Θ and $\tilde{\Theta}_0$. The integrated likelihood ratio statistic is defined as

$$\Lambda(\mathbf{X}^{(n)}) = \frac{\int_\Theta p_n(\mathbf{X}^{(n)}|\theta)\pi(\theta; \mathbf{X}^{(n)})\, d\theta}{\int_{\tilde{\Theta}_0} p_n(\mathbf{X}^{(n)}|\tilde{\theta})\tilde{\pi}(\tilde{\theta}; \mathbf{X}^{(n)})\, d\tilde{\theta}}. \tag{5}$$

3. Asymptotic behavior of FBF

In this section, we consider the general FBF

$$\Lambda_{a,b}(\mathbf{X}^{(n)}) = \frac{z_a(\mathbf{X}^{(n)})}{z_b(\mathbf{X}^{(n)})} \cdot \frac{z_b^*(\mathbf{X}^{(n)})}{z_a^*(\mathbf{X}^{(n)})},$$

where $0 < b < a$. Note that $\Lambda_{2,1}(\mathbf{X}^{(n)})$ is PBF, $\Lambda_{1,1/2}(\mathbf{X}^{(n)})$ is the conventional FBF.

Suppose θ_0 is the true parameter. Let $I_{\theta_0} = P_{\theta_0}\dot{\ell}_{\theta_0}\dot{\ell}_{\theta_0}^T$ be the Fisher information matrix at θ_0 and $\Delta_{n,\theta_0} = \frac{1}{\sqrt{n}}\sum_{i=1}^n I_{\theta_0}^{-1}\dot{\ell}_{\theta_0}(X_i)$ be the 'locally sufficient' statistics. In null space, $\dot{\ell}^* I_{\theta_0}^*$ and Δ_{n,θ_0}^* are defined in the same way. It's easy to see that $\dot{\ell}_{\theta_0}^*$ is the first p_1 coordinates of $\dot{\ell}_{\theta_0}$, $I_{\theta_0}^*$ is the first $p_1 \times p_1$ submatrix of I_{θ_0} and $\Delta_{n,\theta_0}^* = \frac{1}{\sqrt{n}}\sum_{i=1}^n I_{\theta_0}^{*-1}\dot{\ell}_{\theta_0}^*(X_i)$.

Assumption 1. *Parameter θ_0 is an inner point of Θ and is a relative innver point of Θ_0. The function $\theta \mapsto \log p(X|\theta)$ is differentialbe at θ_0 P_0-a.s. with derivative $\dot{\ell}_{\theta_0}(X)$. 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_{θ_0} is positive-definite and as $\theta \to \theta_0$,

$$P_0 \log p(X|\theta) - P_0 \log(X|\theta_0) = -\frac{1}{2}(\theta - \theta_0)^T I_{\theta_0}(\theta - \theta_0) + o(\|\theta - \theta_0\|^2).$$

Assumption 1 is a stand assumption for likelihood. See vaart (1998) and vaart (2012).

Proposition 1. *Under Assumption 1, we have $\|\dot{\ell}_{\theta_0}(X)\| \le m(X)$ P_0-a.s., $P_0\dot{\ell}_{\theta_0}(X) = 0$ and*

$$\sup_{\|h\| \le M}\left|\log\frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)} - \frac{1}{\sqrt{n}}\sum_{i=1}^n h^T\dot{\ell}_{\theta_0}(X_i) + \frac{1}{2}h^T I_{\theta_0}h\right| \xrightarrow{P_0^n} 0.$$

(See der Vaart (2000) Theorem 5.23 or Kleijn and Vaart (2012) Lemma 2.1.)

For $t > 0$, We call $z_a(\mathbf{X}^{(n)})$ consistent if for every $M_n \to \infty$,

$$\frac{\int_{\{\theta:\|\theta-\theta_0\| \le M_n/\sqrt{n}\}}\left[p_n(\mathbf{X}^{(n)}|\theta)\right]^t \pi(\theta)\, d\theta}{\int_\Theta \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^t \pi(\theta)\, d\theta} \xrightarrow{P_{\theta_0}^n} 0.$$

for $t = 1$, this condition is equivalent to the consistency of Posterior distribution.

Theorem 1. *Suppose that Assumption 1 holds, $z_a(\mathbf{X}^{(n)})$, $z_b(\mathbf{X}^{(n)})$, $z_a^*(\mathbf{X}^{(n)})$ and $z_b^*(\mathbf{X}^{(n)})$ are consistent, then for bounded $\{\eta_n\}$,*

$$\log \Lambda_{a,b}(\mathbf{X}^{(n)}) \xrightarrow{P_{\eta_n}^n} .$$

Proof. By contiguity, we have blabla.

