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 we are interested in testing the hypotheses $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$. The well known LRT is defined as

$$\frac{\sup_{\Theta} p_\theta(X)}{\sup_{\Theta_0} p_\theta(X)}, \tag{1}$$

where X is the data, $p_\theta(X)$ is the density function of X with respect to some dominating measure μ. LRT is the most widely used statistical method which

[☆]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.

enjoys many optimal properties. For example, by Neyman-Pearson lemma, it's the most powerful test (MPT) in simple null and simple alternative case. (See Lehmann J. P. R (2005)) In multi-dimensional parameter case, MPT does not exist. However, the LRT is asymptotic optimal in Bahadur efficiency sense. (See 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 maximum. To overcome these difficulties, we propose the following ILRT

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

where $\pi(\theta; X)$ and $\pi^*(\theta; X)$ are weight functions which may rely on data but does not need to equal to posterior density.

Aitkin (1991) proposed posterior Bayes factor

$$\frac{\int_{\Theta} p_\theta(X)\pi(\theta|X)\, d\theta}{\int_{\Theta_0} p_\theta(X)\pi^*(\theta|X)\, d\theta}, \tag{3}$$

where $p(x|\theta)$ is the sample density, $\pi^*(\theta|x)$ and $\pi(\theta|x)$ are the posterior densities under null hypotheses and alternative hypothesis. Gelfand D. K. D (1993) derived it's null distribution. 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 if MLE exists, LRT also exists. 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

Let $X_1, \ldots, X_n$ be a sample from distribution $P_\theta, \theta \in \Theta$. Denote $X = (X_1, \ldots, X_n)$. 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 simplifies notations, we denote $\Theta_0^* = \{(\theta_1, \ldots, \theta_{p_1})^T : (\theta_1, \ldots, \theta_{p_1}, \theta_{0,p_1+1}, \theta_{0,p_2}) \in \Theta_0\}$. And we could use p_1-dimensional vector $\theta^* \in \Theta_0^*$ to represent $\theta \in \Theta_0$ and regard Θ_0^* as null space. Let $\pi(\theta; X)$ and $\pi^*(\theta^*; X)$ be the weight functions in Θ and Θ_0^*. The integrated likelihood ratio statistic is defined as

$$\Lambda(X) = \frac{\int_\Theta p(X|\theta)\pi(\theta; X)\, d\theta}{\int_{\Theta_0^*} p(X|\theta^*)\pi^*(\theta^*; X)\, d\theta^*} \tag{6}$$

3. Main results

We will denote by $\rightsquigarrow$ the weak convergence. $dN(\mu, \Sigma)(h)$ is the density function of a normal distribution with mean μ and variance Σ.

Assume $\theta_0 \in \Theta_0$ is a fixed parameter in null space. We study the asymptotic behavior of integrated likelihood ratio statistics around θ_0. Let the experiment $P_\theta : \theta \in \Theta$ be differentiable in quadratic mean at θ_0. That is, there exists a vector of measurable functions $\dot{\ell}_{\theta_0}$ such that

$$\int \left[\sqrt{p_{\theta_0+h}} - \sqrt{p_{\theta_0}} - \frac{1}{2} h^T \dot{\ell}_{\theta_0} \sqrt{p_{\theta_0}}\right]^2 d\mu = o(\|h\|^2), \quad h \to 0, \tag{7}$$

where μ is the controlling measure of P_θ, $p_\theta(x)$ is the density of $P_\theta(x)$ relative to μ and $\dot{\ell}_{\theta_0}$ is called score function.

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

Let $h = \sqrt{n}(\theta - \theta_0)$ which reparameterizes θ around θ_0 by the scale of $\sqrt{n}$. Obviously, $h = 0$ under null. Under local alternatives, h converges to a constant.

Listed below are the regular conditions we need:

Assumption 1. *The assumptions of Bernstein-von Mises Theorem (see der Vaart (2000) Chapter 10) are met: Θ is a subset of $\mathbb{R}^p$. The experiment $(P_\theta : \theta \in \Theta)$ is differentiable in quadratic mean at $\theta_0 \in \Theta$ with nonsingular Fisher information matrix I_{θ_0}. For every $\epsilon > 0$, there exists a sequence of tests ϕ_n such that*

$$P^n_{\theta_0}\phi_n \to 0, \qquad \sup_{\|\theta-\theta_0\|\geq\epsilon} P^n_\theta(1-\phi_n) \to 0. \tag{8}$$

Let $\pi_n(h;X)$ be a weight function satisfying

$$\|\pi_n(h;X) - dN(\Delta_{n,\theta_0}, I^{-1}_{\theta_0})(h)\| \overset{P^n_{\theta_0}}{\to} 0 \tag{9}$$

