arXiv:2601.09122v1 [math.ST] 14 Jan 2026

# 数据驱动后验温度调整的统计保证

Ruchira Ray MarcoAvella Medina
Cynthia Rush

哥伦比亚大学统计系 *

**摘要**

后验温度调整通过将似然提升至分数次幂 $\alpha$，降低似然在后验分布计算中的影响。由此得到的幂后验——也称为 $\alpha$-后验或分数后验——已被证明具有一些理想性质，包括对模型错设的稳健性和渐近正态性（伯恩斯坦-冯·米塞斯定理）。然而，如何选择温度参数，以及由此得到的幂后验具有何种统计保证，仍是悬而未决的问题。基于交叉验证的调参方法表明，所选 $\alpha$可能呈现有趣的渐近情形：它可能趋于零，也可能表现为一种混合分布，其中包含无穷远处的点质量，而其余质量趋于零。我们对这些情形下幂后验的渐近性质进行了形式化刻画。具体而言，我们给出了以下性质成立的充分条件：（i）幂后验矩的一致性；（ii）幂后验均值的渐近正态性。我们的分析需要建立一种新的拉普拉斯近似，这一结果本身也很有意义，并且是证明存在某个临界阈值 $\alpha \asymp 1/\sqrt{n}$ 、使得后验均值的渐近正态性在该处失效的关键技术工具。我们的结果允许幂以任意方式依赖于数据。

**关键词：** 幂后验、参数调优、交叉验证、伯恩斯坦-冯·米塞斯定理、拉普拉斯近似。

*本研究部分由美国国家科学基金会拨款资助：DMS-2310973（Avella Medina）和 DMS-2413828（Rush）。可在 https://github.com/ rray123/data-dependent-tempering 获取用于以较小规模复现实验和图表的代码。

# 1 引言

后验温度调整在贝叶斯统计中日益受到关注，因为在模型设定错误的情况下，与标准后验分布相比，它能提高稳健性并带来更可靠的推断 [30, 31, 36, 66, 5]。具体而言，后验温度调整通过将似然提升至一个分数指数，降低（可能设定错误的）似然在构造后验分布时的影响。由此得到的后验分布称为幂后验、分数后验 [9], 或 $\alpha$-后验 [82, 27, 第 8 章。6]。近期有一系列研究探讨了 $\alpha$-后验分布的统计性质，其中 $\alpha$ 被设定为一个固定常数 [9, 2, 82, 5, 69]。 [5] 中的作者确立了 $\alpha$-后验分布的渐近正态性，并表明相较于标准后验分布，其极限方差按 $\alpha$ 缩放，而均值保持不变。在类似的正则性条件下， [69] 推广了这一结果，确立了 $\alpha$-后验矩的一致性以及$\alpha$-后验均值的渐近正态性。

在实践中， $\alpha$ 并非一个常数值，因为理论上的设定依赖于未知的总体量 [5], ，而是根据数据进行调参。 $\alpha$ 的最优调参对于$\alpha$-后验分布的性能至关重要。为此，文献中提出了许多方法，用数据 [30, 31, 80, 36, 66, 68, 73, 60, 14, 64]对$\hat{\alpha}_n$ 进行调参。其中一些方法指出，考察 $\alpha$-后验分布性能时，合适的渐近情形并不对应于固定参数 $\alpha$，而应是随样本量 $n$变化的参数。本文将 [5, 69]的渐近分析推广到 $\alpha$ 为随机且依赖于数据的情形。

在第 2 节中，我们通过实证研究数据驱动的 $\alpha$选择如何随 $n$变化。我们在正则化线性回归的背景下，探究通过基于交叉验证的方法调优的 $\hat{\alpha}_n$ 值的渐近行为。我们进行了大量数值实验，结果表明，随着 $n$增加，所选的 $\hat{\alpha}_n$ 可能趋于零；而对于贝叶斯交叉验证和训练集-测试集划分等某些方法，所选值可能服从混合分布。这种混合行为与先前的发现一致，即

这些方法倾向于以某个固定概率 [68, 64, 71]选择 $\hat{}\alpha_n = \infty$ 。总体而言，这些数值结果促使我们提出这样的观点：对数据驱动的 $\hat{}\alpha_n$-后验分布进行统计分析时，应考虑此前尚未探索的情形，即随着样本量增加，$\hat{}\alpha_n$ 依概率趋于零或发散。

我们的主要贡献可概括如下：

(a) 当 $\hat{}\alpha_n$ 是满足 $\frac{1}{n} \ll \hat{}\alpha_n \ll 1$ （依概率）的序列时，我们证明了所有 $\hat{}\alpha_n$-后验矩的一致性（定理 1），表明它们收敛于以最大似然估计量（MLE）为中心的正态分布的矩。我们的结果蕴含了一个关于 $\hat{}\alpha_n$-后验分布的伯恩斯坦-冯·米塞斯（BvM）定理（推论 1），且该定理是尖锐的，也就是说，$\frac{1}{n} \ll \hat{}\alpha_n$ （依概率）是必要条件（命题1）。在 $\hat{}\alpha_n \to \infty$ （依概率）的情形下，我们证明了 $\hat{}\alpha_n$-后验分布的极限分布是位于最大似然估计量处的点质量（命题 3）。我们利用这一结果，在一种混合情形下建立了伯恩斯坦-冯·米塞斯型定理：$\hat{}\alpha_n$ 要么以固定概率趋于无穷大，要么趋于零（定理 3）。据我们所知，文献尚未研究过这些情形中的任何一种，但它们都得到了经验性证据的有力支持。我们的工作推动了对更现实情形下幂后验渐近性质的理解；理论分析还揭示了 $\hat{}\alpha_n$-后验与固定 $\alpha$ 的 $\alpha$-后验在渐近行为上的重要差异，从业者在利用数据构建调节 $\alpha$ 的方法时可能需要考虑这些差异。

(b) 我们证明，当 $\frac{1}{\sqrt{n}} \ll \hat{}\alpha_n \ll 1$ 以概率趋于.....时，$\hat{}\alpha_n$-后验均值具有渐近正态性（推论 2）。我们的证明关键依赖于一项新的拉普拉斯近似结果，该结果本身也很有意义（引理 1）。拉普拉斯近似使我们能够（$i$）准确量化后验均值与最大似然估计之间的距离（定理 2）；（$ii$）证明 $\frac{1}{\sqrt{n}} \ll \hat{}\alpha_n$ 以概率趋于.....是后验均值与最大似然估计渐近等价的必要条件。

—=

## 1.1 相关工作

贝叶斯统计领域有大量与我们这项工作相关的研究。最明显的联系或许是关于贝叶斯后验渐近性质的文献，以及关于温度参数 $\alpha$ 选择方法的文献。具体而言，我们的工作建立在针对模型设定错误的似然 [50, 5]，下标准后验分布的现有贝叶斯—冯·米塞斯定理之上，但我们也注意到，针对广义后验分布 [18, 28, 65, 53, 61] 的结果范围更广。温度参数选择方法包括 [30, 31]，的 SafeBayesian 算法、 [73, 59, 4,62, 81]，频率学派校准方法、 [36]，期望信息匹配、 [14] 期望泛化误差，以及 [64, 53] 贝叶斯预测方法。

$\alpha$后验分布之所以广受欢迎，主要源于它们在理论和实践中均展现出的稳健性。最初，人们在模型设定错误的背景下探讨了 $\alpha$后验分布的稳健性 [31, 36, 66, 5, 56]。此外，也有人提出了多种方法，以增强标准后验构造对离群值的稳健性，例如贝叶斯数据重加权 [80] ，或利用稳健损失 [37, 28,3, 62, 61] 和检验统计量 [6]构造广义后验。最后，我们注意到， $\alpha$ 后验分布也在变分推断 [38, 82] 和变分自编码器 [34, 15]的背景下得到过研究。

我们在附录 E 中对相关工作进行了更详细的讨论。

## 1.2 统计模型与 $\alpha$后验分布

现在，我们介绍本文将采用的统计框架。为对随机样本 $X^n = (X_1, \ldots, X_n) \in \mathcal{X}^n$进行建模，其中 $\mathcal{X} \subset \mathbb{R}^d$ 表示样本空间，令$\mathcal{F}_n = \{f_n(\cdot|\theta) : \theta \in \Theta \subset \mathbb{R}^p\}$ 为一个密度参数族。该模型允许设定错误，即 $f_{0,n}(X^n)$（随机样本 $X^n$的真实密度）可能不属于 $\mathcal{F}_n$。我们将使用伪真参数 $\theta^*$这一概念；若模型设定正确，也就是说 $f_{0,n}(X^n) = f_n(X^n|\theta^*) \in \mathcal{F}_n$，则它就是真实数据生成 $\theta$ 。若模型设定错误，则 $\theta^*$ 在

KL散度意义下最接近真实数据生成过程。我们将（伪）最大似然估计记为 $\theta = \arg\max_\theta f_n(X^n|\theta)$。

对于定义在随机向量 $Y_n$ 上的分布序列 $f_{0,n}$，若对每个 $\epsilon > 0$，都有 $\lim_n \mathbb{P}_{f_{0,n}}(\|Y_n\|_2 > \epsilon) = 0$，则称 $Y_n = o_{f_{0,n}}(1)$。若对每个 $\epsilon > 0$，都存在 $M \equiv M(\epsilon) > 0$，使得 $\mathbb{P}_{f_{0,n}}(\|Y_n\|_2 > M) < \epsilon$，则称 $Y_n$ “在 $f_{0,n}$-概率意义下有界”，或记作 $O_{f_{0,n}}(1)$。

给定统计模型 $f_n(X^n|\theta) \in \mathcal{F}_n$、$\theta$ 的先验密度 $\pi(\theta)$ 以及标量 $\alpha > 0$，$\alpha$-后验分布定义为具有如下密度的分布

$$\pi_{n,\alpha}(\theta|X^n) \equiv \frac{f_n(X^n|\theta)^\alpha \pi(\theta)}{\int_{\mathbb{R}^p} f_n(X^n|\theta)^\alpha \pi(\theta) d\theta}. \tag{1}$$

我们在 $\alpha$ 的下标中加入 $n$，以强调 $\alpha$ 对样本量 $n$ 的依赖性。我们还区分 $\alpha_n$-后验分布（其中序列 $\alpha_n$ 是先验指定的）和 $\hat{}\alpha_n$-后验分布（其中 $\hat{}\alpha_n$ 是从数据中学习得到，或以任何方式依赖于数据）。在全文中，我们假定所有 $\alpha$-后验分布都是正规的（即式 (1) 分母中的归一化常数有限）。事实上，当 $\pi(\theta)$ 是正规的时，情况确实如此，见定理 1]。

## 1.3 符号与一般定义

我们使用 $\phi(\cdot|\mu, \Sigma)$ 表示均值为 $\mu$、协方差为 $\Sigma$ 的多元正态密度。矩阵 $B$ 的行列式为 $|B|$，事件 $A$ 的示性函数为 $\mathbb{1}\{A\}$。对于序列 $a_n$ 和 $b_n$，如果存在 $M > 0$ 和 $N$，使得对于所有 $n > N$ 都有 $|a_n| \leq M b_n$，则称 $a = O(b_n)$。如果 $a_n = O(b_n)$ 且 $b_n = O(a_n)$，则称 $a_n \asymp b_n$。我们使用 $\|v\|_p = (\sum_i |v_i|^p)^{1/p}$ 来表示向量 $v$ 的 $p$-范数。令 $B_v(\delta) \equiv \{\theta : \|\theta - v\|_2 \leq \delta\}$ 表示以向量 $v \in \mathbb{R}^p$ 为中心、半径为 $\delta$ 的闭欧几里得球，令 $B_v(\delta) \equiv \{\theta : \|\theta - v\|_2 < \delta\}$ 表示开欧几里得球。对于任意两个概率密度 $f$ 和 $g$，若它们相对于 $\mathbb{R}^p$ 中的勒贝格测度绝对连续，则它们之间的全变差距离 $d_{\mathrm{TV}}(f, g)$ 为

$$d_{\mathrm{TV}}(f, g) = \frac{1}{2} \int_{\mathbb{R}^p} |f(x) - g(x)| dx. \tag{2}$$

对于定义在带有欧几里得范数的 $\mathbb{R}^p$ 上的任意两个概率测度 $\mu$ 和 $\nu$，它们之间的 $p$-Wasserstein 距离定义为

$$d_{\mathrm{W}_p}(\mu,\nu)=\Big(\inf_{\gamma\in\Gamma(\mu,\nu)}\int_{\mathbb{R}^p\times\mathbb{R}^p}\|x-y\|_2^p d\gamma(x,y)\Big)^{1/p}, \tag{3}$$

其中，$\Gamma(\mu,\ \nu)$ 是 $\mu$ 和 $\nu$ 在乘积空间 $\mathbb{R}^p\times\mathbb{R}^p$ 上的联合分布集合。我们使用张量积记法表示向量的多元多项式，$v\in\mathbb{R}^p$。定义 $\mathrm{k}^{\mathrm{th}}$ 阶张量 $T\in\mathbb{R}^{p\times\cdots\times p}$ 与 k 个向量 $v^{(1)}$, $v^{(2)}$, ……, $v^{(k)}$（全部为 $\in\mathbb{R}^p$）之间的“张量内积”为

$$\left\langle T,v^{(1)}\otimes\ldots\otimes v^{(k)}\right\rangle=\sum_{i_1=1}^{p}\cdots\sum_{i_k=1}^{p}[T]_{i_1,\ldots,i_k}\times v_{i_1}^{(1)}\times\cdots\times v_{i_k}^{(k)}, \tag{4}$$

其中，$[T]_{i_1,\ \ldots\ldots,\ i_k}$ 表示 $T$ 在 $1\le i_1$，……，$i_k\le p$ 处的元素。我们还将向量 $v\in\mathbb{R}^p$ 的 $\mathrm{k}^{\mathrm{th}}$ 阶外积的 1-范数定义为

$$\|v^{\otimes k}\|_1=\sum_{i_1=1}^{p}\cdots\sum_{i_k=1}^{p}\left|v_{i_1}\times\cdots\times v_{i_k}\right|. \tag{5}$$

## 1.4 结构安排

本文其余部分的组织如下。第 2 节详细讨论线性回归背景下 $\alpha$-后验分布与正则化估计之间的联系。这些联系以及大量数值实验为我们的理论分析提供了动机。第3节介绍主要的渐近分析；具体而言，对于第 2 节数值实验中出现的 $\hat{}\alpha_n$ 情形，我们给出 $\hat{}\alpha_n$-后验矩一致性/贝叶斯-冯·米泽斯型结果，以及后验均值的渐近正态性。第 4 节通过讨论总结全文。所有证明、补充讨论和示例均置于附录中。

# 2 一个用于调参的动机示例 $\alpha$

本节考虑岭回归中正则化参数的调参问题，该问题等价地可以表述为贝叶斯线性回归中的温度参数调参问题。我们在岭回归和贝叶斯线性回归这两种情形下，使用各种方法选取调参后的参数值，并对其极限分布进行数值研究。尽管这是一个简单的示例，但它为我们将在本文后续部分研究的一般情形下的渐近分析提供了动机和直观认识。

## 2.1 模型设定、 $\alpha_n$-后验分布与岭回归

我们在以下回归设定中进行数值实验。考虑配对数据$\{x_i, y_i\}_{i=1}^n$，其中对于所有 $i = 1, \ldots, n$， $x_i = [x_{i,1}, x_{i,2}, x_{i,3}]^\top \in \mathbb{R}^3$和 $y_i \in \mathbb{R}$。令 $x_i \sim \mathcal{N}(0, I_3)$ 和 $\epsilon_i \sim \mathcal{N}(0, 1)$ 相互独立。我们根据生成模型对每个样本进行模拟:

$$y_i = x_i^\top \beta^* + \epsilon_i. \tag{6}$$

在实验中，对于 $\epsilon_i \sim \mathcal{N}(0, 1)$ 和 $\tilde{x}_i = [x_{i,2}, x_{i,3}]^\top \in \mathbb{R}^2$，我们考虑以下工作模型:

$$y_i = \tilde{x}_i^\top \tilde{\beta} + \epsilon_i. \tag{7}$$

真实参数设为 $\beta^* = [\beta_1^*, -0.5, 0.1]$，因此，当 $\beta_1^* = 0$ 时，(7) 是“设定正确的”；而当 $\beta_1^* \neq 0$ 时，由于遗漏变量 $x_{i,1}$，它是“设定错误的”。在实验中，我们在设定错误的情况下令 $\beta_1^* = 1$ 。

注意，在斜率参数采用高斯先验的线性回归模型中，使用岭回归估计 $\beta$ 与关于 $\beta$ 的 $\alpha$- 后验推断是等价的。事实上，模型 (6) 给出了在给定系数向量 $\beta$ 时数据 $\{x_i, y_i\}_{i=1}^n$ 的似然。

$\sqrt{n}(\hat{\beta}^{\mathrm{Ridge}} - b(\lambda) - \beta_0)$ 的表现

| 情形 | | $\alpha_n$-后验分布 | $\hat{\beta}^{\mathrm{Ridge}}$（$\alpha_n$-后验均值） | |
|---|---|---|---|---|
| $\lambda_n$ | $\alpha_n$ | 方差 | 偏差 $b(\lambda)$ | $\sqrt{n}(\hat{\beta}^{\mathrm{Ridge}} - \hat{\beta})$ |
| $\frac{1}{\sqrt{n}} \ll \lambda_n \ll 1$ | $\frac{1}{n} \ll \alpha_n \ll \frac{1}{\sqrt{n}}$ | $O_{f_0,\,n}(\frac{1}{\alpha_n n})$ | $O(\lambda_n)$ | 发散 |
| $\lambda_n \asymp \frac{1}{\sqrt{n}}$ | $\alpha_n \asymp \frac{1}{\sqrt{n}}$ | $O_{f_0,\,n}(\frac{1}{\sqrt{n}})$ | $O(\frac{1}{\sqrt{n}})$ | $O_{f_0,\,n}(1)$ |
| $\frac{1}{n} \ll \lambda_n \ll \frac{1}{\sqrt{n}}$ | $\frac{1}{\sqrt{n}} \ll \alpha_n \ll 1$ | $O_{f_0,\,n}(\frac{1}{\alpha_n n})$ | $O(\lambda_n)$ | $o_{f_0,\,n}(1)$ |
| $\lambda_n \asymp \frac{1}{n}$ | $\alpha_n \asymp 1$ | $O_{f_0,\,n}(\frac{1}{n})$ | $O(\frac{1}{n})$ | $o_{f_0,\,n}(1)$ |

表 1：$\hat{\beta}^{\mathrm{Ridge}}$ 与相应 $\alpha$-后验分布的渐近行为。我们用 $b(\lambda)$ 表示岭估计量/$\alpha$-后验均值的偏差。最后三列分别表示 $\alpha_n$-后验分布的方差、岭估计量的偏差，以及 $\sqrt{n}(\hat{\beta}^{\mathrm{Ridge}} - \hat{\beta})$ 的阶，其中 $\beta$ 是 $\beta$ 的最大似然估计（即普通最小二乘估计量）。

Co将此似然与 $\mathcal{N}(0, I_3)$ 先验分布结合，可得到 $\beta$的 $\alpha_n$-后验分布

$$\pi_{n,\alpha_n}(\beta|X,Y) = \phi\Big(\beta \,\Big|\, \hat{\beta}^{\mathrm{Ridge}},\ \frac{1}{\alpha_n n}\Big(\frac{1}{n}X^\top X + \frac{1}{\alpha_n n} I_p\Big)^{-1}\Big), \tag{8}$$

其中，利用 $\lambda_n = \frac{1}{n\alpha_n}$,

$$\hat{\beta}^{\mathrm{Ridge}} = \arg\min_{\beta} \frac{1}{n}\|Y - X\beta\|_2^2 + \lambda_n \|\beta\|_2^2, \tag{9}$$

其中 $X = [x_1 \cdots x_n]^\top$和 $Y = [y_1, \ldots, y_n]$。标准频率学派理论表明，在满足某些正则性条件时，只要 $\lambda_n = o(1)$，$\hat{\beta}^{\mathrm{Ridge}}$ 就是一致的；如果 $\lambda_n = o(1/\sqrt{n})$ [35]，$\hat{\beta}^{\mathrm{Ridge}}$ 则渐近服从正态分布。在表1中，我们比较了这些情形下$\hat{\beta}^{\mathrm{Ridge}}$ 与式（8）中的 $\alpha_n$-后验分布在 $\lambda_n$ 各情形下的渐近行为（以及相应的 $\alpha_n$情形）。表1表明，$1/n \ll \alpha_n \ll 1$ （第1–3行）和$1/\sqrt{n} \ll \alpha_n \ll 1$ （第3行）这两种情形尤为有趣。在 $1/n \ll \alpha_n \ll 1$的情形下，$\alpha_n$-后验分布的方差和岭回归估计量的偏差都趋于零。在 $1/\sqrt{n} \ll \alpha_n \ll 1$的情形下，$\alpha_n$-后验均值——它与岭回归估计量一致——渐近等价于最大似然估计。我们将在第3节看到，这些情形确实最有助于建立伯恩斯坦-冯·米塞斯定理（当 $1/n \ll \alpha_n \ll 1$时）以及 $\alpha$-后验均值估计量的渐近正态性（当

$1/\sqrt{n} \ll \alpha_n \ll 1$）。在介绍我们的理论结果之前，我们先提供经验性证据，说明自然的数据驱动选择 $\alpha$ 也会使我们进入这类$\hat{}\alpha_n$ 趋于零的情形。

## 2.2 数据驱动的温度参数选择

在设定正确和模型设定错误的情形（第 2.1 节）下，我们研究了根据以下常用参数调优方法选择的数据驱动的温度参数 $\hat{}\alpha_n$ 的分布：

1. **贝叶斯交叉验证（BCV）**：我们选择 $\alpha$，以最大化由 [25, 77, 24] 给出的留一法逐点对数预测密度

$$\overline{\mathrm{lppd}}_{\mathrm{LOO}}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \log \int p(x_i, y_i|\beta)\pi_{n,\alpha}\left(\beta|X_{-i}Y_{-i}\right) d\beta, \tag{10}$$

其中，$X_{-i}$ 和 $Y_{-i}$ 分别是移除 $i^{\mathrm{th}}$ 条目后的设计矩阵和因变量。$\mathrm{lppd}_{\mathrm{LOO}}$ 损失函数曾在 [82] 中关于贝叶斯推断中的超参数调优的背景下进行讨论。关于后验分布（8）的 $\mathrm{lppd}_{\mathrm{LOO}}(\alpha)$ 计算，请参见附录 F.1.1。

——

2. **贝叶斯交叉验证 + 变分推断（BCV + 变分推断）**：我们还考虑将（10）中的 $\alpha$-后验分布替换为其高斯平均场变分近似

$$\overline{\mathrm{lppd}}_{\text{LOO-VI}}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \log \int p\left(x_i, y_i|\beta\right) \tilde{\pi}_{n,\alpha}\left(\beta|X_{-i}Y_{-i}\right) d\beta, \tag{11}$$

其中 $\tilde{\pi}_{n,\ \alpha}(\cdot)$ 是高斯平均场变分近似的密度

(8):

$$\tilde{\pi}_{n,\alpha}\left(\beta|X,Y\right) = \phi\Big(\beta\big|\hat{\beta}^{\mathrm{Ridge}},\ \frac{1}{\alpha n}\Big[\mathrm{diag}\Big(\frac{1}{n}X^{\top}X + \frac{1}{\alpha n}I\Big)\Big]^{-1}\Big). \tag{12}$$

关于式（12）中后验分布的 $\overline{\mathrm{lppd}}_{\text{LOO-VI}}(\alpha)$ 计算，请参见附录 F.1.2。

3. **留一交叉验证**：我们选择 $\alpha$，使 Allen 的 PRESS 统计量 [1], 最小，其定义为 $\text{PRESS}(\alpha) = \frac{1}{n}\|B(I_n - H(\alpha))Y\|_2^2$，其中 $B$ 是对角矩阵，其对角元素为 $B_{ii} = (1 - H_{ii})^{-1}$ 和 $H(\alpha) = X(X^\top X + \frac{1}{\alpha} I_p)^{-1} X^\top$。Allen 的 PRESS 统计量等价于 $\hat{}\beta^{\text{Ridge}}$ 的留一交叉验证（LOOCV）平方误差。

4. **训练-测试划分（训练-测试）**：给定划分为训练集和测试集的数据，我们选择 $\alpha$，以最小化基于训练数据集估计的 $\hat{}\beta^{\text{Ridge}}$ 的平方测试误差。我们报告的结果对应于 70-30 的训练-测试数据划分。由于 50-50 和 90-10 划分得到的结果非常相似，因此省略了这些结果。

5. **SafeBayes**：SafeBayesian 算法在 [29] 中提出，并在 [31] 中针对模型设定错误的线性模型进行了研究。该方法选择 $\alpha \in [0, 1]$，以最小化累积后验期望对数损失。由于我们假设噪声方差已知，因此使用 R 平方 Safe-Bayesian 岭回归算法 [33]。

我们使用网格搜索，根据各方法选取最优参数。对于选取 $\lambda$ 的 $\hat{}$ 方法，我们将选定的值（记为 $\lambda_n$）映射为与之对应的 $\hat{}\alpha_n$ 值。这些映射以及所使用的网格见附录 F.2 中的表4。

## 2.3 模拟结果

图1–3展示了在第2.1节所述的模型设定情形下，根据第2.2节讨论的五种方法选取的数据进行1,000次独立重复抽样所得的$\hat{}\alpha_n$分布。为了解$\hat{}\alpha_n$的渐近行为，我们比较了当 $n = 100$、$n = 1,000$以及 $n = 5,000$时$\hat{}\alpha_n$ 的分布。

我们观察到$\hat{}\alpha_n$具有以下几类渐近行为：

1.**常数极限（图1）：**在两种模型设定情形下，留一交叉验证和 SafeBayes 均表明，$\hat{}\alpha_n$ 趋近于一个常数。由此得到的常数 $\alpha$-后验分布的渐近性质已在 [5] 和 [69]中得到研究。具体而言，关于

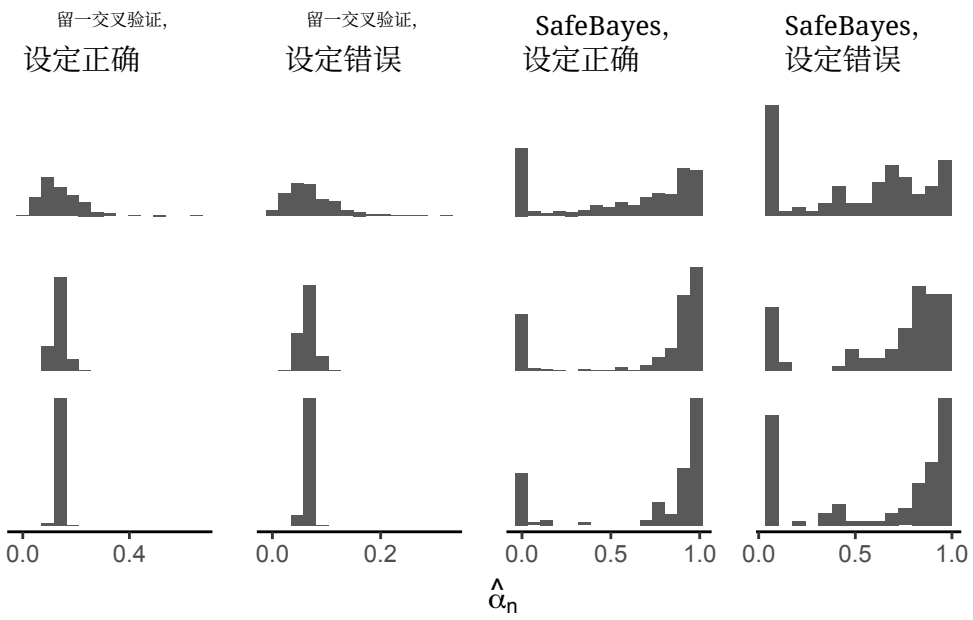



图1： 在 1,000 次重复试验中所选 $\hat{}\alpha_n$ 的分布。* 列标题表示调优方法（留一交叉验证、SafeBayes）和模型设定（设定正确、设定错误）。每列中的各行对应每次重复实验的样本量（$n = 100$, $n = 1$，000, $n = 5$, 000）。
*由于计算限制，SafeBayes 的 $n = 5$，000 图包含 100 次重复实验