By Proposition 1

$$\begin{aligned}\log \Lambda_{a,b}(\mathbf{X}^{(n)}) =& \log \int_{\{\theta:\|\theta-\theta_0\|\le M_n/\sqrt{n}\}} \big[p_n(\mathbf{X}^{(n)}|\theta)\big]^a \pi(\theta)\, d\theta - \log \int_{\{\theta:\|\theta-\theta_0\|\le M_n/\sqrt{n}\}} \big[p_n(\mathbf{X}^{(n)}|\theta)\big]^b \pi(\theta)\, d\theta \\ &+ \log \int_{\{\tilde\theta:\|\tilde\theta-\tilde\theta_0\|\le M_n/\sqrt{n}\}} \big[p_n(\mathbf{X}^{(n)}|\tilde\theta)\big]^b \tilde\pi(\tilde\theta)\, d\tilde\theta - \log \int_{\{\tilde\theta:\|\tilde\theta-\tilde\theta_0\|\le M_n/\sqrt{n}\}} \big[p_n(\mathbf{X}^{(n)}|\tilde\theta)\big]^a \tilde\pi(\tilde\theta)\, d\tilde\theta \\ &+ o_{P_0^n}(1).\end{aligned}$$

□

3.1. Exponential family

We would like to investigate the asymptotic behavior of FBF in exponential family. Exponential family possesses many good properties.

3.2. General case

However, in general case PBF is not good. In the general setting, it seems that FBF can be applied to wider problem. Consider the example

Example 1. *Suppose $X_1, \dots, X_n$ are i.i.d. from the distribution*

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

where $\theta \in \Theta = \mathbb{R}$. We would like to test the hypotheses $H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$. The likelihood is

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

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. To use FBF, we impose a prior $\pi(\theta)$. Suppose that $\pi(\theta)$ is positive for all θ. Then

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

The likelihood will almost surely have no ties and consequently $z_t(\mathbf{X}^{(n)}) = +\infty$ if and only if $t \geq 2$. While FBF is well defined, PBF is not defined.

This example motivates us that FBF is better than PBF. In general, the Assumption 4 can be removed for FBF. Now we give the general theory

4. Integrated likelihood ratio test

4.1. The choice of the weight function

L^1 approximation of posterior by normal.

5. Main results

We study the asymptotic behavior of the ILRT statistic around θ_0. If there exists certain test, Bernstein von Mise theorem will be valid.

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

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

Theorem 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))

Under Assumption 1 and 2, we have

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

See Kleijn and Vaart (2012), Theorem 2.1. However, we may use more general weight function.

Assumption 3. *Let $\pi_n(h; \mathbf{X}^{(n)})$ be a weight function satisfying*

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

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

$$\lim_{n\to\infty} P_{\theta_0}^n(\sup_{\|h\|\geq K\sqrt{n}} (\pi_n(h; \mathbf{X}^{(n)}) - T(h)) \leq 0) \geq 1 - \epsilon \tag{8}$$

$$\lim_{n\to\infty} P_{\theta_0}^n(\sup_{\|h\|\leq K\sqrt{n}} \pi_n(h; \mathbf{X}^{(n)}) \leq A) \geq 1 - \epsilon \tag{9}$$

The condition 8 assume 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. To avoid the bad behavior of the likelihood function when $\sqrt{n}h$ is large, many theoretical works impose assumptions to likelihood. Thanks to the flexibility of weight function, we can impose 8 to weight function instead. The condition 9 is satisfied in most usual case. Condition 8 and 9 will be satisfied, for example, when

$$\pi_n(h; X) = \min(\pi_n(h|X), M) 1_{\|h\|\leq K\sqrt{n}}, \tag{10}$$

where M and K are user-specified constant and $\pi_n(h|\mathbf{X}^{(n)})$ is the posterior density.

