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 $p_\theta(X|\theta)$ 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 satisfied. Aitkin (1991) proposed posterior Bayes factor

$$\frac{\int_{\Theta} p_\theta(X)\pi(\theta|X)\,d\theta}{\int_{\Theta_0} p_\theta(X)\pi^*(\theta|X)\,d\theta}, \tag{2}$$

where $\pi^*(\theta|x)$ and $\pi(\theta|x)$ are the posterior densities under null hypotheses and alternative hypothesis. Gelfand D. K. D (1993) derived it's null distribution. 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.

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

$$\Lambda(X) = \frac{\int_{\Theta} \left[p(\mathbf{X}^{(n)}|\theta)\right]^{\alpha}\pi(\theta;\mathbf{X}^{(n)})\,d\theta}{\int_{\Theta_0} \left[p(\mathbf{X}^{(n)}|\theta)\right]^{\alpha}\pi^*(\theta;\mathbf{X}^{(n)})\,d\theta}, \tag{3}$$

where $\alpha > 0$ 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.

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

$$\Theta_0 = \{\theta \in \Theta : \theta_{p_1+1} = \theta_{0,p_1+1}, \ldots, \theta_{p_2} = \theta_{0,p_2}\}, \tag{4}$$

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

$$H_0 : \theta \in \Theta_0 \quad vs. \quad H_1 : \theta \in \Theta. \tag{5}$$

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 $\tilde{\Theta}_0 = \{(\theta_1, \ldots, \theta_{p_1})^T : (\theta_1, \ldots, \theta_{p_1}, \theta_{0,p_1+1}, \theta_{0,p_2}) \in \Theta_0\}$. 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

$$\Lambda(\mathbf{X}^{(n)}) = \frac{\int_{\Theta} p_n(\mathbf{X}^{(n)}|\theta)\pi(\theta;\mathbf{X}^{(n)})\,d\theta}{\int_{\tilde{\Theta}_0} p_n(\mathbf{X}^{(n)}|\tilde{\theta})\tilde{\pi}(\tilde{\theta};\mathbf{X}^{(n)})\,d\tilde{\theta}}. \tag{6}$$

3. Main results

We study the asymptotic behavior of the ILRT statistic around θ_0.

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

Listed below are the regular conditions we need:

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 $\theta_1, \theta_2 \in V$,*

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

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).

Theorem 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*

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

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

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*

$$P_{\theta_0}^n \phi_n \to 0, \qquad \sup_{\|\theta - \theta_0\| \geq \epsilon} P_\theta^n (1 - \phi_n) \to 0. \tag{7}$$

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*

$$\|\pi_n(h; \mathbf{X}^{(n)}) - \phi(h; \Delta_{n,\theta_0}, I_{\theta_0}^{-1})\| \overset{P_{\theta_0}^n}{\to} 0 \tag{8}$$

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

$$\lim_{n\to\infty} P_{\theta_0}^n (\sup_{\|h\| \geq K\sqrt{n}} (\pi_n(h; \mathbf{X}^{(n)}) - T(h)) \leq 0) \geq 1 - \epsilon \tag{9}$$

$$\lim_{n\to\infty} P_{\theta_0}^n (\sup_{\|h\| \leq K\sqrt{n}} \pi_n(h; \mathbf{X}^{(n)}) \leq A) \geq 1 - \epsilon \tag{10}$$

The condition 9 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 9 to weight function instead. The condition 10 is satisfied in most usual case. Condition 9 and 10 will be satisfied, for example, when

$$\pi_n(h; X) = \min(\pi_n(h|X), M) 1_{\|h\| \leq K\sqrt{n}}, \tag{11}$$

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*

$$\Big| \int_{\mathbb{R}^p} \frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)} \pi_n(h; \mathbf{X}^{(n)})\, dh - 2^{-\frac{p}{2}} \exp\big[\frac{1}{2}\Delta_{n,\theta_0}^T I_{\theta_0} \Delta_{n,\theta_0}\big]\Big| \xrightarrow{P_{\eta_n}^n} 0. \tag{12}$$

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 proof

$$\left| \int_{\|h\| \le M} \frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)} \pi_n(h; \mathbf{X}^{(n)})\, dh - \int_{\|h\| \le M} \exp\big[h^T I_{\theta_0} \Delta_{n,\theta_0} - \frac{1}{2} h^T I_{\theta_0} h\big] \phi(h; \Delta_{n,\theta_0}, I_{\theta_0}^{-1})\, dh \right| \xrightarrow{P_0^n} 0 \tag{13}$$

By Theorem 1,

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

Hence we have

$$\int_{\|h\| \le M} \frac{p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h)}{p_n(\mathbf{X}^{(n)}|\theta_0)} \pi_n(h; \mathbf{X}^{(n)})\, dh = \exp[o_{P_0^n}(1)] \int_{\|h\| \le M} \exp\big[h^T I_{\theta_0} \Delta_{n,\theta_0} - \frac{1}{2} h^T I_{\theta_0} h\big] \pi_n(h; \mathbf{X}^{(n)})\, dh \tag{14}$$

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

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

Combining with (14), we can conclude that (13) holds. This is true for every M and hence also for some $M_n \to \infty$.

