

Integrated likelihood ratio test[☆]

author¹

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

本文针对一般的检验问题，提出了积分似然比检验方法，得到了其在原假设下的分布和局部功效函数，并在较弱的条件下给出了严格证明。我们的定理可以适用于似然函数有界的情形，由于此时似然比检验不存在，所以理论上积分似然比检验可以解决一些似然比检验无法解决的问题。而实际上，混合正态模型正好是这样的例子，积分似然比检验水平控制的很好，而对应的，似然比统计量甚至不存在。在我们的定理无法应用的情形，比如高维情形，积分似然比方法仍然是一种寻找检验的优良方法。我们考虑了高维均值检验问题，此时似然比检验不存在，而积分似然比方法依然存在，所以可以作为似然比检验在高维情形下的一种替代品，它具有类似似然比检验的优良性质，模拟结果显示在一些情形下，它优于已有的检验方法。

Keywords: `elsarticle.cls`, L^AT_EX, Elsevier, template

2010 MSC: 00-01, 99-00

1. 介绍

Aitkin (1991) 针对检验问题 $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ 提出了后验贝叶斯

[☆]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)

¹Since 1880.

因子,形式是

$$\frac{\int_{\Theta} p(X|\theta)\pi(\theta|X) d\theta}{\int_{\Theta_0} p(X|\theta)\pi^*(\theta|X) d\theta} \quad (1)$$

其中 $X = X_1, \dots, X_n$ 是观测数据, $p(x|\theta)$ 是样本密度, $\pi^*(\theta|x)$ 是原假设下的后验密度, $\pi(\theta|x)$ 是被择假设下的后验密度。Alan E Gelfand (1993) 利用了拉普拉斯近似求得了它在原假设的分布。他们的方法并没有显式的给出定理成立需要的条件, 而且证明依赖于拉普拉斯近似, 这就假设了极大似然估计的存在。而当极大似然估计存在时, 似然比检验也存在。所以他们的结论并没有显示出后验贝叶斯因子比似然比检验更为广泛的适用性。

基于后验贝叶斯因子, 我们考虑更一般的积分似然比检验统计量:

$$\Lambda(X) = \frac{\int_{\Theta} p(X|\theta)\pi(\theta; X) d\theta}{\int_{\Theta_0} p(X|\theta)\pi^*(\theta; X) d\theta} \quad (2)$$

其中 $\pi(\theta; X)$ 和 $\pi^*(\theta; X)$ 是基于数据的权函数, 但不必等于后验密度。

似然比检验是被最广泛使用的统计方法, 它具有很多最优性质, 比如由 NP 引理, 它在简单原假设和简单被择假设下是最有效检验, 见 E. L. Lehmann (2005)。在多维情形, 一般是没有最有效检验的, 但是似然比检验是在 Bahadur efficiency 的意义下是渐进最优的, 见 Bahadur (1971)。而我们知道在许多模型中, 即使是一些被广泛使用的模型, 似然也经常是无界的, 见 Cam (1990)。这种情形下, 似然比检验不存在。但此时积分似然比检验是可能存在的, 比如后面我们举的例子。

参考 Bernstein-von Mises 定理的证明, 见 der Vaart (2000), 本文在很一般的假设下证明积分似然比检验的 Wilks 现象及局部功效, 尤其是我们的定理不要求似然函数有界, 这使得我们的方法可以用于一些似然无界存在的模型。并举例说明其广泛的适用性。

2. Integrated likelihood ratio test

Let X_1, \dots, X_n be a sample from $P_\theta, \theta \in \Theta$. Denote $X = (X_1, \dots, X_n)$. 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

$$\Theta_0 = \{\theta \in \Theta : \theta_{p_1+1} = \theta_{0,p_1+1}, \dots, \theta_{p_2} = \theta_{0,p_2}\}, \quad (3)$$

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

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta. \quad (4)$$

The first p_1 parameters are nuisance parameters.