Our first theorem is

Theorem 3. *Under Assumptions 1-3, for bounded real numbers η_n, we have*

$$\Big| \int_{\mathbb{R}^p} \frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)} \pi_n(h; \mathbf{X}^{(n)})\, dh - 2^{-\frac{p}{2}} \exp\big[\frac{1}{2}\Delta_{n,\theta_0}^T I_{\theta_0} \Delta_{n,\theta_0}\big] \Big| \xrightarrow{P_{\eta_n}^n} 0. \tag{11}$$

Proof of Theorem 3. 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

$$\left|\int_{\|h\|\le M}\frac{p_n(\mathbf{X}^{(n)}|\theta_0+n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\pi_n(h;\mathbf{X}^{(n)})\,dh-\int_{\|h\|\le M}\exp\left[h^TI_{\theta_0}\Delta_{n,\theta_0}-\frac{1}{2}h^TI_{\theta_0}h\right]\phi(h;\Delta_{n,\theta_0},I_{\theta_0}^{-1})\,dh\right|\xrightarrow{P_0^n}0 \tag{12}$$

By Theorem 1,

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

Hence we have

$$\int_{\|h\|\le M}\frac{p_n(\mathbf{X}^{(n)}|\theta_0+n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\pi_n(h;\mathbf{X}^{(n)})\,dh=\exp[o_{p_0^n}(1)]\int_{\|h\|\le M}\exp\left[h^TI_{\theta_0}\Delta_{n,\theta_0}-\frac{1}{2}h^TI_{\theta_0}h\right]\pi_n(h;\mathbf{X}^{(n)})\,dh \tag{13}$$

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

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

Combining with (13), we can conclude that (12) holds. This is true for every M and hence also for some $M_n\to\infty$.

In the second part, we prove

$$\psi(M)\overset{def}{=}\frac{\int_{\|h\|>M}p_n(\mathbf{X}^{(n)}|\theta_0+n^{-1/2}h)\pi_n(h;\mathbf{X}^{(n)})\,dh}{\int p_n(\mathbf{X}^{(n)}|\theta_0+n^{-1/2}h)\pi_n(h;\mathbf{X}^{(n)})\,dh}\xrightarrow{P_0^n}0. \tag{14}$$

Let ϕ_n be a test function satisfying the conclusion of Theorem 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

$$\psi(M)(1-\phi_n)\le\frac{\int_{\|h\|>M_n}p_n(\mathbf{X}^{(n)}|\theta_0+n^{-1/2}h)\pi_n(h;\mathbf{X}^{(n)})\,dh}{\int_{\|h\|\le U}p_n(\mathbf{X}^{(n)}|\theta_0+n^{-1/2}h)\pi_n(h;\mathbf{X}^{(n)})\,dh}(1-\phi_n). \tag{15}$$

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

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

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

Finally we have

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

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

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

is bounded in probability. Therefore

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

And J_3 convenges to 0 for trivial reason. □

Based on Theorem 3,the asymptotic distribution of integrated likelihood ratio statistics under null hypothesis can be obtained. It can be used to determine the critical value of the test

Theorem 4. *Suppose the assumptions of 3 are met for both Θ_0 and Θ, the true parameter θ_0 is an interior point of Θ and a relative interior point of Θ_0, then we have*

$$2\log(\Lambda(X)) \overset{P_0^n}{\rightsquigarrow} \chi^2_{p_2-p_1} - (p_2-p_1)\log(2) \tag{16}$$

Proof of Theorem 4. If the null hypothesis is true, the true parameter θ_0 is an interior point of Θ and θ_0 is a relative interior point of Θ_0. Then we can apply Theorem 3 to both the numerator and denominator of integrated likelihood ratio statistics with $\eta_n = 0$. By CLT,

$$I_{\theta_0}\Delta_{n,\theta_0} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \overset{P_0^n}{\rightsquigarrow} \xi, \tag{17}$$

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

$$I^*_{\theta_0}\Delta^*_{n,\theta_0} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}^*_{\theta_0}(X_i) \overset{P_0^n}{\rightsquigarrow} \xi^*, \tag{18}$$

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

$$
\begin{aligned}
\Lambda(X) &= \frac{2^{-\frac{p_2}{2}}\exp\{\frac{1}{2}\Delta_{n,\theta_0}^T I_{\theta_0}\Delta_{n,\theta_0}\} + o_{P_0^n}(1)}{2^{-\frac{p_1}{2}}\exp\{\frac{1}{2}\Delta_{n,\theta_0}^{*T} I^*_{\theta_0}\Delta^*_{n,\theta_0}\} + o_{P_0^n}(1)} \\
&\overset{P_0^n}{\rightsquigarrow} \frac{2^{-\frac{p_2}{2}}\exp\{\frac{1}{2}\xi^T I_{\theta_0}^{-1}\xi\}}{2^{-\frac{p_1}{2}}\exp\{\frac{1}{2}\xi^{*T} I_{\theta_0}^{*-1}\xi^*\}}.
\end{aligned}
\tag{19}
$$

