

积分似然比检验[☆]

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`, \LaTeX , 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. 积分似然比检验

设 X_1, \dots, X_n i.i.d. 服从 $P_{\theta} : \theta \in \Theta$, 设 $X = (X_1, \dots, X_n)$ 为随机样本, 设全空间 Θ 是 \mathbb{R}^{p_2} 的一个开子集, 我们设原空间 Θ_0 是 Θ 的一个 p_1 维子空间, 准确来说, 我们定义

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

也就是说后 $p_2 - p_1$ 维 $(\theta_{0,p_1+1}, \dots, \theta_{0,p_2})$ 固定, 是我们想要检验的参数, 前 p_1 维是讨厌参数。我们的问题是检验

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

由于在 Θ_0 中的 θ 只有前 p_1 个参数变化, Θ_0 可以看成是 \mathbb{R}^{p_1} 空间的一个开子集, 为了记号方便, 我们定义 $\Theta_0^* = \{(\theta_1, \dots, \theta_{p_1})^T : (\theta_1, \dots, \theta_{p_1}, \theta_{0,p_1+1}, \theta_{0,p_2}) \in \Theta_0\}$, 故我们可以用 p_1 维向量 $\theta^* \in \Theta_0^*$ 表示 $\theta \in \Theta_0$ 并把 Θ_0^* 看成原空间。我们下面会根据上下文使用原空间的记号。我们在全空间 Θ 和原空间 Θ_0^* 里分别设置可以依赖于数据的权函数 $\pi(\theta; X)$ 和 $\pi^*(\theta^*; X)$, 则积分似然比统计量定义为:

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

说一说在极限试验的意义下的解释。

3. 主要结果

下面记 \rightsquigarrow 代表弱收敛, $dN(\mu, \Sigma)(h)$ 是一个均值为 μ , 方差为 Σ 的正态密度在 h 处的取值。

设 $\theta_0 \in \Theta_0$ 是原假设中一个固定的参数, 我们研究积分似然比统计量在 θ_0 处 (及其附近) 的渐进性质。设试验 $P_\theta : \theta \in \Theta$ 在 θ_0 处均方可微, 即存在随机 $\dot{\ell}_{\theta_0}$ 使得

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

25 其中 μ 是 P_θ 的控制测度。而 $p_\theta(x)$ 是 $P_\theta(x)$ 相对于 μ 的密度函数。 $\dot{\ell}_{\theta_0}$ 是均方可微意义下的得分函数。

设 $I_{\theta_0} = P_{\theta_0} \dot{\ell}_{\theta_0} \dot{\ell}_{\theta_0}^T$ 是 θ_0 处均方可微意义下的信息阵, 设 “局部充分统计量” $\Delta_{n,\theta_0} = \frac{1}{\sqrt{n}} \sum_{i=1}^n I_{\theta_0}^{-1} \dot{\ell}_{\theta_0}(X_i)$ 。相对应的, 在原空间中 $\dot{\ell}^*$, $I_{\theta_0}^*$, Δ_{n,θ_0}^* 均可同样定义。我们很容易知道, $\dot{\ell}_{\theta_0}^*$ 是 $\dot{\ell}_{\theta_0}$ 的前 p_1 维向量, $I_{\theta_0}^*$ 是 I_{θ_0} 左上角的 $p_1 \times p_1$ 30 维矩阵, $\Delta_{n,\theta_0}^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n I_{\theta_0}^{*-1} \dot{\ell}_{\theta_0}^*(X_i)$

设 $h = \sqrt{n}(\theta - \theta_0)$, h 把参数 θ 在 θ_0 附近以 \sqrt{n} 的比例重参数化。易知在原假设下 $h = 0$, 而研究局部被择下, 一般会让 h 趋于某个常数。

下面是我们定理需要的正规条件:

- (i) 设 Bernstein-von Mises 定理 (见 der Vaart (2000)) 的条件满足: Θ 是 \mathbb{R}^p 的开子集, 试验 $(P_\theta : \theta \in \Theta)$ 在 $\theta_0 \in \Theta$ 处均方可微, 具有非奇异信息阵 I_{θ_0} , 假设对任意的 $\epsilon > 0$, 存在一个相合检验序列:

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

设 $\pi_n(h; X)$ 是一个满足 Bernstein-von Mises 定理结论的权函数:

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

- (ii) $\pi_n(h; X)$ 的支撑在 $\|h\| \leq K\sqrt{n}$ 上, K 是一个固定常数。 $\sup_h |\pi_n(h; X)| \leq A$ 对某个 A 依概率成立。

- (iii) 存在 θ_0 的某个邻域 V 和某个函数 $\dot{\ell}$ 满足 $P_{\theta_0} \dot{\ell}^2 < \infty$, 对 $\forall \theta_1, \theta_2 \in V$, 有

$$|\log p_{\theta_1}(x) - \log p_{\theta_2}(x)| \leq \dot{\ell}(x) \|\theta_1 - \theta_2\|. \quad (9)$$