Assumption 2. *For every $\epsilon > 0$, there's a Lebesgue integrable function $T(h)$, a $K > 0$ and a $A > 0$ such that*

$$\lim_{n\to\infty} P^n_{\theta_0}(\sup_{\|h\|\geq K\sqrt{n}} (\pi_n(h;X) - T(h)) \leq 0) \geq 1 - \epsilon \tag{10}$$

$$\lim_{n\to\infty} P^n_{\theta_0}(\sup_{\|h\|\leq K\sqrt{n}} \pi_n(h;X) \leq A) \geq 1 - \epsilon \tag{11}$$

Assumption 3. *There's a open neighborhood of θ_0 V and a measurable function $\dot{\ell}$ with $P_{\theta_0}\dot{\ell}^2 < \infty$ such that, $\forall \theta_1, \theta_2 \in V$,*

$$|\log p_{\theta_1}(x) - \log p_{\theta_2}(x)| \leq \dot{\ell}(x)\|\theta_1 - \theta_2\|. \tag{12}$$

Assumption 1 makes sure that there exists at least one weight function satisfies 9, that is, the posterior density. The condition 10 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 10 to weight function instead. The condition 11 is satisfied in most usual case. No matter model is, condition 10 and 11 will be satisfied, e.g., when

$$\pi_n(h;X) = \min(\pi_n(h|X), M)1_{\|h\|\leq K\sqrt{n}} \tag{13}$$

where M and M are user-specified constant and $\pi_n(h|X)$ is the posterior density. Assumption 3 is standard in likelihood theory.

Our first theorem is

Theorem 1. *Under the Assumptions (i), (ii) and (iii). For bounded real numbers η_n, we have*

$$\Big|\int_{\mathbb{R}^p} \frac{p_h(X)}{p_0(X)}\pi_n(h;X)\,dh - 2^{-\frac{p}{2}}e^{\frac{1}{2}\Delta_{n,\theta_0}^T I_{\theta_0}\Delta_{n,\theta_0}}\Big| \xrightarrow{P_{\eta_n}^n} 0 \tag{14}$$

Based on Theorem 1,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 2. *Suppose the Assumptions of 1 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*

$$2\log(\Lambda(X)) \overset{P_0^n}{\rightsquigarrow} \chi^2_{p_2-p_1} - (p_2-p_1)\log(2) \tag{15}$$

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

Theorem 3. *Suppose the Assumptions of 2 are met. The true parameter θ satisfies $\eta_n = \sqrt{n}(\theta-\theta_0) \to \eta$. If*

$$I_{\theta_0} = \begin{pmatrix} I_{\theta_0}^* & I_{12} \\ I_{21} & I_{22} \end{pmatrix}, \tag{16}$$

$I_{22\cdot1} = I_{22} - I_{21}I_{\theta_0}^{*-1}I_{12}$, then we have

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

where

$$\delta = \eta^T \begin{pmatrix} 0 & 0 \\ 0 & I_{22\cdot1} \end{pmatrix} \eta \tag{18}$$

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 3.

4. Normal mixture

Normal mixture is the first example of unbounded likelihood given in Cam (1990). In this section, we apply ILRT to this model.

Suppose $X_1, \ldots, X_n$ are i.i.d. distributed as a mixture of normal distributions

$$p_\theta(x) = \frac{1-\alpha}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\mu)^2\right\} + \frac{\alpha}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}, \tag{19}$$

where α is a known constant. Suppose the parameter space is

$$\Theta = \{\theta = (\mu, \sigma^2)^T : \mu \in (\infty, \infty), \sigma^2 \in (0, M)\}, \tag{20}$$

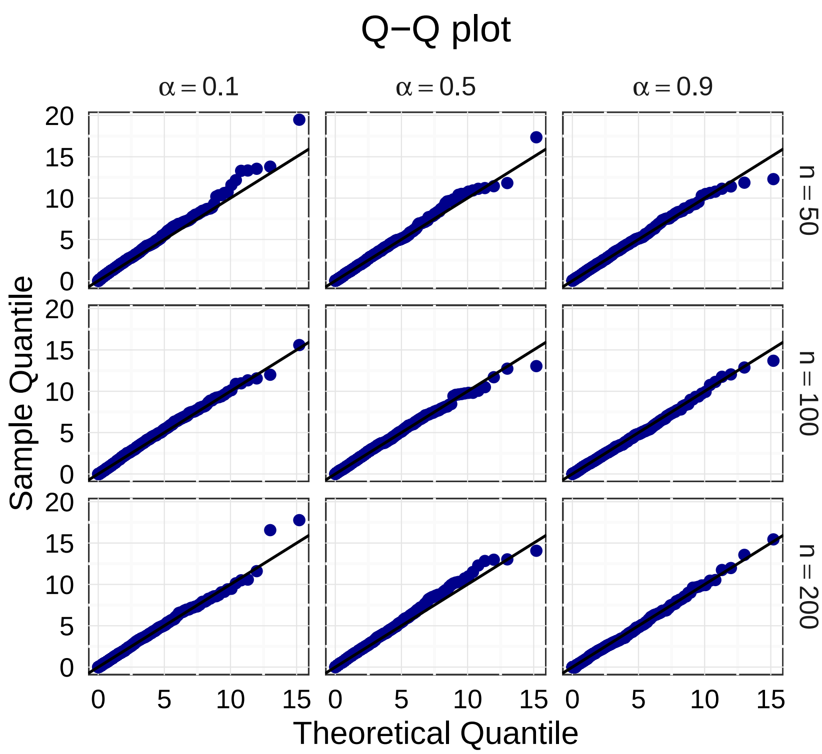
where M is a sufficiently large parameter. Cam (1990) pointed out that the likelihood of the model is unbounded. In fact, let $\mu = X_1$ and let $\sigma^2 \to 0$, then the likelihood tends to infinity.