$\alpha$-后验分布的贝叶斯-冯·米泽斯定理已在 [5] 中得到确立。 $\alpha$-后验分布的矩一致性以及 $\alpha$-后验均值的渐近正态性已在 [69] 中得到确立。

**2. 迅速趋于零的 $\hat{}\boldsymbol{\alpha_n}$ （图2）：** 在模型设定错误的情形下，BCV 和 BCV+ 变分推断表明 $\hat{}\alpha_n$ 趋于零。其收敛速率大致为$\frac{1}{n}$ 的量级，如表 2 所示。在第3.1.2节中，我们研究 $\alpha_n$-后验分布在 $\alpha_n \asymp \frac{1}{n}$ 时的情况，并通过指数族共轭模型给出一个反例（命题1），说明贝叶斯—冯·米泽斯结果通常并不成立。

**3. 混合 $\hat{}\boldsymbol{\alpha_n}$ （图3）：** 其余情形（模型设定正确时的 BCV 和 BCV +VI，以及两种模型设定情形下的训练-测试）表明，$\hat{}\alpha_n$ 的分布在实质上等于 $\alpha = \infty$ 的位置有一个点质量；我们将其称为“角点解”，因为它本质上对应于最大似然估计处的一个质量点，而其余质量似乎趋于零。关于如何将 $\hat{}\alpha_n$值重新编码为 $\infty$ 的更多细节，请参见附录 F.2。表 2 显示了趋于零的收敛速率

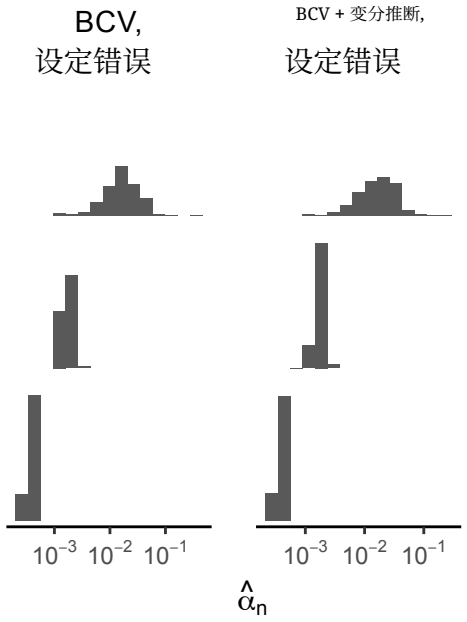



图2： 在 1,000 次重复试验中所选 $\hat{}\alpha_n$ 的分布。* 各列标题表示调优方法（BCV、BCV+变分推断）和模型设定（设定错误）。每列中的各行对应每次重复实验的样本量（$n = 100$、$n = 1,\ 000$、 $n = 5,\ 000$）。

*为更清楚地展示 $\hat{}\alpha_n$ 趋于零的收敛情况，$\hat{}\alpha_n$ 值以对数尺度绘制。$\hat{}\alpha_n$ 值以线性尺度绘制的直方图见附录 F.4.1 中的图8。

分布的非退化分量的估计收敛速度慢于 $\frac{1}{n}$。此外，从渐近角度看，约有一半的 $\hat{}\alpha_n$ – 是角点解（见图 4）。我们将此情形下估计的 $\alpha$ 称为 $\hat{}\alpha_n^{\text{mix}}$，并在第3.3节中研究这类 $\hat{}\alpha^{\text{mix}}$ 后验分布的渐近性质（定理 3）。

我们注意到，数据驱动选择 $\alpha$ 的所有情形都已在先前的研究中出现。SafeBayes 的常数 $\hat{}\alpha_n$ 与 [31, 图 5] 中的结果一致，混合角点解也已在其他研究中观察到。当 $\alpha$ 被选为使期望对数预测密度（一个类似于对数点预测密度的量）达到最优时，[64] 也观察到了位于 $\alpha = \infty$ 的点质量。此外，当岭回归参数$\lambda$ 根据训练-测试划分进行选择时， [71]观察到，渐近地，所选值中约有一半是位于 $\lambda = 0$ 的角点解（即 $\alpha = \infty$），并将这一观察形式化为关于所选 $\lambda$的渐近分布的结果。在第3节中，我们提出了一套一般理论，涵盖了使用可能设定错误的似然模型构造的 $\alpha$后验分布中，数据驱动选择 $\hat{}\alpha_n$ 的所有这些情形。

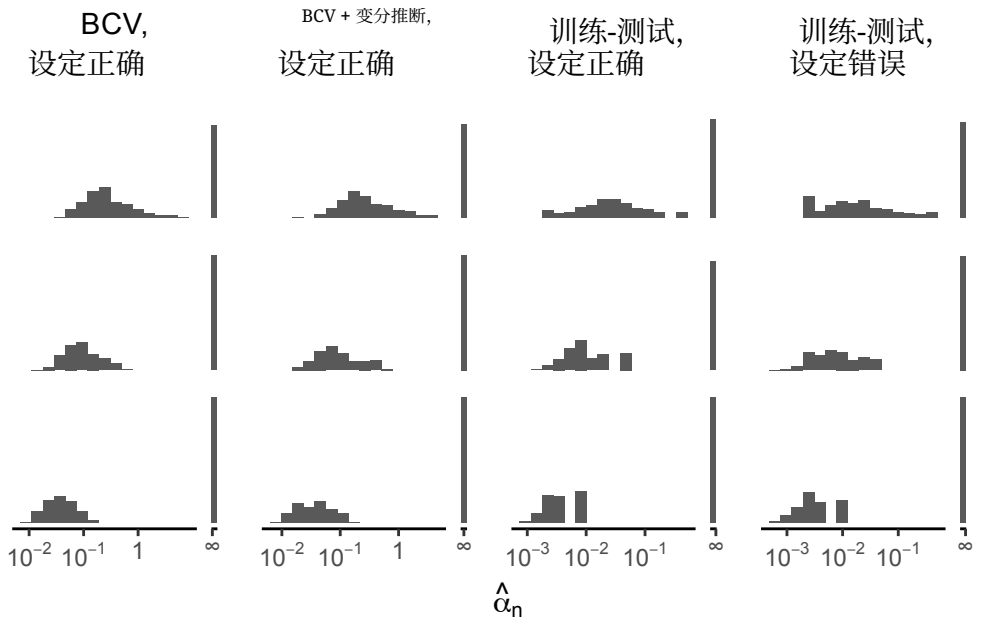



图3： 1,000 次重复试验中所选 $\hat{\alpha}_n$ 的分布。* 列标题表示调优方法（BCV、BCV+VI、训练-测试）和模型设定（设定正确、设定错误）。每列中，各行对应每次重复实验的样本量（$n = 100$、$n = 1{,}000$、$n = 5{,}000$）。 *$\hat{\alpha}_n$ 值以对数尺度绘制，以便更清晰地展示不收敛到零的非角点解 $\hat{\alpha}_n$ 值的收敛情况。有关以线性尺度绘制的 $\hat{\alpha}_n$ 直方图，请参见附录 F.4.1 中的图9。

| 方法 | 模型设定 | 指数，95%置信区间 |
|---|---|---|
| BCV | 设定正确 | -0.63, [-0.66, -0.59] |
| | 设定错误 | -1.04, [-1.05, -1.03] |
| BCV + 变分推断 | 设定正确 | -0.63, [-0.66, -0.59] |
| | 设定错误 | -1.04, [-1.05, -1.02] |
| 训练-测试 | 设定正确 | -0.63, [-0.66, -0.6] |
| | 设定错误 | -0.61, [-0.64, -0.58] |

表 2： $\hat{\alpha}_n$的随机阶估计。我们将$\hat{\alpha}_n$ （非角点解）建模为$\hat{\alpha}_n = Cn^{\gamma}$，并报告 $\gamma$ （指数）的估计值和 95% 置信区间。对应的最佳拟合曲线（图 6）以及$\gamma$的估计详情见附录 F.3。

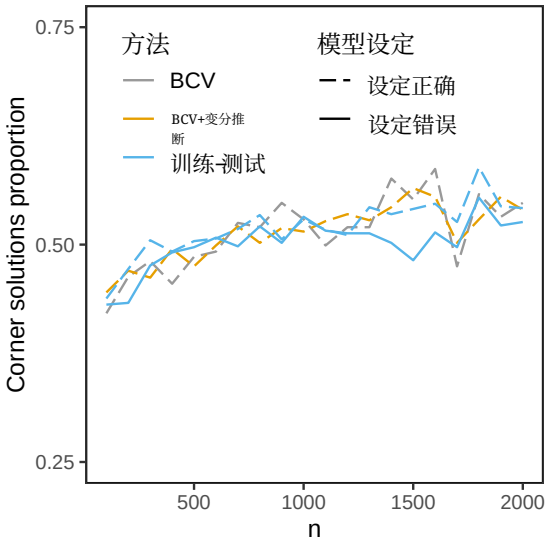



图 4： 角点解比例与样本量（$n$）的关系。

## 2.4 真实数据应用

我们通过考察美国人口普查局发布的1988年3月当前人口调查（CPS）数据，展示模拟中观察到的$\hat{}\alpha_n$ 渐近情形的实际意义。该横截面数据集可通过 R 的 AER包以 CPS1988的形式公开获取。该数据集包含28,155条观测值，记录了年龄在18至70岁之间、年收入超过50美元且既非自雇也无偿工作的男性的工资、教育程度、工作经验和族裔信息 [49]。作为基准，我们从 [49], 中设定的模型开始，并将其称为“完整模型”；具体而言，

$$\log（工资）= \beta_0 + \beta_1 教育程度 + \beta_2 族裔 + \beta_3 工作经验 + \beta_4 工作经验^2 + \epsilon. \tag{13}$$

我们还考虑另外两种从（13）中移除协变量的模型设定。第一种模型移除了族裔变量

$$\log(\text{Wage}) = \beta_0 + \beta_1 \text{Education} + \beta_2 \text{Experience} + \beta_3 \text{Experience}^2 + \epsilon, \tag{14}$$

以及另一个移除了工作经验$^2$ 项的模型

$$\log(\text{Wage}) = \beta_0 + \beta_1 \text{Education} + \beta_2 \text{Ethnicity} + \beta_3 \text{Experience} + \epsilon. \tag{15}$$

我们注意到，去掉工作经验$^2$ 项会极大地影响线性项系数的估计。虽然我们并未声称（13）是正确或真实的模型，但与（14）或（15）相比，它似乎确实是一个更好的模型，或者说更接近正确设定的模型。

首先，我们通过从完整数据集中抽取子样本，重复图1–3和表 2中的实验。这些实验的结果见附录 F.4.2中的图10。请注意，这些子抽样实验并不等同于图1–3和表 2中的实验，因为后者涉及对数据进行独立重复抽样，

而子抽样会产生依赖性。尽管如此，我们发现出现了类似的“渐近”情形（更多详情见图10）。具体而言，($i$) 在所有模型设定下，SafeBayes都表明$\hat{}\alpha_n$ 趋近于一个常数；($ii$) 在任何模型设定下，我们都未观察到“快速趋于零”的情形；($iii$) 根据 BCV $+$ VI（在完整模型和排除族裔的模型下）以及训练-测试（在所有模型设定下）所选$\hat{}\alpha_n$ 的分布，都呈现出模拟示例（表3）中讨论的混合行为。更具体地说，表3报告称，根据 BCV $+$ VI 和训练-测试所选出的非角点解，在所有模型设定下，其估计阶都比 $\frac{1}{n}$ 慢。我们还注意到，在所有模型设定下，留一交叉验证在一定程度上似乎也呈现出相同的混合行为（图10）。然而，我们发现，“非角点”$\hat{}\alpha_n$ 的值似乎保持不变或随 $n$增长，而非缩小（表3；更多详情见附录 F.3中的图 7）。最后，在所有模型设定下，BCV都会选择角点解（图10）。我们在命题 3中形式化描述由此产生的 $\alpha$-后验分布的行为，并将其称为 $\infty$-后验分布。留一交叉验证和BCV是仅有的两种选择方法：在数据示例中，$\hat{}\alpha_n$ 的分布与模拟中$\hat{}\alpha_n$ 的分布不同（图1和图3）。

在所有模型设定下，所有方法都倾向于偏好较高的 $\hat{}\alpha_n$ 值，包括$\alpha = \infty$。Safe Bayes 对 $\alpha \in [0, 1]$ 施加约束，并且在所有模型设定下，压倒性地选择使 $\alpha$ 接近 1 的值。此外，训练-测试表现出与图3中相似的混合行为，其中有一定比例的 $\hat{}\alpha_n$ 值实际上是 $\infty$。

取大小为 $n = 5{,}000$ 的子样本时，鉴于子样本量相对较大，我们可以预期，各子样本中的 $\hat{}\alpha_n$ 值会彼此“吻合”。事实上，在所有方法和模型设定中，根据 BCV、BCV+VI 和 SafeBayes 选出的 $\hat{}\alpha_n$ 值都不受子样本选择的影响。具体而言，我们观察到，BCV 总是选择 $\hat{}\alpha_n$ 等于“无穷大”，而随着 $n$ 变大，BCV+VI 赋予 $\hat{}\alpha_n$ 等于“无穷大”的概率也会增大。然而，按照训练-测试法选出的 $\hat{}\alpha_n$ 的分布，其混合比例会随着 $n$ 增大而保持稳定。按照留一交叉验证选出的 $\hat{}\alpha_n$ 则

| 方法 | 模型设定 | 指数，95%置信区间 |
|---|---|---|
| BCV + 变分推断 | 完整模型 | -0.82, [-0.95，-0.68] |
| | 无族裔 | -0.85, [-0.98, -0.71] |
| 留一交叉验证 | 完整模型 | 0.36, [0.27, 0.44] |
| | 无族裔 | 0.47, [0.3，0.64] |
| | 无工作经验[2] | 0.00, [-0.04，0.04] |
| 训练-测试 | 完整模型 | -0.66, [-0.75, -0.56] |
| | 无族裔 | -0.70, [-0.77，-0.64] |
| | 无工作经验[2] | -0.59, [-0.69，-0.5] |

表3：$\hat{\alpha}_n$ 的随机阶估计。我们将 $\hat{\alpha}_n$ （非角点解）建模为 $\hat{\alpha}_n = Cn^{\gamma}$ ，并报告 $\gamma$ （指数）的估计值和置信区间。我们还注意到，根据留一交叉验证得到的“非角点”解似乎随着 $n$ 增大而增大，而非减小。为求完整起见，我们纳入了这一结果。相应的最佳拟合曲线（图7）以及关于 $\gamma$ 估计的详细信息，请参见附录 F.3。

在完整模型下，它对所选子样本有一定敏感性，但所有选出的 $\hat{\alpha}_n$ 都相对较大（图10）。

当 $n = 5,000$ 时，我们还根据公式 (8) 计算所得 $\alpha$-后验分布的均值（这与根据公式 (12) 计算其变分近似的均值相同）。我们将 $\hat{\alpha}_n$- 后验均值与完整数据集上完整模型系数的最小二乘估计进行比较。我们将这些值称为“完整系数值”。这些实验结果见附录F.4.3 中的图11a–11c。总体而言，结果表明，后验均值估计量在不同随机子样本之间保持稳定。

# 3 渐近结果：关于$\hat{\alpha}_n$-后验分布

本节研究在第 2 节模拟与真实数据实验中观察到的 $\hat{\alpha}_n$ 的三种渐近增长区制下，$\hat{\alpha}_n$-后验分布的渐近性质。我们的第一个结果确立了在 $\frac{1}{n} \ll \hat{\alpha}_n \ll 1$ 以概率收敛的情形下的矩一致性；这属于第 2 节讨论的“混合 $\hat{\alpha}_n$”情形。我们还将讨论

$\frac{1}{n} \ll \hat{\alpha}_n$ 如何成为该结果成立的临界阈值。我们的第二个结果表明，在 $\frac{1}{\sqrt{n}} \ll \hat{\alpha}_n \ll 1$ 以概率收敛的情形下，$\hat{\alpha}_n$-后验均值估计量渐近正态。为了说明 $\frac{1}{\sqrt{n}} \ll \hat{\alpha}_n$ 是该结果成立的临界阈值，我们建立了一种新颖的拉普拉斯近似。最后一个结果则形式化地描述了“混合 $\hat{\alpha}_n$”情形下 $\hat{\alpha}_n$-后验分布的收敛；在该情形中，$\hat{\alpha}_n$ 具有混合分布，其特征是在无穷远处有一个点质量，而剩余质量消失的速度慢于 $\frac{1}{n}$。如前所述，使用贝叶斯交叉验证选择温度参数时，可能会出现这种情形。

## 3.1 BvM 对于$\hat{\alpha}_n$-后验分布及其矩的一致性

本节的主要结果确立了对任意随机序列 $\hat{\alpha}_n$，只要 $\frac{1}{n} \ll \hat{\alpha}_n \ll 1$以趋于 1 的概率成立，$\hat{\alpha}_n$-后验矩就具有一致性。更准确地说，我们证明，$\hat{\alpha}_n$-后验分布的矩（经中心化和重新缩放）与多元高斯分布的矩之间的 $L_1$ 差异依概率收敛到 0。

### 3.1.1 主要结果

我们要求模型满足以下正则性条件。除非另有说明，$\alpha_n$ 表示满足 $\frac{1}{n} \ll \alpha_n \ll 1$ 的任意正序列。

**(A0)** 最大似然估计 $\theta$ 是唯一的，并且存在一个 $\theta^* \in \Theta$，其中 $\Theta$ 是 $\mathbb{R}^p$的开子集，使得 $\theta \to \theta^*$ 在 $f_{0,n}$-概率下成立，且 $\sqrt{n}(\theta - \theta^*)$ 渐近正态。

**(A1)** 先验分布 $\pi(\theta)$ 在 $\theta^*$的一个开球内连续且为正。

**(A2)** 定义 $\Delta_{n,\theta^*} = \sqrt{n}(\theta - \theta^*)$。存在一个正定矩阵 $V_{\theta^*}$，使得对任意 $r > 0$，$\sup_{h \in \bar{B}_0(r)} |R_n(h)| \to 0$ 在 $f_{0,n}$-概率意义下成立，其中

$$R_n(h) \equiv \alpha_n \Big[ \log f_n\Big(X^n \Big| \theta^* + \frac{h}{\sqrt{\alpha_n n}}\Big) - \log f_n(X^n|\theta^*) \Big] - \sqrt{\alpha_n} h^\top V_{\theta^*} \Delta_{n,\theta^*} + \frac{1}{2} h^\top V_{\theta^*} h. \tag{16}$$

**(A3)** 对于所有 $\epsilon > 0$ 和所有 $k_0 \geq 0$，存在常数 $M \equiv M(k_0, \epsilon) < \infty$ 和整数 $N \equiv N(k_0, \epsilon)$，使得对于 $\alpha_n n \geq N$，

$$\mathbb{P}_{f_{0,n}}\Big( \int_{\mathbb{R}^p} \|\sqrt{\alpha_n n}(\theta - \theta^*)\|_2^{k_0} \pi_{n,\alpha_n}(\theta | X^n) d\theta > M \Big) < \epsilon.$$

假设 **(A0)** 和 **(A1)** 分别是关于最大似然估计和先验分布的最小正则性条件。它们是文献中的标准假设，在 [55, 50, 5, 69] 中也有采用。假设 **(A2)** 是 [50, eq. (2.1)], [5, 中标准局部渐近正态性（LAN）条件的一种变体。假设 1], 和 [69, **(A2)**]。我们指出，我们的条件是 LAN 的推广形式，通过令 $\alpha_n = 1$ 即可恢复 LAN。不过，当 $\alpha_n \to 0$ 时，用于检验 LAN 的、涉及对数似然三阶导数的常见正则性条件 [55, 18, 76] 也可用于检验我们这一 LAN 变体。更详细的讨论和正式表述见附录 C.1。假设 **(A3)** 指出，当 $\alpha_n n$ 足够大时，重缩放后的 $\alpha_n$-后验分布的所有矩以高概率有限。该条件与用于建立贝叶斯—冯·米塞斯定理的后验集中假设有关[50, 5, 69]。我们还指出，其他研究对先验分布和对数似然施加了可推出后验集中的条件 [55, 18, 74]。我们补充了关于似然和先验分布的条件；这些条件足以检验 **(A3)** 是否成立，如附录 C.2 所示。我们还需要这些条件来证明随机的 $\hat{}\alpha_n$-后验分布的矩收敛。

**条件 1**。令 $L$, $c_0$, $c_1 > 0$ 和 $c_2 \geq 0$ 为常数。则对于某个 $r > \frac{e^{c_2/c_0}}{c_1}$，存在一个函数 $\gamma$: $[0, \infty) \to [0, \infty)$，其满足性质 $\gamma(v) \geq c_0 \log(1 + c_1 v) - c_2$ 当 $v \geq r$ 时，使得在 $B_{\theta^*}(r)^c$ 上一致地，对于 $f_{0,\,n}$- 概率趋于 1，我们有

$$\log f_n(X^n|\theta) - \log f_n(X^n|\theta^*) \leq -Ln\gamma(\|\theta - \theta^*\|_2). \tag{17}$$

**条件2**。对于条件 1 中定义的 $c_0$、$c_1$、$c_2$ 和 $r$，存在正常数，

$c_3 \equiv c_3(r) > 0$ 和 $c_4 \equiv c_4(r) > 0$，使得在 $B_{\hat{\theta}}(2r)$ 上一致地有 $-nc_3\|\theta - \hat{\theta}\|_2^2 \leq \log f_n(X^n|\theta) - \log f_n(X^n|\hat{\theta}) \leq -nc_4\|\theta - \hat{\theta}\|_2^2$，且以趋于 1 的 $f_{0,n}$-概率成立。

条件 1 要求伪真参数足够"显著"，也就是说，在 $\theta$ 的邻域 $\theta^*$ 之外，取值于 $\theta$ 处的对数似然的次优程度有一个下界，该下界是 $\|\theta - \theta^*\|_2$ 的递增函数 $\gamma$。条件2假设，对数似然以趋于 1 的概率在最大似然估计的某个邻域内局部强凸且光滑。

现在可以陈述我们的第一个主要结果了。证明见附录A.1。

**定理 1.** 假设(**A0**)、(**A1**)、(**A2**)和(**A3**) 成立。那么，对于任意满足 $\frac{1}{n} \ll \alpha_n \ll 1$ 的序列 $\alpha_n$，以及任意整数 $k \geq 0$,

$$\int_{\mathbb{R}^p} \|\sqrt{\alpha_n n}(\theta - \theta^*)\|_2^k \left| \pi_{n,\alpha_n}(\theta|X^n) - \phi\Big(\theta \Big| \hat{\theta}, \frac{1}{\alpha_n n} V_{\theta^*}^{-1}\Big)\right| d\theta \to 0, \tag{18}$$

以 $_n$-概率收敛于 $f_{0,}$。此外，假设对于一个正的、依赖于数据的序列 $\hat{}\alpha_n$，存在一个满足 $\frac{1}{n} \ll \alpha_n \ll 1$ 的正序列 $\alpha_n$，使得 $\hat{} \frac{\alpha_n}{\alpha_n} \to 1$ 以 $_n$-概率收敛于 $f_{0,}$，且 $f_n(X^n|\theta)$ 满足条件1-2。此外，假设对于某个$\kappa > 0$，$\pi(\theta) \leq \kappa$。则将 (18) 中的 $\pi_{n,\alpha_n}(\theta|X^n)$ 替换为 $\pi_{n,\hat{\alpha}_n}(\theta|X^n)$ 后，结果仍然成立。

注意，定理 1 中的 $k = 0$ 立即意味着一个推广 [5,定理 1], 的伯恩斯坦-冯·米塞斯结果，我们在下文陈述该结果。

**推论 y 1.** 假设定理 1 的条件成立。那么，以$f_{0,n}$概率,

$$\mathrm{d_{TV}}\Big(\pi_{n,\hat{\alpha}_n}(\theta|X^n), \phi\Big(\theta \,\Big|\, \hat{\theta}, \frac{1}{\alpha_n n} V_{\theta^*}^{-1}\Big)\Big) \to 0. \tag{19}$$

### 3.1.2 讨论

定理 1 将 [69, 定理 1] 的矩收敛结果推广到一个趋于零的随机序列 $\hat{}\alpha_n$，而推论 1 以类似的

方式推广了 [55, 50, 5]定理 的贝叶斯-冯·米泽斯定理。当 $\hat{}\alpha_n$ 是固定常数 $\alpha$时，我们的结果可还原这些已知结果。由于我们现在假设 $\hat{}\alpha_n n \ll n$ 以高概率成立，因此，我们在极限方差方面的集中速率比标准后验分布 [55, 50, 5]已有结果中的集中速率更慢。

我们强调，条件 $\frac{1}{n} \ll \alpha_n$ 对于 (18) 成立是必要的，因此，如果 $\alpha_n$ 处于第 2 节中观察到的“快速消失”情形，那么对于 $\hat{}\alpha_n$-后验分布，贝叶斯—冯·米泽斯结果并不成立。事实上，当 $\alpha_n \asymp \frac{1}{n}$ 时，我们可以给出明确的反例。考虑统计模型

$$f_n(X^n|\eta) = \exp\left(\eta^\top T(X^n) - nA(\eta) - B(X^n)\right), \tag{20}$$

其中 $\eta \in H$属于某个自然参数空间，并且选择 $B(X^n)$，使得 (20) 的积分为一。此外，考虑先验密度

$$\pi(\eta) = \exp\left(\eta^\top \xi - \nu A(\eta) - \psi(\xi, \nu)\right), \tag{21}$$

其中选择 $\psi(\xi, \nu)$，使得 (21) 的积分为一。简单推导可得，$\alpha_n$-后验分布的密度具有如下形式

$$\pi_{n,\alpha_n}(\eta|X^n) = \exp\left(\eta^\top[\alpha_n T(X^n) + \xi] - [\alpha_n n + \nu]A(\eta) - \psi(\alpha_n T(X^n) + \xi, \alpha_n n + \nu)\right). \tag{22}$$

通过这一简单构造，我们证明，当 $\alpha_n \asymp \frac{1}{n}$ 时，推论 1 中 (19) 所示的贝叶斯—冯·米泽斯结果并不成立。我们在下一个命题中陈述这一点，该命题表明，即使是较弱的依分布收敛概念，在此情形下也不成立。

