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.

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

[☆]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.

example, xxxxxx. Among them, Aitkin (1991) proposed posterior Bayes factor (PBF). 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). 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 parameter space Θ is an open subset of $\mathbb{R}^p$. Let $\theta = (\nu^T, \xi^T)^T$, where ν is a p_0 dimensional subvector, and ξ is a $p - p_0$ dimensional subvector. The null space Θ_0 is a p_0-dimensional subspace of Θ defined as

$$\Theta_0 = \{(\nu^T, \xi^T)^T : (\nu^T, \xi^T)^T \in \Theta, \, \xi = \xi_0\}. \tag{2}$$

We would like to test the hypothesis

$$H_0 : \xi = \xi_0. \tag{3}$$

Let $\tilde{\Theta}_0 = \{\nu : (\nu^T, \xi^T)^T \in \Theta_0\}$.

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

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

$$\Lambda_P(\mathbf{X}^{(n)}) = \frac{L_2(\mathbf{X}^{(n)})}{L_1(\mathbf{X}^{(n)})} \cdot \frac{L_1^*(\mathbf{X}^{(n)})}{L_2^*(\mathbf{X}^{(n)})}.$$

$$\Lambda_F(\mathbf{X}^{(n)}) = \frac{L_1(\mathbf{X}^{(n)})}{L_{1/2}(\mathbf{X}^{(n)})} \cdot \frac{L_{1/2}^*(\mathbf{X}^{(n)})}{L_1^*(\mathbf{X}^{(n)})}.$$

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

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

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.

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

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

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).

Let $\pi(\theta; \mathbf{X}^{(n)})$ and $\pi(\nu; \mathbf{X}^{(n)})$ be the weight functions in Θ and $\tilde{\Theta}_0$. The integrated likelihood ratio statistic is defined as

$$\Lambda(\mathbf{X}^{(n)}) = \frac{\int_\Theta \left[p_n(\mathbf{X}^{(n)}|\theta)\right]^a \pi(\theta; \mathbf{X}^{(n)})\, d\theta}{\int_{\tilde{\Theta}_0} \left[p_n(\mathbf{X}^{(n)}|\nu, \xi_0)\right]^a \pi(\nu; \mathbf{X}^{(n)})\, d\nu}, \tag{5}$$

3. Asymptotic behavior of FBF

In this section, we consider the general FBF

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

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.

We denote by $\rightsquigarrow$ the weak convergence. Let $\mathbf{X}^{(n)}$ denote the data.

Suppose θ_0 is the true parameter.

Denote by P_0 the true distribution of $\mathbf{X}$. Let $p_n(\mathbf{X}^{(n)}|\theta)$ be the density of P_θ^n with respect to measure μ^n.

Let

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

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. The corresponding quantities in the null space are

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

Assumption 1. *The parameter space Θ is an open subset of $\mathbb{R}^p$. Thue null space $\tilde{\Theta}_0$ is an open subset of $\mathbb{R}^{p_1}$. The true 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 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).

Proposition 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 $M > 0$*

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

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

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

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

Define

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

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

Theorem 1. *Suppose that Assumption 1 holds, $L_a(\mathbf{X}^{(n)})$, $L_b(\mathbf{X}^{(n)})$, $L_a^*(\mathbf{X}^{(n)})$ and $L_b^*(\mathbf{X}^{(n)})$ are 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 $\sqrt{n}(\theta_n - \theta_0) \to \eta$,*

$$\log \Lambda_{a,b}(\mathbf{X}^{(n)}) \overset{P_{\theta_n}^n}{\rightsquigarrow} -\frac{p-p_1}{2}\log\frac{a}{b} + \frac{a-b}{2}\chi^2_{p-p_1}(\delta),$$

where $\chi^2_{p-p_1}(\delta)$ is a random variable with chi-squared distribution with $p-p_1$ degrees of freedom and non-centrality parameter $\delta = \eta^T\big(I_{\theta_0} - I_{\theta_0}J(J^T I_{\theta_0} J)^{-1}J^T I_{\theta_0}\big)\eta$ and $J = (I_{p_1}, 0_{p_1\times(p-p_1)})^T$.

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

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

By Proposition 1,

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

Thus

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

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

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

Thus,

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

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

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

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

$$\log L_t^*(\mathbf{X}^{(n)}) = \log \int_{\tilde{\Theta}_0} \left[p_n(\mathbf{X}^{(n)}|\nu, \xi_0)\right]^t \pi(\nu)\, d\nu$$
$$= \frac{p_1}{2}\log\big(\frac{2\pi}{n}\big) - \frac{p_1}{2}\log t - \frac{1}{2}\log|I_{\theta_0}^*| + \log\pi(\nu_0) + \frac{t}{2}\Delta_{n,\theta_0}^{*T} I_{\theta_0}^*\Delta_{n,\theta_0}^* + t\log p_n(\mathbf{X}^{(n)}|\theta_0) + o_{P_{\theta_0}^n}(1).$$

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

$$\begin{aligned}\log\Lambda_{a,b}(\mathbf{X}^{(n)}) =& \log L_a(\mathbf{X}^{(n)}) - \log L_b(\mathbf{X}^{(n)}) - \log L_a^*(\mathbf{X}^{(n)}) + \log L_b^*(\mathbf{X}^{(n)}) \\ =& -\frac{p-p_1}{2}\log\frac{a}{b} + \frac{a-b}{2}\Big(\Delta_{n,\theta_0}^T I_{\theta_0}\Delta_{n,\theta_0} - \Delta_{n,\theta_0}^{*T} I_{\theta_0}^*\Delta_{n,\theta_0}^*\Big) + o_{P_{\theta_n}^n}(1).\end{aligned}$$

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.