Under the model, we consider testing the hypotheses $H_0 : \mu = 0, \sigma = 2$ vs $H_1 : \theta \in \Theta$. Although LRT fails in this model, ILRT can still be used. In fact, we can verify (although cumbersome) that (i), (ii) and (iii) all hold. Hence Theorem 2 and 3 hold. To use ILRT, we need a weight function. First let the weight function be the posterior density of parameters. To make the posterior density bounded from infinity, we consider the so-called zero-avoiding prior (see Andrew Gelman (2013) section 13.2). When $\mu = X_1$, as $\sigma^2 \to 0$, the likelihood tends to infinity at the rate of $1/\sigma$. The rate can be hedged by the density of $\chi^2(3)$. Hence we adopt the following prior distribution

$$dN(0,1)(\mu) \times d\chi^2(3)(\sigma^2), \tag{21}$$

where $d\chi^2(3)(\sigma^2)$ represents the density of χ^2 distribution with freedom 3 taking value at σ^2. Because the σ^2 is limited in $(0, M)$, we also truncate prior of σ^2 at M.

We next verify the Wilks phenomenon by simulation. We take sample size $n = 50, 100, 200$ and $\alpha = 0.1, 0.5, 0.9$. In every combination, we repeat 1000 samples and obtain 1000 ILRT statistics. We expect the empirical distribution is similar to that of $\chi^2(2)$. We plot the QQ-plot of empirical distribution relative to $\chi^2(2)$ distribution, it can be seen that ILRT can be well approximated by $\chi^2(2)$.



We next consider another weight function $\pi(\theta; X) = N(\hat{\theta}, \frac{1}{n}\hat{I}_{\hat{\theta}}^{-1})$. Lt $\hat{\theta}$ be the highest probability density estimator. And

$$\hat{I}_{\theta}^{-1} = \sum_{i=1}^{n} \begin{bmatrix} -\frac{\partial^2 \log p_\theta(x_i)}{\partial \mu^2} & -\frac{\partial^2 \log p_\theta(x_i)}{\partial \mu \partial(\sigma^2)} \\ -\frac{\partial^2 \log p_\theta(x_i)}{\partial \mu \partial(\sigma^2)} & -\frac{\partial^2 \log p_\theta(x_i)}{\partial(\sigma^2)^2} \end{bmatrix}$$