But

$$\xi^T I_{\theta_0}^{-1}\xi - \xi^{*T} I_{\theta_0}^{*-1}\xi^* = (I_{\theta_0}^{-\frac{1}{2}}\xi)^T \left(I_{p_2\times p_2} - I_{\theta_0}^{\frac{1}{2}} \begin{pmatrix} I_{\theta_0}^{*-1} & 0 \\ 0 & 0 \end{pmatrix} I_{\theta_0}^{\frac{1}{2}}\right)(I_{\theta_0}^{-\frac{1}{2}}\xi). \tag{20}$$

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

$$2\log(\Lambda(X)) \overset{P_0^n}{\rightsquigarrow} \chi^2_{p_2-p_1} - (p_2 - p_1)\log(2). \tag{21}$$

□

We can obtain the asymptotic distribution of the integrated likelihood ratio test under local alternatives by Le Cam's third lemma.

Theorem 5. *Suppose the Assumptions of 4 are met. The true parameter θ satisfies $\eta_n = \sqrt{n}(\theta-\theta_0) \to \eta$. If*

$$I_{\theta_0} = \begin{pmatrix} I_{\theta_0}^* & I_{12} \\ I_{21} & I_{22} \end{pmatrix}, \tag{22}$$

$I_{22\cdot 1} = I_{22} - I_{21} I_{\theta_0}^{*-1} I_{12}$*, then we have*

$$2\log(\Lambda(X)) \overset{P_0^n}{\rightsquigarrow} \chi^2_{p_2-p_1}(\delta) - (p_2 - p_1)\log(2) \tag{23}$$

where

$$\delta = \eta^T \begin{pmatrix} 0 & 0 \\ 0 & I_{22\cdot 1} \end{pmatrix} \eta \tag{24}$$

The results can be explained by the limit experiment point of view. As $h_n \to h$, the 'locally sufficient' statistic $\Delta_{n,\theta_0} \rightsquigarrow N(h, I_{\theta_0}^{-1})$. In the limit experiment, we have one observation $X \sim N(h, I_{\theta_0}^{-1})$. In this case, the integrated likelihood ratio test statistics can be calculated easily whose distribution is exactly the same as 5.

Proof of Theorem 5. We note that $h_n = \eta_n$ converges to η. By differentiability in quadratic mean, Lemma **??** and CLT,

$$\begin{aligned} \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \\ \log\frac{p_{\eta_n}(X)}{p_0(X)} \end{pmatrix} &= \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \eta^T\dot{\ell}_{\theta_0}(X_i) - \frac{1}{2}\eta^T I_{\theta_0}\eta \end{pmatrix} + o_{P_0^n}(1) \\ &\overset{P_0^n}{\rightsquigarrow} N(\begin{pmatrix} 0 \\ -\frac{1}{2}\eta^T I_{\theta_0}\eta \end{pmatrix}, \begin{pmatrix} I_{\theta_0} & I_{\theta_0}\eta \\ \eta^T I_{\theta_0} & \eta^T I_{\theta_0}\eta \end{pmatrix}). \end{aligned} \tag{25}$$

Hence by Le Cam's third lemma,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \overset{P_{\eta_n}^n}{\rightsquigarrow} \xi \sim N(I_{\theta_0}\eta, I_{\theta_0}). \tag{26}$$

By Theorem 3, under $P_{\eta_n}^n$, we have (19). Hence

$$2\log(\Lambda(X)) \overset{P_{\eta_n}^n}{\rightsquigarrow} \chi^2_{p_2-p_1}(\delta) - (p_2 - p_1)\log(2), \tag{27}$$

where noncentral parameter δ can be obtained by substituting ξ by $I_{\theta_0}\eta$ in (20):

$$\begin{aligned} \delta &= \eta^T (I_{\theta_0} - I_{\theta_0} \begin{pmatrix} I_{\theta_0}^{*-1} & 0 \\ 0 & 0 \end{pmatrix} I_{\theta_0})\eta \\ &= \eta^T \begin{pmatrix} 0 & 0 \\ 0 & I_{22\cdot 1} \end{pmatrix} \eta. \end{aligned} \tag{28}$$

□

6. New Main Results

We denote by $\rightsquigarrow$ the weak convergence.

Let $\mathbf{X}^{(n)}$ denote the data. Let Θ be an open subset of $\mathbb{R}^p$ parameterising statistical models $\{P_\theta^{(n)} : \theta \in \Theta\}$. Denote by P_0 the true distribution of $\mathbf{X}$. We do not assume that $P_0 \in \{P_\theta^{(n)} : \theta \in \Theta\}$. Let $p_n(x|\theta)$ be the density of $P_\theta^{(n)}$ with respect to a reference measure μ_n.