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

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

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)。

我们的第一个定理是

Theorem 1. 如果条件 (i), (ii), (iii) 满足, 设 η_n 是有界实数序列, 那么我们有

$$\left| \int_{\mathbb{R}^p} \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh - 2^{-\frac{p}{2}} e^{\frac{1}{2} \Delta_{n, \theta_0}^T I_{\theta_0} \Delta_{n, \theta_0}} \right| \xrightarrow{P_{\eta_n}^n} 0 \quad (11)$$

- 40 基于定理4, 我们可以得到积分似然比统计量在原假设的渐进分布, 也就是 wilks 现象。我们可以根据此渐进分布来决定检验的临界值

Theorem 2. 设定理4的条件对原空间 Θ_0 和全空间 Θ 均成立, 真实参数 θ_0 是 Θ 的一个内点, 同时是 Θ_0 的一个相对内点, 那么我们有

$$2 \log(\Lambda(X)) \stackrel{P_0^n}{\rightsquigarrow} \chi_{p_2-p_1}^2 - (p_2 - p_1) \log(2) \quad (12)$$

有了定理4, 我们可以利用 Le Cam 第三引理得到积分似然比统计量在局部被择下的渐近分布:

Theorem 3. 设定理4的条件对原空间 Θ_0 和全空间 Θ 均成立。 θ_0 是 Θ_0 的一个内点, 同时是 Θ_0 的一个相对内点。 真实参数 θ 满足 $\eta_n = \sqrt{n}(\theta - \theta_0) \rightarrow \eta$ 。 若

$$I_{\theta_0} = \begin{pmatrix} I_{\theta_0}^* & I_{12} \\ I_{21} & I_{22} \end{pmatrix} \quad (13)$$

$I_{22 \cdot 1} = I_{22} - I_{21} I_{\theta_0}^{*-1} I_{12}$ 那么我们有

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

其中

$$\delta = \eta^T \begin{pmatrix} 0 & 0 \\ 0 & I_{22 \cdot 1} \end{pmatrix} \eta \quad (15)$$

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

$$\frac{1-\alpha}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(x-\mu)^2 \right\} + \frac{\alpha}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right\} \quad (16)$$

其中 α 是已知的固定参数。参数空间设为

$$\Theta = \{\theta = (\mu, \sigma^2)^T : \mu \in (\infty, \infty), \sigma^2 \in (0, M)\} \quad (17)$$

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

$$\min(\pi_n(h; X), M^*) \quad (18)$$

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。

55 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。

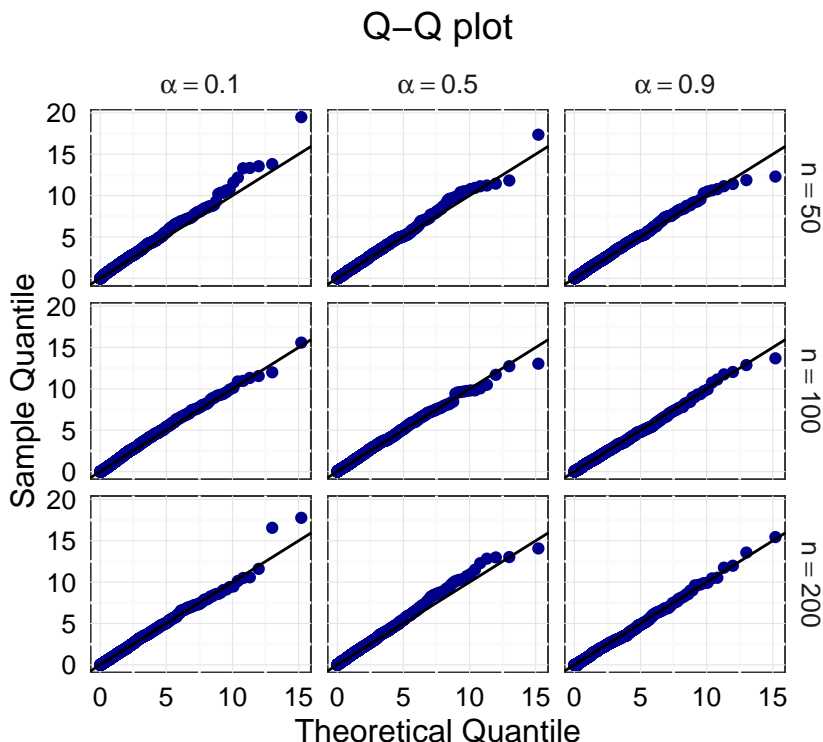
故我们设先验是

$$dN(0, 1)(\mu) \times d\chi^2(3)(\sigma^2) \quad (19)$$

60 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：



5. 高维正态均值检验

在一些问题中，我们可以直接使用积分似然比方法：直接验证条件 (i), (ii), (iii)，然后直接利用 Wilks 现象来求得积分似然比检验的临界值。但是我们在一般性定理在一些情形下不能适用，比如高维问题 ($p \rightarrow \infty$)。但此时积分似然比仍是一种寻找检验统计量的方法，并且我们发现用积分似然比方法得到的统计量仍然具有很多优点。