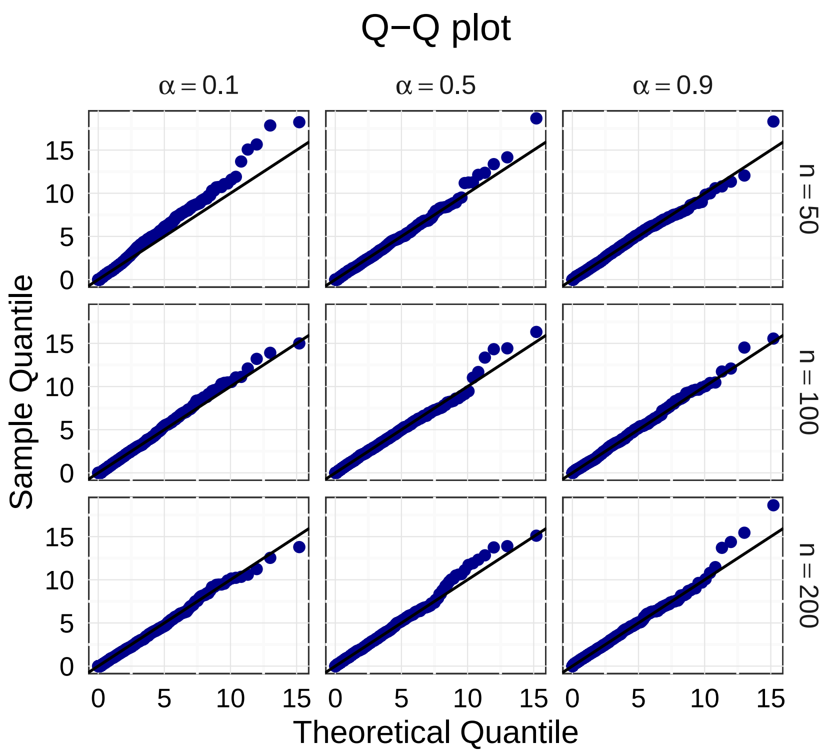
where

$$\frac{\partial^2 \log p_\theta(x)}{\partial \mu^2} = \frac{(1-\alpha)((x-\mu)^2-1)dN(\mu,1)(x)+\alpha((x-\mu)^2/\sigma^4-\sigma^{-2})dN(\mu,\sigma^2)(x)}{p_\theta(x)} - \Big(\frac{(1-\alpha)(x-\mu)dN(\mu,1)(x)+\alpha(x-\mu)/\sigma^2 dN(\mu,\sigma^2)(x)}{p_\theta(x)}\Big)^2, \tag{22}$$

$$\frac{\partial^2 \log p_\theta(x)}{\partial \mu \partial(\sigma^2)} = \frac{(\frac{3\alpha(\mu-x)}{2\sigma^4} - \frac{\alpha(\mu-x)^3}{2\sigma^6})dN(\mu,\sigma^2)(x)}{p_\theta(x)} - \frac{\alpha(\frac{(\mu-x)^2}{2\sigma^4} - \frac{1}{2\sigma^2})dN(\mu,\sigma^2)(x)\big((1-\alpha)(x-\mu)dN(\mu,1)(x)+\alpha(x-\mu)/\sigma^2 d(\mu,\sigma^2)(x)\big)}{p_\theta(x)^2}, \tag{23}$$

$$\frac{\partial^2 \log p_\theta(x)}{\partial (\sigma^2)^2} = \frac{\alpha\big(\frac{3}{4\sigma^4} - \frac{3(x-\mu)^2}{2\sigma^6} + \frac{(x-\mu)^4}{4\sigma^8}\big)dN(\mu,\sigma^2)(x)}{p_\theta(x)} - \Big(\frac{\alpha\big(\frac{(x-\mu)^2}{2\sigma^4} - \frac{1}{2\sigma^2}\big)dN(\mu,\sigma^2)(x)}{p_\theta(x)}\Big)^2. \tag{24}$$

We do the same simulation as above and the QQ-plot is given. It can be seen that the Wilks phenomenon still holds in this case. For mixture model, sampling from posterior distribution is troublesome. The computing burden will be reduce by normal approximation. This is an advantage of normal weight ILRT. From this example, we can see that ILRT is more flexible than posterior Bayes factor.



5. 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 $T_n\colon\ \Omega_n \mapsto \mathbb{R}^k$, we have $T_n \overset{P_n}{\rightsquigarrow} 0 \Leftrightarrow T_n \overset{Q_n}{\rightsquigarrow} 0$.

Lemma 1. *Suppose that Θ is an open subset of $\mathbb{R}^p$ and that the model $(P_\theta : \theta \in \Theta)$ is differentiable in quadratic mean at θ_0. Then $P_{\theta_0}\dot{\ell}_{\theta_0} = 0$ and the Fisher information matrix $I_{\theta_0} = P_{\theta_0}\dot{\ell}_{\theta_0}\dot{\ell}_{\theta_0}^T$ exists. Furthermore, for every converging sequence $h_n \to h$, as $n \to \infty$,*

$$\log \frac{p^n_{h_n}(X)}{p^n_0(X)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h^T \dot{\ell}_{\theta_0}(X_i) - \frac{1}{2} h^T I_{\theta_0} h + o_{P_{\theta_0}}(1), \tag{25}$$

where $p^n_h(X) = \prod_{i=1}^n p_h(X_i)$ is the density of P^n_h relative to $\mu_n = \mu \times \cdots \times \mu$. (See der Vaart (2000) Theorem 7.2.)

Lemma 2. *Suppose the Assumptions of Lemma 1 are met. U is a ball of fixed radius around zero. Then for every random variable sequence $T_n(X)$, $T_n \overset{P_0^n}{\rightsquigarrow} 0 \Leftrightarrow T_n \overset{P_U^n}{\rightsquigarrow} 0$, where*

$$p_U^n(x) = \frac{1}{V(U)} \int_U p_h^n(x) dh, \tag{26}$$

$V(U)$ is the volume of U.