In the second part, we prove

$$\psi(M) \overset{def}{=} \frac{\int_{\|h\| > M} p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h) \pi_n(h; \mathbf{X}^{(n)})\, dh}{\int p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h) \pi_n(h; \mathbf{X}^{(n)})\, dh} \xrightarrow{P_0^n} 0. \tag{15}$$

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

$$\psi(M)(1 - \phi_n) \le \frac{\int_{\|h\| > M_n} p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h) \pi_n(h; \mathbf{X}^{(n)})\, dh}{\int_{\|h\| \le U} p_n(\mathbf{X}^{(n)}|\theta_0 + n^{-1/2}h) \pi_n(h; \mathbf{X}^{(n)})\, dh} (1 - \phi_n). \tag{16}$$

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

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

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

Finally we have

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

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*

$$2\log(\Lambda(X)) \overset{P_0^n}{\rightsquigarrow} \chi^2_{p_2-p_1} - (p_2 - p_1)\log(2) \tag{17}$$

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,

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

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

$$I^*_{\theta_0}\Delta^*_{n,\theta_0} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}^*_{\theta_0}(X_i) \overset{P_0^n}{\rightsquigarrow} \xi^*, \tag{19}$$

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

$$
\begin{aligned}
\Lambda(X) &= \frac{2^{-\frac{p_2}{2}}\exp\{\frac{1}{2}\Delta_{n,\theta_0}^T I_{\theta_0}\Delta_{n,\theta_0}\} + o_{P_0^n}(1)}{2^{-\frac{p_1}{2}}\exp\{\frac{1}{2}\Delta_{n,\theta_0}^{*T} I^*_{\theta_0}\Delta^*_{n,\theta_0}\} + o_{P_0^n}(1)} \\
&\overset{P_0^n}{\rightsquigarrow} \frac{2^{-\frac{p_2}{2}}\exp\{\frac{1}{2}\xi^T I_{\theta_0}^{-1}\xi\}}{2^{-\frac{p_1}{2}}\exp\{\frac{1}{2}\xi^{*T} I_{\theta_0}^{*-1}\xi^*\}}.
\end{aligned}
\tag{20}
$$

But

$$\xi^T I_{\theta_0}^{-1}\xi - \xi^{*T} I_{\theta_0}^{*-1}\xi^* = (I_{\theta_0}^{-\frac{1}{2}}\xi)^T \Big(I_{p_2\times p_2} - I_{\theta_0}^{\frac{1}{2}} \begin{pmatrix} I_{\theta_0}^{*-1} & 0 \\ 0 & 0 \end{pmatrix} I_{\theta_0}^{\frac{1}{2}}\Big)(I_{\theta_0}^{-\frac{1}{2}}\xi). \tag{21}$$

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

$$2\log(\Lambda(X)) \overset{P_0^n}{\rightsquigarrow} \chi^2_{p_2-p_1} - (p_2 - p_1)\log(2). \tag{22}$$

□

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*

$$I_{\theta_0} = \begin{pmatrix} I_{\theta_0}^* & I_{12} \\ I_{21} & I_{22} \end{pmatrix}, \tag{23}$$

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

$$2\log(\Lambda(X)) \overset{P_0^n}{\rightsquigarrow} \chi^2_{p_2-p_1}(\delta) - (p_2 - p_1)\log(2) \tag{24}$$

where

$$\delta = \eta^T \begin{pmatrix} 0 & 0 \\ 0 & I_{22\cdot 1} \end{pmatrix} \eta \tag{25}$$

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,

$$\begin{aligned} \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \\ \log\frac{p_{\eta_n}(X)}{p_0(X)} \end{pmatrix} &= \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \eta^T\dot{\ell}_{\theta_0}(X_i) - \frac{1}{2}\eta^T I_{\theta_0}\eta \end{pmatrix} + o_{P_0^n}(1) \\ &\overset{P_0^n}{\rightsquigarrow} N(\begin{pmatrix} 0 \\ -\frac{1}{2}\eta^T I_{\theta_0}\eta \end{pmatrix}, \begin{pmatrix} I_{\theta_0} & I_{\theta_0}\eta \\ \eta^T I_{\theta_0} & \eta^T I_{\theta_0}\eta \end{pmatrix}). \end{aligned} \tag{26}$$

Hence by Le Cam's third lemma,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \overset{P_{\eta_n}^n}{\rightsquigarrow} \xi \sim N(I_{\theta_0}\eta, I_{\theta_0}). \tag{27}$$

By Theorem 3, under $P_{\eta_n}^n$, we have (20). Hence

$$2\log(\Lambda(X)) \overset{P_{\eta_n}^n}{\rightsquigarrow} \chi^2_{p_2-p_1}(\delta) - (p_2 - p_1)\log(2), \tag{28}$$

where noncentral parameter δ can be obtained by substituting ξ by $I_{\theta_0}\eta$ in (21):

$$\begin{aligned} \delta &= \eta^T (I_{\theta_0} - I_{\theta_0} \begin{pmatrix} I_{\theta_0}^{*-1} & 0 \\ 0 & 0 \end{pmatrix} I_{\theta_0})\eta \\ &= \eta^T \begin{pmatrix} 0 & 0 \\ 0 & I_{22\cdot 1} \end{pmatrix} \eta. \end{aligned} \tag{29}$$

□

4. New Main Results

We denote by $\rightsquigarrow$ the weak convergence.

- One step test. Like one step estimator.
- The key is the proof of results somewhat like the consistency of the posterior distribution. The argument by the existence of certain test can not be applied.

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.