设 $X_1, \dots, X_n \in \mathbb{R}^p$ i.i.d. 服从 $N(\mu, \Sigma)$ ，方差未知， p 随 $n \rightarrow \infty$ 而趋于无穷。我们考虑均值检验问题 $H_0 : \mu = 0$ vs $H_1 : \mu \in \mathbb{R}^p$

这个问题的似然比统计量是著名的 Hotelling T^2 统计量。但是当 $p > n$ 时，由于样本协方差矩阵不可逆，Hotelling T^2 统计量不存在。高维情形下的均值

检验问题被很多学者研究过了，比如Bai and Saranadasa (1996)，他们的提出统计量

$$T_B = \frac{n\|\bar{X}\|^2 - \text{tr}S}{\sqrt{\frac{2(n-1)n}{(n-2)(n+1)}(\text{tr}S^2 - \frac{(\text{tr}S)^2}{n-1})}} \quad (20)$$

其中 $S = \frac{1}{n}(X_i - \bar{X})(X_i - \bar{X})^T$ 是样本协方差矩阵。他们在一些假设下证明了 T_B 在原假设下渐近服从 $N(0, 1)$ 分布。我们知道在这个问题上，由于它是多维的检验问题，所以不存在一个检验一致的优于别的检验。和经典 Hotelling 方法最显著的不同是他们去掉了协方差的信息，这使得他们的统计量不能根据数据自适应的分配功效的方向，而是把所有的方向相同的对待，而这对于数据的尺度不一，或者分量间有相关性的情形会产生严重的功效损失。一般来说，似然比统计量可以克服上述问题。但是当 $p > n$ 时，似然比检验不存在，这是很可惜的。我们发现虽然似然比检验不存在，但是积分似然比检验存在。我们取后验密度为积分似然比的权函数。

设 $W^{-1}(\Psi, m)$ 是参数为 Ψ ，自由度为 m 的逆威沙特分布， $w^{-1}(B|\Psi, m)$ 是对应的密度在 B 处的取值。

在 H_0 下，我们设置先验为 $\Sigma \sim W^{-1}(\Sigma|I_{p \times p}, m)$ ，其中 $I_{p \times p}$ 是 p 维单位矩阵。计算之后，我们得到(5)式分母为一个常数（不依赖于参数与数据）乘以

$$\frac{|I_{p \times p} + \sum_{i=1}^n X_i X_i^T|^{\frac{1}{2}(n+m)}}{|I_{p \times p} + 2 \sum_{i=1}^n X_i X_i^T|^{\frac{1}{2}(2n+m)}} \quad (21)$$

在 H_1 下，我们对 Σ 设置相同的先验，对 μ 设置条件正态分布：

$$\pi(\mu, \Sigma) = dN(0, \Sigma/K)(\mu) \times w^{-1}(\Sigma|I_{p \times p}, m) \quad (22)$$

其中 K 是一常数，则计算得到(5)式的分子为一个常数乘以

$$\frac{|I_{p \times p} + \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T + \frac{nK}{n+K} \bar{X} \bar{X}^T|^{\frac{1}{2}(n+m)}}{|I_{p \times p} + 2 \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T + \frac{2nK}{2n+K} \bar{X} \bar{X}^T|^{\frac{1}{2}(2n+m)}} \quad (23)$$

则我们得到积分似然比统计量为 $\frac{(23)}{(21)}$ ，如果去掉统计中所有的 $I_{p \times p}$ ，则它与 Hotelling 统计量等价。我们发现积分似然比统计量即使当 $p > n$ 时也存在。为了清楚，设 $K = 0$ ，并对统计量进行近似：

$$\begin{aligned} (21) &\approx 2^{-\frac{1}{2}(2n+m)} |I_{p \times p} + \sum_{i=1}^n X_i X_i^T|^{-\frac{n}{2}} \\ (23) &\approx 2^{-\frac{1}{2}(2n+m)} |I_{p \times p} + \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T|^{-\frac{n}{2}} \end{aligned} \quad (24)$$