**命题1**。 假设 (20) 中的模型设定正确，真实参数值为 $\eta^*$，且 $\alpha_n$-后验分布的特征函数关于 $\xi_n$ 和 $\nu_n$ 连续。那么，如果 $\alpha_n n \to \alpha_0 \geq 0$ 和 $\frac{1}{n}T(X^n) \to g(\eta^*)$ 在 $f_{0,n}$-概率下成立，则 $\alpha_n$-后验分布 (22) 弱收敛至一个属于与 (21) 相同分布族的分布。

命题1的证明见附录D.1。对命题1的解读是：如果先验分布 (21) 不是高斯分布，那么 $\alpha_n$- 后验分布将弱收敛到一个

非高斯分布。因此，伯恩斯坦-冯·米塞斯型定理不成立。此外，当 (20) 和 (21) 为高斯分布时，尽管 (22) 及其极限分布均为高斯分布，该极限分布却与推论 1 中给出的极限高斯分布不一致。有关详情以及共轭先验的显式极限计算示例，请参见附录 D.2。

## 3.2 Asy$\hat{}\alpha_n$-后验均值估计量的渐近正态性

本节的主要结果定理 2 量化了$\hat{}\alpha_n$-后验均值与最大似然估计之间的差异，进而推出，对于任意满足 $\frac{1}{\sqrt{n}} \ll \hat{}\alpha_n \ll 1$ 以趋于 1 的概率成立的随机序列$\hat{}\alpha_n$，其$\hat{}\alpha_n$-后验均值具有渐近正态性（推论 2）。

### 3.2.1 主要结果

我们需要以下假设。除非另有说明，$\alpha_n$ 表示满足 $\frac{1}{n} \ll \alpha_n \ll 1$ 的任意正序列。

**(A1')** 先验分布 $\pi(\theta)$在 $\theta^*$的某个开球内为正且二阶连续可微。此外，先验均值有限。

**(A2')** 当 $n$ 充分大时，存在一个以高概率为正定对称的矩阵 $H_n \in \mathbb{R}^{p\times p}$ 、一个张量 $S_n \in \mathbb{R}^{p\times p\times p}$以及一个常数 $M < \infty$ ，使得对所有 $\epsilon > 0$ 及某个 $\delta > 0$,

$$\mathbb{P}_{f_{0,n}}\Big( \sup_{h\in\mathcal{H}_n} \tilde{R}_n(h) - \frac{M}{\alpha_n n}\|h\|_2^4 > 0\Big) < \epsilon, \tag{23}$$

其中 $\mathcal{H}_n \equiv \{h \in \mathbb{R}^p | \frac{\|h\|_2}{\sqrt{\alpha_n n}} \le \delta\}$ 且

$$\tilde{R}_n(h) \equiv \alpha_n\Big[\log f_n\Big(X^n\Big|\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\Big) - \log f_n(X^n|\hat{\theta})\Big] + \frac{1}{2}h^\top H_n h - \frac{1}{6\sqrt{\alpha_n n}}\langle S_n, h^{\otimes 3}\rangle.$$

**(A3')** 对所有 $\epsilon > 0$ 及所有 $\delta' > 0$，均存在一个常数 $c \equiv c(\epsilon, \delta') > 0$ ，使得对于 $n$

充分大时，

$$\mathbb{P}_{f_{0,n}}\Big(\sup_{\theta\in B_{\hat{\theta}}(\delta')^{\mathrm{c}}}\frac{1}{n}\log f_n(X^n|\theta)-\frac{1}{n}\log f_n(X^n|\hat{\theta})>-c\Big)<\epsilon. \tag{24}$$

假设(**A1'**) 在假设(**A1**)的基础上，对先验分布增加了额外的可微性要求。假设(**A2'**) 可以视为一项高阶 LAN 条件。(**A2**) 中的标准 LAN 条件，是以伪真参数为中心对数似然比作二次近似。假设(**A2'**) 则是以最大似然估计为中心对数似然比作三次近似。我们在附录 C.3 中给出了验证(**A2'**) 的充分条件。如下文第 3.2.2 节和第 3.2.3 节所述，假设(**A1'**) 和 (**A2'**) 都是支撑主要结果的拉普拉斯近似所必需的。这类光滑性假设在拉普拉斯近似论证中十分常见[44, 72, 47, 21, 22]。假设(**A3'**) 是关于最大似然估计显著性的条件，与用于证明 M 估计量一致性的标准正则性条件有关[76]。 [55] （第 489 页，假设 (B3)）中也提出了类似条件，用以在经典条件下证明标准后验分布的贝叶斯-冯·米泽斯定理。该条件还被用于证明后验集中（见 [74, 条件 1，式 (8)], [64, 假设 S.1], 和[65, 假设 (2)]）；这进而意味着后验矩依概率有界，正如假设(**A3**)所要求的。更多详情请参见附录 C.2。

现在我们可以陈述本节的主要结果。证明见附录 A.2。**定理 2**。 假设 (**A0**)、(**A1'**)、(**A2'**) 和 (**A3'**) 成立。记 $\alpha_n$-后验均值估计量为 $\hat{}\,\theta^{\mathrm{B}}=[\hat{}\theta_1^{\mathrm{B}},\ \ldots\ldots,\ \hat{}\theta_p^{\mathrm{B}}]^\top$，其中 $\hat{}\,\theta_j^{\mathrm{B}}=\int_{\mathbb{R}^p}\theta_j\pi_{n,\ \alpha_n}(\theta|X^n)d\theta$ 对 $j=1,\ \ldots\ldots,\ p$成立。于是，

$$\hat{\theta}^{\mathrm{B}}-\hat{\theta}=O_{f_{0,n}}\Big(\frac{1}{\alpha_n n}\Big). \tag{25}$$

此外，假设对于一个依赖于数据的正序列 $\hat{}\alpha_n$，存在一个正序列 $\alpha_n\to 0$，满足 $\frac{1}{n}\ll\alpha_n\ll 1$，使得 $\hat{}\underline{\alpha}_n/\alpha_n\to 1$ 属于 $f_{0,\ n}$-概率。于是，

式 (25) 中的结果仍然成立，只需将 $\pi_{n,\ \alpha_n}(\theta|X^n)$ 替换为 $\pi_{n,\ \hat{\alpha}_n}(\theta|X^n)$。

定理 2 的一个直接推论是，当 $\hat{\alpha}_n$ 的衰减足够缓慢时，$\hat{\alpha}_n$-后验均值具有渐近正态性（推论 2）。事实上，定理 2 的结果告诉我们，$\sqrt{n}(\hat{\theta}^{\mathrm{B}} - \theta^*) = \sqrt{n}(\hat{\theta}^{\mathrm{B}} - \hat{\theta}) + \sqrt{n}(\hat{\theta} - \theta^*) = O_{f_{0,\ n}}(\frac{1}{\alpha_n\sqrt{n}}) + \sqrt{n}(\hat{\theta} - \theta^*)$。如果 $\frac{1}{\sqrt{n}} \ll \alpha_n \ll 1$，那么第一项为 $o_{f_{0,\ n}}(1)$，而假设 (**A0**) 保证主导项渐近正态。

**推论 2**。假设定理 2 的条件成立。那么，对于任意满足 $\frac{1}{\sqrt{n}} \ll \alpha_n \ll 1$ 的序列 $\alpha_n$，$\alpha_n$-后验均值估计量与 $\hat{\theta}$ 渐近等价。也就是说，$\sqrt{n}(\hat{\theta}^{\mathrm{B}} - \theta^*) \xrightarrow{d} N(0,\ \tilde{V}_\theta)$，其中 $\tilde{V}_\theta$ 是最大似然估计的渐近方差。此外，这一结果对于定理 2 中所述的 $\hat{\alpha}_n$-后验均值估计量仍然成立。

**备注 1.** 如果 $\log f_n(X^n|\theta^*)$ 在 $\theta^*$ 处可微，则 $\tilde{V}_{\theta^*} = V_{\theta^*}^{-1} M_{\theta^*} V_{\theta^*}^{-1}$，其中 $V_{\theta^*}$ 的定义见 (**A2**)，$M_{\theta^*}$ 是 $f_{0,n}$-概率意义下，当 $n \to \infty$ 时 $\frac{1}{n}\dot{\ell}_{n,\theta^*}\dot{\ell}_{n,\theta^*}^T$ 和 $\dot{\ell}_{n,\theta^*} = \nabla_\theta \log f_n(X^n|\theta)|_{\theta=\theta^*}$ 是模型 $f_n(X^n|\theta)$ 在 $\theta^*$ 处的得分函数。

### 3.2.2 讨论

推论 1 表明，当 $\frac{1}{n} \ll \hat{\alpha}_n \ll 1$ 以趋于 1 的概率成立时，$\hat{\alpha}_n$-后验分布与以最大似然估计为中心的正态分布之间的全变差距离在 $f_{0,n}$-概率意义下趋于零。这表明在此情形下，后验均值可能渐近等价于最大似然估计。然而，只有在更受限的情形下，即 $\frac{1}{\sqrt{n}} \ll \hat{\alpha}_n \ll 1$ 以趋于 1 的概率成立时，才会如此；这恰好是推论 2 所述的结果。

要看出推论 2 是紧的，只需考虑第 2 节中温度参数 $\alpha_n$ 为确定值的贝叶斯线性回归示例。在该示例中，$\alpha_n$-后验均值就是岭回归估计量；当 $\frac{1}{n} \ll \alpha_n \ll 1$ 时它是一致的，但只有在 $\frac{1}{\sqrt{n}} \ll \alpha_n \ll 1$ 时才等价于最大似然估计。我们还在一个无法解析求解的示例中通过数值方式说明了这一事实。我们考虑一个设定正确的贝叶斯

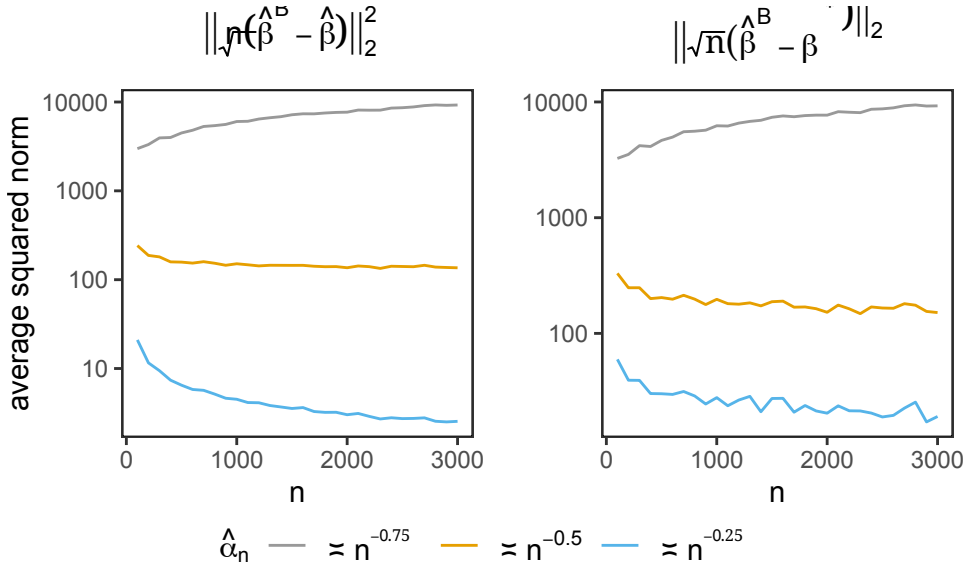



图 5： $\sqrt{n}(\hat{\beta}^{\mathrm{B}}-\hat{\beta})$ 的平方范数（左）和 $\sqrt{n}(\hat{\beta}^{\mathrm{B}}-\beta^*)$（右）在 100 次重复试验中取均值，并表示为样本量的函数。 $\alpha_n$-后验均值由从 $\alpha_n$-后验中抽取的样本获得，其中 $\alpha_n = 0.5n^{-3/4}$ （灰色）、 $\alpha_n = 0.5n^{-1/2}$ （橙色）和$\alpha_n = 0.5n^{-1/4}$ （蓝色）。数据由独立同分布模型 $\mathbb{P}(Y_i = 1|x_i) = [1+\exp(-x_i^\top\beta^*)]^{-1}$ 生成，其中 $\{x_i\} \overset{iid}{\sim} \mathcal{N}(0, I_3)$ 和 $\beta^* = [1, -0.5, 0.1]$。

逻辑回归模型，其回归系数服从均值为零、标准差为 100 的高斯先验。我们在 $\alpha_n$ 的三种设定下从 $\alpha_n$-后验中进行采样：$\alpha_n \asymp n^{-3/4}$、 $\alpha_n \asymp n^{-1/2}$ 和 $\alpha_n \asymp n^{-1/4}$。我们注意到，$\alpha_n \asymp n^{-1/4}$ 与定理 2 所规定的情形相容，而 $\alpha_n \asymp n^{-3/4}$ 则处于该情形之外，其中 $\alpha_n \asymp n^{-1/2}$ 位于边界上。图 5 左图显示，当 $\alpha_n \asymp n^{-3/4}$ 时，后验均值与最大似然估计之间经 $\sqrt{n}$ 缩放的差异发散；当 $\alpha_n \asymp n^{-1/4}$ 时，该差异消失；当 $\alpha_n \asymp n^{-1/2}$ 时，该差异为常数量级。图 5 右图也反映了这些行为，其中展示了 $\alpha_n$-后验均值与真实参数值之间经 $\sqrt{n}$ 重缩放的差异。

通常，给定像定理 1 这样的矩收敛结果后，类似定理 2 的结果只需稍微调整证明技巧即可轻松得到（参见 [55,第 6 章，定理 8.2-8.3] ，其中给出了标准后验分布的一个例子）。然而，当 $\alpha_n \to 0$ 时，这种方法便不再适用，因为它只能得到 $\sqrt{n}(\hat{\theta}^{\mathrm{B}} - \hat{\theta}) = o_{f_{0,n}}(\frac{1}{\sqrt{\alpha_n}})$，这太

不够宽松。事实上，利用定理 1 并令 $k=1$，很容易看出

$$\sqrt{\alpha_n n}\|\hat{\theta}^B-\hat{\theta}\|_1=\sqrt{\alpha_n n}\Big\|\int(\theta-\theta^*)\pi_{n,\alpha_n}(\theta|X^n)d\theta-\int(\theta-\theta^*)\phi\Big(\theta\Big|\hat{\theta},\frac{1}{\alpha_n n}V_{\theta^*}^{-1}\Big)d\theta\Big\|_1$$
$$\leq\int\|\sqrt{\alpha_n n}(\theta-\theta^*)\|_1\Big|\pi_{n,\alpha_n}(\theta|X^n)-\phi\Big(\theta|\hat{\theta},\frac{1}{\alpha_n n}V_{\theta^*}^{-1}\Big)\Big|d\theta=o_{f_{0,n}}(1).$$

相比之下，推论 2 中的相应结果是 $\sqrt{n}(\hat{\theta}^{\mathrm{B}}-\hat{\theta})=O_{f_{0,n}}(\frac{1}{\alpha_n\sqrt{n}})$，如果 $\alpha_n\sqrt{n}\to\infty$（即 $\frac{1}{\sqrt{n}}\ll\alpha_n$），则该结果为 $o_{f_{0,n}}(1)$。

定理 2 证明中的关键在于一种新的拉普拉斯近似。这是一种自然的方法，因为后验均值估计量的各个坐标定义为

$$\hat{\theta}_j^{\mathrm{B}}=\frac{\int_{\mathbb{R}^p}\theta_j\pi(\theta)\exp\left(\alpha_n\log f_n(X^n|\theta)\right)d\theta}{\int_{\mathbb{R}^p}\pi(\theta)\exp\left(\alpha_n\log f_n(X^n|\theta)\right)d\theta},\quad\text{for } j=1,\ldots,p,\tag{26}$$

对于某个函数 $b:\mathbb{R}^p\to\mathbb{R}$，它是两个如下形式的积分之比

$$\int_{\mathbb{R}^p}b(\theta)\exp\Big(\alpha_n\log f_n(X^n|\theta)\Big)d\theta.\tag{27}$$

具体而言，(26) 在分子中使用 $b(\theta)=\theta_j\pi(\theta)$，在分母中使用 $b(\theta)=\pi(\theta)$。拉普拉斯近似是针对 (27) 这类拉普拉斯型积分的渐近展开。在本文中，我们所说的“拉普拉斯近似”是指对拉普拉斯型积分[58, 57, 43, 44] 的近似，而不是伯恩斯坦-冯·米塞斯 [45] 或对后验分布归一化常数的近似 [8, 65]，，尽管这些都是相互关联的概念。

### 3.2.3 拉普拉斯近似

我们针对引理 1 中形如（27）的积分给出一般的拉普拉斯近似。为此，我们要求函数 $b$ 满足以下正则性条件。

**（ALap）**对于函数 $b:\ \mathbb{R}^p\to\mathbb{R}$，假设满足以下条件。**1.** $b$ 在 $\mathbb{R}^p$ 上可积。

**2.** $b(\theta)$ 在 $\theta^*$ 周围的一个开球内是有限的。**3.** 存在一个向量 $v\in\mathbb{R}^p$ 和一个

常数 $M < \infty$，使得对于所有 $\epsilon > 0$ 以及某个 $\delta > 0$,

$$\mathbb{P}_{f_{0,n}}\Big( \sup_{h \in \mathcal{H}_{n,b}} R_b(h) - \frac{M}{\alpha_n n}\|h\|_2^2 > 0 \Big) < \epsilon, \tag{28}$$

其中 $\mathcal{H}_{n,b} \equiv \{h \in \mathbb{R}^p | \frac{\|h\|_2}{\sqrt{\alpha_n n}} \leq \delta\}$ 且 $R_b(h) \equiv [b(\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}) - b(\hat{\theta})] - \frac{v^\top h}{\sqrt{\alpha_n n}}$.

**引理 1.** 假设 **(A0)**、**(A1')**、**(A2')** 和 **(A3')** 均成立。给定函数$q: \mathbb{R}^p \to \mathbb{R}$，定义 $b(\theta) \equiv q(\theta)\pi(\theta)$，其中假定此处选择的 $b$ 满足 **(ALap)**。对于任意序列 $\alpha_n n \to \infty$，当 $n$ 充分大时，以下结论成立:

1. 对任意 $\alpha_n > 0$，令 $H_n$ 如假设 **(A2')** 中所定义，则有以下展开式

$$\int_{\mathbb{R}^p} b(\theta) e^{\alpha_n \log f_n(X^n|\theta)} d\theta = e^{\alpha_n \log f_n(X^n|\hat{\theta})} |H_n|^{-\frac{1}{2}} \Big(\frac{2\pi}{\alpha_n n}\Big)^{\frac{p}{2}} \Big( b(\hat{\theta}) + O_{f_{0,n}}\Big(\frac{1}{\alpha_n n}\Big)\Big). \tag{29}$$

2. 关于 $\alpha_n$-后验分布的 $q(\theta)$ 的期望满足以下近似关系

$$\int_{\mathbb{R}^p} q(\theta)\pi_{n,\alpha_n}(\theta|X^n) d\theta = q(\hat{\theta}) + O_{f_{0,n}}\Big(\frac{1}{\alpha_n n}\Big). \tag{30}$$

此外，假设对于一个依赖于数据的正序列 $\hat{\alpha}_n$，存在一个正序列 $\alpha_n$，使得 $\alpha_n n \to \infty$ 满足 $\frac{\hat{\alpha}_n}{\alpha_n} \to 1$ 以 $f_{0,n}$-概率成立。那么，(30) 中的结果仍成立，只需将 $\pi_{n,\alpha_n}(\theta|X^n)$ 替换为$\pi_{n,\hat{\alpha}_n}(\theta|X^n)$。

引理 1 的证明见附录 A.3。该证明主要沿用了 [44, 定理 1] 中的论证，同时考虑了温度参数 $\alpha_n$，并采用了较少的可微性假设，这些假设已足以满足我们的目的。
[22, 21, 48] 中给出了拉普拉斯型积分的展开，具体针对关于贝叶斯后验的积分推导，见 [58, 57, 43, 72, 47]。值得注意的是，尽管 **(ALap)**看起来与 $b$ 的可微性假设不同，而后者在 [44], 中提出，但这两个假设都可以通过命题 2 中的正则性条件加以验证。

最后，引理 1 中的条件与 [47], 中的条件也大致相似：我们要求先验分布和对数似然具有足够阶数的光滑性，且导数有界（命题 2 和命题 6）；同时还要控制它们在 $\theta$ 的某个邻域 ˆ之外的增长，以确保 (29) 左侧的可积性（这隐含地对应于假设(**A1'**) 和假设 (**A3'**)）。不过，为了在该研究所考虑的高维情形下获得紧致近似， [47] 中的条件需要更仔细地考虑维数依赖性以及导数的增长速率。

我们以以下命题结束本节（证明见附录 C.4）。该命题给出了关于 $q$ 的易于检验的充分条件，并且这些条件蕴含假设（**ALap**）中的条件 3。事实上，为了在定理 2 的证明中应用引理 1 的结果，我们需要验证，对于形如 $b(\theta) = q(\theta)\pi(\theta)$ 的函数，假设（**ALap**）成立。

命题 2. 假设 (A0) 和 (A1') 成立。设 $b : \mathbb{R}^p \to \mathbb{R}$ 具有如下形式： $b(\theta) = q(\theta)\pi(\theta)$，其中 $q : \mathbb{R}^p \to \mathbb{R}$ ；并且 $q$ 在 $B_{\theta^*}(\gamma)$ 上二阶连续可微，其中 ˆ $\gamma > 0$。那么，对于足够大的 $\alpha_n n$ ，假设（ALap）中的条件 3 成立，其中 $\delta = \gamma/2$, $v = \nabla b(\theta) = \pi(\hat{\theta})\nabla q(\hat{\theta}) + q(\hat{\theta})\nabla\pi(\hat{\theta})$，且 $M = p\sup_{\theta\in B_{\theta^*}(\gamma)} \max_{1\le i,j\le p} |[\nabla^2 b(\theta)]_{i,j}|$ 。

## 3.3 $\hat{\alpha}_n^{\text{mix}}$后验分布的伯恩斯坦-冯·米塞斯型定理

第 2 节的实证研究中有一项观察结果：在某些情形下，以数据驱动方式选取的 $\alpha$ 的极限分布将是混合分布，其特征是包含一个无穷远处的点质量，而剩余概率质量收敛至零的速度比 $1/n$ 慢（见图3、表 2 和附录 F.3 中的图 6）。本节旨在形式化由此得到的后验分布的极限行为，我们将其称为 $\hat{\alpha}_n^{\text{mix}}$-后验分布。我们感兴趣的对象是 $\pi_{n,\,\hat{\alpha}_n^{\text{mix}}}(\theta|X^n)$，其中令$q_n \equiv q + o_{f_{0,\,n}}(1)$，对某个 $q \in [0,1]$，记作

$$\pi_{n,\hat{\alpha}_n^{\text{mix}}}(\theta|X^n) = q_n\pi_{n,\hat{\alpha}_n}(\theta|X^n) + (1-q_n)\pi_{n,\hat{\gamma}_n}(\theta|X^n), \tag{31}$$

其中，对所有 $\epsilon > 0$、所有 $B > 0$，以及一个序列 $\frac{1}{n} \ll \alpha_n \ll 1$，

$$\mathbb{P}_{f_{0,n}}\left(1 - \epsilon < \frac{\hat{\alpha}_n}{\alpha_n} < 1 + \epsilon\right) \to 1 \quad \text{and} \quad \mathbb{P}_{f_{0,n}}\left(\hat{\gamma}_n > B\right) \to 1. \tag{32}$$

本质上，（31）研究的是混合后验分布 $\pi_{n,\ \hat{\alpha}_n^{\mathrm{mix}}}(\theta|X^n)$，其中温度参数 $\alpha$ 以概率 $1 - q_n$ 为无穷远处的点质量，以概率 $q_n$缓慢收缩。在下文中，我们将考察混合后验分布 $\pi_{n,\ \hat{\alpha}_n^{\mathrm{mix}}}(\theta|X^n)$的极限分布。

我们首先需要建立一个关键的中间结果，以形式化刻画 $\infty$-后验分布的行为，其定义为 $\pi_{n,\ \infty}(\theta|X^n) \equiv \lim_{n\to\infty} \pi_{n,\ \hat{\gamma}_n}(\theta|X^n)$，其中 $\hat{}\gamma_n$ 依概率发散。由 [5, 定理 1] 给出的回火后验的极限分布表明，当温度参数趋于无穷大时，所得的 $\infty$-后验分布应收敛到位于 $\theta$的一个 $\hat{}$点质量，我们将其记为 $\delta_{\hat{\theta}}$。文献中曾非正式地提出过类似结果，例如 [64], ，或者针对吉布斯后验以弱收敛的形式证明过此类结果 [39]。这表明，（31）中的混合后验分布 $\pi_{n,\ \hat{\alpha}_n^{\mathrm{mix}}}(\theta|X^n)$ 的极限分布也将是一个混合分布：以概率 $1 - q$ 在最大似然估计处有一个点质量，以概率 $q$为一个极限高斯分布（由第 3.1 节的伯恩斯坦-冯·米塞斯结果刻画；见推论 1）。下文的定理 3 将对此结果作出形式化表述。

不同于前几节中的收敛结果，如果对于每个 $n$，$\pi_{n,\hat{\gamma}_n}(\theta|X^n)$ 相对于勒贝格测度是绝对连续的，则无法实现全变差收敛。事实上，在这种情况下，对于每个固定的 $n$， $d_{\mathrm{TV}}(\pi_{n,\ \hat{\gamma}_n}(\theta|X^n)\ \delta_{\hat{\theta}}) = 1$。尽管 $\infty$-后验分布到 $\delta_{\hat{\theta}}$ 的收敛不以全变差距离成立，但它在 $p$-Wasserstein 距离下成立。我们在下面的结果中将这一点形式化。

**命题 3.** 假设(**A0**)、(**A1**)和(**A3'**) 成立。此外，假设 $(i)$ $f_n(X^n|\theta)$ 在 $\hat{}\theta$ 处连续，且 $(ii)$ $\int \|\theta\|^p \pi(\theta) d\theta < \infty$。那么，对于任意 $t > 0$ 以及任意确定性的

序列 $\gamma_n \to \infty$，当 $n \to \infty$ 时，以 $f_{0,n}$-概率，

$$\mathrm{d}_{\mathrm{W}_p}\left(\pi_{n,\gamma_n}, \delta_{\hat{\theta}}\right) \to 0. \tag{33}$$

此外，若随机序列 $\hat{}\gamma_n$ 是增长的（即满足（32）），则将 $\gamma_n$ 替换为 $\hat{}\gamma_n$ 后，（33）中的结果仍然成立。

命题 3 的证明见附录A.4。现在陈述关于 $\hat{}\alpha_n^{\mathrm{mix}}$-后验分布的极限分布的主要结果，证明见附录A.5。

**定理 3.** 考虑（31）和（32）中所描述的混合后验分布 $\pi_{n,\hat{\alpha}_n^{\mathrm{mix}}}(\theta|X^n)$。假设定理 1 的条件成立（即，将 $\pi_{n,\alpha_n}(\theta|X^n)$ 替换为 $\pi_{n,\hat{\alpha}_n}(\theta|X^n)$ 后，（18）中的结果仍成立），且命题 3 的条件成立。进一步假设，对于某个 $m > p$，有 $\int \|\theta\|^m \pi(\theta) d\theta < \infty$。那么，令 $\phi_n := \phi(\theta|\hat{}\theta, \frac{1}{\alpha_n n} V_{\theta^*}^{-1})$，我们有，在 $f_{0,n}$-概率意义下