Proof. In fact, we only need to prove

$$\int_{A_n} p_0^n(x)\, d\mu \to 0 \Leftrightarrow \int_{A_n} \frac{1}{V(U)} \int_U p_h^n(x) dh\, d\mu \to 0, \tag{27}$$

or

$$\int_{A_n} p_0^n(x)\, d\mu \to 0 \Leftrightarrow \int_U \int_{A_n} p_h^n(x) d\mu\, dh \to 0. \tag{28}$$

Under the assumptions of 1, for every bounded sequence h_n, $P_{h_n}^n \triangleleft\triangleright P_0^n$, that is

$$\int_{A_n} p_0^n(x)\, d\mu \to 0 \Leftrightarrow \int_{A_n} p_{h_n}^n(x) d\mu \to 0. \tag{29}$$

On the other hand, there exists sequence $\overline{h}_n$ such that

$$\int_U \int_{A_n} p_h^n(x) d\mu\, dh \leq V(U) \sup_{h \in U} \int_{A_n} p_h^n(x) d\mu \leq V(U)(\int_{A_n} p_{\overline{h}_n}^n(x) d\mu + 1/n). \tag{30}$$

We have similar lower bound. Hence,

$$V(U)(\int_{A_n} p_{\underline{h}_n}^n(x) d\mu + 1/n) \leq \int_U \int_{A_n} p_h^n(x) d\mu\, dh \leq V(U)(\int_{A_n} p_{\overline{h}_n}^n(x) d\mu + 1/n) \tag{31}$$

The (28) follows from (29) and (31). $\square$

Lemma 3. *Suppose the assumptions of Lemma 1 are met. Suppose that for every $\epsilon > 0$ there exists a sequence of tests ϕ_n such that*

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

Then there exists for every $M_n \to \infty$ a sequence of tests ϕ_n and a constant $c > 0$ such that, for every sufficiently large n and every $\|\theta - \theta_0\| \geq M_n/\sqrt{n}$,

$$P_{\theta_0}^n \phi_n \to 0, \quad P_\theta^n (1 - \phi_n) \leq e^{-cn(\|\theta - \theta_0\|^2 \wedge 1)}.$$

(See der Vaart (2000) Lemma 10.3.)

Lemma 4. *Suppose the assumptions of Lemma 1 are met. Further more, suppose there is an open neighborhood V of θ_0 and a function $m(x)$ with $P_{\theta_0} m^2 < \infty$ such that for all $\forall \theta_1, \theta_2 \in V$:*

$$|\log p_{\theta_1}(x) - \log p_{\theta_2}(x)| \leq m(x)\|\theta_1 - \theta_2\|. \tag{32}$$

Then for every $M > 0$,

$$\sup_{\|h\| \leq M} \left| \log \frac{p^n_{h_n}(X)}{p^n_0(X)} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h^T \dot{\ell}_{\theta_0}(X_i) + \frac{1}{2} h^T I_{\theta_0} h \right| \xrightarrow{P^n_0} 0. \tag{33}$$

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

Proof of Theorem 1. By contiguity, we only need to proof the convergence in P^n_0.

The proof consists of two steps. In the first part of the proof, let C be the ball of fixed radius M around zero. We proof

$$\left| \int_C \frac{p^n_h(X)}{p^n_0(X)} \pi_n(h; X)\, dh - \int_C e^{h^T I_{\theta_0} \Delta_{n,\theta_0} - \frac{1}{2} h^T I_{\theta_0} h} dN(\Delta_{n,\theta_0}, I^{-1}_{\theta_0})(h)\, dh \right| \xrightarrow{P^n_0} 0 \tag{34}$$

By Lemma 4, for every fixed M,

$$\sup_{\|h\| \leq M} |\log \frac{p^n_h(X)}{p^n_0(X)} - h^T I_{\theta_0} \Delta_{n,\theta_0} + \frac{1}{2} h^T I_{\theta_0} h| \xrightarrow{P^n_0} 0 \tag{35}$$

Hence we have

$$\int_C \frac{p^n_h(X)}{p^n_0(X)} \pi_n(h; X)\, dh = e^{o_{p^n_0}(1)} \int_C e^{h^T I_{\theta_0} \Delta_{n,\theta_0} - \frac{1}{2} h^T I_{\theta_0} h} \pi_n(h; X)\, dh \tag{36}$$