如果这样近似，积分似然比统计量等价于

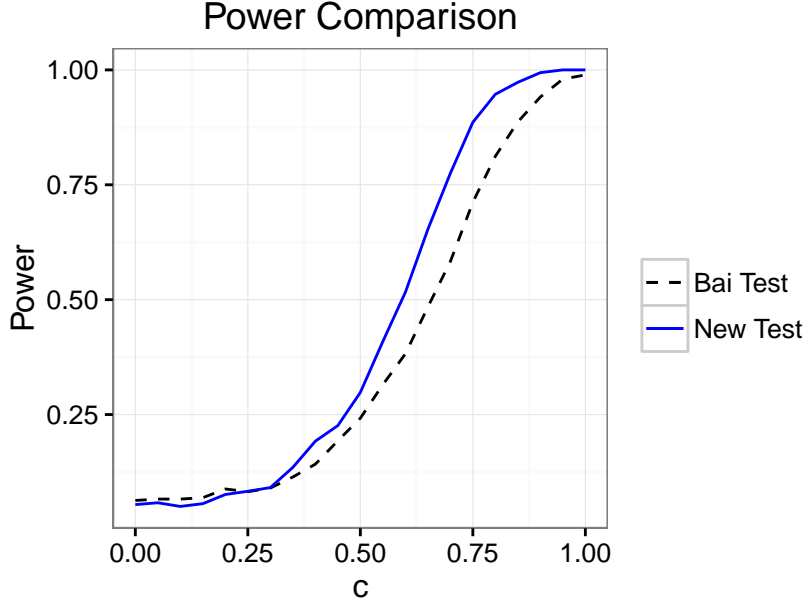
$$\bar{X}^T \left(\frac{1}{n} I_{p \times p} + S \right)^{-1} \bar{X} \quad (25)$$

其中 $S = \frac{1}{n} (X_i - \bar{X})(X_i - \bar{X})^T$ 是样本协方差矩阵。

由于在高维情形下我们的定理2失效。所以我们用Zhao and Xu (2016) 中提出的 randomization procedure 决定检验的临界值。

Bai and Saranadasa (1996) 的统计量中没有对协方差矩阵进行估计，只估计很少的参数使得他们的统计量方差很小，从而功效较好。但是这样做也优缺点：首先，他们在 Hotelling 检验中用单位阵代替了协方差阵的逆的估计，所以他们必须假定真实的协方差阵与单位阵差别不大，否则他们的定理失效。其次，似然比统计量把功效分配在了比较合理的方向，不估计协方差阵就只能自己事先分配一个功效的方向（他们的统计量里，这个方向是 I ），而这在很多情况下会丧失大量信息，因而不合理的。

我们考虑如下的情形： $\mu = c(1, 0.01, \dots, 0.01)^T$, c 是一个实数。 Σ 是对角阵， $\text{diag}(\Sigma) = (0.01, 1, 1, \dots, 1)$ 。由于第一个分量方差很小，我们显然应该更在意第一个分量方向，只要第一个分量方向的均值稍微远离 0 一点，我们就应该拒绝，所以理想的功效函数对 c 应该很敏感。下面是在功效函数的模拟结果， $n = 40$, $p = 100$ ，对于白的检验，我们使用渐近正态性确定临界值，对于积分似然比统计量(25)，我们用Zhao and Xu (2016) 的方法重复抽样了 $B = 100$ 次得到检验 P -值。我们模拟了 1000 次样本来计算模拟功效函数。我们的模拟结果显示积分似然比检验在这个功效方向上表现很好，而白、陈的统计量则表现没那么好。



6. Appendix

对于两个测度序列 P_n 和 Q_n , $P_n \ll Q_n$ 代表 P_n 和 Q_n 是 mutually contiguous 的, 即对任意的随即变量序列 T_n , 有 $T_n \xrightarrow{P_n} 0 \Leftrightarrow T_n \xrightarrow{Q_n} 0$

Lemma 1. 假设 Θ 是一个 \mathbb{R}^p 的一个开子集并且试验 $(P_\theta : \theta \in \Theta)$ 在 θ_0 处均方可微。那么 $P_{\theta_0} \dot{\ell}_{\theta_0} = 0$, 并且 Fisher 信息阵 $I_{\theta_0} = P_{\theta_0} \dot{\ell}_{\theta_0} \dot{\ell}_{\theta_0}^T$ 存在。进一步, 对于每个收敛序列 $h_n \rightarrow h$, 随着 $n \rightarrow \infty$, 有

$$\log \frac{p_{h_n}^n(X)}{p_0^n(X)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n h^T \dot{\ell}_{\theta_0}(X_i) - \frac{1}{2} h^T I_{\theta_0} h + o_{P_{\theta_0}}(1) \quad (26)$$