—

$$\mathrm{d}_{\mathrm{W}_p}\left(\pi_{n,\hat{\alpha}_n^{\mathrm{mix}}},\, q\phi_n + (1-q)\delta_{\hat{\theta}}\right) \to 0. \tag{34}$$

# 4 讨论

后验温度调整的效用取决于是否能够最优地调节温度参数 $\alpha$。尽管已有研究提出了根据数据调节 $\hat{}\alpha_n$ 的实用建议，但对于由此得到的温度化 $\hat{}\alpha_n$-后验分布，目前尚无统计保证可指导推断。在本文中，我们证明了：（i）$\hat{}\alpha_n$-后验矩的一致性（定理 1）；（ii）$\hat{}\alpha_n$-后验均值的渐近正态性（定理 2/推论 2）；以及（iii）混合 $\hat{}\alpha_n$-后验的伯恩斯坦-冯·米塞斯型定理（定理 3），这些结果适用于此前对温度化后验的理论分析尚未考虑过的 $\hat{}\alpha_n$ 渐近情形。

这些新颖情形——趋于零的随机序列、发散随机序列和混合分布——通过我们在模拟数据和真实世界数据上开展的数值实验得到了验证。通过在这些情形下建立保证，我们弥合了理论与实践之间的差距

。我们的结果表明，$\hat{}\alpha_n$-后验分布的表现与具有固定非随机 $\alpha$的$\alpha$-后验分布不同。对于任意固定的 $\alpha > 0$， $\alpha$-后验分布渐近正态 [5]。相比之下，经验中出现的某些$\hat{}\alpha_n$ 情形下，$\hat{}\alpha_n$-后验分布并非渐近正态（例如，定理 3、命题1）。我们在这些情形下的结果揭示了可以基于此类$\hat{}\alpha_n$-后验分布进行何种推断。

虽然我们的渐近保证适用于温度化后验，但我们认为，我们的证明技巧可用于将结果扩展到现代后验构造，例如广义后验，或对 $\hat{}\alpha_n$-后验的变分近似。后一种扩展尤其重要，因为变分推断所近似的 $\hat{}\alpha_n$-后验无法以闭式形式计算。

我们的研究为若干有趣的未来方向奠定了基础。我们针对贝叶斯线性回归模型开展的实证研究发现，贝叶斯交叉验证会选择一个服从混合分布的温度参数；[64] 在正态位置模型中也报告了这一发现。一条研究路径是从理论上理解这种混合分布为何会出现，这或许能揭示其他交叉验证不稳定的统计模型。另一个方向是将我们关于 $\hat{}\alpha_n$-后验均值的结果推广到风险函数最小化的点估计量（例如，$\hat{}\alpha_n$-后验中位数）。我们在引理 1 中提出的拉普拉斯近似可用于此目的。最后，我们的结果适用于固定维度的参数模型。值得进一步研究的是，在参数维度随样本量增加的高维参数设置中，或在半参数情形下，$\hat{}\alpha_n$-后验的渐近行为。

# 参考文献

[1] David M. Allen. “将预测均方误差作为变量选择的准则”。载于：*Technometrics* 13.3（1971），第469–475页。[2] Pierre Alquier 和 James Ridgway。“温度化后验及其变分近似的集中性”。载于：*Annals of Statistics* 48.3（2020），第1475–1497页。

[3] Matias Altamirano、Francois-Xavier Briol 和 Jeremias Knoblauch。“稳健且共轭的高斯过程回归”。载于：第*41*届国际机器学习会议论文集。2024年，第1155–1185页。[4] Matias Altamirano、Fran¸cois-Xavier Briol 和 Jeremias Knoblauch。“稳健且可扩展的贝叶斯在线变点检测”。载于：国际机器学习会议（*ICML*）。2023年，第642–663页。[5] Marco Avella Medina、Jos´e Luis Montiel Olea、Cynthia Rush 和 Amilcar Velez。“关于 $\alpha$-后验及其变分近似对模型设定错误的稳健性”。载于：机器学习研究期刊 23.147（2022），第1–51页。[6] Yannick Baraud。“从稳健检验到类贝叶斯后验分布”。载于：概率论及相关领域188.1（2024），第159–234页。[7] Andrew R. Barron 和 Thomas M. Cover。“最小复杂度密度估计”。载于：*IEEE*信息论汇刊 37.4（1991），第1034–1054页。[8] Anirban Bhattacharya 和 Debdeep Pati。“模型错设下的非渐近拉普拉斯近似”。载于：*arXiv*预印本 *arXiv:2005.07844*（2020）。[9] Anirban Bhattacharya、Debdeep Pati 和 Yun Yang。“贝叶斯分数后验分布”。载于：*Annals of Statistics* 47.1（2019），第39–66页。[10] Pier G. Bissiri、Chris C. Holmes 和 Stephen G. Walker。“信念分布更新的一般框架”。载于：皇家统计学会期刊*B*辑（统计方法学）78.5（2016），第1103–1130页。[11] DavidM. Blei、AlpKucukelbir 和 Jon D. McAuliffe。“变分推断：统计学家综述”。载于：美国统计学会会刊 112.518（2017），第859–877页。[12] Dominique Bontemps。“回归量数量递增时高斯回归的伯恩斯坦–冯·米塞斯定理”。载于：*Annals of Statistics* 39.5（2011），第2557–2584页。[13] St´ephane Boucheron、G´abor Lugosi 和 Pascal Massart。集中不等式：独立性的非渐近理论。牛津大学出版社，2013年。[14] Yuheng Bu。“迈向吉布斯算法的最优逆温度”。载于：*2024*年*IEEE*国际信息论研讨会（*ISIT*）。2024年，第2257–2262页。[15] Christopher P. Burgess、Irina Higgins、Arka Pal、Loic Matthey、Nick Watters、Guillaume Desjardins 和 Alexander Lerchner。“理解 $\beta$-VAE中的解耦”。载于：*2017*年*NIPS*解耦表示学习研讨会。2017年，第1–11页。[16] Luiz Max Carvalho 和 Joseph G. Ibrahim。“关于归一化幂先验”。载于：医学统计 40.24（2021），第5251–5275页。[17] Isma¨el Castillo 和 Judith Rousseau。“半参数模型中光滑泛函的伯恩斯坦-冯·米塞斯定理”。载于：*Annals of Statistics* 43.6（2015），第2353–2383页。[18] Victor Chernozhukov 和 Han Hong。“一种经典估计的马尔可夫链蒙特卡罗方法”。载于：计量经济学杂志 115.2（2003），第293–346页。

[19] Sinho Chewi、Jonathan Niles-Weed 和 Philippe Rigollet。《统计最优传输》。Springer，2025。[20] A. Dalalyan 和 Alexandre B. Tsybakov。“指数加权聚合、尖锐 PAC-贝叶斯界与稀疏性”。载于：《Machine Learning》72.1（2008），第39–61页。[21] Nicolaas Govert De Bruijn。《分析中的渐近方法》。Courier Corporation，2014。[22] Arthur Erd´elyi。《渐近展开》。Courier Corporation，1956。[23] Gerald B Folland。《实分析：现代技术及其应用》。John Wiley & Sons，1999。[24] Andrew Gelman、John B. Carlin、Hal S. Stern、Aki Vehtari、David B. Dunson 和 Donald B. Rubin。《贝叶斯数据分析》。Chapman and Hall/CRC，2013。[25] Andrew Gelman、Jessica Hwang 和 Aki Vehtari。“理解贝叶斯模型的预测信息准则”。载于：《Statistics and Computing》24.6（2014），第997–1016页。[26] Subhashis Ghosal。“高维线性模型中后验分布的渐近正态性”。载于：Bernoulli 5.2（1999），第315–331页。[27] Subhashis Ghosal 和 Aad van der Vaart。《非参数贝叶斯推断基础》。剑桥大学出版社，2017。[28] Abhik Ghosh 和 Ayanendranath Basu。“利用密度幂散度进行稳健贝叶斯估计”。载于：Annals of the Institute of Statistical Mathematics 68.2（2016），第413–437页。[29] Peter Gr¨unwald。“安全学习：通过经验凸性弥合贝叶斯、最小描述长度与统计学习理论之间的差距”。载于：第24届学习理论年度会议论文集。2011，第397–420页。[30] Peter Gr¨unwald。“安全贝叶斯”。载于：算法学习理论。2012，第169–183页。[31] Peter Gr¨unwald 和 Thijs van Ommen。“贝叶斯推断在模型设定错误的线性模型中的不一致性及其修正方案”。载于：Bayesian Analysis 12.4（2017），第1069–1103页。[32] Ethan Harvey、Mikhail Petrov 和 Michael C Hughes。“通过数据强调变分目标学习超参数”。载于：arXiv预印本 arXiv:2502.01861（2025）。[33] Rianne de Heide。SafeBayes：广义且安全的贝叶斯岭回归与套索回归。R 软件包版本1.1。2016。URL: https://CRAN.R-project.org/package=SafeBayes。[34] Irina Higgins、Loic Matthey、Arka Pal、Christopher Burgess、Xavier Glorot、Matthew Botvinick、Shakir Mohamed 和 Alexander Lerchner。“$\beta$-VAE：通过受约束变分框架学习基本视觉概念”。载于：国际学习表征会议。2017，第1–13页。[35] Arthur E Hoerl 和 Robert W Kennard。“岭回归：针对非正交问题的有偏估计”。载于：Technometrics 12.1（1970），第55–67页。

[36] Chris C. Holmes 和 Stephen G. Walker。“为一般贝叶斯模型中的幂似然赋值”。载于：*Biometrika* 104.2（2017），第497–503页。[37] Giles Hooker 和 Anand N. Vidyashankar。“通过差异度实现贝叶斯模型稳健性”。载于：*Test* 23.3（2014），第556–584页。[38] Chin-Wei Huang、Shawn Tan、Alexandre Lacoste 和 Aaron C. Courville。“利用退火变分目标提升变分推断的可探索性”。载于：神经信息处理系统进展。2018，第9724–9734页。[39] Chii-Ruey Hwang。“重新审视拉普拉斯方法：概率测度的弱收敛”。载于：概率年刊 8.6（1980），第1177–1182页。[40] Joseph G. Ibrahim 和 Ming-Hui Chen。“回归模型的幂先验分布”。载于：《统计科学》 15.1（2000），第46–60页。[41] Joseph G. Ibrahim、Ming-Hui Chen、Yeongjin Gwon 和 Fang Chen。“幂先验：理论与应用”。载于：医学统计 34.28（2015），第3724–3749页。[42] Michael I. Jordan、Zoubin Ghahramani、Tommi S. Jaakkola 和 Lawrence K. Saul。“图模型变分方法导论”。载于：*Machine Learning*37.2（1999），第183–233页。[43] Robert E. Kass 和 Duane Steffey。“条件独立层级模型中的近似贝叶斯推断（参数经验贝叶斯模型）”。载于：美国统计学会会刊 84.407（1989），第717–726页。[44] Robert E. Kass、Luke. Tierney 和 Joseph B. Kadane。“基于拉普拉斯方法的后验展开的有效性”。载于：《献给 *George Bernard* 的论文集》，编者：*S. Geisser*、*J.S. Hodges*、*S.J. Press* 和 *A. Zellner*。1990，第473–488页。[45] Anya Katsevich。“通过新的拉普拉斯近似界改善伯恩斯坦–冯·米塞斯定理中的维数依赖性”。载于：信息与推断：*IMA*期刊 14.3（2025），iaaf020。[46] Anya Katsevich。“高维中的拉普拉斯近似精度：精细分析与新的偏斜调整”。载于：*arXiv*预印本 *arXiv:2306.07262* （2024）。[47] Anya Katsevich。“高维中的拉普拉斯渐近展开”。载于：*arXiv*预印本 *arXiv:2406.12706* （2025）。[48] William D Kirwin。“拉普拉斯近似的高阶渐近”。载于：渐近分析 70.3-4（2010），第231–248页。[49] Christian Kleiber 和 Achim Zeileis。*R*应用计量经济学。Springer，2008。[50] Bas J. K. Kleijn 和 Aad van der Vaart。“模型设定错误条件下的伯恩斯坦–冯·米塞斯定理”。载于：电子统计学期刊 6（2012），第354–381页。[51] Jeremias Knoblauch、JackJewson 和 TheodorosDamoulas。“贝叶斯法则的以优化为中心的视角：变分推断的综述与推广”。载于：机器学习研究期刊 23.132（2022），第1–109页。

[52] Alice L'Huillier、Luke Travis、Isma¨el Castillo 和 Kolyan Ray。"使用分数后验进行半参数推断"。载于：机器学习研究期刊 24.389 (2023)，第 1–61 页。[53] Jeong Eun Lee、Sitong Liu 和 Geoff K. Nicholls。"广义贝叶斯推断中的学习率贝叶斯推断"。载于：*arXiv* 预印本 *arXiv:2506.12532* (2025)。[54] SeunghyunLee、Nabarun Deb 和 SumitMukherjee。"低信噪比下高维贝叶斯线性回归的中心极限定理"。载于：*arXiv* 预印本 *arXiv:2507.23285* (2025)。[55] Erich Leo Lehmann 和 George Casella。点估计理论。Springer，1998。[56] Xi Lin、Jens Magelund Tarp 和 Robin J. Evans。"通过幂似然结合实验数据与观测数据"。载于：*Biometrics* 81.1 (2025)，ujaf008。[57] Dennis V. Lindley。"近似贝叶斯方法"。载于：*Trabajos de Estadistica Yde Investigacion Operativa* 31.1 (1980)，第 223–245 页。[58] Dennis V. Lindley。"统计推断与决策中先验概率分布的使用"。载于：伯克利数学统计与概率第四届研讨会论文集，第*1*卷：统计理论贡献。1961，第 453–469 页。[59] S P Lyddon、C C Holmes 和 S G Walker。"广义贝叶斯更新与损失似然自助法"。载于：*Biometrika* 106.2 (2019)，第 465–478 页。[60] Ryan Martin 和 Nicholas Syring。"第1章——直接对风险最小化器进行吉布斯后验推断：构造、集中与校准"。载于：统计学手册。2022，第 1–41 页。[61] Juraj Marusic、Marco Avella Medina 和 Cynthia Rush。"M-后验的理论框架：频率学保证与稳健性"。载于：*arXiv* 预印本 *arXiv:2510.01358* (2025)。[62] Takuo Matsubara、Jeremias Knoblauch、Fran¸cois-Xavier Briol 和 Chris. J. Oates。"离散不可处理似然的广义贝叶斯推断"。载于：美国统计学会会刊 119.547 (2024)，第 2345–2355 页。[63] David A McAllester。"PAC-贝叶斯随机模型选择"。载于：*Machine Learning*51.1 (2003)，第 5–21 页。[64] Yann McLatchie、Edwin Fong、David T. Frazier 和 Jeremias Knoblauch。"幂后验的预测性能"。载于：*Biometrika* 112.3 (2025)，asaf034。[65] Jeffrey W. Miller。"广义后验的渐近正态性、集中性与覆盖率"。载于：机器学习研究期刊 22.168 (2021)，第 1–53 页。[66] Jeffrey W. Miller 和 David B. Dunson。" 通过粗化实现稳健贝叶斯推断"。载于：美国统计学会会刊 114.527 (2019)，第 1113–1125 页。[67] Anthony O'Hagan。"用于模型比较的分数贝叶斯因子"。载于：《皇家统计学会会刊*B*辑：统计方法学》57.1 (1995)，第 99–138 页。[68] Lucie Perrotta。"吉布斯后验中温度参数的实用校准"。载于：*arXiv* 预印本 *arXiv:2004.10522* (2020)。

[69] Ruchira Ray、Marco Avella Medina 和 Cynthia Rush。“幂后验均值估计的渐近性质”。载于：*2023*年第*59*届阿勒顿通信、控制与计算年会（*Allerton*）。2023，第 1–8 页。[70] Philippe Rigollet 和 Alexandre B. Tsybakov。“指数加权稀疏估计”。载于：《统计科学》27.4 (2012)，第 558–575 页。[71] Nandana Sengupta 和 Fallaw Sowell。“使用训练样本和测试样本的岭回归估计量的渐近分布”。载于：《计量经济学》8.4 (2020)，第 1–25 页。[72] Zhenming Shun 和 Peter McCullagh。“ 高维积分的拉普拉斯近似 ”。载于：《皇家统计学会会刊*B*辑：统计方法学》57.4 (1995)，第 749–760 页。[73] Nicholas Syring 和 Ryan Martin。“广义后验可信区域的校准”。载于：*Biometrika* 106.2 (2019)，第 479–486 页。[74] Nicholas Syring 和 Ryan Martin。“次指数型损失下的吉布斯后验集中速率”。载于：*Bernoulli* 29.2 (2023)，第 1080–1108 页。[75] B.S. Tsirelson、L.A. Ibragimov 和 V.N. Sudakov。“高斯样本函数的范数”。载于：第三届日苏概率论研讨会论文集。《数学讲义》。1976，第 20–41 页。[76] Aad van der Vaart。《渐近统计学》。剑桥大学出版社，1998。[77] Aki Vehtari、Andrew Gelman 和 Jonah Gabry。“使用留一交叉验证和广泛适用信息准则进行实用贝叶斯模型评估”。载于：*Statistics and Computing* 27.5 (2017)，第 1413–1432 页。[78] Volodimir G. Vovk。“聚合策略”。载于：第三届计算学习理论年度研讨会论文集。1990，第 371–386 页。[79] Martin J.Wainwright 和 MichaelI. Jordan。“图模型、指数族与变分推断”。载于：*Foundations and Trends*® 机器学习 1.1–2 (2008)，第 1–305 页。[80] Yixin Wang、Alp Kucukelbir 和 David M. Blei。“利用贝叶斯数据重加权进行稳健概率建模”。载于：第*34*届国际机器学习会议论文集——第*70*卷。2017，第 3646–3655 页。[81] Pei-Shien Wu 和 Ryan Martin。“广义贝叶斯推断中学习率选择方法的比较”。载于：*Bayesian Analysis* 18.1 (2023)，第 105–132 页。[82] Yun Yang、Debdeep Pati 和 Anirban Bhattacharya。“具有统计保证的$\alpha$-变分推断”。载于：*Annals of Statistics* 48.2 (2020)，第 886–905 页。[83] Tong Zhang。“从ε-熵到KL熵：最小信息复杂度密度估计分析”。载于：*Annals of Statistics* 34.5 (2006)，第 2180–2210 页。[84] Tong Zhang。“统计估计的信息论上界与下界”。载于：*IEEE*信息论汇刊 52.4 (2006)，第 1307–1321 页。

# A 主要结果的证明

## A.1 定理 1 的证明

证明。 我们的证明方法与 [69, 定理 1], 的证明非常接近，该方法改编自 BvM 结果中的论证，即 [5, 定理 1] 和 [50, 定理 2.1], ，旨在证明后验矩的收敛。令 $Z$ 表示我们试图控制的积分，即:

$$Z \equiv \int_{\mathbb{R}^p} \|\sqrt{\alpha_n n}(\theta-\theta^*)\|_2^k \Big|\pi_{n,\alpha_n}(\theta|X^n) - \phi\Big(\theta\Big|\hat{\theta}, \frac{1}{\alpha_n n} V_{\theta^*}^{-1}\Big)\Big| d\theta = \int_{\mathbb{R}^p} \|h\|_2^k |\pi_{n,\alpha_n}^{LAN}(h) - \phi_n(h)| dh, \tag{35}$$

其中，(35) 根据变量替换 $h = \sqrt{\alpha_n n}(\theta - \theta^*)$ 以及变换后的密度得出

$$\pi_{n,\alpha_n}^{\mathrm{LAN}}(h) \equiv \frac{1}{(\alpha_n n)^{\frac{p}{2}}} \pi_{n,\alpha_n}\Big(\theta^* + \frac{h}{\sqrt{\alpha_n n}}\Big|X^n\Big) \text{ and } \phi_n(h) \equiv \frac{1}{(\alpha_n n)^{\frac{p}{2}}} \phi\left(h\Big|\sqrt{\alpha_n n}(\hat{\theta}-\theta^*), V_{\theta^*}^{-1}\right). \tag{36}$$

要证明 (18)，只需证明对于任意 $t > 0$, $\mathbb{P}_{f_{0,n}}(Z > t)$ 收敛于 0。

令 $\{x\}^+ = \max\{0, x\}$。对于向量 $g, h \in \mathbb{R}^p$，定义随机变量

$$f_n(g,h) = \Big\{1 - \frac{\phi_n(h)\, \pi_{n,\alpha_n}^{\mathrm{LAN}}(g|X^n)}{\pi_{n,\alpha_n}^{\mathrm{LAN}}(h|X^n)\, \phi_n(g)}\Big\}^+. \tag{37}$$

对于任意的 $\epsilon > 0$ 和 $\eta > 0$，定义事件

$$\mathcal{A} \equiv \mathcal{A}_n = \Big\{ \sup_{g,h \in \bar{B}_0(r_n)} f_n(g,h) \le \eta \Big\}, \tag{38}$$

其中， $r_n$ 是引理 3 中的增长序列，存在一个 $N_0 \equiv N(\eta, \epsilon)$，使得对于 $\alpha_n n > N_0$ $f_n$ $g$ $h$ $1 - \frac{\epsilon}{2}$，随机变量 (，)– 以至少为 $B_0(r_n)$ 的概率良定义。

注意到

$$\mathbb{P}_{f_{0,n}}(Z > t) \le \mathbb{P}_{f_{0,n}}(\{Z > t\} \cap \mathcal{A}) + \mathbb{P}_{f_{0,n}}(\mathcal{A}^{\mathsf{c}}), \tag{39}$$

因此，为完成证明，我们旨在证明（39）中的第一项收敛于 0，因为根据引理 3，有 $\mathbb{P}_{f_{0,\ n}}(\mathcal{A}^{\mathsf{c}}) < \epsilon$ 对所有 $\alpha_n n > N_0$ 成立。下文中，我们聚焦于（39）的第一项。在整个证明过程中，对于稍后指定的常数 $0 < M_1 < \infty$ 和 $0 < M_2 < \infty$，令

$$\eta \equiv \eta(t, M_1, M_2) = \frac{t}{8} \min\Big\{\frac{1}{M_1}, \frac{1}{M_2}\Big\}. \tag{40}$$

注意，对于任意 $t > 0$, $\eta$ 都是正的，因此可以应用引理 3；我们将证明， $\eta$ 的这一选择导致对于任意 $t > 0$, $\mathbb{P}_{f_{0,\ n}}(\{Z > t\} \cap \mathcal{A}) \to 0$。

根据引理 3，密度 $\pi_{n}^{\mathrm{LAN}}{}_{,\ \alpha_n}$ 以至少 $1 - \frac{\epsilon}{2}$ 的概率在 $\bar{B}_0(r_n)$ 上一致为正，只要 $\alpha_n n > N_0$。此外， $\phi_n(h)$ 处处为正，因为它是一个高斯分布

密度。因此，只要 $\alpha_n n \geq N_0$，我们就可以将引理 4 应用于 $s(h) = \|h\|_2^k$ 以及 $K = B_0(r_n)$——对于任意固定的 $\alpha_n n$，后者都是紧集——从而得到

$$
\begin{aligned}
Z &= \int_{\mathbb{R}^p} \|h\|_2^k \left|\pi_{n,\alpha_n}^{\mathrm{LAN}}(h|X^n) - \phi_n(h)\right| dh \\
&\leq \Big[\sup_{g',h' \in \bar{B}_0(r_n)} f_n(g',h')\Big] \int_{\mathbb{R}^p} \|h\|_2^k \pi_{n,\alpha_n}^{\mathrm{LAN}}(h|X^n) dh + \Big[\sup_{g',h' \in \bar{B}_0(r_n)} f_n(g',h')\Big] \int_{\mathbb{R}^p} \|h\|_2^k \phi_n(h) dh \\
&\quad + \int_{\bar{B}_0(r_n)^{\mathrm{c}}} \|h\|_2^k \pi_{n,\alpha_n}^{\mathrm{LAN}}(h|X^n) dh + \int_{\bar{B}_0(r_n)^{\mathrm{c}}} \|h\|_2^k \phi_n(h) dh.
\end{aligned} \tag{41}
$$

将 (41) 中的四项依次标记为 $T_1$、$T_2$、$T_3$ 和 $T_4$。由

$$
\{T_1 + T_2 + T_3 + T_4 > t\} \cap \mathcal{A} \quad \subseteq \quad \cup_{j=1}^4 \{T_j > \frac{t}{4}\} \cap \mathcal{A} \quad \subseteq \quad \cup_{j=1}^4 \Big\{\{T_j > \frac{t}{4}\} \cap \mathcal{A}\Big\},
$$

再结合并集界，我们得到

$$
\begin{aligned}
\mathbb{P}_{f_{0,n}}(\{Z > t\} \cap \mathcal{A}) \leq \mathbb{P}_{f_{0,n}}(\{T_1 + T_2 + T_3 + T_4 > t\} \cap \mathcal{A}) &\leq \mathbb{P}_{f_{0,n}}\Big(\cup_{j=1}^4 \Big\{\{T_j > \frac{t}{4}\} \cap \mathcal{A}\Big\}\Big) \\
&\leq \sum_{j=1}^4 \mathbb{P}_{f_{0,n}}\Big(\{T_j > \frac{t}{4}\} \cap \mathcal{A}\Big).
\end{aligned} \tag{42}
$$

令 $P_j \equiv \mathbb{P}_{f_{0,n}}(\{T_j > \frac{t}{4}\} \cap \mathcal{A})$。下文将论证，对于任意 $\epsilon > 0$ 和任意 $t > 0$，每一项都可以被 $\epsilon/4$ （可能要求 $\alpha_n n$ 足够大）上界控制。

**(42) 中的项 $\mathbf{P_1}$** 。我们有

$$
P_1 = \mathbb{P}_{f_{0,n}}\Big(\Big\{\Big[\sup_{g',h' \in \bar{B}_0(r_n)} f_n(g',h')\Big] \int_{\mathbb{R}^p} \|h\|_2^k \pi_{n,\alpha_n}^{\mathrm{LAN}}(h|X^n) dh > \frac{t}{4}\Big\} \cap \mathcal{A}\Big). \tag{43}
$$

注意，(43) 中的上确界被 $\frac{t}{8} \min\{\frac{1}{M_1}, \frac{1}{M_2}\}$ 所界定，条件是事件 $\mathcal{A}$ 发生。因此，

$$
\begin{aligned}
P_1 &\leq \mathbb{P}_{f_{0,n}}\Big(\Big[\frac{t}{8} \min\Big\{\frac{1}{M_1}, \frac{1}{M_2}\Big\}\Big] \int_{\mathbb{R}^p} \|h\|_2^k \pi_{n,\alpha_n}^{\mathrm{LAN}}(h|X^n) dh > \frac{t}{4}\Big) \\
&\leq \mathbb{P}_{f_{0,n}}\Big(\int_{\mathbb{R}^p} \|h\|_2^k \pi_{n,\alpha_n}^{\mathrm{LAN}}(h|X^n) dh > 2\max\{M_1, M_2\}\Big) \leq \mathbb{P}_{f_{0,n}}\Big(\int_{\mathbb{R}^p} \|h\|_2^k \pi_{n,\alpha_n}^{\mathrm{LAN}}(h|X^n) dh > M_1\Big),
\end{aligned}
$$