4.1. Examples

Posterior Bayes factor, proposed by Aitkin (1991), is an alternative of the Bayes factor. Posterior Bayes factor is defined as

$$B_{10} = \frac{\int_\Theta p_n(\mathbf{X}^{(n)}|\theta)\pi(\theta|\mathbf{X}^{(n)})\, d\theta}{\int_{\tilde{\Theta}_0} p_n(\mathbf{X}^{(n)}|\tilde{\theta})\tilde{\pi}(\tilde{\theta}|\mathbf{X}^{(n)})\, d\tilde{\theta}}.$$

By some algebra, we have

$$B_{10} = \frac{\int_\Theta \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^2 \pi(\theta)\, d\theta}{\int_{\tilde{\Theta}_0} \left[p_n(\mathbf{X}^{(n)}|\tilde{\theta})\right]^2 \tilde{\pi}(\tilde{\theta})\, d\tilde{\theta}} \frac{\int_{\tilde{\Theta}_0} p_n(\mathbf{X}^{(n)}|\tilde{\theta})\tilde{\pi}(\tilde{\theta})\, d\tilde{\theta}}{\int_\Theta p_n(\mathbf{X}^{(n)}|\theta)\pi(\theta)\, d\theta}.$$

We would like to derive the asymptotic behavior of

$$\int_{\mathbb{R}^p} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta^* + \delta_n h)}{p_n(\mathbf{X}^{(n)}|\theta^*)}\right]^k \pi^*(h)\, dh.$$

For $M > 0$, define $K(M) = \{h : \|h\| \leq M\}$. We have

$$\int_{\mathbb{R}^p} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta^* + \delta_n h)}{p_n(\mathbf{X}^{(n)}|\theta^*)}\right]^k \pi^*(h)\, dh = \int_{K(M)} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta^* + \delta_n h)}{p_n(\mathbf{X}^{(n)}|\theta^*)}\right]^k \pi^*(h)\, dh + \int_{K(M)^C} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta^* + \delta_n h)}{p_n(\mathbf{X}^{(n)}|\theta^*)}\right]^k \pi^*(h)\, dh.$$

We expect that the second term is a smaller term of the first term. Define

$$\epsilon_2(M) = \frac{\int_{K(M)^C} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta^* + \delta_n h)}{p_n(\mathbf{X}^{(n)}|\theta^*)}\right]^k \pi^*(h)\, dh}{\int_{K(M)} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta^* + \delta_n h)}{p_n(\mathbf{X}^{(n)}|\theta^*)}\right]^k \pi^*(h)\, dh}.$$

Hence

$$\int_{\mathbb{R}^p} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta^* + \delta_n h)}{p_n(\mathbf{X}^{(n)}|\theta^*)}\right]^k \pi^*(h)\, dh = (1 + \epsilon_2(M)) \int_{K(M)} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta^* + \delta_n h)}{p_n(\mathbf{X}^{(n)}|\theta^*)}\right]^k \pi^*(h)\, dh.$$

But

$$
\begin{aligned}
& e^{-k\epsilon_{1,n}(K)} \min_{h\in K} \frac{\pi^*(h)}{\pi^*(0)}\pi^*(0) \int_{K(M)} \big[\frac{p_n^*(\mathbf{X}^{(n)}|\theta^*+\delta_n h)}{p_n(\mathbf{X}^{(n)}|\theta^*)}\big]^k \, dh \\
\le & \int_{K(M)} \big[\frac{p_n(\mathbf{X}^{(n)}|\theta^*+\delta_n h)}{p_n(\mathbf{X}^{(n)}|\theta^*)}\big]^k \pi^*(h) \, dh \\
\le & e^{k\epsilon_{1,n}(K)} \max_{h\in K} \frac{\pi^*(h)}{\pi^*(0)}\pi^*(0) \int_{K(M)} \big[\frac{p_n^*(\mathbf{X}^{(n)}|\theta^*+\delta_n h)}{p_n(\mathbf{X}^{(n)}|\theta^*)}\big]^k \, dh
\end{aligned}
$$

So the key is to bound $\epsilon_2(M)$.
Exponential family.

Theorem 6 (Asymptotic normality of posterior distribution)**.** *Suppose Assumption ?? holds. Let C be a quantity satisfying $C \gg \sqrt{p}$. Suppose that for large n, $\{\theta : \|\mathbf{J}(\theta-\theta_0)\| \le n^{-1/2}C\} \subset \Theta$. Suppose that $\frac{1}{3}\big(\frac{1}{n^{1/2}}CB_{1n}(0) + \frac{1}{n}C^2B_{2n}(n^{-1/2}C)\big) \le 1/2$ for sufficiently large n. Then for any $\epsilon > 0$, for sufficiently large n, with probability larger than $1-\epsilon$,*

$$
\begin{aligned}
& \int |\pi^*(u) - \phi_p(u;\Delta_n,\mathbf{I}_p)| \, du \\
\le & \Big| \exp\Big\{\frac{1}{6}\big(\frac{1}{n^{1/2}}C^3B_{1n}(0) + \frac{1}{n}C^4B_{2n}(n^{-1/2}C)\big)\Big\} - 1\Big| \sup_{\|\mathbf{J}(\theta-\theta_0)\|\le n^{-1/2}C} \frac{\pi(\theta)}{\pi(\theta_0)} \\
& + \sup_{\|\mathbf{J}(\theta-\theta_0)\|\le n^{-1/2}C} \Big|\frac{\pi(\theta)}{\pi(\theta_0)} - 1\Big| \\
& + \exp\Big\{\frac{p}{2}\log\frac{n}{2\pi} + \frac{1}{2}\log|\psi''(\theta_0)|\Big\} \int_{\|\mathbf{J}(\theta-\theta_0)\|>n^{-1/2}C} \exp\Big\{-\frac{\sqrt{n}}{4}C\|\mathbf{J}(\theta-\theta_0)\|\Big\} \frac{\pi(\theta)}{\pi(\theta_0)} \, d\theta \\
& + \exp\big(-\frac{1}{4}(C - (1/\sqrt{\epsilon}+1)\sqrt{p})^2\big).
\end{aligned}
$$