So we only need to consider $\int_C e^{h^T I_{\theta_0} \Delta_{n,\theta_0} - \frac{1}{2} h^T I_{\theta_0} h} \pi_n(h; X)\, dh$. Notice that C is a bounded region, hence Δ_{n,θ_0} converges to a normal distribution. Therefore, $\sup_{h \in C} e^{h^T I_{\theta_0} \Delta_{n,\theta_0} - \frac{1}{2} h^T I_{\theta_0} h}$ is bounded in probability. So for every $\delta > 0$, there exists M such that, with probability $1 - \delta$,

$$\begin{aligned} &\int_C e^{h^T I_{\theta_0} \Delta_{n,\theta_0} - \frac{1}{2} h^T I_{\theta_0} h} |\pi_n(h; X) - dN(\Delta_{n,\theta_0}, I^{-1}_{\theta_0})(h)|\, dh \\ &\leq M \int_C |\pi_n(h; X) - dN(\Delta_{n,\theta_0}, I^{-1}_{\theta_0})(h)|\, dh \xrightarrow{P^n_0} 0 \end{aligned} \tag{37}$$

Combining with (36), we can conclude that (34) holds.

This is true for every ball C of fixed radius M and hence also for some $M_n \to \infty$.

In the second part, we proof

$$\frac{\int_{C_n^c} p_h^n(X)\pi_n(h;X)\,dh}{\int p_h^n(X)\pi_n(h;X)\,dh} \xrightarrow{P_0^n} 0, \tag{38}$$

where C_n is a ball with radius M_n, for any $M_n \to \infty$.

Let ϕ_n be the test satisfies condition (i), we have

$$\frac{\int_{C_n^C} p_h^n(x)\pi(h|x)\,dh}{\int p_h^n(x)\pi(h|x)\,dh} = \frac{\int_{C_n^C} p_h^n(x)\pi(h|x)\,dh}{\int p_h^n(x)\pi(h|x)\,dh}\phi_n + \frac{\int_{C_n^C} p_h^n(x)\pi(h|x)\,dh}{\int p_h^n(x)\pi(h|x)\,dh}(1-\phi_n) \tag{39}$$

Since (38) ≤ 1,

$$\frac{\int_{C_n^C} p_h^n(X)\pi_n(h;X)\,dh}{\int p_h^n(X)\pi_n(h;X)\,dh}\phi_n \leq \phi_n \xrightarrow{P_0^n} 0 \tag{40}$$

It's enough to proof

$$\frac{\int_{C_n^C} p_h^n(X)\pi_n(h;X)\,dh}{\int p_h^n(X)\pi_n(h;X)\,dh}(1-\phi_n) \xrightarrow{P_0^n} 0 \tag{41}$$

Fix a ball U around zero. Then

$$(41) \leq \frac{\int_{C_n^C} p_h^n(X)\pi_n(h;X)\,dh}{\int_U p_h^n(X)\pi_n(h;X)\,dh}(1-\phi_n) \tag{42}$$

By the Assumption (ii) and the fact that Δ_{n,θ_0} is uniformly tight, we can assume $\sup_h(\pi_n(h;X) - T(h)) \leq 0$ and $|\Delta_{n,\theta_0}| \leq M$ for some M without loss of generality since the probability they don't hold will be eventually smaller than any prespecified constant.

There exists an $m > 0$ such that

$$\inf_{h\in U} dN(\Delta_{n,\theta_0}, I_{\theta_0}^{-1})(h) \geq m. \tag{43}$$

Let $D_n(X)$ be the set $\{h : |\pi_n(h;X) - n(\Delta_{n,\theta_0}, I_{\theta_0}^{-1})| \geq \frac{m}{2}\}$. We have

$$\begin{aligned}(42) \leq& \frac{\int_{C_n^C} p_h^n(X)\pi_n(h;X)\,dh}{\int_{U/D_n(X)} p_h^n(X)\pi_n(h;X)\,dh}(1-\phi_n) \\ \leq& \frac{\int_{C_n^C} p_h^n(X)\pi_n(h;X)\,dh}{\frac{m}{2}\int_{U/D_n(X)} p_h^n(X)\,dh}(1-\phi_n)\end{aligned} \tag{44}$$

Next we proof

$$\frac{\int_{U/D_n(X)} p_h^n(X)\,dh}{\int_U p_h^n(X)\,dh} \xrightarrow{P_0^n} 1. \tag{45}$$

By Assumption (i), Bernstein-von Mises Theorem holds. That is

$$\int |\pi_n(h;X) - n(\Delta_{n,\theta_0}, I_{\theta_0}^{-1})|\,dh \xrightarrow{P_0^n} 0 \tag{46}$$

(46) implies $\int 1_{D_n(x)}\,dh \xrightarrow{P_0^n} 0$. Similar to the proof in step 1, we have

$$\begin{aligned}
(45) &= \frac{\int_{U/D_n(X)} \frac{p_h^n(X)}{p_0^n(X)}\,dh}{\int_U \frac{p_h^n(X)}{p_0^n(X)}\,dh} \\
&= \frac{e^{o_{P_0^n}(1)} \int_{U/D_n(X)} e^{h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0} h}\,dh}{e^{o_{P_0}^n(1)} \int_U e^{h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0} h}\,dh} \\
&\xrightarrow{P_0^n} 1
\end{aligned} \tag{47}$$

since $h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2}h^T I_{\theta_0} h$ is bounded.