取 $2\max\{M_1, M_2\} \geq 2M_1 \geq M_1$。根据假设(**A3**)，对于所有 $\epsilon > 0$，存在常数$M_1 \equiv M_1(k) < \infty$ 和整数 $N_1 \equiv N_1(M_1, \epsilon)$，使得对于所有 $\alpha_n n > N_1$，

$$
\mathbb{P}_{f_{0,n}}\Big(\int_{\mathbb{R}^p} \|h\|_2^k \pi_{n,\alpha_n}^{\mathrm{LAN}}(h|X^n) dh > M_1\Big) < \frac{\epsilon}{4}.
$$

Ch在我们根据式 (40) 定义 $\eta$ 时选择这个 $M_1$，则当 $\alpha_n n > N_1$ 时有 $P_1 \leq \frac{\epsilon}{4}$。

**项 $\mathbf{P_2}$ 的(42)**。回顾 $P_2$ 的定义：

$$
P_2 = \mathbb{P}_{f_{0,n}}\Big(\Big\{\Big[\sup_{g',h' \in \bar{B}_0(r_n)} f_n(g',h')\Big] \int_{\mathbb{R}^p} \|h\|_2^k \phi_n(h) dh > \frac{t}{4}\Big\} \cap \mathcal{A}\Big). \tag{44}
$$

与我们对 $P_1$ 的界类似，注意到，在事件 $\mathcal{A}$ 下，(44) 中的上确界由 $\frac{t}{8}\min\{\frac{1}{M_1},\frac{1}{M_2}\}$ 界定。因此，$P_2 \le \mathbb{P}_{f_{0,n}}(\int_{\mathbb{R}^p}\|h\|_2^k\phi_n(h)dh > M_2) \le \frac{\epsilon}{4}$，其中，最终不等式由假设 (**A0**) 和引理 5 得出；根据该引理，对于所有 $\epsilon > 0$，存在一个常数 $M_2 \equiv M_2(k) < \infty$ 和一个整数 $N_2 \equiv N_2(M_2, \epsilon)$，使得

$$\mathbb{P}_{f_{0,n}}\Big(\int_{\mathbb{R}^p}\|h\|_2^k\phi_n(h)dh > M_2\Big) < \frac{\epsilon}{4}, \tag{45}$$

对所有 $\alpha_n n > N_2$ 均成立。这就是我们在 (40) 中定义 $\eta$ 时所用的 $M_2$。

**项 $\mathbf{P_3}$ 的(42)。** 我们通过应用引理 6 来界定 $P_3$。在假设 (**A3**) 下，对任意 $\gamma > 0$，令 $k_0 = k(1+\gamma)$，则有 $\int_{\mathbb{R}^p}\|h\|_2^{k(1+\gamma)}\pi_{n\,\alpha_n}^{\mathrm{LAN}}(h|X^n)dh = O_{f_{0,n}}(1)$。因此，我们可以对 $f_Z(z)$ 应用引理 6，其中 $f_Z(z)$ 为 $\pi_{n,\alpha_n}^{\mathrm{LAN}}(z|X^n)$，并得出存在 $N_3 \equiv N_3(\gamma, k, t, \epsilon)$，使得对所有 $\alpha_n n > N_3$，

$$\begin{aligned} P_3 &= \mathbb{P}_{f_{0,n}}\Big(\Big\{\int_{\bar{B}_0(r_n)^{\mathsf{c}}}\|h\|_2^k\pi_{n,\alpha_n}^{\mathrm{LAN}}(h|X^n)dh > \frac{t}{4}\Big\}\cap\mathcal{A}\Big) \\ &\le \mathbb{P}_{f_{0,n}}\Big(\int_{\bar{B}_0(r_n)^{\mathsf{c}}}\|h\|_2^k\pi_{n,\alpha_n}^{\mathrm{LAN}}(h|X^n)dh > \frac{t}{4}\Big) < \frac{\epsilon}{4}. \end{aligned}$$

**项 $\mathbf{P_4}$ 的(42)。** 回顾

$$P_4 = \mathbb{P}_{f_{0,n}}\Big(\Big\{\int_{\bar{B}_0(r_n)^{\mathsf{c}}}\|h\|_2^k\phi_n(h)dh > \frac{t}{4}\Big\}\cap\mathcal{A}\Big) \le \mathbb{P}_{f_{0,n}}\Big(\int_{\bar{B}_0(r_n)^{\mathsf{c}}}\|h\|_2^k\phi_n(h)dh > \frac{t}{4}\Big). \tag{46}$$

注意，根据 Hölder 不等式，

$$\begin{aligned} \int_{\bar{B}_0(r_n)^{\mathsf{c}}}\|h\|_2^k\phi_n(h)dh &= \int_{\mathbb{R}^p}\|h\|_2^k(1-\mathbb{1}\{\|h\|_2 \le r_n\})\phi_n(h)dh \\ &\le \Big(\int_{\mathbb{R}^p}\|h\|_2^{k(1+\gamma)}\phi_n(h)dh\Big)^{\frac{1}{1+\gamma}}\Big(\int_{\mathbb{R}^p}(1-\mathbb{1}\{\|h\|_2 \le r_n\})\phi_n(h)dh\Big)^{\frac{\gamma}{1+\gamma}}. \end{aligned} \tag{47}$$

为界定 $P_4$，定义如下事件 $\mathcal{E}_4 = \{\int_{\mathbb{R}^p}\|h\|_2^{k(1+\gamma)}\phi_n(h)dh < M_4^{1+\gamma}\}$，其中 $M_4$ 将在后文指定。根据假设 (**A0**) 以及应用于 $k = k(1+\gamma)$ 的引理 5，对所有 $\epsilon > 0$，存在常数 $M_4 \equiv M_4(k,\gamma) < \infty$ 和整数 $N_4 \equiv N_4(\epsilon, M_4)$，使得对所有 $\alpha_n n > N_4$，

$$\mathbb{P}_{f_{0,n}}(\mathcal{E}_4^{\mathsf{c}}) = \mathbb{P}_{f_{0,n}}\Big(\int_{\mathbb{R}^p}\|h\|_2^{k(1+\gamma)}\phi_n(h)dh > M_4^{1+\gamma}\Big) < \frac{\epsilon}{4}.$$

此外，注意到对于所有 $h \in \mathbb{R}^p$，$r_n \to \infty$ 和 $\phi_n(h) \le |V_{\theta^*}|^{1/2}$，因此可知，对于任意固定的 $h$，$(1-\mathbb{1}\{\|h\|_2 \le r_n\})\phi_n(h) \to 0$。此外，$(1-\mathbb{1}\{\|h\|_2 \le r_n\})\phi_n(h) \le \phi_n(h)$，这是一个概率密度，因此可积。于是，根据控制收敛定理，存在 $N_5 \equiv N_5(t, M_4, \gamma)$，使得对于所有 $\alpha_n n > \max(1, N_5)$，

$$\int_{\mathbb{R}^p}(1-\mathbb{1}\{\|h\|_2 \le r_n\})\phi_n(h)dh < \Big(\frac{t}{8}\frac{1}{M_4}\Big)^{\frac{1+\gamma}{\gamma}}. \tag{48}$$

因此，根据 (46)、(47) 和 (48)，我们可以如下界定 $P_4$：当 $\alpha_n n > \max(1, N_4, N_5)$ 时，

$$
\begin{aligned}
P_4 &\le \mathbb{P}_{f_{0,n}}\Big(\Big\{\int_{\bar{B}_0(r_n)^{\mathsf{c}}} \|h\|_2^k \phi_n(h)dh > \frac{t}{4}\Big\} \cap \mathcal{E}_4\Big) + \mathbb{P}_{f_{0,n}}(\mathcal{E}_4^{\mathsf{c}}) \\
&\le \mathbb{P}_{f_{0,n}}\Big(\Big\{\Big(\int_{\mathbb{R}^p} \|h\|_2^{k(1+\gamma)} \phi_n(h)dh\Big)^{\frac{1}{1+\gamma}} \Big(\int_{\mathbb{R}^p} (1 - \mathbb{1}\{\|h\|_2 \le r_n\})\phi_n(h)dh\Big)^{\frac{\gamma}{1+\gamma}} > \frac{t}{4}\Big\} \cap \mathcal{E}_4\Big) + \frac{\epsilon}{4} \\
&\le \mathbb{P}_{f_{0,n}}\Big(M_4\Big(\frac{t}{8}\frac{1}{M_4}\Big) > \frac{t}{4}\Big) + \frac{\epsilon}{4} = \frac{\epsilon}{4}
\end{aligned}
$$

现在我们已经对各项 $P_1 - P_4$ 给出了界，利用 (39) 可得方程 (18) 中的结果。为证明将 $\pi_{n,\hat{\alpha}_n}(\theta|X^n)$ 替换为 $\pi_{n,\alpha_n}(\theta|X^n)$ 后，(18) 中的结果仍成立，我们将 $\hat{\alpha}_n$-后验分布改写如下：

$$
\pi_{n,\hat{\alpha}_n}(\theta|X^n) = \frac{f_n(X^n|\theta)^{\hat{\alpha}_n}\pi(\theta)}{\int f_n(X^n|\theta)^{\hat{\alpha}_n}\pi(\theta)d\theta} = \frac{[f_n(X^n|\theta)^{\frac{\hat{\alpha}_n}{\alpha_n}}]^{\alpha_n}\pi(\theta)}{\int [f_n(X^n|\theta)^{\frac{\hat{\alpha}_n}{\alpha_n}}]^{\alpha_n}\pi(\theta)d\theta} = \frac{\tilde{f}_n(X^n|\theta)^{\alpha_n}\pi(\theta)}{\int \tilde{f}_n(X^n|\theta)^{\alpha_n}\pi(\theta)d\theta},
$$

其中 $\tilde{f}_n(X^n|\theta)^{\alpha_n} = f_n(X^n|\theta)^{\frac{\alpha_n}{\alpha_n}}$。因此，对定理 1 应用 $f_n(X^n|\theta)$ 和 $\pi(\theta)$，即可得到所需结果；由此，下式依概率 $f_{0,n}$ 收敛到 0：

$$
\int_{\mathbb{R}^p} \|\sqrt{\alpha_n n}(\theta - \theta^*)\|_2^k \left| \frac{\tilde{f}_n(X^n|\theta)^{\alpha_n}\pi(\theta)}{\int \tilde{f}_n(X^n|\theta)^{\alpha_n}\pi(\theta)d\theta} - \phi\Big(\theta\Big|\hat{\theta}, \frac{1}{\alpha_n n} V_{\theta^*}^{-1}\Big)\right| d\theta. \tag{49}
$$

为应用定理 1，我们需要验证 $f_n(X^n|\theta)$ 和 $\pi(\theta)$ 满足假设 (**A0**)–(**A3**)。注意，(**A1**) 已假定成立。首先注意到，(49) 中出现了对象 $\theta$、$\theta^*$ 和 $V_{\theta^*}$。因此，我们不仅需要证明 (**A0**) 和 (**A2**) 对 $f_n(X^n|\theta)$ 成立，还要证明它们在与 $f_n(X^n|\theta)$ 满足假设 (**A0**) 和 (**A2**) 时相同的 $\theta$、$\theta^*$ 和 $V_{\theta^*}$ 下也成立。我们在命题 7 中验证这一点。根据命题 8，假设 (**A3**) 对 $f_n(X^n|\theta)$ 成立（因为我们已假设 $f_n(X^n|\theta)$ 和 $\pi(\theta)$ 满足命题 5 的条件）。

□

## A.2 定理 2 的证明

证明。回顾一下，后验均值为 $\hat{\theta}^{\mathrm{B}} = [\hat{\theta}_1^{\mathrm{B}}, \ldots, \hat{\theta}_p^{\mathrm{B}}]^\top$ 其中 $\hat{\theta}_j^{\mathrm{B}} = \int_{\mathbb{R}^p} \theta_j \pi_{n,\alpha_n}(\theta|X^n)d\theta$。暂且假设引理 1 的条件成立（稍后我们会验证这些条件）。于是，对每个 $j = 1, \ldots, p$，根据引理 1 的第二个结论，我们有 $\hat{\theta}_j^{\mathrm{B}} = \hat{\theta}_j + O_{f_{0,n}}(\frac{1}{\alpha_n n})$。这意味着存在一个常数 $M_j > 0$ 和一个整数 $N_j \equiv N_j(\epsilon, p)$，使得对于 $\alpha_n n > N_j$，

$$
\mathbb{P}_{f_{0,n}}\Big(\Big|\hat{\theta}_j^{\mathrm{B}} - \hat{\theta}_j\Big| > \frac{M_j}{\alpha_n n}\Big) < \frac{\epsilon}{p}. \tag{50}
$$

我们将利用 (50) 证明

$$
\hat{\theta}^{\mathrm{B}} - \hat{\theta} = O_{f_{0,n}}\Big(\frac{1}{\alpha_n n}\Big), \tag{51}
$$

或者换句话说，存在常数 $M$ 和一个整数 $N$，使得对于 $\alpha_n n > N$，我们有 $\mathbb{P}_{f_{0,n}}(\|\hat{\theta}^{\mathrm{B}} - \hat{\theta}\|_2 > \frac{M}{\alpha_n n}) < \epsilon$. 要证明这一点，请注意 $\|\hat{\theta}^{\mathrm{B}} - \hat{\theta}\|_2^2 = \sum_{j=1}^p (\hat{\theta}_j^{\mathrm{B}} - \hat{\theta}_j)^2 \le$

$p$最大值$_{j\in[p]}\{(\hat{\theta}_j^{\mathrm{B}} - \hat{\theta}_j)^2\}$. 因此，根据并集界，我们得到

$$\begin{aligned}\mathbb{P}_{f_{0,n}}\Big(\|\hat{\theta}^{\mathrm{B}} - \hat{\theta}\|_2 > \frac{M}{\alpha_n n}\Big) &\le \mathbb{P}_{f_{0,n}}\Big(\max_{j\in[p]}\{(\hat{\theta}_j^{\mathrm{B}} - \hat{\theta}_j)^2\} > \frac{M^2}{p(\alpha_n n)^2}\Big) \\ &= \mathbb{P}_{f_{0,n}}\Big(\cup_{j=1}^{p}\Big\{(\hat{\theta}_j^{\mathrm{B}} - \hat{\theta}_j)^2 > \frac{M^2}{p(\alpha_n n)^2}\Big\}\Big) \\ &\le \sum_{j=1}^{p}\mathbb{P}_{f_{0,n}}\Big(\Big|\hat{\theta}_j^{\mathrm{B}} - \hat{\theta}_j\Big| > \frac{M}{\sqrt{p}\alpha_n n}\Big). \end{aligned} \tag{52}$$

现在，令 $M = \sqrt[-]{p}\max\{M_1, \ldots, M_p\}$，其中 $M_j$来自 (50)。

根据 (52) 和 (50)，对于 $\alpha_n n > \max(N_1, \ldots, N_p)$，我们有

$$\begin{aligned}\mathbb{P}_{f_{0,n}}\Big(\|\hat{\theta}^{\mathrm{B}} - \hat{\theta}\|_2 > \frac{M}{\alpha_n n}\Big) &\le \sum_{j=1}^{p}\mathbb{P}_{f_{0,n}}\Big(\Big|\hat{\theta}_j^{\mathrm{B}} - \hat{\theta}_j\Big| > \frac{M}{\sqrt{p}\alpha_n n}\Big) \\ &= \sum_{j=1}^{p}\mathbb{P}_{f_{0,n}}\Big(\Big|\hat{\theta}_j^{\mathrm{B}} - \hat{\theta}_j\Big| > \frac{\sqrt{p}\max\{M_1, \ldots, M_p\}}{\sqrt{p}\alpha_n n}\Big) \\ &\le \sum_{j=1}^{p}\mathbb{P}_{f_{0,n}}\Big(\Big|\hat{\theta}_j^{\mathrm{B}} - \hat{\theta}_j\Big| > \frac{M_j}{\alpha_n n}\Big) < \sum_{j=1}^{p}\frac{\epsilon}{p} = \epsilon. \end{aligned}$$

剩下的工作是验证引理 1 的条件。我们已假设**(A0)**、**(A1')**、**(A2')**和**(A3')** 成立，因此只需验证对于每个 $j = 1, \ldots, p$，假设**(ALap)** 对 $b(\theta) = \theta_j\pi(\theta)$成立：

1. 我们需要 验证 $\int_{\mathbb{R}^p}|\theta_j\pi(\theta)|d\theta < \infty$。这由假设**(A1')**推出。事实上，

$$\int_{\mathbb{R}^p}|\theta_j\pi(\theta)|d\theta = \int_{\mathbb{R}^p}|\theta_j|\pi(\theta)d\theta \le \int_{\mathbb{R}^p}\|\theta\|_1\pi(\theta)d\theta < \infty.$$

2. 我们需要 验证存在一个 $\delta > 0$，使得 $\theta_j\pi(\theta)$ 在 $B_{\theta^*}(\delta)$ 上有限。事实上，

$$|\theta_j\pi(\theta)| \le |\theta_j|\pi(\theta) \le \|\theta\|_1\pi(\theta) \le \sqrt{p}\|\theta\|_2\pi(\theta) \le \sqrt{p}(\|\theta - \theta^*\|_2 + \|\theta^*\|_2)\pi(\theta).$$

根据假设 **(A1')**，存在一个 $\delta > 0$，使得 $\pi(\theta)$ 在 $B_{\theta^*}(\delta)$ 上有限；因此，上述右侧有限。事实上，$\|\theta - \theta^*\|_2 < \delta$ 在 $B_{\theta^*}(\delta)$ 上，且对于所有 $\theta^* \in \mathbb{R}^p$，$\|\theta^*\|_2$ 有限。

3. 我们通过应用命题 2 并取 $q(\theta) = \theta_j$ 来验证条件 (28)。这一选择的$q$ 处处二阶可微；因此，假设 **(ALap)** 成立，其中的 $\delta$ 取自上文。

根据引理 1，将 $\pi_{n,\alpha_n}(\theta|X^n)$ 替换为 $\pi_{n,\hat{\alpha}_n}(\theta|X^n)$ 后，(30) 中的结果仍然成立。因此，采用与刚才类似的论证，将 $\pi_{n,\alpha_n}(\theta|X^n)$ 替换为 $\pi_{n,\hat{\alpha}_n}(\theta|X^n)$ 后，(25) 中的结果也成立。

□

## A.3 拉普拉斯近似的证明，引理 1

证明。 我们先假设 (29) 成立，证明 (30) 中的近似结果，然后再证明 (29)。首先，我们针对 $\pi_{n,\alpha_n}$ 得出结果，再验证这些结果对 $\pi_{n,\hat{\alpha}_n}$ 也成立。注意，(30) 中的期望是形如 (29) 的积分之比，因为

$$\int_{\mathbb{R}^p} q(\theta)\pi_{n,\alpha_n}(\theta|X^n)d\theta = \frac{\int_{\mathbb{R}^p} q(\theta)\pi(\theta)e^{\alpha_n \log f_n(X^n|\theta)}d\theta}{\int_{\mathbb{R}^p} \pi(\theta)e^{\alpha_n \log f_n(X^n|\theta)}d\theta}. \tag{53}$$

首先，我们对 (53) 的分子应用近似式 (29)，取 $b(\theta) = q(\theta)\pi(\theta)$，并假设其满足假设(**ALap**)的条件。这样得到

$$\int_{\mathbb{R}^p} q(\theta)\pi(\theta)e^{\alpha_n \log f_n(X^n|\theta)}d\theta = e^{\alpha_n \log f_n(X^n|\hat{\theta})}|H_n|^{-\frac{1}{2}}\Big(\frac{2\pi}{\alpha_n n}\Big)^{\frac{p}{2}}\Big[q(\hat{\theta})\pi(\hat{\theta}) + O_{f_{0,n}}\Big(\frac{1}{\alpha_n n}\Big)\Big]. \tag{54}$$

接下来，我们对 (53) 的分母应用近似式 (29)，取 $b(\theta) = \pi(\theta)$，得到

$$\int_{\mathbb{R}^p} \pi(\theta)e^{\alpha_n \log f_n(X^n|\theta)}d\theta = e^{\alpha_n \log f_n(X^n|\hat{\theta})}|H_n|^{-\frac{1}{2}}\Big(\frac{2\pi}{\alpha_n n}\Big)^{\frac{p}{2}}\Big[\pi(\hat{\theta}) + O_{f_{0,n}}\Big(\frac{1}{\alpha_n n}\Big)\Big]. \tag{55}$$

注意到，$b(\theta) = \pi(\theta)$ 满足假设(**ALap**)的三个条件，因此可以应用 (55) 中的近似式：

1. 我们需要验证 $\int_{\mathbb{R}^p} |\pi(\theta)|d\theta < \infty$。事实上，$\pi$ 是一个密度；因此处处非负，且积分为一。

2. 根据假设(**A1'**)，存在 $r > 0$，使得 $\pi$ 在 $B_{\theta^*}(r)$ 上有限。
3. 注意，在 $\int_{\mathbb{R}^p} \pi(\theta)e^{\alpha_n \text{对数} f_n(X^n|\theta)}d\theta$ 中，我们有 $b(\theta) = 1 \times \pi(\theta)$。应用命题 2，令 $q(\theta) = 1$，它处处二次连续可微。因此，假设 (**ALap**) 的条件 3 成立。

在验证（或假设）式 (53) 的分子和分母均满足假设 (**ALap**) 后，我们将式 (54) 和 (55) 应用于式 (53)，得到：

$$\begin{aligned}\int_{\mathbb{R}^p} q(\theta)\pi_{n,\alpha_n}(\theta|X^n)d\theta &= \frac{\int_{\mathbb{R}^p} q(\theta)\pi(\theta)e^{\alpha_n \log f_n(X^n|\theta)}d\theta}{\int_{\mathbb{R}^p} \pi(\theta)e^{\alpha_n \log f_n(X^n|\theta)}d\theta} \\ &= \frac{e^{\alpha_n \log f_n(X^n|\hat{\theta})}|H_n|^{-\frac{1}{2}}(\frac{2\pi}{\alpha_n n})^{\frac{p}{2}}[q(\hat{\theta})\pi(\hat{\theta}) + O_{f_{0,n}}(\frac{1}{\alpha_n n})]}{e^{\alpha_n \log f_n(X^n|\hat{\theta})}|H_n|^{-\frac{1}{2}}(\frac{2\pi}{\alpha_n n})^{\frac{p}{2}}[\pi(\hat{\theta}) + O_{f_{0,n}}(\frac{1}{\alpha_n n})]} = \frac{q(\hat{\theta})\pi(\hat{\theta}) + O_{f_{0,n}}(\frac{1}{\alpha_n n})}{\pi(\hat{\theta}) + O_{f_{0,n}}(\frac{1}{\alpha_n n})}.\end{aligned} \tag{56}$$

ˆ 我们将说明，在高概率下，(56) 是良定义的（即，$0 < \pi(\theta) < \infty$，因此不会出现除以 0 的情况），具体如下。首先，根据假设 (**A1'**) 和假设 (**A0**)，$\pi(\theta)$ 在 0 的某个邻域内连续且为正；对于所有 $\epsilon > 0$，我们可以选择某个 $N'' \equiv N''(\epsilon)$

ˆ 足够大，使得 $\theta$ 以至少 $1-\epsilon$ 的概率属于 **(A1')** 所规定的邻域，只要 $\alpha_n n > N''$。接下来，注意到

$$\frac{O_{f_{0,n}}(\frac{1}{\alpha_n n})}{\pi(\hat{\theta}) + O_{f_{0,n}}(\frac{1}{\alpha_n n})} = O_{f_{0,n}}\Big(\frac{1}{\alpha_n n}\Big),$$

并对函数 $f(z) = \frac{a}{a+z}$ 在 $z=0$附近进行泰勒展开,—

$$\frac{q(\hat{\theta})\pi(\hat{\theta})}{\pi(\hat{\theta}) + O_{f_{0,n}}(\frac{1}{\alpha_n n})} = q(\hat{\theta}) - \frac{q(\hat{\theta})O_{f_{0,n}}(\frac{1}{\alpha_n n})}{\pi(\hat{\theta})} = q(\hat{\theta}) + O_{f_{0,n}}\Big(\frac{1}{\alpha_n n}\Big).$$

因此，我们得到了所需结论：利用 (56),

$$\begin{aligned}\int_{\mathbb{R}^p} q(\theta)\pi_{n,\alpha_n}(\theta|X^n)d\theta = \frac{q(\hat{\theta})\pi(\hat{\theta}) + O_{f_{0,n}}(\frac{1}{\alpha_n n})}{\pi(\hat{\theta}) + O_{f_{0,n}}(\frac{1}{\alpha_n n})} &= \frac{q(\hat{\theta})\pi(\hat{\theta})}{\pi(\hat{\theta}) + O_{f_{0,n}}(\frac{1}{\alpha_n n})} + \frac{O_{f_{0,n}}(\frac{1}{\alpha_n n})}{\pi(\hat{\theta}) + O_{f_{0,n}}(\frac{1}{\alpha_n n})} \\ &= q(\hat{\theta}) + O_{f_{0,n}}\Big(\frac{1}{\alpha_n n}\Big).\end{aligned}$$

现在我们将说明，将 $\pi_{n,\alpha_n}(\theta|X^n)$ 替换为 $\pi_{n,\hat{\alpha}_n}(\theta|X^n)$ 后，(30) 中的结果仍然成立。首先，我们用 $\pi_{n,\hat{\alpha}_n}$ 替代 $\pi_{n,\alpha_n}$，重新写出期望:

$$\begin{aligned}\int q(\theta)\pi_{n,\hat{\alpha}_n}(\theta|X^n)d\theta = \frac{\int q(\theta)\pi(\theta)f_n(X^n|\theta)^{\hat{\alpha}_n}d\theta}{\int \pi(\theta)f_n(X^n|\theta)^{\hat{\alpha}_n}d\theta} &= \frac{\int q(\theta)\pi(\theta)\big[f_n(X^n|\theta)^{\frac{\hat{\alpha}_n}{\alpha_n}}\big]^{\alpha_n}d\theta}{\int \pi(\theta)\big[f_n(X^n|\theta)^{\frac{\hat{\alpha}_n}{\alpha_n}}\big]^{\alpha_n}d\theta} \\ &= \frac{\int q(\theta)\pi(\theta)\exp(\alpha_n \log \tilde{f}_n(X^n|\theta))d\theta}{\int \pi(\theta)\exp(\alpha_n \log \tilde{f}_n(X^n|\theta))d\theta},\end{aligned} \tag{57}$$

其中 $\tilde{f}_n(X^n|\theta) = f_n(X^n|\theta)^{\frac{\alpha_n}{\alpha_n}}$。注意，上式是形如 (29) 的积分之比，其中 $f_n(X^n|\theta)$ 被替换为 $f_n(X^n|\theta)$。因此，根据 (56) 附近的论证以及 (57) 中的等式，若将 (29) 中的 $f_n(X^n|\theta)$ 替换为 $f_n(X^n|\theta)$，所需结果便成立。我们将在下文证明这一点。

接下来，我们将证明式 (29) 中的近似。证明分为两部分。首先，我们确定近似中的主导阶项，然后分析余项，证明当 $\alpha_n n$ 足够大时，余项以高概率为 $\frac{1}{\alpha_n n}$ 阶。为此，我们采用与定理 1 证明中类似的方法，在 $\theta$ 的一个（不断扩大的）邻域及其补集上分析余项。然而，与定理 1 证明中的论证不同，本证明中的不断扩大的邻域以 $\theta$ 为中心，而不是以 $\theta^*$ 为中心。我们改变中心点，以利用假设（**A2'**）中的对数似然表示，并得到所需的近似误差阶数。