Proof. Let $\tilde{Z}_n(u) = \exp[\Delta_n^T u - \frac{1}{2}\|u\|^2]$. Note that $\phi_p(u;\Delta_n,\mathbf{I}_p) = \tilde{Z}_n(u)\pi(\theta_0)/\int \tilde{Z}_n(u)\pi(\theta_0)\,du$. We have

$$
\begin{aligned}
& \int |\pi^*(u) - \phi_p(u;\Delta_n,\mathbf{I}_p)| \, du = \int \Big| \frac{Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)}{\int Z_n(w)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}w)\,dw} - \frac{\tilde{Z}_n(u)\pi(\theta_0)}{\int \tilde{Z}_n(w)\pi(\theta_0)\,dw}\Big| \, du \\
\le & \int \Big| \frac{Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)}{\int Z_n(w)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}w)\,dw} - \frac{Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)}{\int \tilde{Z}_n(w)\pi(\theta_0)\,dw}\Big| \, du \\
& + \int \Big| \frac{Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)}{\int \tilde{Z}_n(w)\pi(\theta_0)\,dw} - \frac{\tilde{Z}_n(u)\pi(\theta_0)}{\int \tilde{Z}_n(w)\pi(\theta_0)\,dw}\Big| \, du \\
= & \Big| 1 - \frac{\int Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)\,du}{\int \tilde{Z}_n(u)\pi(\theta_0)\,du}\Big| + \frac{\int \big|Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u) - \tilde{Z}_n(u)\pi(\theta_0)\big|\,du}{\int \tilde{Z}_n(u)\pi(\theta_0)\,du} \\
\le & \Big| 1 - \frac{\int Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)\,du}{\int \tilde{Z}_n(u)\pi(\theta_0)\,du}\Big| + \frac{\int \big|Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u) - \tilde{Z}_n(u)\pi(\theta_0)\big|\,du}{\int \tilde{Z}_n(u)\pi(\theta_0)\,du} \\
\le & 2\frac{\int \big|Z_n(u)\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u) - \tilde{Z}_n(u)\pi(\theta_0)\big|\,du}{\int \tilde{Z}_n(u)\pi(\theta_0)\,du} \\
= & 2\int \Big| \exp\Big\{\log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2\Big\} \frac{\pi(\theta_0+n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)} - \phi_p(u;\Delta_n,\mathbf{I}_p)\Big| \, du
\end{aligned}
$$

We split the integral into the region $\|u\| \le C$ and $\|u\| > C$, where C will be specified latter. Then

$$
\begin{aligned}
&\int \Big| \exp\big\{ \log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 \big\} \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)} - \phi_p(u;\Delta_n,\mathbf{I}_p) \Big|\, du \\
\le &\int_{\|u\|\le C} \Big| \exp\big\{ \log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 \big\} \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)} - \phi_p(u;\Delta_n,\mathbf{I}_p) \Big|\, du \qquad (30)\\
&+ \int_{\|u\|>C} \exp\big\{ \log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 \big\} \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)}\, du + \int_{\|u\|>C} \phi_p(u;\Delta_n,\mathbf{I}_p)\, du
\end{aligned}
$$

We deal the three terms of (**??**) separately. Consider the first term. For $\|u\| \le C$, we have

$$
\begin{aligned}
&\log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 = -\frac{p}{2}\log(2\pi) - \frac{1}{2}\|u-\Delta_n\|^2 \\
&- n\Big(\frac{1}{6n^{3/2}} \mathrm{E}_{\theta_0} \big(u^T\mathbf{J}^{-1}(U - \mathrm{E}_{\theta_0} U)\big)^3 + O(1)\frac{1}{n^2}\|u\|^4 B_{2n}(n^{-1/2}\|u\|) \Big).
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\Big| \Big(\log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 \Big) - \Big(-\frac{p}{2}\log(2\pi) - \frac{1}{2}\|u-\Delta_n\|^2 \Big) \Big| \\
\le &\frac{1}{6}\Big(\frac{1}{n^{1/2}}\|u\|^3 B_{1n}(0) + \frac{1}{n}\|u\|^4 B_{2n}(n^{-1/2}\|u\|) \Big) \qquad (31)\\
\le &\frac{1}{6}\Big(\frac{1}{n^{1/2}} C^3 B_{1n}(0) + \frac{1}{n} C^4 B_{2n}(n^{-1/2}C) \Big).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\int_{\|u\|\le C} \Big| \exp\big\{ \log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 \big\} \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)} - \phi_p(u;\Delta_n,\mathbf{I}_p) \Big|\, du \\
\le &\int_{\|u\|\le C} \Big| \exp\big\{ \log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 \big\} - \phi_p(u;\Delta_n,\mathbf{I}_p) \Big| \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)}\, du \\
&+ \int_{\|u\|\le C} \Big| \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)} - 1 \Big| \phi_p(u;\Delta_n,\mathbf{I}_p)\, du \\
\le &\Big| \exp\Big\{ \frac{1}{6}\big(\frac{1}{n^{1/2}} C^3 B_{1n}(0) + \frac{1}{n} C^4 B_{2n}(n^{-1/2}C) \big) \Big\} - 1 \Big| \int_{\|u\|\le C} \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)} \phi_p(u;\Delta_n,\mathbf{I}_p)\, du \\
&+ \int_{\|u\|\le C} \Big| \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)} - 1 \Big| \phi_p(u;\Delta_n,\mathbf{I}_p)\, du \\
\le &\Big| \exp\Big\{ \frac{1}{6}\big(\frac{1}{n^{1/2}} C^3 B_{1n}(0) + \frac{1}{n} C^4 B_{2n}(n^{-1/2}C) \big) \Big\} - 1 \Big| \sup_{\|\mathbf{J}(\theta-\theta_0)\|\le n^{-1/2}C} \frac{\pi(\theta)}{\pi(\theta_0)} \\
&+ \sup_{\|\mathbf{J}(\theta-\theta_0)\|\le n^{-1/2}C} \Big| \frac{\pi(\theta)}{\pi(\theta_0)} - 1 \Big|.
\end{aligned}
$$