Θ_0 can be regarded as a open subset of \mathbb{R}^{p_1} . To simplify notations, we denote $\Theta_0^* = \{(\theta_1, \dots, \theta_{p_1})^T : (\theta_1, \dots, \theta_{p_1}, \theta_{0,p_1+1}, \theta_{0,p_2}) \in \Theta_0\}$. 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

$$\Lambda(X) = \frac{\int_{\Theta} p(X|\theta) \pi(\theta; X) d\theta}{\int_{\Theta_0^*} p(X|\theta^*) \pi^*(\theta^*; X) d\theta^*} \quad (5)$$

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

$$\int [\sqrt{p_{\theta_0+h}} - \sqrt{p_{\theta_0}} - \frac{1}{2} h^T \dot{\ell}_{\theta_0} \sqrt{p_{\theta_0}}]^2 d\mu = o(\|h\|^2), \quad h \rightarrow 0, \quad (6)$$

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 reparameterize 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*

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

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

$$\|\pi_n(h; X) - dN(\Delta_{n, \theta_0}, I_{\theta_0}^{-1})(h)\| \xrightarrow{P_{\theta_0}^n} 0 \quad (8)$$

³⁵ **Assumption 2.** *For every $\epsilon > 0$, there's a Lebesgue integrable function $T(h)$, a $K > 0$ and a $A > 0$ such that*

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

$$\lim_{n \rightarrow \infty} P_{\theta_0}^n \left(\sup_{\|h\| \leq K\sqrt{n}} \pi_n(h; X) \leq A \right) \geq 1 - \epsilon \quad (10)$$

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

$$|\log p_{\theta_1}(x) - \log p_{\theta_2}(x)| \leq \dot{\ell}(x) \|\theta_1 - \theta_2\|. \quad (11)$$

Assumption 1 makes sure that there exists at least one weight function satisfies 8, that is, the posterior density. The condition 9 prevents the bad behavior weight function's tail. The condition 10 is satisfied in most usual case. 条件 (i) 是很弱的条件。条件 (ii) 中的支撑条件可以等价的表述为 $\pi_n(h; X)$ 作为 θ 的函数, 其支撑有界。对于一个统计模型, 我们通常认为似然函数在很远处的取值没有太大意义, 所以不分配权是合理的。因为我们不该某个 θ 附近分配过大的权, 所以权函数有界的条件也是合理的。事实上, 我们总可以把后验密度 $\pi_n(h|X)$ 改造成满足条件 (ii) 的权函数 $\pi_n(h; X)$, 比如我们选取一个充分大的 M , 则容易证明

$$\pi_n(h; X) = \min(\pi_n(h|X), M) 1_{\|h\| \leq K\sqrt{n}} \quad (12)$$

同时满足条件 (i), (ii)。而且我们有理由相信这样得到的权函数比后验密度收敛的更快一点。均方可微条件使得似然可以在 θ_0 的邻域内展开, 我们在证明一般性的定理时需要把似然函数在 θ_0 的邻域内一致展开, 所以我们需要条件 (iii)。

40 我们的第一个定理是

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

$$\left| \int_{\mathbb{R}^p} \frac{p_n(X)}{p_0(X)} \pi_n(h; X) dh - 2^{-\frac{p}{2}} e^{\frac{1}{2} \Delta_{n, \theta_0}^T I_{\theta_0} \Delta_{n, \theta_0}} \right| \xrightarrow{P_{\eta_n}^n} 0 \quad (13)$$

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*

$$2 \log(\Lambda(X)) \xrightarrow{P_0^n} \chi_{p_2-p_1}^2 - (p_2 - p_1) \log(2) \quad (14)$$

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

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

$$I_{\theta_0} = \begin{pmatrix} I_{\theta_0}^* & I_{12} \\ I_{21} & I_{22} \end{pmatrix}, \quad (15)$$

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

$$2 \log(\Lambda(X)) \xrightarrow{P_0^n} \chi_{p_2-p_1}^2(\delta) - (p_2 - p_1) \log(2) \quad (16)$$

where

$$\delta = \eta^T \begin{pmatrix} 0 & 0 \\ 0 & I_{22 \cdot 1} \end{pmatrix} \eta \quad (17)$$

我们的结果和似然比检验结果很类似。这样的结果可以从试验的收敛的观点来解释。从渐进的角度来讲，我们在 n 个样本空间中的统计问题近似于只观测到一个量 Δ_{n,θ_0} 。而在 $h_n \rightarrow h$ 时， $\Delta_{n,\theta_0} \rightsquigarrow N(h, I_{\theta_0}^{-1})$ 。所以问题可以近似为只有一个观测 $X \sim N(h, I_{\theta_0}^{-1})$ 的问题。在这种情形下，我们可以直接计算似然比统计量，正是我们定理的结论。这样我们定理的结论就没什么奇怪的了。