Now, to obtain (44) $\xrightarrow{P_0^n} 0$, we only need to prove

$$\frac{\int_{C_n^C} p_h^n(X)(A\mathbf{1}_{M_n \le \|h\| \le K\sqrt{n}} + T(h)\mathbf{1}_{\|h\| > K\sqrt{n}})\,dh}{\int_U p_h^n(X)\,dh}(1-\phi_n) \xrightarrow{P_0^n} 0 \tag{48}$$

By Lemma 2, we only need to prove (48) $\xrightarrow{P_U^n} 0$. To prove that, we only need to prove (48) $\xrightarrow{L^1_{P_U^n}} 0$, that is

$$\int \frac{\int_{C_n^C} p_h^n(x)(A\mathbf{1}_{M_n \le \|h\| \le K\sqrt{n}} + T(h)\mathbf{1}_{\|h\| > K\sqrt{n}})\,dh}{\int_U p_h^n(x)\,dh}(1-\phi_n)\Big(\int_U p_h^n(x)dh\Big)\,d\mu \to 0 \tag{49}$$

We note that

$$(49) = \int_{C_n^C} \Big(\int (1-\phi_n)p_h^n(x)d\mu\Big)(A\mathbf{1}_{M_n \le \|h\| \le K\sqrt{n}} + T(h)\mathbf{1}_{\|h\| > K\sqrt{n}})\,dh \tag{50}$$

By Lemma 3, there automatically exist tests ϕ_n such that for sufficiently large n and $\|h\| \ge M_n$,

$$\int (1-\phi_n)p_h^n(x)d\mu \le e^{-c(\|h\|^2 \wedge n)} \tag{51}$$

If $K < 1$, then $\|h\|^2 \wedge n \ge \|h\|^2 \wedge K^2 n$. If $K \ge 1$, then

$$\|h\|^2 \wedge n = \frac{1}{K^2}(K^2\|h\|^2 \wedge K^2 n) \ge \frac{1}{K^2}(\|h\|^2 \wedge K^2 n) \tag{52}$$

Let $c^* = c\min(1, 1/K^2)$, then

$$\int (1-\phi_n) p_h^n(x) d\mu \le e^{-c^*(\|h\|^2 \wedge K^2 n)} \tag{53}$$

Splitting the integral into the domains $M_n \le \|h\| \le K\sqrt{n}$ and $\|h\| \ge K\sqrt{n}$, we see that

$$(50) \le \int_{\|h\|\ge M_n} e^{-c^*\|h\|^2}\, dh + e^{-c^* K^2 n} \int_{\|h\|>K\sqrt{n}} T(h)\, dh \to 0 \tag{54}$$

Finally we have

$$\begin{aligned}
&\left| \int \frac{p_h(X)}{p_0(X)} \pi_n(h;X)\, dh - 2^{-\frac{p}{2}} e^{\frac{1}{2}\Delta_{n,\theta_0}^T I_{\theta_0} \Delta_{n,\theta_0}} \right| \\
&= \left| \int \frac{p_h(X)}{p_0(X)} \pi_n(h;X)\, dh - \int_{C_n} \frac{p_h(X)}{p_0(X)} \pi_n(h;X)\, dh \right| \\
&+ \left| \int_{C_n} \frac{p_h(X)}{p_0(X)} \pi_n(h;X)\, dh - \int_{C_n} e^{h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2} h^T I_{\theta_0} h} dN(\Delta_{n,\theta_0}, I_{\theta_0}^{-1})(h)\, dh \right| \\
&+ \left| \int_{C_n} e^{h^T I_{\theta_0}\Delta_{n,\theta_0} - \frac{1}{2} h^T I_{\theta_0} h} n(\Delta_{n,\theta_0}, I_{\theta_0}^{-1})\, dh - 2^{-\frac{p}{2}} e^{\frac{1}{2}\Delta_{n,\theta_0}^T I_{\theta_0} \Delta_{n,\theta_0}} \right| \\
&= J_1 + J_2 + J_3
\end{aligned} \tag{55}$$

By the first step of the proof, we have $J_2 \xrightarrow{P_0^n} 0$. Hence $\int_{C_n} \frac{p_h(X)}{p_0(X)} \pi_n(h;X)\, dh$ is bounded in probability. Therefore

$$\begin{aligned}
J_1 &= \int_{C_n} \frac{p_h(X)}{p_0(X)} \pi_n(h;X)\, dh \left| \frac{\int \frac{p_h(X)}{p_0(X)} \pi_n(h;X)\, dh}{\int_{C_n} \frac{p_h(X)}{p_0(X)} \pi_n(h;X)\, dh} - 1 \right| \\
&= O_{P_0^n}(1) o_{P_0^n}(1)
\end{aligned} \tag{56}$$