There are many works give Bernstein-von Mises type theorems, which assert that the posterior distribution of h converges to a normal distribution with mean Δ_{n,θ^*} and variance $\mathbf{V}_{\theta^*}^{-1}$. However, most existing work consider the convergence under the total variation distance, that is

$$\int_{\mathbb{R}^p} \left|\pi^*(h|\mathbf{X}^{(n)}) - \phi(h|\Delta_{n,\theta^*}, \mathbf{V}_{\theta^*}^{-1})\right| dh \xrightarrow{P} 0.$$

Or Hellinger distance.

The following lemma is adapted from Ghosal et al. (2000). For two parameter θ_1 and θ_2, let

$$D_{KL}(\theta_1||\theta_2) = P_{\theta_1} \log \frac{p(X|\theta_1)}{p(X|\theta_2)}, \quad V(\theta_1||\theta_2) = \mathrm{Var}_{\theta_1} \Big(\log \frac{p(X|\theta_1)}{p(X|\theta_2)} \Big).$$

Denote by $\Pi(\cdot)$ the prior measure defined as $\Pi(A) = \int_A \pi(\theta)\, d\theta$. For positive ϵ, let

$$A_\epsilon = \{\theta : D_{KL}(\theta_0, \theta) \le \epsilon,\, V(\theta_0||\theta) \le \epsilon\}.$$

Lemma 1. *For every $\epsilon > 0$ and $C > 0$, we have*

$$P_0^n\Big(\int_\Theta \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^\alpha \pi(\theta)\, d\theta < \Pi(A_\epsilon) \exp\big(-(1+C)n\epsilon\big)\Big) \le \frac{\alpha^2}{C^2 n\epsilon}.$$

Proof. Without loss of generality, we assume $\Pi(A_\epsilon) > 0$. Then

$$\begin{aligned}
&P_0^n\Big(\int_\Theta \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^\alpha \pi(\theta)\, d\theta < \Pi(A_\epsilon) \exp\big(-(1+C)n\epsilon\big)\Big) \\
=&P_0^n\Big(\log \int_\Theta \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^\alpha \pi_\epsilon(\theta)\, d\theta < -(1+C)n\epsilon\Big),
\end{aligned}$$

where $\pi_\epsilon(\cdot)$ is a probability density on Θ defined as $\pi_\epsilon(\theta) = \pi(\theta)/\Pi(A_\epsilon)$. By Jensen's inequality, we have

$$\begin{aligned}
P_0^n\Big(\log \int_\Theta \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^\alpha \pi_\epsilon(\theta)\, d\theta < -(1+C)n\epsilon\Big) &\le P_0^n\Big(\int_\Theta \alpha \log \frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)} \pi_\epsilon(\theta)\, d\theta < -(1+C)n\epsilon\Big) \\
&= P_0^n\Big(\sum_{i=1}^n \int_\Theta \log \frac{p(X_i|\theta)}{p(X_i|\theta_0)} \pi_\epsilon(\theta)\, d\theta < -(1+C)n\epsilon/\alpha\Big).
\end{aligned}$$

Note that the expectation of $\log\big(p(X_i|\theta)/p(X_i|\theta_0)\big)$ is $-D_{KL}(\theta_0, \theta)$ and $D_{KL}(\theta_0, \theta) \le \epsilon$ for $\theta \in A_\epsilon$. We have

$$\begin{aligned}
&P_0^n\Big(\sum_{i=1}^n \int_\Theta \log \frac{p(X_i|\theta)}{p(X_i|\theta_0)} \pi_\epsilon(\theta)\, d\theta < -(1+C)n\epsilon/\alpha\Big) \\
\le&P_0^n\Big(\sum_{i=1}^n \int_\Theta \Big(\log \frac{p(X_i|\theta)}{p(X_i|\theta_0)} + D_{KL}(\theta_0, \theta)\Big) \pi_\epsilon(\theta)\, d\theta < -Cn\epsilon/\alpha\Big) \\
\le&\frac{\alpha^2}{C^2 n^2 \epsilon^2} n P_0\Big(\int_\Theta \Big(\log \frac{p(X_1|\theta)}{p(X_1|\theta_0)} + D_{KL}(\theta_0||\theta)\Big) \pi_\epsilon(\theta)\, d\theta\Big)^2 \\
\le&\frac{\alpha^2}{C^2 n\epsilon^2} P_0 \int_\Theta \Big(\log \frac{p(X_1|\theta)}{p(X_1|\theta_0)} + D_{KL}(\theta_0||\theta)\Big)^2 \pi_\epsilon(\theta)\, d\theta \\
=&\frac{\alpha^2}{C^2 n\epsilon^2} \int_\Theta V(\theta_0||\theta) \pi_\epsilon(\theta)\, d\theta \le \frac{\alpha^2}{C^2 n\epsilon},
\end{aligned}$$

where the second inequality follows from Markov inequality and the third inequality follows from Jensen's inequality. $\square$

This work is done by Walker and Hjort (2001).