4. 混合正态分布

Cam (1990) 给出了一些似然无界的例子。混合正态分布是其中给出的第一个例子。这一节通过这个例子来说明积分似然比方法的优越性。

设 X_1, \dots, X_n i.i.d. 服从混合正态分布

$$\frac{1-\alpha}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\mu)^2\right\} + \frac{\alpha}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} \quad (18)$$

其中 α 是已知的固定参数。参数空间设为

$$\Theta = \{\theta = (\mu, \sigma^2)^T : \mu \in (\infty, \infty), \sigma^2 \in (0, M)\} \quad (19)$$

Cam (1990) 指出，在这个模型下，似然函数是无界的。事实上，令 $\mu = X_1$ ，然后令 $\sigma^2 \rightarrow 0$ ，则似然函数趋于无穷。

现在我们在这个模型下考虑检验问题 $H_0 : \mu = 0, \sigma = 2$ vs $H_1 : \theta \in \Theta$ 。似然比检验显然已经失效了，但此时积分似然比统计量仍然可以使用。事实上，我们可以验证（虽然很繁琐）条件 (i),(ii),(iii) 均成立，所以定理2和3均成立。当然要想使用积分似然比统计量，我们需要寻找权函数，在这个例子中我们考虑用后验密度。这里我们可以看出，假设我们先验设置成无信息先验，那么后验密度与似然函数成正比，所以是无界的。无界的权函数不仅会使得我们的定理条件 (ii) 无法满足，更重要的是，在积分似然比统计量中，它会给似然函数在没意义的地方设置过多的权重，而我们实际上期望权函数集中在真实参数附近，所以无界的权函数使得统计量的性质变得很复杂并且收敛速度变慢。一种解决这个问题的方法是按照前面提到的，选取一个合适的充分大的 M^* ，定义新的权函数为

$$\min(\pi_n(h; X), M^*) \quad (20)$$

则新的有界的权函数依然满足条件 (i)，甚至表现的更好。而条件 (ii) 也满足了。

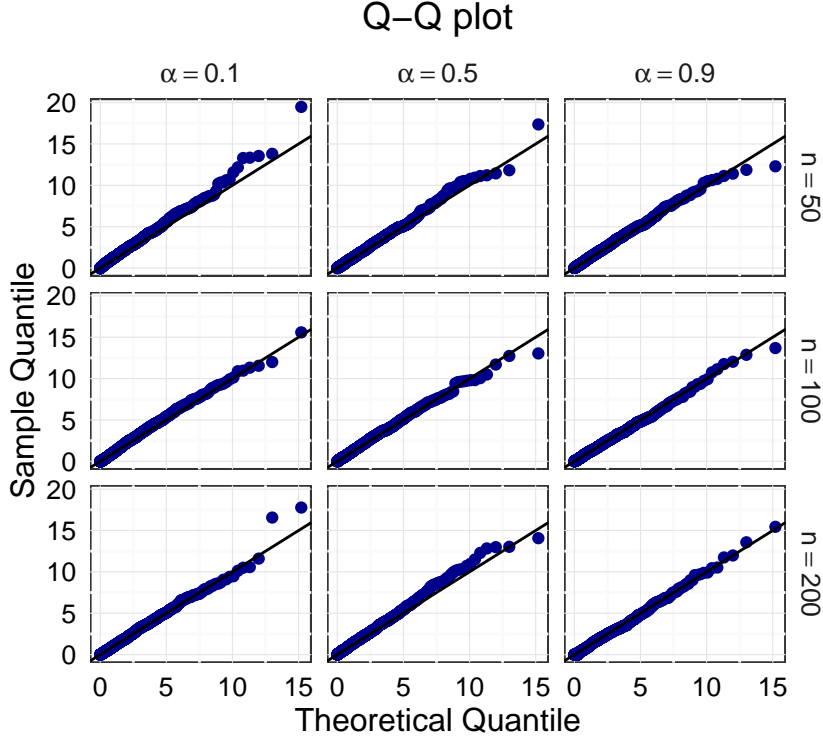
在我们这个例子中，我们并不需要这样解决权函数无界的问题。因为我们只要设置合适的先验分布，后验密度就会有界，这样我们仍然可以使用后验密度作为权函数。我们发现当 $\mu = X_1, \sigma^2 \rightarrow 0$ 时，似然趋于无穷的速度正比例于 $\frac{1}{\sigma}$ 。所以我们设置合适的先验对冲掉这个量。容易看出， $\chi^2(3)$ 的密度函数正好可以做到这一点。