Next we deal with the last term of (**??**). Note that $\mathrm{E}\,\Delta_n = \mathbf{0}_p$ and $\mathrm{Var}\,\Delta_n = \mathbf{I}_p$. By Chebyshev's inequality, for $\epsilon > 0$, there is an $M = 1/\sqrt{\epsilon}$ such that

$$
\sup_n \Pr(\|\Delta_n\| \ge M\sqrt{p}) < \epsilon.
$$

Denote $\mathcal{A} = \{\|\Delta_n\| \le M\sqrt{p}\}$. On the event $\mathcal{A}$, for $M_1 > 0$,

$$
\int_{\|u\|>(M+1)\sqrt{p}+M_1} \phi_p(u;\Delta_n,\mathbf{I}_p)\, du \le \int_{\|u\|>M_1+\sqrt{p}} \phi_p(u;\mathbf{0}_p,\mathbf{I}_p)\, du \le \exp\big(-\frac{1}{4}M_1^2 \big).
$$

Hence for large n such that $C > (M+1)\sqrt{p}$, we have

$$\int_{\|u\|>C} \phi_p(u; \Delta_n, \mathbf{I}_p)\, du \le \exp\big(-\frac{1}{4}\big(C - (M+1)\sqrt{p}\big)^2\big).$$

Now we deal with the second term of (**??**). For $\|u\| \ge C$, by the concavity of $\log Z_n(u)$, we have that

$$(1 - \frac{C}{\|u\|}) \log Z_n(0) + \frac{C}{\|u\|} \log Z_n(u) \le \log Z_n\big(\frac{C}{\|u\|} u\big).$$

Hence

$$\log Z_n(u) \le \frac{\|u\|}{C} \log Z_n(\frac{C}{\|u\|} u).$$

This, combined with (**??**), yields

$$\begin{aligned}
\log Z_n(u) \le & \frac{\|u\|}{C}\Big(-\frac{1}{2}\big\|\frac{C}{\|u\|}u - \Delta_n\big\|^2 + \frac{1}{2}\|\Delta_n\|^2 + \frac{1}{6}\big(\frac{1}{n^{1/2}} C^3 B_{1n}(0) + \frac{1}{n} C^4 B_{2n}(n^{-1/2}C)\big)\Big) \\
= & -\frac{1}{2} C\|u\| + \Delta_n^T u + \frac{\|u\|}{C}\frac{1}{6}\big(\frac{1}{n^{1/2}} C^3 B_{1n}(0) + \frac{1}{n} C^4 B_{2n}(n^{-1/2}C)\big).
\end{aligned}$$

Hence on the event $\mathcal{A}$, for sufficiently large n, we have

$$\begin{aligned}
& \log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2 \\
\le & -\frac{p}{2}\log(2\pi) - \frac{1}{2}C\|u\| + M\sqrt{p}\|u\| + \frac{\|u\|}{C}\frac{1}{6}\big(\frac{1}{n^{1/2}} C^3 B_{1n}(0) + \frac{1}{n} C^4 B_{2n}(n^{-1/2}C)\big) \\
= & -\frac{p}{2}\log(2\pi) - \frac{1}{2}C\|u\|\Big(1 - \frac{2M}{C}\sqrt{p} - \frac{1}{3}\big(\frac{1}{n^{1/2}} C B_{1n}(0) + \frac{1}{n} C^2 B_{2n}(n^{-1/2}C)\big)\Big) \\
\le & -\frac{p}{2}\log(2\pi) - \frac{1}{4}C\|u\|.
\end{aligned}$$

Hence the second term of (**??**) can be bounded by

$$\begin{aligned}
& \int_{\|u\|>C} \exp\big\{\log Z_n(u) - \frac{p}{2}\log(2\pi) - \frac{1}{2}\|\Delta_n\|^2\big\} \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)}\, du \\
\le & \int_{\|u\|>C} \exp\Big\{-\frac{p}{2}\log(2\pi) - \frac{1}{4}C\|u\|\Big\} \frac{\pi(\theta_0 + n^{-1/2}\mathbf{J}^{-1}u)}{\pi(\theta_0)}\, du \\
= & \int_{\|\mathbf{J}(\theta-\theta_0)\|>n^{-1/2}C} \exp\Big\{-\frac{p}{2}\log(2\pi) - \frac{\sqrt{n}}{4}C\|\mathbf{J}(\theta-\theta_0)\|\Big\} \frac{\pi(\theta)}{\pi(\theta_0)} n^{p/2}|\mathbf{J}|\, d\theta \\
= & \exp\Big\{\frac{p}{2}\log\frac{n}{2\pi} + \frac{1}{2}\log|\psi''(\theta_0)|\Big\} \int_{\|\mathbf{J}(\theta-\theta_0)\|>n^{-1/2}C} \exp\Big\{-\frac{\sqrt{n}}{4}C\|\mathbf{J}(\theta-\theta_0)\|\Big\} \frac{\pi(\theta)}{\pi(\theta_0)}\, d\theta.
\end{aligned}$$

This proves the theorem. □

4.2. fractional posterior Bayes factor

Here k may be less than 1. And the assumptions can be weakened.