**步骤 1：确定主导项**。将近似式 (29) 中的乘法常数定义为

$$L \equiv e^{\alpha_n \log f_n(X^n|\hat{\theta})}|H_n|^{-\frac{1}{2}}\Big(\frac{2\pi}{\alpha_n n}\Big)^{\frac{p}{2}}. \tag{58}$$

利用 (58)，我们将 (29) 中的积分改写为

$$\begin{aligned} I &= \int_{\mathbb{R}^p} b(\theta) e^{\alpha_n \log f_n(X^n|\theta)} d\theta \\ &= L \int_{\mathbb{R}^p} |H_n|^{\frac{1}{2}} \Big(\frac{2\pi}{\alpha_n n}\Big)^{-\frac{p}{2}} b(\theta) e^{\alpha_n [\log f_n(X^n|\theta) - \log f_n(X^n|\hat{\theta})]} d\theta \\ &= L \int_{\mathbb{R}^p} |H_n|^{\frac{1}{2}} \Big(\frac{2\pi}{\alpha_n n}\Big)^{-\frac{p}{2}} b(\hat{\theta}) e^{-\frac{1}{2}(\theta-\hat{\theta})^\top (\alpha_n n H_n)(\theta-\hat{\theta})} d\theta \\ &\quad + L \int_{\mathbb{R}^p} |H_n|^{\frac{1}{2}} \Big(\frac{2\pi}{\alpha_n n}\Big)^{-\frac{p}{2}} \Big[ b(\theta) e^{\alpha_n [\log f_n(X^n|\theta) - \log f_n(X^n|\hat{\theta})]} - b(\hat{\theta}) e^{-\frac{1}{2}(\theta-\hat{\theta})^\top (\alpha_n n H_n)(\theta-\hat{\theta})} \Big] d\theta \\ &\equiv L(T_1 + T_2). \end{aligned} \tag{59}$$

式 (29) 中近似的主导项将来自 $T_1$，因为

$$\begin{aligned} T_1 &= \int_{\mathbb{R}^p} |H_n|^{\frac{1}{2}} \Big(\frac{2\pi}{\alpha_n n}\Big)^{-\frac{p}{2}} b(\hat{\theta}) e^{-\frac{1}{2}(\theta-\hat{\theta})^\top (\alpha_n n H_n)(\theta-\hat{\theta})} d\theta \\ &= b(\hat{\theta}) \int_{\mathbb{R}^p} (2\pi)^{-\frac{p}{2}} |\alpha_n n H_n|^{\frac{1}{2}} e^{-\frac{1}{2}(\theta-\hat{\theta})^\top (\alpha_n n H_n)(\theta-\hat{\theta})} d\theta = b(\hat{\theta}). \end{aligned}$$

上面的最后一个等式成立，是因为多元高斯密度的积分为 1。

**步骤 2：分析余项 $\mathbf{T_2}$。** 我们将证明，对于充分大的 $\alpha_n n$，以高概率而言，式 (59) 中的余项 $T_2$ 由阶为 $\frac{1}{\alpha_n n}$ 或更低的项构成。为此，我们在集合 $K$及其补集 $K^{\mathsf{c}} = \mathbb{R}^p \setminus K$上分析积分 $T_2$。我们将在下文详细说明 $K$ 的选择。首先注意到

$$\begin{aligned} T_2 &= \int_{K \cup K^{\mathsf{c}}} |H_n|^{\frac{1}{2}} \Big(\frac{2\pi}{\alpha_n n}\Big)^{-\frac{p}{2}} \Big[ b(\theta) e^{\alpha_n [\log f_n(X^n|\theta) - \log f_n(X^n|\hat{\theta})]} - b(\hat{\theta}) e^{-\frac{1}{2}(\theta-\hat{\theta})^\top (\alpha_n n H_n)(\theta-\hat{\theta})} \Big] d\theta \\ &= \int_{K^{\mathsf{c}}} |H_n|^{\frac{1}{2}} \Big(\frac{2\pi}{\alpha_n n}\Big)^{-\frac{p}{2}} b(\theta) e^{\alpha_n [\log f_n(X^n|\theta) - \log f_n(X^n|\hat{\theta})]} d\theta \\ &\quad - \int_{K^{\mathsf{c}}} |H_n|^{\frac{1}{2}} \Big(\frac{2\pi}{\alpha_n n}\Big)^{-\frac{p}{2}} b(\hat{\theta}) e^{-\frac{1}{2}(\theta-\hat{\theta})^\top (\alpha_n n H_n)(\theta-\hat{\theta})} d\theta \\ &\quad + \int_{K} |H_n|^{\frac{1}{2}} \Big(\frac{2\pi}{\alpha_n n}\Big)^{-\frac{p}{2}} \Big[ b(\theta) e^{\alpha_n [\log f_n(X^n|\theta) - \log f_n(X^n|\hat{\theta})]} - b(\hat{\theta}) e^{-\frac{1}{2}(\theta-\hat{\theta})^\top (\alpha_n n H_n)(\theta-\hat{\theta})} \Big] d\theta. \end{aligned} \tag{60}$$

我们将上式右侧的三项分别记为 $T_{21}$、$T_{22}$ 和 $T_{23}$。证明接下来的目标是说明这些项各自均为 $O_{f_{0,n}}(\frac{1}{\alpha_n n})$。在此之前，我们先介绍并讨论 $K$的选择。

**选择 K：** 令 $\delta_1$ 和 $\delta_2$ 分别表示使假设 **(ALap)** 第 (3) 点和 **(A2')** 成立的 $\delta$ 取值。定义 $\delta \equiv \min(\delta_1, \delta_2)$。我们将证明，在式 (60) 中选择$K = B_{\hat{\theta}}(\delta)$可得到所需的近似式 (29)。

为此，我们 定义以下事件，其中常数 $c > 0$ 将在后文给出 ：

$$\begin{aligned} \mathcal{A}_K &= \Big\{ \sup_{\theta \in K^{\mathsf{c}}} \Big[ \frac{1}{n} \log f_n(X^n|\theta) - \frac{1}{n} \log f_n(X^n|\hat{\theta}) \Big] < -c \Big\}, \\ \mathcal{B} &= \{ H_n \text{ is symmetric and } \lambda_{\min}(H_n) < -c \}. \end{aligned} \tag{61}$$

本节将证明这些事件以高概率发生。具体而言，我们将证明，对于每个事件，其补集的概率不超过 $\epsilon/4$，且这对任意的 $\epsilon > 0$都成立。首先， $K^{\mathsf{c}} = B_{\hat{\theta}}(\delta)^{\mathsf{c}}$，因此

$$\mathbb{P}_{f_{0,n}}(\mathcal{A}_K^{\mathsf{c}}) = \mathbb{P}_{f_{0,n}}\Big( \sup_{\theta \in B_{\hat{\theta}}(\delta)^{\mathsf{c}}} \Big[ \frac{1}{n} \log f_n(X^n|\theta) - \frac{1}{n} \log f_n(X^n|\hat{\theta}) \Big] > -c \Big). \tag{62}$$

根据假设(**A3'**)，存在常数 $c \equiv c(\epsilon,\ \delta)$ 和 $N_0 \equiv N_0(\epsilon,\ \delta)$，使得当 $\alpha_n n > N_0$ 时，其上界为 $\epsilon/4$。我们在 (61) 中 $\mathcal{A}_K$ 的定义里使用这个 $c$。

接下来，由于 $\mathcal{B}$ 对应于 $H_n$ 为对称正定矩阵，根据假设(**A2'**)，存在 $N_1 \equiv N_1(\epsilon)$，使得当 $\alpha_n n > N_1$ 时，$\mathbb{P}_{f_{0,n}}(\mathcal{B}^{\mathsf{c}}) < \frac{\epsilon}{4}$。由此可知，(61) 中的事件以高概率成立。此外，根据并集界，对于任意$\epsilon > 0$以及 $\alpha_n n > \max(N_0,\ N_1)$，有 $\mathbb{P}_{f_{0,n}}((\mathcal{A}_K \cap \mathcal{B})^{\mathsf{c}}) \le \mathbb{P}_{f_{0,n}}(\mathcal{A}_K^{\mathsf{c}}) + \mathbb{P}_{f_{0,n}}(\mathcal{B}^{\mathsf{c}}) \le \frac{\epsilon}{2}$.利用上述结果，注意到对于任意事件 $\mathcal{C}$，根据全概率公式，

$$\mathbb{P}_{f_{0,n}}(\mathcal{C}) \le \mathbb{P}_{f_{0,n}}(\mathcal{C} \,|\, \mathcal{A}_K \cap \mathcal{B}) + \mathbb{P}_{f_{0,n}}((\mathcal{A}_K \cap \mathcal{B})^{\mathsf{c}}) \le \mathbb{P}_{f_{0,n}}(\mathcal{C} \,|\, \mathcal{A}_K \cap \mathcal{B}) + \frac{\epsilon}{2}. \tag{63}$$

我们现在可以转而关注 (60) 中的三项。

**对项 $\mathbf{T_{21}}$ 的分析**(60)**：** 注意，对于 $p \ge 1$,

$$\begin{aligned}
|T_{21}| &= \Big| \int_{K^{\mathsf{c}}} |H_n|^{\frac{1}{2}} \Big( \frac{2\pi}{\alpha_n n} \Big)^{-\frac{p}{2}} b(\theta) e^{\alpha_n [\log f_n(X^n|\theta) - \log f_n(X^n|\hat{\theta})]} d\theta \Big| \\
&\le |H_n|^{\frac{1}{2}} (\alpha_n n)^{\frac{p}{2}} \exp\Big( \alpha_n n \sup_{\theta \in K^{\mathsf{c}}} \Big[ \frac{1}{n} \log f_n(X^n|\theta) - \frac{1}{n} \log f_n(X^n|\hat{\theta}) \Big] \Big) \int_{\mathbb{R}^p} |b(\theta)| d\theta \\
&\le |H_n|^{\frac{1}{2}} (\alpha_n n)^{\frac{p}{2}} \exp(-c\alpha_n n) \int_{\mathbb{R}^p} |b(\theta)| d\theta,
\end{aligned}$$

其中最后一个不等式由对 $\mathcal{A}_K$ 的条件化得到。对于任意给定的 $M_0 > 0$，由 (63)，我们有：对于 $\alpha_n n > \max(N_0,\ N_1)$,

$$\mathbb{P}_{f_{0,n}}\Big( |T_{21}| > \frac{M_0}{\alpha_n n} \Big) \le \mathbb{P}_{f_{0,n}}\Big( |T_{21}| > \frac{M_0}{\alpha_n n} \,\Big|\, \mathcal{A}_K \cap \mathcal{B} \Big) + \frac{\epsilon}{2}.$$

此外，存在 $N_2 \equiv N_2(M_0,\ \delta)$，使得只要 $\alpha_n n > N_2$，便有 $\mathbb{P}_{f_{0,n}}(|T_{21}| > \frac{M_0}{\alpha_n n} \,|\, \mathcal{A}_K \cap \mathcal{B}) = 0$，这是因为根据 $b(\theta)$ 的假定可积性， $T_{21}$ 在事件 $\mathcal{A}_K$ 上（关于 $\alpha_n n$）指数级快速地递减至零。对于任意 $\epsilon > 0$ 以及 $\alpha_n n > \max(N_0,\ N_1,\ N_2)$，我们得出$\mathbb{P}_{f_{0,n}}(|T_{21}| > \frac{M_0}{\alpha_n n}) < \frac{\epsilon}{2}$，由此可得 $T_{21} = O_{f_{0,n}}(\frac{1}{\alpha_n n})$。

**$\mathbf{T_{22}}$ 的分析 (60)：** 考虑变量替换 $h = \sqrt{\alpha_n n}(\theta - \hat{\theta})$。我们注意到，当且仅当 $h \in K \equiv B_0(\delta\sqrt{\alpha_n n})$ 时， $\theta \in K \equiv B_{\hat{\theta}}(\delta)$。特别地，由此可得$K^{\mathsf{c}} = B_0(\delta\sqrt{\alpha_n n})^{\mathsf{c}}$，因此通过变量替换,

$$|T_{22}| = b(\hat{\theta}) \int_{\tilde{K}^{\mathsf{c}}} |H_n|^{\frac{1}{2}} \Big( \frac{2\pi}{\alpha_n n} \Big)^{-\frac{p}{2}} (\alpha_n n)^{-\frac{p}{2}} e^{-\frac{1}{2} h^\top H_n h} dh = b(\hat{\theta}) \int_{\tilde{K}^{\mathsf{c}}} |H_n|^{\frac{1}{2}} (2\pi)^{-\frac{p}{2}} e^{-\frac{1}{2} h^\top H_n h} dh. \tag{64}$$

我们注意到，在事件 $\mathcal{B}$ 下， $H_n$ 是对称正定的。因此，(64) 中的积分给出 $X \sim \mathcal{N}(0,\ H_n^{-1})$

$$|T_{22}| = b(\hat{\theta})\mathbb{P}(X \in \tilde{K}^{\mathsf{c}}) = b(\hat{\theta})\mathbb{P}(X \in B_0(\delta\sqrt{\alpha_n n})^{\mathsf{c}}) = b(\hat{\theta})\mathbb{P}(\|X\| > \delta\sqrt{\alpha_n n}). \tag{65}$$

将引理 8 应用于 $s = \delta\sqrt{\alpha_n n}$ (65)，得到

$$|T_{22}| \le 2b(\hat{\theta}) \exp\Big( - \frac{(\delta\sqrt{\alpha_n n} - \sqrt{\mathrm{tr}(H_n^{-1})})^2}{2\|H_n^{-1}\|_{\mathrm{op}}}\Big), \tag{66}$$

其中， $\|H_n^{-1}\|_{\mathrm{op}} \equiv \sup_{\|v\|_2=1} \|H_n^{-1}v\|_2$ 表示 $H_n^{-1}$的算子范数。对于任意给定的$M_1 > 0$, 由 (63) 可知，当 $\alpha_n n > \max(N_0, N_1)$ 时，有 $\mathbb{P}_{f_0,n}(|T_{22}| > \frac{M_1}{\alpha_n n}) \le \mathbb{P}_{f_0,n}(|T_{22}| > \frac{M_1}{\alpha_n n} \mid \mathcal{A}_K \cap \mathcal{B}) + \frac{\epsilon}{2}$. 此外， $\|H_n^{-1}\|_{\mathrm{op}} = \lambda_{\min}(H_n)^{-1}$，由于 $H_n$ 在事件 $\mathcal{B}$ 上正定，因此该值有限。因此, (66) 在事件 $\mathcal{B}$上随 $\alpha_n n$ 指数衰减，所以存在 $N_3 \equiv N_3(M_1, \delta)$，使得当 $\alpha_n n > N_3$ 时, $\mathbb{P}_{f_0,n}(|T_{22}| > \frac{M_1}{\alpha_n n} \mid \mathcal{A}_K \cap \mathcal{B}) = 0$。由此我们得出，当 $\alpha_n n > \max(N_0, N_1, N_3)$ 时，有 $\mathbb{P}_{f_0,n}(|T_{22}| > \frac{M_1}{\alpha_n n}) < \frac{\epsilon}{2}$，进而可知 $T_{22} = O_{f_0,n}(\frac{1}{\alpha_n n})$。

**项 T 的分析$_{23}$ 的 (60)：** 我们首先分析 $T_{23}$ 的被积函数。回忆一下,

$$T_{23} = \int_K |H_n|^{\frac{1}{2}} \Big(\frac{2\pi}{\alpha_n n}\Big)^{-\frac{p}{2}} \Big[ b(\theta) e^{\alpha_n[\log f_n(X^n|\theta) - \log f_n(X^n|\hat{\theta})]} - b(\hat{\theta}) e^{-\frac{1}{2}(\theta-\hat{\theta})^\top(\alpha_n n H_n)(\theta-\hat{\theta})} \Big] d\theta.$$

令 $\tilde{T}_{23}(\theta) \equiv b(\theta) e^{\alpha_n[\text{对数 } f_n(X^n|\theta) - \text{对数 } f_n(X^n|\hat{\theta})]}$表示 $T_{23}$的 ˆ ˜括号中的第一项被积项。我们对 $T_{23}(\theta)$ 应用变量替换 $h = \sqrt{\alpha_n n}(\theta - \theta)$，并使用假设(**A2'**) 中 $R_n(h)$ 的˜表示，得到:

$$\begin{aligned} \tilde{T}_{23}^h(h) &= b\Big(\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\Big) e^{\alpha_n[\log f_n(X^n|\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}) - \log f_n(X^n|\hat{\theta})]} \\ &= b\Big(\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\Big) e^{-\frac{1}{2}h^\top H_n h} e^{\frac{1}{6\sqrt{\alpha_n n}}\langle S_n, h^{\otimes 3}\rangle + \tilde{R}_n(h)}, \end{aligned} \tag{67}$$

其中 $\langle S_n,\ h^{\otimes 3}\rangle$ 使用了式 (4) 中定义的张量积记法。此外，定义$A(h) \equiv -\frac{1}{6\sqrt{\alpha_n n}}\langle S_n,\ h^{\otimes 3}\rangle - \tilde{R}_n(h)$，我们看到，泰勒展开给出

$$\exp(-A(h)) = 1 - A(h) + \frac{1}{2}A(h)^2 - \frac{\exp(-\tilde{A})}{6}A(h)^3, \tag{68}$$

其中 $\tilde{A} \in (0,\ A(h))$. 我们使用假设(**ALap**) 中 $R_b(h)$ 的表示，得到

$$b\Big(\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\Big) = b(\hat{\theta}) + \frac{v^\top h}{\sqrt{\alpha_n n}} + R_b(h). \tag{69}$$

结合式 (67)–(69) 可得

$$\begin{aligned} \tilde{T}_{23}^h(h) &= e^{-\frac{1}{2}h^\top H_n h}\Big[1 - A(h) + \frac{1}{2}A(h)^2 - \frac{\exp(-\tilde{A})}{6}A(h)^3\Big]\Big[b(\hat{\theta}) + \frac{1}{\sqrt{\alpha_n n}}v^\top h + R_b(h)\Big] \\ &= e^{-\frac{1}{2}h^\top H_n h} B_n(h), \end{aligned} \tag{70}$$

其中

$$B_n(h) \equiv \Big[1 - A(h) + \frac{1}{2}A(h)^2 - \frac{\exp(-\tilde{A})}{6}A(h)^3\Big]\Big[b(\hat{\theta}) + \frac{v^\top h}{\sqrt{\alpha_n n}} + R_b(h)\Big].$$

应用变量替换 $h = \sqrt{\alpha_n n}(\theta - \theta)$ 后，我们利用式 (70) 重写 $T_{23}$。将 $\mathcal{N}(0, H_n^{-1})$ 密度记为 $\rho_H(h) = |H_n|^{\frac{1}{2}}(2\pi)^{-\frac{p}{2}}\exp(-\frac{1}{2}h^\top H_n h)$。由于 $\mathcal{B}$ 对应于 $H_n$ 为对称正定矩阵，在给定 $\mathcal{B}$ 的情况下，该密度是良定义的。由此可得

$$\begin{aligned} T_{23} &= \int_K |H_n|^{\frac{1}{2}}\Big(\frac{2\pi}{\alpha_n n}\Big)^{-\frac{p}{2}}\Big[\tilde{T}_{23}(\theta) - b(\hat{\theta})e^{-\frac{1}{2}(\theta-\hat{\theta})^\top(\alpha_n n H_n)(\theta-\hat{\theta})}\Big]d\theta \\ &= \int_{\tilde{K}} |H_n|^{\frac{1}{2}}\Big(\frac{2\pi}{\alpha_n n}\Big)^{-\frac{p}{2}}\Big[\tilde{T}^h_{23}(h) - b(\hat{\theta})e^{-\frac{1}{2}h^\top H_n h}\Big](\alpha_n n)^{-\frac{p}{2}}dh = \int_{\tilde{K}} \rho_H(h)[B_n(h) - b(\hat{\theta})]dh. \end{aligned} \tag{71}$$

接下来，回顾 $A(h) \equiv -\frac{1}{6\sqrt{\alpha_n n}}\langle S_n, h^{\otimes 3}\rangle - \tilde{R}_n(h)$。令

$$\begin{aligned} R_{1,n}(h) &\equiv \tilde{R}_n(h) + \frac{1}{2}A(h)^2 - \frac{\exp(-\tilde{A})}{6}A(h)^3 \\ &= \tilde{R}_n(h) + \frac{1}{2}\Big(-\frac{\langle S_n, h^{\otimes 3}\rangle}{6\sqrt{\alpha_n n}} - \tilde{R}_n(h)\Big)^2 - \frac{\exp(-\tilde{A})}{6}\Big(-\frac{\langle S_n, h^{\otimes 3}\rangle}{6\sqrt{\alpha_n n}} - \tilde{R}_n(h)\Big)^3, \end{aligned} \tag{72}$$

我们可见

$$\begin{aligned} B_n(h) - b(\hat{\theta}) &= \Big[1 + \frac{\langle S_n, h^{\otimes 3}\rangle}{6\sqrt{\alpha_n n}} + R_{1,n}(h)\Big]\Big[b(\hat{\theta}) + \frac{v^\top h}{\sqrt{\alpha_n n}} + R_b(h)\Big] - b(\hat{\theta}) \\ &= \frac{\langle S_n, h^{\otimes 3}\rangle \cdot b(\hat{\theta})}{6\sqrt{\alpha_n n}} + \frac{v^\top h}{\sqrt{\alpha_n n}} + R_{n,n}(h), \end{aligned} \tag{73}$$

其中 $R_{n,n}(h)$ 包含依赖于 $R_b(h)$、$R_{1,n}(h)$ 以及 $h^{\otimes s}$（其中 $s \geq 4$）；具体而言，

$$R_{n,n}(h) = R_{1,n}(h)\Big[b(\hat{\theta}) + \frac{v^\top h}{\sqrt{\alpha_n n}}\Big] + R_{1,n}(h)R_b(h) + R_b(h)\Big[1 + \frac{\langle S_n, h^{\otimes 3}\rangle}{6\sqrt{\alpha_n n}}\Big] + \frac{\langle S_n, h^{\otimes 3}\rangle}{6\alpha_n n} \cdot v^\top h.$$

我们利用式 (71) 和 (73) 重写 $T_{23}$：

$$\begin{aligned} T_{23} &= \int_{\tilde{K}} \rho_H(h)[B_n(h) - b(\hat{\theta})]dh \\ &= \int_{\tilde{K}} \rho_H(h)R_{n,n}(h)dh + \int_{\tilde{K}} \rho_H(h)\Big[\frac{v^\top h}{\sqrt{\alpha_n n}} + \frac{\langle S_n, h^{\otimes 3}\rangle}{6\sqrt{\alpha_n n}} \cdot b(\hat{\theta})\Big]dh \equiv \tilde{V}_1 + \tilde{V}_2. \end{aligned}$$

我们证明，$\tilde{V}_1$ 和 $\tilde{V}_2$ 的阶均为 $\frac{1}{\alpha_n n}$ 或更低阶，这意味着 $T_{23} = O_{f_0,n}(\frac{1}{\alpha_n n})$。 —

**分析 $\tilde{\mathbf{V}}_1$：** 首先，根据三角不等式，我们有

$$\begin{aligned} |R_{n,n}(h)| \leq &\Big|R_{1,n}(h)\Big[b(\hat{\theta}) + \frac{v^\top h}{\sqrt{\alpha_n n}}\Big]\Big| + |R_{1,n}(h)R_b(h)| \\ &+ \Big|R_b(h)\Big[1 + \frac{\langle S_n, h^{\otimes 3}\rangle}{6\sqrt{\alpha_n n}}\Big]\Big| + \Big|\frac{\langle S_n, h^{\otimes 3}\rangle v^\top h}{6\alpha_n n}\Big|. \end{aligned} \tag{74}$$

˜我们使用式（74）中的不等式对 $|V_1|$ 进行上界估计。然后，对所得的每一项应用柯西-施瓦茨不等式。

$$
\begin{aligned}
|\tilde{V}_1| \le \int_{\tilde{K}} \rho_H(h)|R_{n,n}(h)|dh \le & \Big[\int_{\tilde{K}} \rho_H(h) R_{1,n}(h)^2 dh\Big]^{\frac{1}{2}} \Big[\int_{\tilde{K}} \rho_H(h)\Big|b(\hat{\theta}) + \frac{v^\top h}{\sqrt{\alpha_n n}}\Big|^2 dh\Big]^{\frac{1}{2}} \\
& + \Big[\int_{\tilde{K}} \rho_H(h) R_{1,n}(h)^2 dh\Big]^{\frac{1}{2}} \Big[\int_{\tilde{K}} \rho_H(h) R_b(h)^2 dh\Big]^{\frac{1}{2}} \\
& + \Big[\int_{\tilde{K}} \rho_H(h) R_b(h)^2 dh\Big]^{\frac{1}{2}} \Big[\int_{\tilde{K}} \rho_H(h)\Big|1 + \frac{\langle S_n, h^{\otimes 3}\rangle}{6\sqrt{\alpha_n n}}\Big|^2 dh\Big]^{\frac{1}{2}} \\
& + \frac{1}{\alpha_n n}\int_{\tilde{K}} \rho_H(h)\Big|\frac{1}{6}\langle S_n, h^{\otimes 3}\rangle \cdot v^\top h\Big| dh.
\end{aligned}
\tag{75}
$$

（75）的前三项是以下表达式的倍数

$$
\Big(\int_{\tilde{K}} \rho_H(h) R_b(h)^2 dh\Big)^{\frac{1}{2}}, \quad \text{and} \quad \Big(\int_{\tilde{K}} \rho_H(h) R_{1,n}(h)^2 dh\Big)^{\frac{1}{2}}. \tag{76}
$$

˜给定 $\mathcal{B}$，并注意到 $K \subseteq \mathbb{R}^p$，则（75）前三项中除（76）所含项以外的其他项，都可以用协方差为 $H_n^{-1}$ 的零均值高斯分布下多项式的期望作为上界。这些项的阶为常数或更低。此外，（75）的最后一项为 $O_{f_{0,\,n}}(\frac{1}{\alpha_n n})$。因此，只需证明（76）中的各项为 $O_{f_{0,\,n}}(\frac{1}{\alpha_n n})$，即可得出 $\tilde{V}_1 = O_{f_{0,\,n}}(\frac{1}{\alpha_n n})$。

我们分析（76）中的第一项。令 $M_2 > 0$ 为假设 **(ALap)** 第（3）点中的常数，并定义事件 $\mathcal{E}_0 = \{R_b(h) \le \frac{M_2}{\alpha_n n}\|h\|_2^2$，对所有 $h \in \tilde{K}\}$. 均成立。回忆假设 **(ALap)** 中集合 $\mathcal{H}_{n,\,b} \equiv \{h \in \mathbb{R}^p | \frac{\|h\|_2}{\sqrt{\alpha_n n}} \le \delta_1\}$ 的定义，由于根据定义（见（61）上方）有 $\tilde{K} = B_0(\delta\sqrt{\alpha_n n})$ 且 $\delta \le \delta_1$，因此 $K \subseteq \mathcal{H}_{n,\,b}$。