For two parameters θ_1 and θ_2, let $H(\theta_1, \theta_2)$ denote the Hellinger distance between P_{θ_1} and P_{θ_2}:

$$H(\theta_1, \theta_2) = \Big(\int_{\mathcal{X}} \big(\sqrt{p(X|\theta_1)} - \sqrt{p(X|\theta_2)} \big)^2 \, d\mu \Big)^{1/2}.$$

Define $\rho(\theta_1, \theta_2) = \int_{\mathcal{X}} \sqrt{p(X|\theta_1) p(X|\theta_2)} \, d\mu$, then $H(\theta_1, \theta_2) = \sqrt{2 - 2\rho(\theta_1, \theta_2)}$.

Define

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

Assumption 4. *There exist positive constancts δ, ϵ, C_1 and C_2 such that for $\|\theta - \theta_0\| \leq \delta$,*

$$D_{KL}(p_{\theta_0} || p_\theta) \leq C_1 \|\theta - \theta_0\|^2, \quad V(p_{\theta_0} || p_\theta) \leq C_1 \|\theta - \theta_0\|^2, \quad C_2 \|\theta - \theta_0\| \leq H(\theta, \theta_0),$$

for $\|\theta - \theta_0\| > \delta$, $H(\theta, \theta_0) \geq \epsilon$.

Theorem 6. *Suppose θ_0 is an interior of Θ, $\pi(\theta)$ is continuous at θ_0 and $\pi(\theta_0) > 0$. Under Assumptions 1,2 and 4, for any $M_n \to \infty$,*

$$P_0^n \left\{ \frac{Z_{1/2}(\{\theta : \|\theta - \theta_0\| \geq \frac{M_n}{\sqrt{n}}\})}{Z_{1/2}(\Theta)} \right\} \to 0.$$

Proof. Without loss of generality, we assume $\frac{M_n}{\sqrt{n}} \to 0$, otherwise we replace M_n by a smaller one. Note that

$$\frac{Z_{1/2}(\{\theta : \|\theta - \theta_0\| \geq \frac{M_n}{\sqrt{n}}\})}{Z_{1/2}(\Theta)} = \frac{\int_{\{\theta : \|\theta - \theta_0\| \geq \frac{M_n}{\sqrt{n}}\}} \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)} \Big]^{1/2} \pi(\theta) \, d\theta}{\int_\Theta \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)} \Big]^{1/2} \pi(\theta) \, d\theta}. \tag{29}$$

Consider the expactation of the numerator of 29. It follows from Fubini's theorem that

$$\begin{aligned}
& P_0^n \int_{\{\theta : \|\theta - \theta_0\| \geq \frac{M_n}{\sqrt{n}}\}} \Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)} \Big]^{1/2} \pi(\theta) \, d\theta \\
&= \int_{\{\theta : \|\theta - \theta_0\| \geq \frac{M_n}{\sqrt{n}}\}} \Big\{ \int_{\mathcal{X}^n} \Big[p_n(\mathbf{X}^{(n)}|\theta) p_n(\mathbf{X}^{(n)}|\theta_0) \Big]^{1/2} d\mu^n \Big\} \pi(\theta) \, d\theta \\
&= \int_{\{\theta : \|\theta - \theta_0\| \geq \frac{M_n}{\sqrt{n}}\}} \big[\rho(\theta, \theta_0) \big]^n \pi(\theta) \, d\theta \\
&= \int_{\{\theta : \|\theta - \theta_0\| \geq \frac{M_n}{\sqrt{n}}\}} \big[1 - \frac{1}{2} H(\theta, \theta_0)^2 \big]^n \pi(\theta) \, d\theta \\
&\leq \int_{\{\theta : \|\theta - \theta_0\| \geq \frac{M_n}{\sqrt{n}}\}} \exp \big[- \frac{n}{2} H(\theta, \theta_0)^2 \big] \pi(\theta) \, d\theta.
\end{aligned}$$