见 *der Vaart (2000)*

Lemma 2. 设引理1条件满足。假设 T_n 是一个统计量序列, U 是一个以原点为中心的球。则对任意的随机变量序列 T_n , $T_n \xrightarrow{P_0^n} 0 \Leftrightarrow T_n \xrightarrow{P_U^n} 0$, 其中

$$p_U^n(x) = \frac{1}{V(U)} \int_U p_h^n(x) dh \quad (27)$$

$V(U)$ 是球 U 的体积。

Proof. 事实上, 我们只需证明

$$\int_{A_n} p_0^n(x) d\mu \rightarrow 0 \Leftrightarrow \int_{A_n} \frac{1}{V(U)} \int_U p_h^n(x) dh d\mu \rightarrow 0 \quad (28)$$

或者

$$\int_{A_n} p_0^n(x) d\mu \rightarrow 0 \Leftrightarrow \int_U \int_{A_n} p_h^n(x) d\mu dh \rightarrow 0 \quad (29)$$

在引理1的条件下, 对于每个有界序列 h_n , 我们有 $P_{h_n}^n \triangleleft P_0^n$, 也就是

$$\int_{A_n} p_0^n(x) d\mu \rightarrow 0 \Leftrightarrow \int_{A_n} p_{h_n}^n(x) d\mu \rightarrow 0 \quad (30)$$

我们又知道, 存在 \bar{h}_n 使得

$$\int_U \int_{A_n} p_h^n(x) d\mu dh \leq V(U) \sup_h \int_{A_n} p_h^n(x) d\mu \leq V(U) \left(\int_{A_n} p_{\bar{h}_n}^n(x) d\mu + 1/n \right) \quad (31)$$

类似的我们也有下界。综上

$$V(U) \left(\int_{A_n} p_{\bar{h}_n}^n(x) d\mu + 1/n \right) \leq \int_U \int_{A_n} p_h^n(x) d\mu dh \leq V(U) \left(\int_{A_n} p_{\bar{h}_n}^n(x) d\mu + 1/n \right) \quad (32)$$

由(30) 和(32), (29) 成立 \square

Lemma 3. 设试验 $(P_\theta : \theta \in \Theta)$ 在 θ_0 处均方可微, 具有非奇异的 Fisher 信息阵 I_{θ_0} , 假设对每个 $\epsilon > 0$, 存在一个检验序列 ϕ_n 使得

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

那么, 设 $M_n \rightarrow \infty$, 则存在一个检验序列 ϕ_n 和一个常数 $c > 0$ 使得对每个充分大的 n 和每个 θ 使得 $\|\theta - \theta_0\| \geq M_n/\sqrt{n}$, 我们有

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

见 *der Vaart (2000)*

Lemma 4. 设试验 $(P_\theta : \theta \in \Theta)$ 在 θ_0 处均方可微, 具有非奇异的 Fisher 信息阵 I_{θ_0} , 进一步, 假设存在 θ_0 的某个邻域 V 和某个函数 $\dot{\ell}$ 满足 $P_{\theta_0} \dot{\ell}^2 < \infty$, 对 $\forall \theta_1, \theta_2 \in V$, 有

$$|\log p_{\theta_1}(x) - \log p_{\theta_2}(x)| \leq \dot{\ell}(x) \|\theta_1 - \theta_2\|. \quad (33)$$

则对每个 M , 有

$$\sup_{\|h\| \leq M} \left| \log \frac{p_{h_n}^n(X)}{p_0^n(X)} - \frac{1}{\sqrt{n}} \sum_{i=1}^n h^T \dot{\ell}_{\theta_0}(X_i) + \frac{1}{2} h^T I_{\theta_0} h \right| \xrightarrow{P_0^n} 0. \quad (34)$$

(i) 设 Bernstein-von Mises 定理 (见 der Vaart (2000)) 的条件满足: Θ 是 \mathbb{R}^p 的开子集, 试验 $(P_\theta : \theta \in \Theta)$ 在 $\theta_0 \in \Theta$ 处均方可微, 具有非奇异信息阵 I_{θ_0} , 假设对任意的 $\epsilon > 0$, 存在一个相合检验序列:

$$P_{\theta_0}^n \phi_n \rightarrow 0, \quad \sup_{\|\theta - \theta_0\| \geq \epsilon} P_\theta^n (1 - \phi_n) \rightarrow 0. \quad (35)$$

设 $\pi_n(h; X)$ 是一个满足 Bernstein-von Mises 定理结论的权函数:

$$\|\pi_n(h; X) - dN(\Delta_{n, \theta_0}, I_{\theta_0}^{-1})(h)\| \xrightarrow{P_{\theta_0}^n} 0 \quad (36)$$

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

$$\lim_{n \rightarrow \infty} P_{\theta_0}^n \left(\sup_{\|h\| \geq K\sqrt{n}} (\pi_n(h; X) - T(h)) \leq 0 \right) \geq 1 - \epsilon \quad (37)$$

$$\lim_{n \rightarrow \infty} P_{\theta_0}^n \left(\sup_{\|h\| \leq K\sqrt{n}} \pi_n(h; X) \leq A \right) \geq 1 - \epsilon \quad (38)$$