故我们设先验是

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

其中 $d\chi^2(3)(\sigma^2)$ 代表自由度为 3 的 χ^2 分布的密度函数在 σ^2 处的取值。由于参数空间的限制， σ^2 只在 $(0, M)$ 上取值，所以这个先验并不是个概率密度，但这并不妨碍我们求得后验密度并计算积分似然比统计量。

我们下面通过数值模拟验证 Wilks 现象。我们取样本量 $n = 50, 100, 200$ 和超参数 $\alpha = 0.1, 0.5, 0.9$ 。每种情况重复抽取 1000 次样本，得到 1000 个积分似然比统计量。理想情况应该是它们的样本与 $\chi^2(2)$ 的分布很接近。我们画出样本分布相对于 $\chi^2(2)$ 分布的 QQ 图，模拟显示样本分布与 $\chi^2(2)$ 分布很接近，见下图：



70

5. Appendix

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

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 $h_n \rightarrow h$, as $n \rightarrow \infty$,*

$$\log \frac{p_{h_n}^n(X)}{p_0^n(X)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n h^T \dot{\ell}_{\theta_0}(X_i) - \frac{1}{2} h^T I_{\theta_0} h + o_{P_{\theta_0}}(1), \quad (22)$$

75 where $p_h^n(X) = \prod_{i=1}^n p_h(X_i)$ is the density of P_h^n 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 $T_n(X)$, $T_n \xrightarrow{P_0^n} 0 \Leftrightarrow T_n \xrightarrow{P_U^n} 0$, where*

$$p_U^n(x) = \frac{1}{V(U)} \int_U p_h^n(x) dh, \quad (23)$$

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

Proof. In fact, we only need to prove

$$\int_{A_n} p_0^n(x) d\mu \rightarrow 0 \Leftrightarrow \int_{A_n} \frac{1}{V(U)} \int_U p_h^n(x) dh d\mu \rightarrow 0, \quad (24)$$

or

$$\int_{A_n} p_0^n(x) d\mu \rightarrow 0 \Leftrightarrow \int_U \int_{A_n} p_h^n(x) d\mu dh \rightarrow 0. \quad (25)$$

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

$$\int_{A_n} p_0^n(x) d\mu \rightarrow 0 \Leftrightarrow \int_{A_n} p_{h_n}^n(x) d\mu \rightarrow 0. \quad (26)$$

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

$$\int_U \int_{A_n} p_h^n(x) d\mu dh \leq V(U) \sup_{h \in U} \int_{A_n} p_h^n(x) d\mu \leq V(U) \left(\int_{A_n} p_{\bar{h}_n}^n(x) d\mu + 1/n \right). \quad (27)$$

We have similar lower bound. Hence,

$$V(U) \left(\int_{A_n} p_{\bar{h}_n}^n(x) d\mu + 1/n \right) \leq \int_U \int_{A_n} p_h^n(x) d\mu dh \leq V(U) \left(\int_{A_n} p_{\bar{h}_n}^n(x) d\mu + 1/n \right) \quad (28)$$

The (25) follows from (26) and (28). \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*

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

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

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

(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 $\forall \theta_1, \theta_2 \in V$:

$$|\log p_{\theta_1}(x) - \log p_{\theta_2}(x)| \leq m(x) \|\theta_1 - \theta_2\|. \quad (29)$$

Then for every $M > 0$,

$$\sup_{\|h\| \leq M} \left| \log \frac{p_h^n(X)}{p_0^n(X)} - \frac{1}{\sqrt{n}} \sum_{i=1}^n h^T \dot{\ell}_{\theta_0}(X_i) + \frac{1}{2} h^T I_{\theta_0} h \right| \xrightarrow{P_0^n} 0. \quad (30)$$

(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_0^n .