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

$$\begin{aligned}
& \int_{\{\theta : \|\theta - \theta_0\| \geq \frac{M_n}{\sqrt{n}}\}} \exp \big[- \frac{n}{2} H(\theta, \theta_0)^2 \big] \pi(\theta) \, d\theta \\
&= \int_{\{\theta : \frac{M_n}{\sqrt{n}} \leq \|\theta - \theta_0\| \leq \delta\}} \exp \big[- \frac{n}{2} H(\theta, \theta_0)^2 \big] \pi(\theta) \, d\theta + \int_{\{\theta : \|\theta - \theta_0\| > \delta\}} \exp \big[- \frac{n}{2} H(\theta, \theta_0)^2 \big] \pi(\theta) \, d\theta \\
&\leq \max_{\|\theta - \theta_0\| \leq \delta} \pi(\theta) \int_{\big\{\theta : \|\theta - \theta_0\| \geq \frac{M_n}{\sqrt{n}}\big\}} \exp \big[- \frac{n}{2} C_2^2 \|\theta - \theta_0\|^2 \big] \, d\theta + \exp \big[- \frac{n}{2} \epsilon^2 \big].
\end{aligned}$$

Note that

$$
\begin{aligned}
&\int_{\left\{\theta:\|\theta-\theta_0\|\geq \frac{M_n}{\sqrt{n}}\right\}} \exp\big[-\frac{n}{2}C_2^2\|\theta-\theta_0\|^2\big]\,d\theta\\
=&(\sqrt{n}C_2)^{-p}\int_{\left\{t:\|t\|\geq C_2M_n\right\}} \exp\big[-\frac{t^2}{2}\big]\,d\theta = \Big(\frac{\sqrt{2\pi}}{\sqrt{n}C_2}\Big)^p P\{\chi_p^2\geq C_2^2M_n^2\},
\end{aligned}
$$

where χ_p^2 is a chi-squared random variable with p degrees of freedom. By Laurent and Massart (2000) Lemma 1, for $M_n > \sqrt{2p}/C_2$,

$$
P\{\chi_p^2\geq C_2^2M_n^2\}\leq \exp\Big[-\big(\frac{\sqrt{2}}{2}C_2M_n-\sqrt{p}\big)^2\Big].
$$

Thus, for large n,

$$
\begin{aligned}
&P_0^n\int_{\{\theta:\|\theta-\theta_0\|\geq\frac{M_n}{\sqrt{n}}\}}\Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^{1/2}\pi(\theta)\,d\theta\\
\leq&\Big[\max_{\|\theta-\theta_0\|\leq\delta}\pi(\theta)\Big]\Big(\frac{\sqrt{2\pi}}{\sqrt{n}C_2}\Big)^p\exp\Big[-\big(\frac{\sqrt{2}}{2}C_2M_n-\sqrt{p}\big)^2\Big]+\exp\big[-\frac{n}{2}\epsilon^2\big].
\end{aligned}
\tag{30}
$$

Now we consider the denominator of 29. Lemma 1 implies that for every $\epsilon'>0$, there is a set $B_{\epsilon'}$ with $P_0^nB_{\epsilon'}>1-1/(4C^2n\epsilon')$ on which

$$
\int_\Theta\Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^{1/2}\pi(\theta)\,d\theta\geq\Pi(A_{\epsilon'})\exp\big(-(1+C)\epsilon'n\big).\tag{31}
$$

We take

$$
\epsilon'=\frac{\big(\frac{\sqrt{2}}{2}C_2M_n-\sqrt{p}\big)^2}{2(1+C)n}.
$$

It can be seen that $\epsilon'\to0$. Hence for sufficiently large n, we have

$$
\begin{aligned}
\Pi(A_{\epsilon'})=&\Pi\big(\{\theta:D_{KL}(\theta_0,\theta)\leq\epsilon',\,V(\theta_0||\theta)\leq\epsilon'\}\big)\\
\geq&\Pi\Big(\big\{\theta:\|\theta-\theta_0\|^2\leq\frac{\epsilon'}{C_1}\big\}\Big)\\
=&\int_{\left\{\theta:\|\theta-\theta_0\|^2\leq\frac{\epsilon'}{C_1}\right\}}\pi(\theta)\,d\theta\\
\geq&\Big(\min_{\|\theta-\theta_0\|\leq\sqrt{\frac{\epsilon'}{C_1}}}\pi(\theta)\Big)\int_{\left\{\theta:\|\theta-\theta_0\|\leq\sqrt{\frac{\epsilon'}{C_1}}\right\}}d\theta\\
=&\Big(\min_{\|\theta-\theta_0\|\leq\sqrt{\frac{\epsilon'}{C_1}}}\pi(\theta)\Big)\big(\frac{\epsilon'}{C_1}\big)^{p/2}\frac{2\pi^{p/2}}{\Gamma(p/2)}\\
=&\Big(\min_{\|\theta-\theta_0\|\leq\sqrt{\frac{\epsilon'}{C_1}}}\pi(\theta)\Big)\epsilon'^{p/2}\frac{2\pi^{p/2}}{(2(1+C)C_1)^{p/2}\Gamma(p/2)}\cdot\frac{\big(\frac{\sqrt{2}}{2}C_2M_n-\sqrt{p}\big)^p}{n^{p/2}}\\
\asymp&\frac{M_n^p}{n^{p/2}}.
\end{aligned}
$$