For the models more general than the exponential families, the tail behavior of the likelihood is hard to control. As a result, Bayes consistency is not trivial. We consider the general case. Suppose that we observe a random sample $X_1, \ldots, X_n$ from a distribution P_0 with densitu p relative to some reference measure μ on the sample space $(\mathbb{X}, \mathbb{A})$. Let P_0^n denote the expectation with respect to $X_1, \ldots, X_n$. Let μ^n denote the n-fold product measure of μ. Let p^n denote the density of P^n with respect to μ^n. Let $\mathbf{X}^n = (X_1, \ldots, X_n)$ be the pooled data. Suppose the model space is $\mathcal{P}$. Given some prior distribution Π on the set $\mathcal{P}$, the posterior distribution is the random measure given by

$$\Pi(B|X_1,\ldots,X_n) = \frac{\int_B \prod_{i=1}^n p(X_i)\, d\Pi_n(P)}{\int \prod_{i=1}^n p(X_i)\, d\Pi_n(P)}. \tag{32}$$

To prove the consistency result, i.e., the posterior probability of $\{P : d(P_0, P) > \epsilon\}$ ($d(\cdot,\cdot)$ is certain distance) tends to 0, we need to lower bound the denominator of (**??**) and upper bound the numerator of (**??**). There is a commonly used method for lower bounding the denominator. The following lemma is adapted from **?** and **?**. Let

$$D_{KL}(P||Q) = P\log(dP/dQ), \quad V(P||Q) = \mathrm{Var}_P\big(\log(dP/dQ)\big).$$

Lemma 1. *Let $\alpha > 0$ and $\epsilon > 0$. Let*

$$A_\epsilon = \{P : D_{KL}(P_0, P) \le \epsilon,\, V(P_0||P) \le \epsilon\}.$$

Then for every prior probability measure Π and every $C > 0$, we have

$$P_0^n\Big(\int_{\mathscr{P}} \Big[\frac{p^n}{p_0^n}(\mathbf{X}^n)\Big]^\alpha d\Pi(P) < \Pi(A_\epsilon)\exp\big(-(1+C)n\epsilon\big)\Big) \le \frac{\alpha^2}{C^2 n\epsilon}$$

Proof. Without loss of generality, we assume $\Pi(A_\epsilon) > 0$. Let Π_ϵ be the restriction of Π on A_ϵ. Then

$$\begin{aligned}
&P_0^n\Big(\int_{\mathscr{P}} \Big[\frac{p^n}{p_0^n}(\mathbf{X}^n)\Big]^\alpha d\Pi(P) < \Pi(A_\epsilon)\exp\big(-(1+C)n\epsilon\big)\Big)\\
\le& P_0^n\Big(\log\int_{\mathscr{P}} \Big[\frac{p^n}{p_0^n}(\mathbf{X}^n)\Big]^\alpha d\Pi_\epsilon(P) < -(1+C)n\epsilon\Big)\\
\le& P_0^n\Big(\int \alpha\log\frac{p^n}{p_0^n}(\mathbf{X}^n)\, d\Pi_\epsilon(P) < -(1+C)n\epsilon\Big)\\
=& P_0^n\Big(\sum_{i=1}^n \int \log\frac{p}{p_0}(X_i)\, d\Pi_\epsilon(P) < -(1+C)n\epsilon/\alpha\Big)\\
\le& P_0^n\Big(\sum_{i=1}^n \int \log\frac{p}{p_0}(X_i) + D_{KL}(P_0||P)\, d\Pi_\epsilon(P) < -Cn\epsilon/\alpha\Big)\\
\le& \frac{\alpha^2}{C^2n^2\epsilon^2} nP_0\Big(\int \log\frac{p}{p_0} + D_{KL}(P_0||P)\, d\Pi_\epsilon(P)\Big)^2\\
\le& \frac{\alpha^2}{C^2n\epsilon^2} P_0\int\Big(\log\frac{p}{p_0} + D_{KL}(P_0||P)\Big)^2 d\Pi_\epsilon(P)\\
=& \frac{\alpha^2}{C^2n\epsilon^2}\int P_0\Big(\log\frac{p}{p_0} + D_{KL}(P_0||P)\Big)^2 d\Pi_\epsilon(P)\\
=& \frac{\alpha^2}{C^2n\epsilon^2}\int V(P_0||P)\, d\Pi_\epsilon(P) \le \frac{\alpha^2}{C^2n\epsilon}.
\end{aligned}$$

□

The hard part is the numerator. **?** directly upper bound $p^n/p_0^n(X)$ to upper bound the numerator. **?** imposed a test condition to upper bound the numerator. If no additional assumption is adopted, the numerator can not be bounded. In fact, there are counterexamples, see **?**.