$\square$

The key assumption of 1 is the consistency of $L_a(\mathbf{X}^{(n)})$, $L_b(\mathbf{X}^{(n)})$, $L_a^*(\mathbf{X}^{(n)})$ and $L_b^*(\mathbf{X}^{(n)})$. Next we consider the consistency of $L_t(\mathbf{X}^{(n)})$.

Proposition 2. *Under Assumption 1, for $0 < b < a$, if L_a is consistent, then L_b is consistent.*

We would like to investigate the asymptotic behavior of FBF in exponential family. Exponential family possesses many good properties. It can be seen that the full-rank exponential family fullfill Assumption 1.

Proposition 3. *Suppose* $p(X|\theta) = \exp\big[\theta^T T(X) - A(\theta)\big]$, 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* $t > 0$.

Proof. 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

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

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

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

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,

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

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

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

Thus,

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

□

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

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 $H_1:\theta\neq0$. 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, $i = 1, \ldots, n$. 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

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

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

There are many works give Bernstein-von Mises type theorems, which assert that the posterior distribution of $h = \sqrt{n}(\theta - \theta_0)$ converges to $N(\Delta_{n,\theta_0}, I_{\theta_0}^{-1})$, the normal distribution with mean Δ_{n,θ_0} and variance $I_{\theta_0}^{-1}$. However, most existing work consider the convergence under the total variation distance, that is

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

Or Hellinger distance. Blabla.

The consistency of L_1 is equivalent to the consistency of posteriror distribution. There have been substantial work on the consistency of posterior distribution. See xxx. The consistency of posterior distribution needs further assumptions in addition to Assumption 1. A popular assumption which is adopted by Vaart et. al. is the existence of a consistent test.

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

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

The following proposition is adapted from Vaart.

Proposition 4. *Suppose θ_0 is an interior of Θ, $\pi(\theta)$ is continuous at θ_0 and $\pi(\theta_0) > 0$. Under Assumptions 1 and 2, L_1 is consistent.*

The consistency of L_t ($t < 1$) is different from the consistency of posterior distribution. XXX considered the consistency of $L_{1/2}$. However, they didn't track the convergence rate. We shall prove the consisency of L_t for $0 < t < 1$.

To show our results, some notations are needed. For two parameters θ_1 and θ_2, the Rényi divergence of order α ($0 < \alpha < 1$) 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.

We impose the following assumption on the family $\{P_\theta : \theta \in \Theta\}$.

Assumption 3. *There exist positive constancts δ, ϵ and C such that, $D_t(\theta||\theta_0) \geq C\|\theta - \theta_0\|^2$ for $\|\theta - \theta_0\| \leq \delta$ and $D_t(\theta||\theta_0) \geq \epsilon$ for $\|\theta - \theta_0\| > \delta$.*

Assumption 3 is reasonable. For example, if $\Theta = \mathbb{R}^p$ and $X_i \sim N_p(\mu, I_p)$

Theorem 2. *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.*

Proof. Note that

$$\frac{L_t(\{\theta : \|\theta-\theta_0\| \geq \frac{M_n}{\sqrt{n}}\})}{L_t} = \frac{\int_{\left\{\theta:\|\theta-\theta_0\|\geq \frac{M_n}{\sqrt{n}}\right\}} \left[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^t \pi(\theta)\, d\theta}{\int_\Theta \left[\frac{p_n(\mathbf{X}^{(n)}|\theta)}{p_n(\mathbf{X}^{(n)}|\theta_0)}\right]^t \pi(\theta)\, d\theta}. \tag{7}$$

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

Consider the expactation of the numerator of 7. It follows from Fubini's theorem that

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

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

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

Now we consider the denominator of (7).

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

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

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

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

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

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

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

Since ϵ is arbitrary, the theorem follows. □

Proposition 5. *If $a=1$, then blabla*

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.

Theorem 3. *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 would like to consider more general weight functions.

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

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

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

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

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

The condition 9 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. 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 4. *Under Assumptions 1-4, 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 4. 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

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

By Theorem 1,

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

Hence we have

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

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

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

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

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

Let ϕ_n be a test function satisfying the conclusion of Theorem 3. 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 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. □

Based on Theorem 4,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 5. *Suppose the assumptions of 4 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 5. 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 4 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

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

But

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

$I^{-\frac{1}{2}}_{\theta_0}\xi$ 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 6. *Suppose the Assumptions of 5 are met. The true parameter θ satisfies $\eta_n = \sqrt{n}(\theta - \theta_0) \to \eta$. If*

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

$I_{22\cdot 1} = I_{22} - I_{21}I^{*-1}_{\theta_0}I_{12}$*, 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 $\Delta_{n,\theta_0} \rightsquigarrow N(h, I^{-1}_{\theta_0})$. In the limit experiment, we have one observation $X \sim N(h, I^{-1}_{\theta_0})$. In this case, the integrated likelihood ratio test statistics can be calculated easily whose distribution is exactly the same as 6.

Proof of Theorem 6. We note that $h_n = \eta_n$ converges to η. By Proposition 1 and CLT,

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

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 4, under $P^n_{\eta_n}$, we have (20). Hence

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

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

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

□

5.1. 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.

6. 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/`.