Then it follows from (31) that on the set $B_{\epsilon'}$,

$$
\int_\Theta\Big[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\Big]^{1/2}\pi(\theta)\,d\theta\gtrsim\frac{M_n^p}{n^{p/2}}\exp\Big[-\frac{1}{2}\big(\frac{\sqrt{2}}{2}C_2M_n-\sqrt{p}\big)^2\Big].\tag{32}
$$

$$P_0^n\left\{\frac{Z_{1/2}(\{\theta:\|\theta-\theta_0\|\geq\frac{M_n}{\sqrt{n}}\})}{Z_{1/2}(\Theta)}\right\}\leq P_0^n\left\{\mathbf{1}_{B_{\epsilon'}}\frac{Z_{1/2}(\{\theta:\|\theta-\theta_0\|\geq\frac{M_n}{\sqrt{n}}\})}{Z_{1/2}(\Theta)}\right\}+P_0^nB_{\epsilon'}^C.$$

Since

$$P_0^nB_{\epsilon'}^C=1-P_0^nB_{\epsilon'}<\frac{1}{4C^2n\epsilon'}=\frac{1+C}{2C^2\big(\frac{\sqrt{2}}{2}C_2M_n-\sqrt{p}\big)^2}\to 0,$$

we only need to upper bound the first term. Combine (29), (30) and (32), for sufficiently large n we have

$$\begin{aligned}
&P_0^n\left\{\mathbf{1}_{B_{\epsilon'}}\frac{Z_{1/2}(\{\theta:\|\theta-\theta_0\|\geq\frac{M_n}{\sqrt{n}}\})}{Z_{1/2}(\Theta)}\right\}\\
\lesssim&\frac{1}{M_n^p}\exp\Big[-\frac{1}{2}\big(\frac{\sqrt{2}}{2}C_2M_n-\sqrt{p}\big)^2\Big]+\frac{n^{p/2}}{M_n^p}\exp\Big[-\frac{n}{2}\epsilon^2+\frac{1}{2}\big(\frac{\sqrt{2}}{2}C_2M_n-\sqrt{p}\big)^2\Big]\\
\leq&\frac{1}{M_n^p}\exp\Big[-\frac{1}{2}\big(\frac{\sqrt{2}}{2}C_2M_n-\sqrt{p}\big)^2\Big]+\frac{n^{p/2}}{M_n^p}\exp\big[-\frac{n}{4}\epsilon^2\big]\to 0.
\end{aligned}$$

This completes the proof. □

6.1. Normal mixture

Although posterior Bayes factor can be used for some prior Aitkin et al. (1996). Posterior Bayes estimator can not be defined for certain prior distribution.

Fractional posterior Bayes factor (O'Hagan, 1995) can be defined.

7. Appendix

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

References

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

Aitkin M, Finch S, Mendell N, Thode H. A new test for the presence of a normal mixture distribution based on the posterior bayes factor. Statistics and Computing 1996;6(2):121–5. URL: `https://doi.org/10.1007/BF00162522`. doi:`10.1007/BF00162522`.

Bahadur R. Some Limit Theorems in Statistics. volume 4 of *Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics*. Philadelphia, Pa: Society for Industrial and Applied Mathematics, 1971.

Cam L. Maximum likelihood: An introduction. International Statistical Review 1990;58(2):153–71.

Gelfand D. K. D AE. Bayesian model choice: Asymptotics and exact calculations. 1993.

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`.

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

Laurent B, Massart P. Adaptive estimation of a quadratic functional by model selection. Annals of Statistics 2000;28(5):1302–38.

Lehmann J. P. R E. Testing Statistical Hypotheses. New York: Springer, 2005.

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

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

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/`.