4.3. The work of **?**

While the numerator of the posterior distribution is hard to control, a variation of posterior distribution is easier to control. This work is done by **?**.

For density f_1 and f_2, let

$$H(f_1, f_2) = \Big(\int (\sqrt{f_1} - \sqrt{f_2})^2 \, d\mu\Big)^{1/2} = \Big(2 - 2\int \sqrt{f_1 f_2} \, d\mu\Big)^{1/2},$$

the Hellinger distance of f_1 and f_2. For $0 < \alpha < 1$, Hellinger integral is defined as

$$\rho_\alpha(f_1, f_2) = \int_{\mathcal{X}} f_1^\alpha f_2^{1-\alpha} \, d\mu.$$

For $0 < \alpha < 1$, define the pseudoposterior distribution Q based on Π as

$$Q^n(A) = \frac{\int_A \big[p^n(\mathbf{X}^n)\big]^\alpha \, d\Pi_n(P)}{\int_{\mathscr{P}} \big[p^n(\mathbf{X}^n)\big]^\alpha \, d\Pi_n(P)}.$$

Theorem 7. *Suppose* $\Pi_n\Big(A_\epsilon\Big) > 0$ *for every* $\epsilon > 0$*, where*

$$A_\epsilon = \{P : D_{KL}(P_0, P) \le \epsilon,\ V(P_0||P) \le \epsilon\}.$$

Then for every $\epsilon > 0$ *and* $C > 0$

$$P_0^n\Big\{Q^n\big(\rho_\alpha(P, P_0) \le 1 - \epsilon\big)\Big\} \le \frac{1}{\Pi(A_{\frac{\epsilon}{2(1+C)}})} \exp(-\frac{1}{2}\epsilon n) + \frac{2(1+C)\alpha^2}{C^2 n\epsilon}.$$

Proof. Consider the expactation of the numerator,

$$\begin{aligned}
&P_0^n \int_{\rho_\alpha(P,P_0)\le 1-\epsilon} \Big[\frac{p^n}{p_0^n}(\mathbf{X}^n)\Big]^\alpha \, d\Pi_n(P)\\
=&\int_{\rho_\alpha(P,P_0)\le 1-\epsilon} \int_{\mathcal{X}^n} \Big[\frac{p^n}{p_0^n}(\mathbf{X}^n)\Big]^\alpha p_0^n(\mathbf{X}^n) \, d\mu^n \, d\Pi_n(P)\\
=&\int_{\rho_\alpha(P,P_0)\le 1-\epsilon} \int_{\mathcal{X}^n} \Big[p^n(\mathbf{X}^n)\Big]^\alpha \Big[p_0^n(\mathbf{X}^n)\Big]^{1-\alpha} \, d\mu^n \, d\Pi_n(P)\\
=&\int_{\rho_\alpha(P,P_0)\le 1-\epsilon} \big(\rho_\alpha(p, p_0)\big)^n \, d\Pi_n(P)\\
\le&\exp(-\epsilon n).
\end{aligned}$$

Consider the denominator. From Lemma **??**, on a set B with $P_0^n(B) > 1 - \alpha^2/(C^2 n\epsilon')$, we have

$$\int_{\mathscr{P}} \Big[\frac{p^n}{p_0^n}(\mathbf{X}^n)\Big]^\alpha \, d\Pi(P) \ge \Pi(A_{\epsilon'}) \exp\big(-(1+C)\epsilon' n\big).$$

Hence

$$\begin{aligned}
&P_0^n\Big\{Q^n\big(\rho_\alpha(P, P_0) \le 1 - \epsilon\big)\Big\}\\
\le&P_0^n\Big\{\mathbf{1}_B Q^n\big(\rho_\alpha(P, P_0) \le 1 - \epsilon\big)\Big\} + P_0^n(B^C)\\
\le&\frac{1}{\Pi(A_{\epsilon'})} \exp(-\epsilon n + (1+C)\epsilon' n) + \frac{\alpha^2}{C^2 n\epsilon'}.
\end{aligned}$$

Let $\epsilon' = \frac{\epsilon}{2(1+C)}$, we have

$$P_0^n\Big\{Q^n\big(\rho_\alpha(P, P_0) \le 1 - \epsilon\big)\Big\} \le \frac{1}{\Pi(A_{\frac{\epsilon}{2(1+C)}})} \exp(-\frac{1}{2}\epsilon n) + \frac{2(1+C)\alpha^2}{C^2 n\epsilon}.$$

□

One deficit of the theorem is that it does not satisfactorily cover finite-dimensional models. When applied to such models, it would yield the rate $1/\sqrt{n}$ times a logarithmic factor rather than $1/\sqrt{n}$ itself.

Next we consider finite-dimensional models. Let $\{p_\theta : \theta \in \Theta\}$ be a family of densities parametrized by a Euclidean parameter θ running through an open set $\Theta \subset \mathbb{R}^p$. Assume that for every $\theta, \theta_1, \theta_2 \in \Theta$ and some $\alpha > 0$, there exists positive constants C_1, C_2, C_3, C_4, such that

$$
\begin{aligned}
&D_{KL}(p_{\theta_0}||p_\theta) \le C_1\|\theta-\theta_0\|^{2\alpha}\\
&V(p_{\theta_0}||p_\theta) \le C_1\|\theta-\theta_0\|^{2\alpha}\\
&C_3\|\theta_1-\theta_2\|^{2\alpha} \le 1-\rho_\alpha(p_\theta,p_{\theta_0}) \le C_4\|\theta_1-\theta_2\|^{2\alpha}
\end{aligned}
$$

The C_4 seems useless, and we can assume that the third inequality only holds locally. The proof of the following theorem is similar to the corresponding nonparametric one.