(iii) 存在 θ_0 的某个邻域 V 和某个函数 $\dot{\ell}$ 满足 $P_{\theta_0} \dot{\ell}^2 < \infty$, 对 $\forall \theta_1, \theta_2 \in V$, 有

$$|\log p_{\theta_1}(x) - \log p_{\theta_2}(x)| \leq \dot{\ell}(x) \|\theta_1 - \theta_2\|. \quad (39)$$

Theorem 4. Under the Assumptions (i), (ii) and (iii). For bounded real numbers η_n , we have

$$\left| \int_{\mathbb{R}^p} \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh - 2^{-\frac{p}{2}} e^{\frac{1}{2} \Delta_{n, \theta_0}^T I_{\theta_0} \Delta_{n, \theta_0}} \right| \xrightarrow{P_{\eta_n}^n} 0 \quad (40)$$

Proof. 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

$$\left| \int_C \frac{p_h^n(X)}{p_0^n(X)} \pi_n(h; X) dh - \int_C e^{h^T I_{\theta_0} \Delta_{n, \theta_0} - \frac{1}{2} h^T I_{\theta_0} h} dN(\Delta_{n, \theta_0}, I_{\theta_0}^{-1})(h) dh \right| \xrightarrow{P_0^n} 0 \quad (41)$$

By Lemma 4, for every fixed M ,

$$\sup_{\|h\| \leq M} \left| \log \frac{p_h^n(X)}{p_0^n(X)} - h^T I_{\theta_0} \Delta_{n, \theta_0} + \frac{1}{2} h^T I_{\theta_0} h \right| \xrightarrow{P_0^n} 0 \quad (42)$$

Hence we have

$$\int_C \frac{p_h^n(X)}{p_0^n(X)} \pi_n(h; X) dh = e^{o_{P_0^n}(1)} \int_C e^{h^T I_{\theta_0} \Delta_{n, \theta_0} - \frac{1}{2} h^T I_{\theta_0} h} \pi_n(h; X) dh \quad (43)$$

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 $1 - \delta$,

$$\begin{aligned} & \int_C e^{h^T I_{\theta_0} \Delta_{n, \theta_0} - \frac{1}{2} h^T I_{\theta_0} h} |\pi_n(h; X) - dN(\Delta_{n, \theta_0}, I_{\theta_0}^{-1})(h)| dh \\ & \leq M \int_C |\pi_n(h; X) - dN(\Delta_{n, \theta_0}, I_{\theta_0}^{-1})(h)| dh \xrightarrow{P_0^n} 0 \end{aligned} \quad (44)$$

120 Combining with (43), we can conclude that (41) 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

$$\frac{\int_{C_n^c} p_h^n(X) \pi_n(h; X) dh}{\int p_h^n(X) \pi_n(h; X) dh} \xrightarrow{P_0^n} 0, \quad (45)$$

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

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

$$\frac{\int_{C_n^c} p_h^n(x) \pi(h|x) dh}{\int p_h^n(x) \pi(h|x) dh} = \frac{\int_{C_n^c} p_h^n(x) \pi(h|x) dh}{\int p_h^n(x) \pi(h|x) dh} \phi_n + \frac{\int_{C_n^c} p_h^n(x) \pi(h|x) dh}{\int p_h^n(x) \pi(h|x) dh} (1 - \phi_n) \quad (46)$$

Since (45) ≤ 1 ,

$$\frac{\int_{C_n^c} p_h^n(X) \pi_n(h; X) dh}{\int p_h^n(X) \pi_n(h; X) dh} \phi_n \leq \phi_n \xrightarrow{P_0^n} 0 \quad (47)$$

It's enough to proof

$$\frac{\int_{C_n^c} p_h^n(X) \pi_n(h; X) dh}{\int p_h^n(X) \pi_n(h; X) dh} (1 - \phi_n) \xrightarrow{P_0^n} 0 \quad (48)$$

Fix a ball U around zero. Then

$$(48) \leq \frac{\int_{C_n^C} p_h^n(X) \pi_n(h; X) dh}{\int_U p_h^n(X) \pi_n(h; X) dh} (1 - \phi_n) \quad (49)$$

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

$$\inf_{h \in U} dN(\Delta_{n, \theta_0}, I_{\theta_0}^{-1})(h) \geq m. \quad (50)$$

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

$$(49) \leq \frac{\int_{C_n^C} p_h^n(X) \pi_n(h; X) dh}{\int_{U/D_n(X)} p_h^n(X) \pi_n(h; X) dh} (1 - \phi_n) \leq \frac{\int_{C_n^C} p_h^n(X) \pi_n(h; X) dh}{\frac{m}{2} \int_{U/D_n(X)} p_h^n(X) dh} (1 - \phi_n) \quad (51)$$