And J_3 convenges to 0 for trivial reasion. $\square$

Proof of Theorem 2. 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 1 to both the numerator and denominator of integrated likelihood ratio statistics with $\eta_n = 0$. By CLT,

$$I_{\theta_0}\Delta_{n,\theta_0} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\ell}_{\theta_0}(X_i) \overset{P_0^n}{\rightsquigarrow} \xi, \tag{57}$$

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

$$I^*_{\theta_0}\Delta^*_{n,\theta_0} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}^*_{\theta_0}(X_i) \overset{P_0^n}{\rightsquigarrow} \xi^*, \tag{58}$$

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

$$\begin{aligned}\Lambda(X) &= \frac{2^{-\frac{p_2}{2}}\exp\{\frac{1}{2}\Delta^T_{n,\theta_0} I_{\theta_0}\Delta_{n,\theta_0}\} + o_{P_0^n}(1)}{2^{-\frac{p_1}{2}}\exp\{\frac{1}{2}\Delta^{*T}_{n,\theta_0} I^*_{\theta_0}\Delta^*_{n,\theta_0}\} + o_{P_0^n}(1)} \\ &\overset{P_0^n}{\rightsquigarrow} \frac{2^{-\frac{p_2}{2}}\exp\{\frac{1}{2}\xi^T I^{-1}_{\theta_0}\xi\}}{2^{-\frac{p_1}{2}}\exp\{\frac{1}{2}\xi^{*T} I^{*-1}_{\theta_0}\xi^*\}}.\end{aligned} \tag{59}$$

But

$$\xi^T I^{-1}_{\theta_0}\xi - \xi^{*T} I^{*-1}_{\theta_0}\xi^* = (I_{\theta_0}^{-\frac{1}{2}}\xi)^T \Big(I_{p_2\times p_2} - I_{\theta_0}^{\frac{1}{2}}\begin{pmatrix} I^{*-1}_{\theta_0} & 0 \\ 0 & 0\end{pmatrix} I_{\theta_0}^{\frac{1}{2}}\Big)(I_{\theta_0}^{-\frac{1}{2}}\xi). \tag{60}$$

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

$$2\log(\Lambda(X)) \overset{P_0^n}{\rightsquigarrow} \chi^2_{p_2-p_1} - (p_2 - p_1)\log(2). \tag{61}$$

$\square$

Proof of Theorem 3. We note that $h_n = \eta_n$ converges to η. By differentiability in quadratic mean, Lemma 1 and CLT,

$$\begin{aligned}\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \\ \log\frac{p_{\eta_n}(X)}{p_0(X)}\end{pmatrix} &= \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \eta^T\dot{\ell}_{\theta_0}(X_i) - \frac{1}{2}\eta^T I_{\theta_0}\eta\end{pmatrix} + o_{P_0^n}(1) \\ &\overset{P_0^n}{\rightsquigarrow} N(\begin{pmatrix} 0 \\ -\frac{1}{2}\eta^T I_{\theta_0}\eta\end{pmatrix}, \begin{pmatrix} I_{\theta_0} & I_{\theta_0}\eta \\ \eta^T I_{\theta_0} & \eta^T I_{\theta_0}\eta\end{pmatrix}).\end{aligned} \tag{62}$$

Hence by Le Cam's third lemma,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \overset{P_{\eta_n}^n}{\rightsquigarrow} \xi \sim N(I_{\theta_0}\eta, I_{\theta_0}). \tag{63}$$

By Theorem 1, under $P^n_{\eta_n}$, we have (59). Hence

$$2\log(\Lambda(X)) \overset{P_{\eta_n}^n}{\rightsquigarrow} \chi^2_{p_2-p_1}(\delta) - (p_2 - p_1)\log(2), \tag{64}$$

where noncentral parameter δ can be obtained by substituting ξ by $I_{\theta_0}\eta$ in (60):

$$\begin{aligned}\delta &= \eta^T (I_{\theta_0} - I_{\theta_0} \begin{pmatrix} I_{\theta_0}^{*-1} & 0 \\ 0 & 0 \end{pmatrix} I_{\theta_0})\eta \\ &= \eta^T \begin{pmatrix} 0 & 0 \\ 0 & I_{22\cdot 1} \end{pmatrix} \eta.\end{aligned} \tag{65}$$

□

References

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

Andrew Gelman John B. Carlin HSSDBDAVDBR. Bayesian Data Analysis. Chapman and Hall/CRC, 2013.

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.

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.