由此以及 假设 **(ALap)**，存在 $N_4 \equiv N_4(\epsilon, \delta_1)$，使得对于 $\alpha_n n > N_4$ ，

$$
\begin{aligned}
&\mathbb{P}_{f_{0,n}}(\mathcal{E}_0^{\mathsf{c}}) = \mathbb{P}_{f_{0,n}}\Big(R_b(h) > \frac{M_2}{\alpha_n n}\|h\|_2^2 \text{ for some } h \in \tilde{K}\Big) \\
&\le \mathbb{P}_{f_{0,n}}\Big(R_b(h) > \frac{M_2}{\alpha_n n}\|h\|_2^2 \text{ for some } h \in \mathcal{H}_{n,b}\Big) = \mathbb{P}_{f_{0,n}}\Big(\sup_{h \in \mathcal{H}_{n,b}} R_b(h) - \frac{M_2}{\alpha_n n}\|h\|_2^2 > 0\Big) < \frac{\epsilon}{2}.
\end{aligned}
\tag{77}
$$

˜现在，给定一个稍后指定的常数 $M_2 > 0$，对于 $\alpha_n n > \max(N_0, N_1, N_4)$，由（63）

$$
\begin{aligned}
&\mathbb{P}_{f_{0,n}}\Big(\Big(\int_{\tilde{K}} \rho_H(h) R_b(h)^2 dh\Big)^{\frac{1}{2}} > \frac{\tilde{M}_2}{\alpha_n n}\Big) \\
&\quad \le \mathbb{P}_{f_{0,n}}\Big(\Big(\int_{\tilde{K}} \rho_H(h) R_b(h)^2 dh\Big)^{\frac{1}{2}} > \frac{\tilde{M}_2}{\alpha_n n} \,\Big|\, \mathcal{A}_K \cap \mathcal{B}\Big) + \frac{\epsilon}{2} \\
&\quad \le \mathbb{P}_{f_{0,n}}\Big(\frac{M_2}{\alpha_n n}\Big(\int_{\tilde{K}} \rho_H(h)\|h\|_2^4 dh\Big)^{\frac{1}{2}} > \frac{\tilde{M}_2}{\alpha_n n} \,\Big|\, \mathcal{A}_K \cap \mathcal{B} \cap \mathcal{E}_0\Big) + \mathbb{P}_{f_{0,n}}(\mathcal{E}_0^{\mathsf{c}}) + \frac{\epsilon}{2} \\
&\quad \le \mathbb{P}_{f_{0,n}}\Big(\frac{M_2}{\alpha_n n}\Big(\int_{\mathbb{R}^p} \rho_H(h)\|h\|_2^4 dh\Big)^{\frac{1}{2}} > \frac{\tilde{M}_2}{\alpha_n n} \,\Big|\, \mathcal{A}_K \cap \mathcal{B} \cap \mathcal{E}_0\Big) + \mathbb{P}_{f_{0,n}}(\mathcal{E}_0^{\mathsf{c}}) + \frac{\epsilon}{2} < \epsilon.
\end{aligned}
\tag{78}
$$

（78）中的第二个不等式由对 $\mathcal{E}_0$ 进行条件化得到，最后一个不等式则由（77）并选择 $\tilde{M}_2 = 2M_2(\int_{\mathbb{R}^p} \rho_H(h)\|h\|_2^4 dh)^{\frac{1}{2}}$ 使得

$$\mathbb{P}_{f_{0,n}}\Big(\frac{M_2}{\alpha_n n}\Big(\int_{\mathbb{R}^p} \rho_H(h)\|h\|_2^4 dh\Big)^{\frac{1}{2}} > \frac{\tilde{M}_2}{\alpha_n n} \,\Big|\, \mathcal{A}_K \cap \mathcal{B} \cap \mathcal{E}_0\Big) = 0.$$

我们注意到，由于在给定 $\mathcal{B}$ 的条件下，积分项是关于高斯分布的多项式期望，因此 $M_2$ 存在（即为有限值）。因此，我们已经证明，当 $\alpha_n n$ 足够大时，$(\int_{\tilde{K}} \rho_H(h) R_b(h)^2 dh)^{\frac{1}{2}}$ 是 $O_{f_{0,n}}(\frac{1}{\alpha_n n})$。

我们对（76）中的第二项，即 $(\int_{\tilde{K}} \rho_H(h) R_{1,n}(h)^2 dh)^{\frac{1}{2}}$，采用类似的论证。考察（72）中 $R_{1,n}(h)$ 的定义，我们看到，其关于 $\alpha_n n$ 的最高阶项为 $(\frac{\langle S_n, h^{\otimes 3}\rangle}{6\sqrt{\alpha_n n}})^2$（关于 $\rho_H(h)$ 积分时，其阶数为 $\frac{1}{\alpha_n n}$）以及一个关于 $\tilde{R}_n(h)$ 的多项式。依赖于 $\tilde{R}_n(h)$ 的最高阶项是 $(\int_{\tilde{K}} \rho_H(h)\, \tilde{R}_n(h)^2 dh)^{\frac{1}{2}}$（这意味着，对于 $p > 2$，$\tilde{R}_n(h)^p$ 在 $\alpha_n n$ 中更快衰减至零），我们将证明该项为 $O_{f_0,n}(\frac{1}{\alpha_n n})$。我们指出，交叉项 $\frac{\langle S_n, h^{\otimes 3}\rangle}{6\sqrt{\alpha_n n}} \tilde{R}_n(h)$ 也可以用类似方法处理。

我们采用与上述 $R_b(h)$ 类似的论证。令 $M_3 > 0$ 为假设 **(A2')** 所规定的常数，并定义事件 $\mathcal{E}_1 = \{\tilde{R}_n(h) \le \frac{M_3}{\alpha_n n}\|h\|_2^4$ 对所有 $h \in \tilde{K}\}$。回顾假设 **(A2')** 中集合 $\mathcal{H}_n \equiv \{h \in \mathbb{R}^p | \frac{\|h\|_2}{\sqrt{\alpha_n n}} \le \delta_2\}$ 的定义，其中 $\delta_2$ 是使 **(A2')** 成立的 $\delta$ 的值。与（77）附近的论证相同，利用 $K \subseteq \mathcal{H}_n$ 和假设 **(A2')**，存在 $N_5 \equiv N_5(\epsilon, \delta_2)$，使得对于 $\alpha_n n > N_5$，

$$\begin{aligned}\mathbb{P}_{f_{0,n}}(\mathcal{E}_1^{\mathsf{c}}) &= \mathbb{P}_{f_{0,n}}\Big(\tilde{R}_n(h) - \frac{M_3}{\alpha_n n}\|h\|_2^4 > 0 \text{ for some } h \in \tilde{K}\Big) \\ &\le \mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\mathcal{H}_n} \tilde{R}_n(h) - \frac{M_3}{\alpha_n n}\|h\|_2^4 > 0\Big) < \frac{\epsilon}{2}.\end{aligned} \tag{79}$$

现在，给定一个稍后指定的常数 $M_3 > 0$，我们采用类似的论证，并利用（63）得出对于 $\alpha_n n > \max(N_0, N_1, N_5)$，

$$\begin{aligned}&\mathbb{P}_{f_{0,n}}\Big(\Big(\int_{\tilde{K}} \rho_H(h)\tilde{R}_n(h)^2 dh\Big)^{\frac{1}{2}} > \frac{\tilde{M}_3}{\alpha_n n}\Big) \\ &\le \mathbb{P}_{f_{0,n}}\Big(\frac{M_3}{\alpha_n n}\Big(\int_{\tilde{K}} \rho_H(h)\|h\|_2^8\, dh\Big)^{\frac{1}{2}} > \frac{\tilde{M}_3}{\alpha_n n} \,\Big|\, \mathcal{A}_K \cap \mathcal{B} \cap \mathcal{E}_1\Big) + \mathbb{P}_{f_{0,n}}(\mathcal{E}_1^{\mathsf{c}}) + \frac{\epsilon}{2} < \epsilon.\end{aligned} \tag{80}$$

**(80)** 中的第一个不等式由对 $\mathcal{E}_1$ 进行条件化得到，最后一个不等式则由 **(79)** 以及选择 $\tilde{M}_3 = 2M_3(\int_{\mathbb{R}^p} \rho_H(h)\|h\|_2^8\, dh)^{\frac{1}{2}}$，得出；该值有限，因为积分项是关于高斯分布（以 $\mathcal{B}$ 为条件）的多项式的期望。此 $M_3$ 的选择意味着

$$\mathbb{P}_{f_{0,n}}\Big(\frac{M_3}{\alpha_n n}\Big(\int_{\mathbb{R}^p} \rho_H(h)\|h\|_2^8 dh\Big)^{\frac{1}{2}} > \frac{\tilde{M}_3}{\alpha_n n} \,\Big|\, \mathcal{A}_K \cap \mathcal{B} \cap \mathcal{E}_1\Big) = 0.$$

我们已经证明 $(\int_{\tilde{K}} \rho_H(h)\, \tilde{R}_n(h)^2 dh)^{\frac{1}{2}}$ 对于充分大的 $\alpha_n n$ 而言是 $O_{f_{0,n}}(\frac{1}{\alpha_n n})$。因此，$\tilde{V}_1 = O_{f_{0,n}}(\frac{1}{\alpha_n n})$。

**$\tilde{V}_2$ 的分析：** 我们将 $V_2$ 写作，

$$\tilde{V}_2 = \int_{\tilde{K}} \rho_H(h)\Big[\frac{v^\top h}{\sqrt{\alpha_n n}} + \frac{\langle S_n, h^{\otimes 3}\rangle}{6\sqrt{\alpha_n n}} \cdot b(\hat{\theta})\Big]dh = -\int_{\tilde{K}^{\mathsf{c}}} \rho_H(h)\Big[\frac{v^\top h}{\sqrt{\alpha_n n}} + \frac{\langle S_n, h^{\otimes 3}\rangle}{6\sqrt{\alpha_n n}} \cdot b(\hat{\theta})\Big]dh, \tag{81}$$

其中我们使用了 $\int_{\mathbb{R}^p} \rho_H(h)[\frac{v^\top h}{\sqrt{\alpha_n n}} + \frac{\langle S_n, h^{\otimes 3}\rangle}{6\sqrt{\alpha_n n}} \cdot b(\hat{\theta})]dh = 0$。这是因为，给定 $\mathcal{B}$ 后，该积分是关于协方差为 $H_n^{-1}$ 的零均值高斯密度的 $h$ 多项式的期望，因此奇数阶矩为零。接下来，根据柯西-施瓦茨不等式

$$\begin{aligned}
|\tilde{V}_2| &\le \int_{\tilde{K}^{\mathsf{c}}} \rho_H(h)\Big|\frac{v^\top h}{\sqrt{\alpha_n n}} + \frac{\langle S_n, h^{\otimes 3}\rangle}{6\sqrt{\alpha_n n}} \cdot b(\hat{\theta})\Big|dh \\
&\le \Big[\int_{\tilde{K}^{\mathsf{c}}} \rho_H(h)\Big|\frac{v^\top h}{\sqrt{\alpha_n n}} + \frac{\langle S_n, h^{\otimes 3}\rangle}{6\sqrt{\alpha_n n}} \cdot b(\hat{\theta})\Big|^2 dh\Big]^{\frac{1}{2}}\Big[\int_{\tilde{K}^{\mathsf{c}}} \rho_H(h)dh\Big]^{\frac{1}{2}} \\
&\le \Big[\int_{\mathbb{R}^p} \rho_H(h)\Big|\frac{v^\top h}{\sqrt{\alpha_n n}} + \frac{\langle S_n, h^{\otimes 3}\rangle}{6\sqrt{\alpha_n n}} \cdot b(\hat{\theta})\Big|^2 dh\Big]^{\frac{1}{2}}\Big[\mathbb{P}\Big(X \in \bar{B}_0(\delta\sqrt{\alpha_n n})^{\mathsf{c}}\Big)\Big]^{\frac{1}{2}},
\end{aligned}$$

其中，在最后一个不等式中，$X \sim \mathcal{N}(0, H_n^{-1})$，并且我们使用了 $\tilde{K}^{\mathsf{c}} = B_0(\delta\sqrt{\alpha_n n})^{\mathsf{c}}$。注意，给定 $\mathcal{B}$ 后，上式右侧的第一个积分有一个常数上界，因为它是一个高斯多项式的期望。最后，我们将如同在 (66) 中那样应用引理 8 来对上式进行界定，从而得到对于某个常数 $C > 0$,

$$|\tilde{V}_2| \le C\Big[\exp\Big(-\frac{(\delta\sqrt{\alpha_n n} - \sqrt{\mathrm{tr}(H_n^{-1})})^2}{2\|H_n^{-1}\|_{\mathrm{op}}}\Big)\Big]^{\frac{1}{2}}. \tag{82}$$

利用 (63)，对于任意给定的 $M_5 > 0$，我们有，当 $\alpha_n n > \max(N_0, N_1)$ 时，

$$\mathbb{P}_{f_{0,n}}\Big(|\tilde{V}_2| > \frac{M_5}{\alpha_n n}\Big) \le \mathbb{P}_{f_{0,n}}\Big(|\tilde{V}_2| > \frac{M_5}{\alpha_n n}\,\Big|\,\mathcal{A}_K \cap \mathcal{B}\Big) + \frac{\epsilon}{2}.$$

此外，存在一个 $M_5 > 0$ $N_7 \equiv N_7\ \epsilon\ \delta\ M_5$ $\mathbb{P}_{f_{0\,n}}|\tilde{V}_2| > \frac{M_5}{\alpha_n n}$ $\mathcal{A}_K \cap$ 和 $\mathcal{B} = 0$ $\alpha_n n > N_7$ 82 $\alpha_n n$ $\mathcal{B}$ (，，)，使得 .（）；这是因为自变量（）在事件 E 上呈指数下降，论证与（66）附近的论证类似。因此，通过与界定$|T_{21}|$和 $|T_{22}|$时类似的论证，我们有 $\tilde{V}_2 = O_{f_{0,\,n}}(\frac{1}{\alpha_n n})$ 以及 $T_{21} = \tilde{V}_1 + \tilde{V}_2 = O_{f_{0,\,n}}(\frac{1}{\alpha_n n})$。

既然我们已经证明了（29），现在将证明：将 $f_n(X^n|\theta)$ 替换为 $f_n(X^n|\theta)$ 后，该结果仍然成立。为此，我们验证 $f_n(X^n|\theta)$满足假设 **(A0)**、**(A2')** 和 **(A3')**。假设 **(A1')** 和 **(ALap)** 是关于先验分布和 $q(\theta)$ 的假设，因此已经假定它们成立。我们如下检查其余假设：

**假设（A0）：** 注意到，最大似然估计 $\theta$ 出现在式（29）中。因此，我们必须证明对于 $f_n$（$X^n|\theta$）和 $f_n$（$X^n|\theta$），最大似然估计相同。我们在命题 7 中证明这一点。

**假设 (A2') 和 (A3')：** 根据命题 9，假设(**A2'**) 和(**A3'**) 对 $f_n(X^n|\theta)$成立（因为我们已假设它们对 $f_n(X^n|\theta)$成立）。 □

## A.4 命题 3 的证明

证明。我们首先证明 (33)。我们的目标是证明：对于任意 $t > 0$ 和任意 $\epsilon > 0$，存在一个 $N > 0$，使得对于所有 $n > N$，都有 $\mathbb{P}_{f_{0,n}}(\mathrm{d}_{\mathrm{W}_p}(\pi_{n,\gamma_n}, \delta_{\hat{\theta}}) > t) < \epsilon$。回忆一下，对于狄拉克测度，$\mathrm{d}_{\mathrm{W}_p}(\pi_{n,\gamma_n}, \delta_{\hat{\theta}})^p = \int_{\mathbb{R}^p} \|\theta - \hat{\theta}\|_2^p \, \pi_{n,\gamma_n}(\theta|X_n) d\theta$. 固定一个任意足够小的 $r > 0$，并考虑最大似然估计附近的一个球，记作 $B_{\hat{\theta}}(r)$。注意，根据(**A0**)，似然有唯一的极大值点；也就是说，存在唯一的 $\theta \in \Theta$，使得 $\hat{f}_n(X^n|\theta) = \sup_{\theta\in\Theta} f_n(X^n|\theta) =: f_{\max}$。

根据 **(A3')** 我们有，对于所有 $r > 0$，一个常数 $c \equiv c(r, \epsilon)$，且当 $n$ 足够大时，

$$\mathbb{P}_{f_{0,n}}\Big( \sup_{\theta\in B_{\hat{\theta}}(r)^{\mathsf{c}}} f_n(X^n|\theta) > f_{\max} \Big) \le \mathbb{P}_{f_{0,n}}\Big( \sup_{\theta\in B_{\hat{\theta}}(r)^{\mathsf{c}}} \frac{1}{n} \log f_n(X^n|\theta) - \frac{1}{n} \log f_{\max} > -c \Big) < \epsilon.$$

这意味着存在一个 $\rho(r) \in [0, 1)$ 和一个 $N_0 > 0$，使得当 $n > N_0$ 时，

$$\mathbb{P}_{f_{0,n}}\Big( \sup_{B_{\hat{\theta}}(r)^{\mathsf{c}}} f_n(X^n|\theta) \le \rho(r) \, f_{\max} \Big) \ge 1 - \frac{\epsilon}{3}. \tag{83}$$

将事件 $\mathcal{E}_r := \{\sup_{B_{\hat{\theta}}(r)^{\mathsf{c}}} f_n(X^n|\theta) \le \rho(r) \, f_{\max}\}$ 标记为该事件，以上结论意味着，只要 $n > N_0$，就有 $\mathbb{P}_{f_{0,n}}(\mathcal{E}_r^{\mathsf{c}}) < \epsilon/3$。

由似然在 $\theta$ 处的连续性可知，对于任意足够小的 $r > 0$，存在 $c(r) \in (0, 1]$，使得 $\inf_{\theta\in\bar{B}_{\hat{\theta}}(r)} f_n(X^n|\theta) \ge c(r) \, f_{\max}$。注意，当 $r \to 0$ 时，$c(r) \to 1$。

最后，我们根据假设（**A0**）和先验支持假设（**A1**），论证当 $r_0 > 0$ 足够小时，$\pi(B_{\hat{\theta}}(r_0)) > 0$。令 $\delta > 0$ 表示以 $\theta^*$ 为中心的（**A1**）中的球，其中 $\pi$ 为正，并取 $r_0 \le \delta/2$。具体而言，定义事件 $\mathcal{E}_{r_0} := \{\|\theta^* - \theta\|_2 < r_0\}$，并注意到（**A0**）意味着存在 $N_1$，使得只要 $n > N_1$，就有 $\mathbb{P}_{f_0,n}(\mathcal{E}_{r_0}^{\mathsf{c}}) < \epsilon/3$。因此，在事件 $\mathcal{E}_{r_0}$ 上，我们有 $B_{\hat{\theta}}(r_0) \subset B_{\theta^*}(\delta)$，即 $\pi$ 为正的球，并且 $\theta^* \in B_{\hat{\theta}}(r_0)$。特别地，存在一个 $r_0' \le r_0$，使得 $B_{\theta^*}(r_0') \subset B_{\hat{\theta}}(r_0)$，因此 $\pi(B_{\hat{\theta}}(r_0)) > \pi(B_{\theta^*}(r_0'))$。

于是，

$$\begin{aligned} \mathrm{d}_{\mathrm{W}_p}\big(\pi_{n,\gamma_n}, \delta_{\hat{\theta}}\big)^p &= \int_{\mathbb{R}^p} \|\theta - \hat{\theta}\|_2^p \, \pi_{n,\gamma_n}(\theta|X_n) d\theta \\ &= \int_{B_{\hat{\theta}}(r)} \|\theta - \hat{\theta}\|_2^p \, \pi_{n,\gamma_n}(\theta|X_n) d\theta + \int_{B_{\hat{\theta}}(r)^{\mathsf{c}}} \|\theta - \hat{\theta}\|_2^p \, \pi_{n,\gamma_n}(\theta|X_n) d\theta \\ &\le r^p + \frac{\int_{B_{\hat{\theta}}(r)^{\mathsf{c}}} \|\theta - \hat{\theta}\|_2^p f_n(X^n|\theta)^{\gamma_n} \pi(\theta) d\theta}{\int_{\mathbb{R}^p} f_n(X^n|\theta)^{\gamma_n} \pi(\theta) d\theta}. \end{aligned} \tag{84}$$

我们注意到，在给定事件 $\mathcal{E}_M = \{\|\theta\|_2^p < M\}$ 的条件下，对于某个 $M > 0$，根据假设 (ii)，有 $\int_{B_{\hat{\theta}}(r)^{\mathsf{c}}} \|\theta - \hat{\theta}\|_2^p \, \pi(\theta) d\theta \le M_p(r) < \infty$，其中 $M_p(r)$ 有限。根据假设(**A0**)，对于任意 $\epsilon > 0$，以及某个 $N_2$，有 $\mathbb{P}_{f_{0,n}}(\mathcal{E}_M^{\mathsf{c}}) < \epsilon/3$，只要 $n > N_2$ 成立。接下来，对于某个固定的 $r_0 > 0$（选取它以确保我们位于假设 **(A1)** 给出的球内），

且在给定 $\mathcal{E} \equiv \mathcal{E}_r \cap \mathcal{E}_{r_0} \cap \mathcal{E}_M$ 的条件下，

$$\begin{aligned}\frac{\int_{B_{\hat{\theta}}(r)^{\mathrm{c}}} \|\theta-\hat{\theta}\|_2^p f_n(X^n|\theta)^{\gamma_n}\pi(\theta)d\theta}{\int_{\mathbb{R}^p} f_n(X^n|\theta)^{\gamma_n}\pi(\theta)d\theta} &\le \frac{\int_{B_{\hat{\theta}}(r)^{\mathrm{c}}} \|\theta-\hat{\theta}\|_2^p f_n(X^n|\theta)^{\gamma_n}\pi(\theta)d\theta}{\int_{\bar{B}_{\hat{\theta}}(r_0)} f_n(X^n|\theta)^{\gamma_n}\pi(\theta)d\theta} \\ &\le \frac{(\rho(r)f_{\max})^{\gamma_n}\int_{B_{\hat{\theta}}(r)^{\mathrm{c}}} \|\theta-\hat{\theta}\|_2^p \pi(\theta)d\theta}{(c(r_0)f_{\max})^{\gamma_n}\pi(\bar{B}_{\hat{\theta}}(r_0))} \\ &\le \frac{M_p(r)}{\pi(B_{\theta^*}(r_0'))}\Big(\frac{\rho(r)}{c(r_0)}\Big)^{\gamma_n}.\end{aligned} \tag{85}$$

我们已经证明$\mathbb{P}_{f_{0,\ n}}(\mathcal{E}^{\mathrm{c}}) \le 3\times(\epsilon/3) = \epsilon$对任意 $\epsilon > 0$，当 $n > \max(N_0,\ N_1,\ N_2)$时成立。因此，

$$\begin{aligned}&\mathbb{P}_{f_{0,n}}(\mathrm{d}_{\mathrm{W}_p}(\pi_{n,\gamma_n},\delta_{\hat{\theta}}) > t) \\ &\quad\le \mathbb{P}_{f_{0,n}}\Big(\Big\{r^p + \frac{\int_{B_{\hat{\theta}}(r)^{\mathrm{c}}} \|\theta-\hat{\theta}\|_2^p f_n(X^n|\theta)^{\gamma_n}\pi(\theta)d\theta}{\int_\Theta f_n(X^n|\theta)^{\gamma_n}\pi(\theta)d\theta} > t^p\Big\}\cap\mathcal{E}\Big) + \mathbb{P}_{f_{0,n}}(\mathcal{E}^{\mathrm{c}}) \\ &\quad\le \mathbb{I}\Big\{r^p + \frac{M_p(r)}{\pi(B_{\theta^*}(r_0'))}\Big(\frac{\rho(r)}{c(r_0)}\Big)^{\gamma_n} > t^p\Big\} + \epsilon.\end{aligned}$$

现在，由于对任意 $r>0$，都有 $\rho(r)<1$，我们可以选取足够小的 $0<r_0\le\delta/2$，使得当 $c(r)\to 1$ 随 $r$ 收缩时，$\rho(r) < c(r_0)$。

例如，令 $r = t/\sqrt[p]{2}$。因此，

$$\mathbb{P}_{f_{0,n}}\Big(\mathrm{d}_{\mathrm{W}_p}\big(\pi_{n,\gamma_n},\delta_{\hat{\theta}}\big) > t\Big) \le \mathbb{I}\Big\{\frac{M_p(r)}{\pi(B_{\theta^*}(r_0'))}\Big(\frac{\rho(r)}{c(r_0)}\Big)^{\gamma_n} > \frac{1}{2}t^p\Big\} + \epsilon, \tag{86}$$

由于 $M_p(r)$和 $\pi(B_{\theta^*}(r_0'))$均为常数，且

$$\mathbb{I}\Big(\frac{M_p(r)}{\pi(B_{\theta^*}(r_0'))}\Big(\frac{\rho(r)}{c(r_0)}\Big)^{\gamma_n} > \frac{1}{2}t^p\Big) \overset{n\to\infty}{\to} 0 \text{ as } \Big(\frac{\rho(r)}{c(r_0)}\Big)^{\gamma_n} \overset{n\to\infty}{\to} 0. \tag{87}$$

现在，我们要证明满足(32)的$\hat{}\gamma_n$所对应的(33)中的结论。首先，利用任意一个满足以下条件的值 $N_\gamma$ 定义集合 $\mathcal{A} = [N_\gamma, \infty)$

$$\frac{M_p(r)}{\pi(B_{\theta^*}(r_0'))}\Big(\frac{\rho(r)}{c(r_0)}\Big)^{N_\gamma} \le \frac{1}{2}t^p \quad\Longrightarrow\quad N_\gamma > \frac{\log\Big(\frac{t^p\pi(B_{\theta^*}(r_0'))}{2M_p(r)}\Big)}{\log\Big(\frac{\rho(r)}{c(r_0)}\Big)}.$$

根据(32)，对任意 $\epsilon>0$，都存在一个 $N^*$，使得对于 $n>N^*$，有

$$\mathbb{P}_{f_{0,n}}(\hat{\gamma}_n \notin \mathcal{A}) = \mathbb{P}_{f_{0,n}}(\hat{\gamma}_n < N_\gamma) \le \epsilon.$$

接下来，由（84）和（85）可知，在给定 $\mathcal{E} \equiv \mathcal{E}_r \cap \mathcal{E}_{r_0} \cap \mathcal{E}_M$ 的条件下，

$$\sup_{\gamma\in\mathcal{A}} \mathrm{d}_{\mathrm{W}_p}\big(\pi_{n,\gamma},\delta_{\hat{\theta}}\big)^p \le r^p + \sup_{\gamma\in\mathcal{A}} \frac{M_p(r)}{\pi(B_{\theta^*}(r_0'))}\Big(\frac{\rho(r)}{c(r_0)}\Big)^{\gamma} \le r^p + \frac{1}{2}t^p, \tag{88}$$

其中，最后一步用到了集合 $\mathcal{A}$ 的定义。我们已证明，对于任意 $\epsilon > 0$，当 $n > \max(N_0, N_1, N_2)$ 时，$\mathbb{P}_{f_{0,n}}(\mathcal{E}^c) \le 3\times(\epsilon/3) = \epsilon$。因此，当 $n > \max(N_0, N_1, N_2)$ 时，

$$\begin{aligned}\mathbb{P}_{f_{0,n}}\Big(\sup_{\gamma\in\mathcal{A}} \mathrm{d}_{\mathrm{W}_p}(\pi_{n,\gamma}, \delta_{\hat{\theta}}) > t\Big) &\le \mathbb{P}_{f_{0,n}}\Big(\Big\{\sup_{\gamma\in\mathcal{A}} \mathrm{d}_{\mathrm{W}_p}(\pi_{n,\gamma}, \delta_{\hat{\theta}})^p > t^p\Big\} \cap \mathcal{E}\Big) + \epsilon \\ &\le \mathbb{P}_{f_{0,n}}\Big(r^p > \frac{1}{2}t^p \cap \mathcal{E}\Big) + \epsilon = \epsilon,\end{aligned}$$

再次通过取足够小的 $r$。

我们的目标是证明，对于任意 $t > 0$ 和任意 $\epsilon > 0$，存在一个 $N > 0$，使得对于所有 $n > N$，都有 $\mathbb{P}_{f_{0,n}}(\mathrm{d}_{\mathrm{W}_p}(\pi_{n,\hat{\gamma}_n}, \delta_{\hat{\theta}}) > t) < 2\epsilon$。注意，

$$\mathbb{P}_{f_{0,n}}(\mathrm{d}_{\mathrm{W}_p}(\pi_{n,\hat{\gamma}_n}, \delta_{\hat{\theta}}) > t) \le \mathbb{P}_{f_{0,n}}(\mathrm{d}_{\mathrm{W}_p}(\pi_{n,\hat{\gamma}_n}, \delta_{\hat{\theta}}) > t) \cap \hat{\gamma}_n \in \mathcal{A}) + \mathbb{P}_{f_{0,n}}(\hat{\gamma}_n \notin \mathcal{A}). \tag{89}$$

当 $n > N^*$ 足够大时，上式右侧的第二项有上界 $\epsilon$。对于第一项，注意当 $n > \max(N_0, N_1, N_2)$ 时，

$$\begin{aligned}\mathbb{P}_{f_{0,n}}(\mathrm{d}_{\mathrm{W}_p}(\pi_{n,\hat{\gamma}_n}, \delta_{\hat{\theta}}) > t \cap \hat{\gamma}_n \in \mathcal{A}) &\le \mathbb{P}_{f_{0,n}}\Big(\sup_{\gamma\in\mathcal{A}} \mathrm{d}_{\mathrm{W}_p}(\pi_{n,\gamma}, \delta_{\hat{\theta}}) > t \cap \hat{\gamma}_n \in \mathcal{A}\Big) \\ &\le \mathbb{P}_{f_{0,n}}\Big(\sup_{\gamma\in\mathcal{A}} \mathrm{d}_{\mathrm{W}_p}(\pi_{n,\gamma}, \delta_{\hat{\theta}}) > t\Big) \le \epsilon,\end{aligned} \tag{90}$$

这就完成了证明，其中 $N = \max(N^*, N_0, N_1, N_2)$。 □

## A.5 定理 3 的证明

在开始证明定理 3 之前，我们先陈述并证明两个辅助结果，用全变差距离加上矩来界定 Wasserstein 距离，以便我们能够引用第 3.1 节中关于全变差收敛的结果。

**引理 2.** 设 $(X, d)$ 为一个度量空间，并固定 $p \ge 1$ 和 $m > p$。对于任意 $x_0 \in X$ 及定义在 $X$ 上、具有有限 $m$ 阶矩的概率测度 $\mu, \nu$，令 $\delta := \mathrm{d}_{\mathrm{TV}}(\mu, \nu) := \sup_A |\mu(A) - \nu(A)|$，并且

$$M_m(\mu, \nu) := \int d(x, x_0)^m \mu(dx) + \int d(y, x_0)^m \nu(dy). \tag{91}$$

则 $\mathrm{d}_{\mathrm{W}_p}(\mu, \nu) \le 2^{1+\frac{1}{p}} M_m(\mu, \nu)^{\frac{1}{m}} \delta^{\frac{1}{p}-\frac{1}{m}}$.