Next we proof

$$\frac{\int_{U/D_n(X)} p_h^n(X) dh}{\int_U p_h^n(X) dh} \xrightarrow{P_0^n} 1. \quad (52)$$

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

$$\int |\pi_n(h; X) - n(\Delta_{n, \theta_0}, I_{\theta_0}^{-1})| dh \xrightarrow{P_0^n} 0 \quad (53)$$

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

$$(52) = \frac{\int_{U/D_n(X)} \frac{p_h^n(X)}{p_0^n(X)} dh}{\int_U \frac{p_h^n(X)}{p_0^n(X)} dh} = \frac{e^{o_{P_0^n}(1)} \int_{U/D_n(X)} e^{h^T I_{\theta_0} \Delta_{n, \theta_0} - \frac{1}{2} h^T I_{\theta_0} h} dh}{e^{o_{P_0^n}(1)} \int_U e^{h^T I_{\theta_0} \Delta_{n, \theta_0} - \frac{1}{2} h^T I_{\theta_0} h} dh} \xrightarrow{P_0^n} 1 \quad (54)$$

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

Now, to obtain (51) $\xrightarrow{P_0^n} 0$, we only need to prove

$$\frac{\int_{C_n^C} p_h^n(X) (A \mathbf{1}_{M_n \leq \|h\| \leq K\sqrt{n}} + T(h) \mathbf{1}_{\|h\| > K\sqrt{n}}) dh}{\int_U p_h^n(X) dh} (1 - \phi_n) \xrightarrow{P_0^n} 0 \quad (55)$$

By Lemma 2, we only need to prove (55) $\xrightarrow{P_U^n} 0$. To prove that, we only need to prove (55) $\xrightarrow{L_{P_U^n}^1} 0$, that is

$$\int \frac{\int_{C_n^C} p_h^n(x) T(h) dh}{\int_U p_h^n(x) dh} (1 - \phi_n) \int_U p_h^n(x) dh d\mu \rightarrow 0 \quad (56)$$

We note that

$$(56) = \int_{C_n^C} \left(\int (1 - \phi_n) p_h^n(x) d\mu \right) T(h) dh \quad (57)$$

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

$$\int (1 - \phi_n) p_h^n(x) d\mu \leq e^{-c(\|h\|^2 \wedge n)} \quad (58)$$

Splitting the integral into the domains $M_n \leq \|h\| \leq \sqrt{n}$ and $\|h\| \geq \sqrt{n}$ for $D \leq 1$, we see that

$$(57) \leq \int_{\|h\| \geq M_n} e^{-c^* \|h\|^2} dh + \int \rightarrow 0 \quad (59)$$

Finally we have

$$\begin{aligned} & \left| \int \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh - 2^{-\frac{p}{2}} e^{\frac{1}{2} \Delta_{n, \theta_0}^T I_{\theta_0} \Delta_{n, \theta_0}} \right| \\ &= \left| \int \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh - \int_{C_n} \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh \right| \\ &+ \left| \int_{C_n} \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh - \int_{C_n} e^{h^T I_{\theta_0} \Delta_{n, \theta_0} - \frac{1}{2} h^T I_{\theta_0} h} dN(\Delta_{n, \theta_0}, I_{\theta_0}^{-1})(h) dh \right| \\ &+ \left| \int_{C_n} e^{h^T I_{\theta_0} \Delta_{n, \theta_0} - \frac{1}{2} h^T I_{\theta_0} h} n(\Delta_{n, \theta_0}, I_{\theta_0}^{-1}) dh - 2^{-\frac{p}{2}} e^{\frac{1}{2} \Delta_{n, \theta_0}^T I_{\theta_0} \Delta_{n, \theta_0}} \right| \\ &= J_1 + J_2 + J_3 \end{aligned} \quad (60)$$

由证明的第一步, 我们有 $J_2 \xrightarrow{P_0^n} 0$, 所以我们知道 $\int_{C_n} \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh$ 依概率有界。所以

$$\begin{aligned} J_1 &= \int_{C_n} \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh \left| \frac{\int \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh}{\int_{C_n} \frac{p_h(X)}{p_0(X)} \pi_n(h; X) dh} - 1 \right| \\ &= O_{P_0^n}(1) o_{P_0^n}(1) \end{aligned} \quad (61)$$

And J_3 convenges to 0 for trivial reason. \square

Proof of Theorem 1. 如果原假设成立, 即真实参数 θ_0 是 Θ 的一个内点, 同时 θ_0 是 Θ_0 的一个相对内点, 那么对积分似然比统计量(5)的分子和分母, 我们都可以应用定理4, 并取定理中的 $\eta_n = 0$ 。而由中心极限定理, 我们有

$$I_{\theta_0} \Delta_{n, \theta_0} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \overset{P_0^n}{\rightsquigarrow} \xi, \quad (62)$$