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

$$\left| \int_C \frac{p_h^n(X)}{p_0^n(X)} \pi_n(h; X) dh - \int_C e^{h^T I_{\theta_0} \Delta_{n, \theta_0} - \frac{1}{2} h^T I_{\theta_0} h} dN(\Delta_{n, \theta_0}, I_{\theta_0}^{-1})(h) dh \right| \xrightarrow{P_0^n} 0 \quad (31)$$

By Lemma 4, for every fixed M ,

$$\sup_{\|h\| \leq M} \left| \log \frac{p_h^n(X)}{p_0^n(X)} - h^T I_{\theta_0} \Delta_{n, \theta_0} + \frac{1}{2} h^T I_{\theta_0} h \right| \xrightarrow{P_0^n} 0 \quad (32)$$

Hence we have

$$\int_C \frac{p_h^n(X)}{p_0^n(X)} \pi_n(h; X) dh = e^{o_{P_0^n}(1)} \int_C e^{h^T I_{\theta_0} \Delta_{n, \theta_0} - \frac{1}{2} h^T I_{\theta_0} h} \pi_n(h; X) dh \quad (33)$$

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 $1 - \delta$,

$$\begin{aligned} & \int_C e^{h^T I_{\theta_0} \Delta_{n, \theta_0} - \frac{1}{2} h^T I_{\theta_0} h} |\pi_n(h; X) - dN(\Delta_{n, \theta_0}, I_{\theta_0}^{-1})(h)| dh \\ & \leq M \int_C |\pi_n(h; X) - dN(\Delta_{n, \theta_0}, I_{\theta_0}^{-1})(h)| dh \xrightarrow{P_0^n} 0 \end{aligned} \quad (34)$$

Combining with (33), we can conclude that (31) holds.

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

In the second part, we proof

$$\frac{\int_{C_n^c} p_h^n(X) \pi_n(h; X) dh}{\int p_h^n(X) \pi_n(h; X) dh} \xrightarrow{P_0^n} 0, \quad (35)$$

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

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

$$\frac{\int_{C_n^c} p_h^n(x) \pi(h|x) dh}{\int p_h^n(x) \pi(h|x) dh} = \frac{\int_{C_n^c} p_h^n(x) \pi(h|x) dh}{\int p_h^n(x) \pi(h|x) dh} \phi_n + \frac{\int_{C_n^c} p_h^n(x) \pi(h|x) dh}{\int p_h^n(x) \pi(h|x) dh} (1 - \phi_n) \quad (36)$$

Since (35) ≤ 1 ,

$$\frac{\int_{C_n^c} p_h^n(X) \pi_n(h; X) dh}{\int p_h^n(X) \pi_n(h; X) dh} \phi_n \leq \phi_n \xrightarrow{P_0^n} 0 \quad (37)$$

It's enough to proof

$$\frac{\int_{C_n^c} p_h^n(X) \pi_n(h; X) dh}{\int p_h^n(X) \pi_n(h; X) dh} (1 - \phi_n) \xrightarrow{P_0^n} 0 \quad (38)$$

Fix a ball U around zero. Then

$$(38) \leq \frac{\int_{C_n^c} p_h^n(X) \pi_n(h; X) dh}{\int_U p_h^n(X) \pi_n(h; X) dh} (1 - \phi_n) \quad (39)$$

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

95 any prespecified constant.

There exists an $m > 0$ such that

$$\inf_{h \in U} dN(\Delta_{n, \theta_0}, I_{\theta_0}^{-1})(h) \geq m. \quad (40)$$

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

$$\begin{aligned} (39) &\leq \frac{\int_{C_n^c} p_h^n(X) \pi_n(h; X) dh}{\int_{U/D_n(X)} p_h^n(X) \pi_n(h; X) dh} (1 - \phi_n) \\ &\leq \frac{\int_{C_n^c} p_h^n(X) \pi_n(h; X) dh}{\frac{m}{2} \int_{U/D_n(X)} p_h^n(X) dh} (1 - \phi_n) \end{aligned} \quad (41)$$

Next we proof

$$\frac{\int_{U/D_n(X)} p_h^n(X) dh}{\int_U p_h^n(X) dh} \xrightarrow{P_0^n} 1. \quad (42)$$

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

$$\int |\pi_n(h; X) - n(\Delta_{n, \theta_0}, I_{\theta_0}^{-1})| dh \xrightarrow{P_0^n} 0 \quad (43)$$

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

$$\begin{aligned} (42) &= \frac{\int_{U/D_n(X)} \frac{p_h^n(X)}{p_0^n(X)} dh}{\int_U \frac{p_h^n(X)}{p_0^n(X)} dh} \\ &= \frac{e^{o_{P_0^n}(1)} \int_{U/D_n(X)} e^{h^T I_{\theta_0} \Delta_{n, \theta_0} - \frac{1}{2} h^T I_{\theta_0} h} dh}{e^{o_{P_0^n}(1)} \int_U e^{h^T I_{\theta_0} \Delta_{n, \theta_0} - \frac{1}{2} h^T I_{\theta_0} h} dh} \\ &\xrightarrow{P_0^n} 1 \end{aligned} \quad (44)$$