Theorem 8. *Under the conditions listed previously and θ_0 interior to Θ, then for any $M_n \to \infty$,*

$$P_0^n\Big\{Q^n\big(\|\theta-\theta_0\| \ge \frac{M_n}{n^{\frac{1}{2\alpha}}}\big)\Big\} \to 0.$$

Proof. Without loss of generality, we assume $\frac{M_n}{n^{\frac{1}{2\alpha}}} \to 0$, otherwise we replace M_n by a smaller one. Consider the expactation of the numerator,

$$
\begin{aligned}
&P_0^n\int_{\|\theta-\theta_0\|\ge\frac{M_n}{n^{\frac{1}{2\alpha}}}}\Big[\frac{p_\theta^n}{p_0^n}(\mathbf{X}^n)\Big]^\alpha d\Pi_n(\theta)\\
=&\int_{\|\theta-\theta_0\|\ge\frac{M_n}{n^{\frac{1}{2\alpha}}}}\int_{\mathcal{X}^n}\Big[\frac{p_\theta^n}{p_0^n}(\mathbf{X}^n)\Big]^\alpha p_0^n(\mathbf{X}^n)\,d\mu^n\,d\Pi_n(\theta)\\
=&\int_{\|\theta-\theta_0\|\ge\frac{M_n}{n^{\frac{1}{2\alpha}}}}\int_{\mathcal{X}^n}\Big[p_\theta^n(\mathbf{X}^n)\Big]^\alpha\Big[p_0^n(\mathbf{X}^n)\Big]^{1-\alpha} d\mu^n\,d\Pi_n(\theta)\\
=&\int_{\|\theta-\theta_0\|\ge\frac{M_n}{n^{\frac{1}{2\alpha}}}}\big(\rho_\alpha(p_\theta,p_{\theta_0})\big)^n d\Pi_n(\theta)\\
=&\sum_{j=1}^{+\infty}\int_{\frac{jM_n}{n^{\frac{1}{2\alpha}}}\le\|\theta-\theta_0\|\le\frac{(j+1)M_n}{n^{\frac{1}{2\alpha}}}}\big(\rho_\alpha(p_\theta,p_{\theta_0})\big)^n d\Pi_n(\theta)\\
\le&\sum_{j=1}^{+\infty}\int_{\frac{jM_n}{n^{\frac{1}{2\alpha}}}\le\|\theta-\theta_0\|\le\frac{(j+1)M_n}{n^{\frac{1}{2\alpha}}}}\big(1-C_3(\frac{jM_n}{n^{\frac{1}{2\alpha}}})^{2\alpha}\big)^n d\Pi_n(\theta)\\
\lesssim&\sum_{j=1}^{+\infty}\Big[\frac{(j+1)M_n}{n^{\frac{1}{2\alpha}}}\Big]^p\exp[-C_3\Big(\frac{jM_n}{n^{\frac{1}{2\alpha}}}\Big)^{2\alpha}n]\\
=&\sum_{j=1}^{+\infty}\Big[\frac{(j+1)M_n}{n^{\frac{1}{2\alpha}}}\Big]^p\exp[-C_3\Big(jM_n\Big)^{2\alpha}]
\end{aligned}
$$

Consider the denominator. From Lemma **??**, on a set B with $P_0^n(B) > 1-\alpha^2/(C^2n\epsilon')$, we have

$$\int_\Theta\Big[\frac{p_\theta^n}{p_0^n}(\mathbf{X}^n)\Big]^\alpha d\Pi(\theta) \ge \Pi(A_{\epsilon'})\exp\big(-(1+C)\epsilon' n\big) \ge \Pi\Big(\Big\{\theta : \|\theta-\theta_0\| \le (\epsilon'/C_1)^{\frac{1}{2\alpha}}\Big\}\Big)\exp\big(-(1+C)\epsilon' n\big).$$

Let $\epsilon' = \frac{C_3M_n^{2\alpha}}{2(1+C)n}$, we have

$$\int_\Theta\Big[\frac{p_\theta^n}{p_0^n}(\mathbf{X}^n)\Big]^\alpha d\Pi(\theta) \gtrsim \Big[\frac{M_n}{n^{\frac{1}{2\alpha}}}\Big]^p\exp\big(-\frac{1}{2}C_3M_n^{2\alpha}\big).$$

Hence

$$
\begin{aligned}
&P_0^n\Big\{Q^n\big(\|\theta-\theta_0\| \geq \frac{M_n}{n^{\frac{1}{2\alpha}}}\big)\Big\} \\
\leq &P_0^n\Big\{\mathbf{1}_B Q^n\big(\|\theta-\theta_0\| \geq \frac{M_n}{n^{\frac{1}{2\alpha}}}\big)\Big\} + P_0^n(B^C) \\
\leq &\sum_{j=1}^{+\infty}(j+1)^p \exp[-\frac{1}{2}C_3\Big(jM_n\Big)^{2\alpha}] + \frac{2(1+C)\alpha^2}{C^2C_3M_n^{2\alpha}} \to 0.
\end{aligned}
$$

□

4.4. The choice of the weight function

L^1 approximation of posterior by normal.

5. Appendix

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: $\Omega_n \mapsto \mathbb{R}^k$, 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.

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.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. Chapman and Hall/CRC, 2013.

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

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

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