其中 $\xi \sim N(0, I_{\theta_0})$ 。

$$I_{\theta_0}^* \Delta_{n, \theta_0}^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_{\theta_0}^*(X_i) \overset{P_0^n}{\rightsquigarrow} \xi^*, \quad (63)$$

其中 ξ^* 是 ξ 的前 p_1 维向量。

所以

$$\begin{aligned} \Lambda(X) &= \frac{2^{-\frac{p_2}{2}} \exp\{\frac{1}{2} \Delta_{n, \theta_0}^T I_{\theta_0} \Delta_{n, \theta_0}\} + o_{P_0^n}(1)}{2^{-\frac{p_1}{2}} \exp\{\frac{1}{2} \Delta_{n, \theta_0}^{*T} I_{\theta_0}^* \Delta_{n, \theta_0}^*\} + o_{P_0^n}(1)} \\ &\overset{P_0^n}{\rightsquigarrow} \frac{2^{-\frac{p_2}{2}} \exp\{\frac{1}{2} \xi^T I_{\theta_0}^{-1} \xi\}}{2^{-\frac{p_1}{2}} \exp\{\frac{1}{2} \xi^{*T} I_{\theta_0}^{*-1} \xi^*\}} \end{aligned} \quad (64)$$

而

$$\xi^T I_{\theta_0}^{-1} \xi - \xi^{*T} I_{\theta_0}^{*-1} \xi^* = (I_{\theta_0}^{-\frac{1}{2}} \xi)^T (I_{p_2 \times p_2} - I_{\theta_0}^{\frac{1}{2}} \begin{pmatrix} I_{\theta_0}^{*-1} & 0 \\ 0 & 0 \end{pmatrix} I_{\theta_0}^{\frac{1}{2}}) (I_{\theta_0}^{-\frac{1}{2}} \xi) \quad (65)$$

易知 $I_{\theta_0}^{-\frac{1}{2}} \xi$ 是 p_2 维标准正态分布, 而中间乘的矩阵是个秩为 $p_2 - p_1$ 的投影矩阵, 所以我们得到

$$2 \log(\Lambda(X)) \overset{P_0^n}{\rightsquigarrow} \chi_{p_2 - p_1}^2 - (p_2 - p_1) \log(2) \quad (66)$$

Proof of Theorem 2. 由于 $h_n = \eta_n$ 收敛到 η , 由均方可微性, 由引理1及中心极限定理有:

$$\begin{aligned} \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \\ \log \frac{p_{\eta_n}(X)}{p_0(X)} \end{pmatrix} &= \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta^T \dot{\ell}_{\theta_0}(X_i) - \frac{1}{2} \eta^T I_{\theta_0} \eta \end{pmatrix} + o_{P_0^n}(1) \\ &\overset{P_0^n}{\rightsquigarrow} N\left(\begin{pmatrix} 0 \\ -\frac{1}{2} \eta^T I_{\theta_0} \eta \end{pmatrix}, \begin{pmatrix} I_{\theta_0} & I_{\theta_0} \eta \\ \eta^T I_{\theta_0} & \eta^T I_{\theta_0} \eta \end{pmatrix}\right) \end{aligned} \quad (67)$$

故根据 Le Cam 第三引理, 我们有

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) \overset{P_0^n}{\rightsquigarrow} \xi \sim N(I_{\theta_0} \eta, I_{\theta_0}) \quad (68)$$

由定理4, 在 $P_{\eta_n}^n$ 下, 我们依然有渐近等式(64) 所以有

$$2\log(\Lambda(X)) \stackrel{P_{\eta_n}^n}{\rightsquigarrow} \chi_{p_2-p_1}^2(\delta) - (p_2 - p_1) \log(2) \quad (69)$$

其中非中心参数 δ 可以在(65)中把 ξ 替换成 $I_{\theta_0}\eta$ 得到:

$$\begin{aligned} \delta &= \eta^T (I_{\theta_0} - I_{\theta_0} \begin{pmatrix} I_{\theta_0}^{*-1} & 0 \\ 0 & 0 \end{pmatrix} I_{\theta_0}) \eta \\ &= \eta^T \begin{pmatrix} 0 & 0 \\ 0 & I_{22 \cdot 1} \end{pmatrix} \eta \end{aligned} \quad (70)$$

References

- 135 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;.
- 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.
- 140 Bai Z, Saranadasa H. Effect of high dimension: by an example of a two sample problem. Statistica Sinica 1996;6(2):311–29.
- 145 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.
- der Vaart AV. Asymptotic statistics. Cambridge university press, 2000.
- 150 Zhao J, Xu X. A generalized likelihood ratio test for normal mean when p is greater than n. Computational Statistics & Data Analysis 2016;.