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

Now, to obtain (41) $\xrightarrow{P_0^n} 0$, we only need to prove

$$\frac{\int_{C_n^C} p_h^n(X) (A \mathbf{1}_{M_n \leq \|h\| \leq K\sqrt{n}} + T(h) \mathbf{1}_{\|h\| > K\sqrt{n}}) dh}{\int_U p_h^n(X) dh} (1 - \phi_n) \xrightarrow{P_0^n} 0 \quad (45)$$

By Lemma 2, we only need to prove (45) $\xrightarrow{P_U^n} 0$. To prove that, we only need to prove (45) $\xrightarrow{L_{P_U^n}^1} 0$, that is

$$\int \frac{\int_{C_n^C} p_h^n(x) (A \mathbf{1}_{M_n \leq \|h\| \leq K\sqrt{n}} + T(h) \mathbf{1}_{\|h\| > K\sqrt{n}}) dh}{\int_U p_h^n(x) dh} (1 - \phi_n) \left(\int_U p_h^n(x) dh \right) d\mu \rightarrow 0 \quad (46)$$

We note that

$$(46) = \int_{C_n^C} \left(\int (1 - \phi_n) p_h^n(x) d\mu \right) (A \mathbf{1}_{M_n \leq \|h\| \leq K\sqrt{n}} + T(h) \mathbf{1}_{\|h\| > K\sqrt{n}}) dh \quad (47)$$

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

$$\int (1 - \phi_n) p_h^n(x) d\mu \leq e^{-c(\|h\|^2 \wedge n)} \quad (48)$$

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

$$\|h\|^2 \wedge n = \frac{1}{K^2} (K^2 \|h\|^2 \wedge K^2 n) \geq \frac{1}{K^2} (\|h\|^2 \wedge K^2 n) \quad (49)$$

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

$$\int (1 - \phi_n) p_h^n(x) d\mu \leq e^{-c^*(\|h\|^2 \wedge K^2 n)} \quad (50)$$

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

$$(47) \leq \int_{\|h\| \geq M_n} e^{-c^*\|h\|^2} dh + e^{-c^*K^2n} \int_{\|h\| > K\sqrt{n}} T(h) dh \rightarrow 0 \quad (51)$$

Finally we have

$$\begin{aligned} & \left| \int \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh - 2^{-\frac{p}{2}} e^{\frac{1}{2} \Delta_{n, \theta_0}^T I_{\theta_0} \Delta_{n, \theta_0}} \right| \\ &= \left| \int \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh - \int_{C_n} \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh \right| \\ &+ \left| \int_{C_n} \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh - \int_{C_n} e^{h^T I_{\theta_0} \Delta_{n, \theta_0} - \frac{1}{2} h^T I_{\theta_0} h} dN(\Delta_{n, \theta_0}, I_{\theta_0}^{-1})(h) dh \right| \\ &+ \left| \int_{C_n} e^{h^T I_{\theta_0} \Delta_{n, \theta_0} - \frac{1}{2} h^T I_{\theta_0} h} n(\Delta_{n, \theta_0}, I_{\theta_0}^{-1}) dh - 2^{-\frac{p}{2}} e^{\frac{1}{2} \Delta_{n, \theta_0}^T I_{\theta_0} \Delta_{n, \theta_0}} \right| \\ &= J_1 + J_2 + J_3 \end{aligned} \quad (52)$$

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

$$\begin{aligned} J_1 &= \int_{C_n} \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh \left| \frac{\int \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh}{\int_{C_n} \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh} - 1 \right| \\ &= O_{P_0^n}(1) o_{P_0^n}(1) \end{aligned} \quad (53)$$

And J_3 converges to 0 for trivial reason. \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,

$$I_{\theta_0} \Delta_{n, \theta_0} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \overset{P_0^n}{\rightsquigarrow} \xi, \quad (54)$$

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

$$I_{\theta_0}^* \Delta_{n, \theta_0}^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_{\theta_0}^*(X_i) \overset{P_0^n}{\rightsquigarrow} \xi^*, \quad (55)$$

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

$$\begin{aligned} \Lambda(X) &= \frac{2^{-\frac{p_2}{2}} \exp\{\frac{1}{2} \Delta_{n, \theta_0}^T I_{\theta_0} \Delta_{n, \theta_0}\} + o_{P_0^n}(1)}{2^{-\frac{p_1}{2}} \exp\{\frac{1}{2} \Delta_{n, \theta_0}^{*T} I_{\theta_0}^* \Delta_{n, \theta_0}^*\} + o_{P_0^n}(1)} \\ &\overset{P_0^n}{\rightsquigarrow} \frac{2^{-\frac{p_2}{2}} \exp\{\frac{1}{2} \xi^T I_{\theta_0}^{-1} \xi\}}{2^{-\frac{p_1}{2}} \exp\{\frac{1}{2} \xi^{*T} I_{\theta_0}^{*-1} \xi^*\}}. \end{aligned} \quad (56)$$

But

$$\xi^T I_{\theta_0}^{-1} \xi - \xi^{*T} I_{\theta_0}^{*-1} \xi^* = (I_{\theta_0}^{-\frac{1}{2}} \xi)^T \left(I_{p_2 \times p_2} - I_{\theta_0}^{-\frac{1}{2}} \begin{pmatrix} I_{\theta_0}^{*-1} & 0 \\ 0 & 0 \end{pmatrix} I_{\theta_0}^{-\frac{1}{2}} \right) (I_{\theta_0}^{-\frac{1}{2}} \xi). \quad (57)$$

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

$$2 \log(\Lambda(X)) \overset{P_0^n}{\rightsquigarrow} \chi_{p_2 - p_1}^2 - (p_2 - p_1) \log(2). \quad (58)$$

□

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

$$\begin{aligned} \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \\ \log \frac{p_{\eta_n}(X)}{p_0(X)} \end{pmatrix} &= \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta^T \dot{\ell}_{\theta_0}(X_i) - \frac{1}{2} \eta^T I_{\theta_0} \eta \end{pmatrix} + o_{P_0^n}(1) \\ &\overset{P_0^n}{\rightsquigarrow} N \left(\begin{pmatrix} 0 \\ -\frac{1}{2} \eta^T I_{\theta_0} \eta \end{pmatrix}, \begin{pmatrix} I_{\theta_0} & I_{\theta_0} \eta \\ \eta^T I_{\theta_0} & \eta^T I_{\theta_0} \eta \end{pmatrix} \right). \end{aligned} \quad (59)$$

Hence by Le Cam's third lemma,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \overset{P_{\eta_n}^n}{\rightsquigarrow} \xi \sim N(I_{\theta_0} \eta, I_{\theta_0}). \quad (60)$$

By Theorem 1, under $P_{\eta_n}^n$, we have (56). Hence

$$2 \log(\Lambda(X)) \overset{P_{\eta_n}^n}{\rightsquigarrow} \chi_{p_2 - p_1}^2(\delta) - (p_2 - p_1) \log(2), \quad (61)$$

where noncentral parameter δ can be obtained by substituting ξ by $I_{\theta_0}\eta$ in (57):

$$\begin{aligned}\delta &= \eta^T (I_{\theta_0} - I_{\theta_0} \begin{pmatrix} I_{\theta_0}^{*-1} & 0 \\ 0 & 0 \end{pmatrix} I_{\theta_0}) \eta \\ &= \eta^T \begin{pmatrix} 0 & 0 \\ 0 & I_{22 \cdot 1} \end{pmatrix} \eta.\end{aligned}\tag{62}$$

105

□

References

- Aitkin M. Posterior bayes factors. journal of the royal statistical society series b-methodological 1991;53(1):111–42.
- Alan E Gelfand DKD. Bayesian model choice: Asymptotics and exact calculations 1993;.
- 110 Bahadur RR. Some limit theorems in statistics. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1971. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 4.
- 115 Cam LL. Maximum likelihood: An introduction. International Statistical Review 1990;58(2):153–71.
- E. L. Lehmann JPR. Testing Statistical Hypotheses. Springer New York, 2005. doi:10.1007/0-387-27605-X.
- Kleijn BJK, Vaart AWVD. The bernstein-von-mises theorem under misspecification. Electronic Journal of Statistics 2012;6(1):354–81.
- 120 der Vaart AV. Asymptotic statistics. Cambridge university press, 2000.