证明。令 $(X, Y)$ 为 $(\mu, \nu)$ 的最大耦合，使得 $\mathbb{P}(X \ne Y) = \delta$。对于任意 $x_0 \in X$，有 $d(X, Y) \le d(X, x_0) + d(Y, x_0)$。此外，$(a+b)^p \le 2^{p-1}(a^p + b^p)$ 且 $\mathbf{1}\{\max(a, b) > c\} \le \mathbf{1}\{a > c\} + \mathbf{1}\{b > c\}$。利用这些结果，我们看到，对于任意 $R > 0$ 以及任意 $x_0 \in X$，

$$\begin{aligned}d(X, Y)^p &= d(X, Y)^p \mathbf{1}\{X \ne Y, d(X, x_0) \le R, d(Y, x_0) \le R\} \\ &\quad + d(X, Y)^p \mathbf{1}\{\max(d(X, x_0), d(Y, x_0)) > R\} \\ &\le (2R)^p \mathbf{1}\{X \ne Y\} + 2^{p-1}(d(X, x_0)^p + d(Y, x_0)^p)\mathbf{1}\{\max(d(X, x_0), d(Y, x_0)) > R\} \\ &\le (2R)^p \mathbf{1}\{X \ne Y\} + 2^p d(X, x_0)^p \mathbf{1}\{d(X, x_0) > R\} + 2^p d(Y, x_0)^p \mathbf{1}\{d(Y, x_0) > R\}.\end{aligned} \tag{92}$$

因此，根据（92），我们有

$$\begin{aligned}\mathbb{E}[d(X,Y)^p] &\le (2R)^p\mathbb{P}(X\neq Y) + 2^p\mathbb{E}\left[d(X,x_0)^p\mathbf{1}\{d(X,x_0)>R\} + d(Y,x_0)^p\mathbf{1}\{d(Y,x_0)>R\}\right]\\ &\le (2R)^p\delta + 2^pR^{p-m}M_m(\mu,\nu),\end{aligned} \tag{93}$$

其中，最后一步使用了关于 $m>p$ 的尾界 $\int_{\{d>R\}} d^p d\mu \le R^{p-m}\int d^m d\mu$ （$\nu$ 也同样如此）。通过平衡两项来选择 $R$：$(2R)^p\delta = 2^pR^{p-m}M_m$，从而得到 $R^m = M_m/\delta$。然后，(93) 中的界给出 $\mathbb{E}[d(X,\ Y)^p] \le 2(2R)^p\delta = 2^{p+1}M_m(\mu,\ \nu)^{\frac{p}{m}}\delta^{1-\frac{p}{m}}$. 取 $p$ 次方根即可得到所要证明的结论：

$$\mathrm{d}_{\mathrm{W}_p}(\mu,\nu) = \Big(\inf_{(X,Y):X\sim\mu;\,Y\sim\nu}\mathbb{E}[d(X,Y)^p]\Big)^{\frac{1}{p}} \le 2^{1+\frac{1}{p}}M_m(\mu,\nu)^{\frac{1}{m}}\delta^{\frac{1}{p}-\frac{1}{m}}.$$

□

下面的推论考虑随机测度的情形，这正是我们的设定所需要的。

**推论 3。** 设 $\{\widehat{\mu}_n\},\ \{\widehat{\nu}_n\}$ 是 $(X,\ d)$ 上的随机概率测度，并固定 $m>p\ge 1$。如果 $\mathrm{d}_{\mathrm{TV}}(\widehat{\mu}_n,\ \widehat{\nu}_n) \xrightarrow{\mathbb{P}} 0$ 以及 $M_m(\widehat{\mu}_n,\ \widehat{\nu}_n) = O_{\mathbb{P}}$（(91) 中定义的 1）对某个 $x_0\in X$ 成立，则 $\mathrm{d}_{\mathrm{W}_p}(\widehat{\mu}_n,\ \widehat{\nu}_n) \xrightarrow{\mathbb{P}} 0$.

证明。 对 $(\widehat{\mu}_n,\widehat{\nu}_n)$ 逐点应用引理 2，我们得到

$$\mathrm{d}_{\mathrm{W}_p}(\widehat{\mu}_n,\widehat{\nu}_n) \le 2^{1+\frac{1}{p}}M_m(\widehat{\mu}_n,\widehat{\nu}_n)^{\frac{1}{m}}\mathrm{d}_{\mathrm{TV}}(\widehat{\mu}_n,\widehat{\nu}_n)^{\frac{1}{p}-\frac{1}{m}}.$$

由于 $M_m(\widehat{\mu}_n,\widehat{\nu}_n) = O_{\mathbb{P}}(1)$，且 $\mathrm{d}_{\mathrm{TV}}(\widehat{\mu}_n,\widehat{\nu}_n) \xrightarrow{p} 0$，右侧依概率趋于 0。

□

现在我们转而证明定理 3。

证明。 首先注意到，三角不等式给出

$$\begin{aligned}\mathrm{d}_{\mathrm{W}_p}\Big(\pi_{n,\hat{\alpha}_n^{\mathrm{mix}}},\ q\phi_n + (1-q)\delta_{\hat{\theta}}\Big) &= \mathrm{d}_{\mathrm{W}_p}\Big(q_n\pi_{n,\hat{\alpha}_n} + (1-q_n)\pi_{n,\hat{\gamma}_n},\ q\phi_n + (1-q)\delta_{\hat{\theta}}\Big)\\ &\le \mathrm{d}_{\mathrm{W}_p}\Big(q_n\pi_{n,\hat{\alpha}_n} + (1-q_n)\pi_{n,\hat{\gamma}_n},\ q\pi_{n,\hat{\alpha}_n} + (1-q)\pi_{n,\hat{\gamma}_n}\Big) \quad (94)\\ &\quad + \mathrm{d}_{\mathrm{W}_p}\Big(q\pi_{n,\hat{\alpha}_n} + (1-q)\pi_{n,\hat{\gamma}_n},\ q\phi_n + (1-q)\delta_{\hat{\theta}}\Big). \quad (95)\end{aligned}$$

我们将证明项 (94) 和 (95) 均为 $o_{f_{0,n}}(1)$，由上述界即可得出所需结果。

**项 (94)：** 为证明 (94) 在 $f_{0,n}$-概率下收敛至零，我们将对 $\widehat{\mu}_n = q_n\pi_{n,\hat{\alpha}_n} + (1-q_n)\pi_{n,\hat{\gamma}_n}$ 和 $\widehat{\nu}_n = q\pi_{n,\hat{\alpha}_n} + (1-q)\pi_{n,\hat{\gamma}_n}$ 应用推论 3。为此，我们将在下文中证明以下命题，分别称为断言 1 和断言 2：

1. $\mathrm{d}_{\mathrm{TV}}(\widehat{\mu}_n,\ \widehat{\nu}_n) = o_{f_{0,\ n}}(1)$。

2. 对于某个 $m>p$，$M_m(\widehat{\mu}_n,\widehat{\nu}_n) = O_{f_{0,n}}(1)$，其中 $M_m$ 定义见 (91) .

**主张 1：** 根据全变差距离的定义，我们有

$$\begin{aligned}\mathrm{d_{TV}}(\widehat{\mu}_n,\widehat{\nu}_n) &= \mathrm{d_{TV}}\left(q_n\pi_{n,\hat{\alpha}_n}(\theta|X^n)+(1-q_n)\pi_{n,\hat{\gamma}_n}(\theta|X^n),\ q\pi_{n,\hat{\alpha}_n}(\theta|X^n)+(1-q)\pi_{n,\hat{\gamma}_n}(\theta|X^n)\right)\\ &= \frac{1}{2}\int_{\mathbb{R}^p}\left|(q_n-q)\pi_{n,\hat{\alpha}_n}(\theta|X^n)-(q_n-q)\pi_{n,\hat{\gamma}_n}(\theta|X^n)\right|d\theta\\ &\le \frac{|q_n-q|}{2}\Big(\int_{\mathbb{R}^p}\pi_{n,\hat{\alpha}_n}(\theta|X^n)d\theta+\int_{\mathbb{R}^p}\pi_{n,\hat{\gamma}_n}(\theta|X^n)d\theta\Big)=|q_n-q|.\end{aligned}$$

最后，由于 $q_n\to q$ 以 $f_{0,n}$ 的概率成立，上述右侧收敛于 0，因此我们得出 $\mathrm{d_{TV}}(\widehat{\mu}_n,\widehat{\nu}_n)$ 收敛于 0。

**主张 2：** 给定 $m>p$，并令 $x_0=\theta^*$ 代入 (91)，我们可以如下界定 $M_m(\cdot,\ \cdot)$:

$$\begin{aligned}\frac{M_m(\widehat{\mu}_n,\widehat{\nu}_n)}{2} &= \frac{q_n+q}{2}\int_{\mathbb{R}^p}\|\theta-\theta^*\|_2^m\pi_{n,\hat{\alpha}_n}(\theta|X^n)d\theta+\frac{2-(q_n+q)}{2}\int_{\mathbb{R}^p}\|\theta-\theta^*\|_2^m\pi_{n,\hat{\gamma}_n}(\theta|X^n)d\theta\\ &\le \int_{\mathbb{R}^p}\|\theta-\theta^*\|_2^m\pi_{n,\hat{\alpha}_n}(\theta|X^n)d\theta+\int_{\mathbb{R}^p}\|\theta-\theta^*\|_2^m\pi_{n,\hat{\gamma}_n}(\theta|X^n)d\theta=:T_1+T_2.\quad(96)\end{aligned}$$

注意到我们已经证明了 $\frac{1}{2}M_m(\widehat{\mu}_n,\ \widehat{\nu}_n)=T_1+T_2=M_m(\pi_{n,\ \hat{\alpha}_n},\ \pi_{n,\ \hat{\gamma}_n})$，而要证明该主张，只需论证上述量为 $O_{f_{0,\ n}}(1)$。接下来我们将看到，事实上 $T_1$ 和 $T_2$ 都是 $o_{f_{0,\ n}}(1)$。事实上，变量替换 $h=\sqrt{\alpha_n n}(\theta-\theta^*)$，以及定理 1 和 $(\alpha_n n)^{-m/2}\to 0$ 表明

$$\begin{aligned}T_1 &\le \int_{\mathbb{R}^p}\|\theta-\theta^*\|_2^m\phi\Big(\theta\Big|\hat{\theta},\frac{1}{\alpha_n n}V_{\theta^*}^{-1}\Big)d\theta+\int_{\mathbb{R}^p}\|(\theta-\theta^*)\|_2^m\Big|\pi_{n,\hat{\alpha}_n}(\theta|X^n)-\phi\Big(\theta\Big|\hat{\theta},\frac{1}{\alpha_n n}V_{\theta^*}^{-1}\Big)\Big|d\theta\\ &= (\alpha_n n)^{-\frac{m}{2}}\int_{\mathbb{R}^p}(\alpha_n n)^{-\frac{p}{2}}\|h\|_2^m\phi\Big(h\ \Big|\ \sqrt{\alpha_n n}(\hat{\theta}-\theta^*),V_{\theta^*}^{-1}\Big)dh+o_{f_{0,n}}(1)=o_{f_{0,n}}(1),\qquad(97)\end{aligned}$$

最后一个等式由引理 5 和 $(\alpha_n n)^{-m/2}\to 0$得出。

We 现在使用界 $\|\theta-\theta^*\|_2^m\le 2^{m-1}\|\theta-\hat{\theta}\|_2^m+2^{m-1}\|\hat{\theta}-\theta^*\|_2^m$ 得到

$$\frac{T_2}{2^{m-1}}\le\int_{\mathbb{R}^p}\left(\|\theta-\hat{\theta}\|_2^m+\|\hat{\theta}-\theta^*\|_2^m\right)\pi_{n,\hat{\gamma}_n}(\theta|X^n)d\theta=d_{\mathrm{W_m}}(\pi_{n,\hat{\gamma}_n},\delta_{\hat{\theta}})^m+\|\hat{\theta}-\theta^*\|_2^m.$$

由于我们假设对于某个 $m>p$，命题 3 的条件成立，因此根据命题 3，上式的第一项依 $f_{0,n}$-概率收敛于 0。此外，根据定理 1 的 (A0)，第二项为 $o_{f_{0,n}}(1)$。因此，$T_2=o_{f_{0,n}}(1)$。

**Term** (95)**:** B由 $d_{\mathrm{W}_p}(P,Q)^p$ 的凸性（引理 9）和命题 3，可得

$$\begin{aligned}\mathrm{d}_{\mathrm{W}_p}(q\pi_{n,\hat{\alpha}_n}+(1-q)\pi_{n,\hat{\gamma}_n},\ q\phi_n+(1-q)\delta_{\hat{\theta}})^p &\le q\mathrm{d}_{\mathrm{W}_p}(\pi_{n,\hat{\alpha}_n},\phi_n)^p+(1-q)\mathrm{d}_{\mathrm{W}_p}(\pi_{n,\hat{\gamma}_n},\delta_{\hat{\theta}})^p\\ &= q\mathrm{d}_{\mathrm{W}_p}(\pi_{n,\hat{\alpha}_n},\phi_n)^p+o_{f_{0,n}}(1).\qquad(98)\end{aligned}$$

根据推论 3，结合 $\hat{}\mu_n=\pi_{n,\hat{\alpha}_n}$ 和 $\hat{}\nu_n=\phi_n$，式 (98) 中的剩余项依 $f_{0,n}$-概率收敛于零。事实上，推论 1 表明 $\mathrm{d_{TV}}(\widehat{\mu}_n,\widehat{\nu}_n)\to 0$ 依 $f_{0,n}$-概率成立；因此，要应用推论 3，我们需要证明对于 (91) 中定义的 $M_m$，$M_m(\widehat{\mu}_n,\widehat{\nu}_n)=O_{f_{0,n}}(1)$, 。注意到

$$M_m(\widehat{\mu}_n,\widehat{\nu}_n)=\int\|\theta-\theta^*\|_2^m\pi_{n,\hat{\alpha}_n}(\theta|X^n)d\theta+\int\|\theta-\theta^*\|_2^m\phi\Big(\theta|\hat{\theta},\frac{1}{\alpha_n n}V_{\theta^*}^{-1}\Big)d\theta.$$

上式右侧的第一项是 (96) 的 $T_1$，第二项与 (97) 的第一项相同；如上所示，两者均为 $o_{f_{0,\ n}}(1)$。 □

# B 技术引理

**引理 3**。 假设(A0)–(A2) 成立。那么，对于任意 $\eta > 0$ 和 $\epsilon > 0$，存在一个整数 $N(\eta, \epsilon)$ 和一个序列 $r_n \equiv r_n(\eta, \epsilon) \to \infty$，使得对于任意满足 $\alpha_n n > N(\eta, \epsilon)$ 的固定 $n$，以下结论成立：

1. 在 $\bar{B}_0(r_n)$ 上，(36) 中给出的密度 $\pi_n^{\mathrm{LAN}}{}_{,\alpha_n}(\cdot|X^n)$ 为正，并且 (37) 中给出的随机变量 $f_n(g,h)$ 以至少 $1-\epsilon/2$ 的概率良定义。

2. 我们有

$$\mathbb{P}_{f_{0,n}}\Big(\sup_{g,h\in\bar{B}_0(r_n)} f_n(g,h) > \eta\Big) < \epsilon. \tag{99}$$

证明。 我们的证明分为三个步骤，并改编自 [5, 引理 5] 的论证。步骤 1 中，我们验证 $f_n(g,h)$ 在半径固定（在 $\alpha_n n$ 中）的球上良定义。步骤 2 中，我们证明 (99)，其中上确界取遍半径固定为 $r > 0$ 的球。步骤 3 中，我们构造一个序列 $r_n \to \infty$，使得两项结果对 $g, h \in B_0(r_n)$ 均成立（即，步骤 1 和步骤 2 的结果均成立，从而得到最终结果）。

**步骤 1：检查 $f_n(g,h)$ 在固定半径的球上是良定义的。** 我们证明，对于足够大的 $\alpha_n n$（稍后说明其取值），函数 $f_n(g,h)$ 以高概率在 $B_0(r)$ 上良定义。回顾 $f_n(g,h)$ 的定义：

$$f_n(g,h) = \left\{1 - \frac{\phi_n(h)\pi^{\mathrm{LAN}}_{n,\alpha_n}(g|X^n)}{\pi^{\mathrm{LAN}}_{n,\alpha_n}(h|X^n)\phi_n(g)}\right\}^+.$$

我们称 $f_n(g,h)$ 良定义，若对于 $g,h \in \bar{B}_0(r)$，$\pi^{LAN}_{n\,\alpha}{}_{,}(h|X^n)$ 和 $\phi_n(g)$ 均为正（即分母不为 0）。

回顾式 (1) 中 $\alpha_n$-后验分布以及式 (36) 中缩放后的密度的定义。于是，

$$\frac{\pi^{\mathrm{LAN}}_{n,\alpha_n}(g|X^n)}{\pi^{\mathrm{LAN}}_{n,\alpha_n}(h|X^n)} = \left[\frac{f_n(X^n|\theta^* + \frac{g}{\sqrt{\alpha_n n}})}{f_n(X^n|\theta^* + \frac{h}{\sqrt{\alpha_n n}})}\right]^{\alpha_n} \frac{\pi_n(g)}{\pi_n(h)} = \frac{\exp\left(\alpha_n \log f_n\left(X^n|\theta^* + \frac{g}{\sqrt{\alpha_n n}}\right)\right)}{\exp\left(\alpha_n \log f_n\left(X^n|\theta^* + \frac{h}{\sqrt{\alpha_n n}}\right)\right)}\frac{\pi_n(g)}{\pi_n(h)},$$

变换 $\sqrt{\alpha_n n}(\theta - \theta^*)$ 的先验分布密度记为

$$\pi_n(h) \equiv (\alpha_n n)^{-p/2}\pi(\theta^* + h/\sqrt{\alpha_n n}). \tag{100}$$

根据式 (37) 的定义，令

$$s_n(h) = \exp\left(\alpha_n \log\left(\frac{f_n(X^n|\theta^* + \frac{h}{\sqrt{\alpha_n n}})}{f_n(X^n|\theta^*)}\right)\right),$$

我们有

$$f_n(g,h) = \left\{1 - \frac{\phi_n(h)\pi^{\mathrm{LAN}}_{n,\alpha_n}(g|X^n)}{\pi^{\mathrm{LAN}}_{n,\alpha_n}(h|X^n)\phi_n(g)}\right\}^+ = \left\{1 - \frac{\phi_n(h)s_n(g)\pi_n(g)}{\phi_n(g)s_n(h)\pi_n(h)}\right\}^+. \tag{101}$$

为了证明 (101) 在球上是良定义的，我们说明 $\phi_n(\cdot)$、 $s_n(\cdot)$ 和 $\pi_n(\cdot)$ 均为正。这对于 $\phi_n(\cdot)$ 显然成立，因为它是高斯密度。

首先，我们考察 $s_n(\cdot)$ 的正性，这可由 (**A2**) 得出。对于任意 $h \in B_0(r)$,

$$\begin{aligned} s_n(h) &= \exp\left(\alpha_n\left[\log f_n\Big(X^n\Big|\theta^* + \frac{h}{\sqrt{\alpha_n n}}\Big) - \log f_n(X^n|\theta^*)\right]\right) \\ &= \exp\left(\alpha_n\left[\log f_n\Big(X^n\Big|\theta^* + \frac{h}{\sqrt{\alpha_n n}}\Big) - \log f_n(X^n|\theta^*)\right] - R_n(h)\right)\exp\left(R_n(h)\right) \\ &= \exp\left(\sqrt{\alpha_n}h^\top V_{\theta^*}\Delta_{n,\theta^*} - \frac{1}{2}h^\top V_{\theta^*}h\right)\exp\left(R_n(h)\right), \end{aligned} \tag{102}$$

其中，(102) 是由假设 (**A2**) 中 $R_n(h)$的定义得出的。要得出$s_n(h)$在该球上以高概率为正，只需证明 (102) 中 exp(·) 内的两个项在 $h \in B_0(r)$ 上均不会一致发散至 $-\infty$。

由 (A0) 可知 $\Delta_{n,\theta^*} = O_{f_0,n}(1)$，且由 (A2) 可知 $V_{\theta^*}$ 为正定矩阵。因此，我们有 $\sup_{h\in\bar{B}_0(r)} |h^\top V_{\theta^*}\Delta_{n,\theta^*}| = O_{f_0,n}(1)$，进而 $\sup_{h\in\bar{B}_0(r)} \sqrt{\alpha_n}|h^\top V_{\theta^*}\Delta_{n,\theta^*}| = o_{f_0,n}(1)$。此外，由 (A2) 可知，$\sup_{h\in\bar{B}_0(r)} |R_n(h)| = o_{f_0,n}(1)$。因此，对于任意$\epsilon > 0$，存在 $t(\epsilon) > 0$ 以及 $N_0 \equiv N_0(\epsilon, t(\epsilon)\ r)$，使得以下事件

$$\mathcal{E} = \Big\{ \inf_{h\in\bar{B}_0(r)} \sqrt{\alpha_n}h^\top V_{\theta^*}\Delta_{n,\theta^*} - \frac{1}{2}h^\top V_{\theta^*}h > -t \quad \cap \quad \inf_{h\in\bar{B}_0(r)} R_n(h) > -t\Big\},$$

满足，对于 $\alpha_n n > N_0$,

$$\mathbb{P}_{f_{0,n}}(\mathcal{E}) \geq 1 - \epsilon/2. \tag{103}$$

尚需证明 (100) 中给出的 $\pi_n(\cdot)$在该球上为正。我们验证，对于任意 $g,\ h \in B_0(r)$，$\theta^* + h/\sqrt{\alpha_n n}$和 $\theta^* + g/\sqrt{\alpha_n n}$均属于球 $B_{\theta^*}(\delta)$，其中 $\pi(\theta)$ 根据假设 (**A1**) 连续且为正。事实上，请注意，对于任意 $r > 0$，存在一个整数 $N_1 \equiv N_1(r,\ \delta) \equiv \lceil \frac{4r^2}{\delta^2} \rceil > 0$，使得当 $h \in \bar{B}_0(r)$ 且 $\alpha_n n \geq N_1(r,\ \delta)$ 时，$\theta^* + h/\sqrt{\alpha_n}n \in B_{\theta^*}(\delta)$。为此，注意若 $\|h\|_2 \leq r$ 且 $\alpha_n n \geq \frac{4r^2}{\delta^2}$，则

$$\frac{\|h\|_2}{\sqrt{\alpha_n n}} \leq \frac{r}{\sqrt{\alpha_n n}} \leq \frac{r}{\sqrt{\frac{4r^2}{\delta^2}}} = \frac{\delta}{2} < \delta.$$

因此，当 $\|h\|_2/\sqrt{\alpha_n}n < \delta$ 和 $\|g\|_2/\sqrt{\alpha_n n} < \delta$时，我们有 $\theta^* + h/\sqrt{\alpha_n}n$ 和 $\theta^* + g/\sqrt{\alpha_n}n$属于球 $B_{\theta^*}(\delta)$。由此可知，当 $\alpha_n n > N_1$时，$\pi_n(\cdot)$在该球上为正显而易见。

根据这一结论和 (103)，对于 $\alpha_n n > \max(N_0,\ N_1)$，(101) 在 $B_0(r)$ 上是良定义的，且其 $f_{0,\ n}$-概率至少为 $1 - \epsilon/2$。

请注意，我们已经证明了一个与我们希望在结果 1 中证明的结果类似的结论，但针对的是 $g,\ h \in B_0(r)$，其中 $r$ 在 $\alpha_n n$中固定。在步骤 3 中，我们将证明，对于某个特定的增长序列，当 $g,\ h \in B_0(r_n)$ 时，也可以证明这些相同的步骤，从而完成结果 1 的证明。

**步骤 2：证明 (99) 中针对固定 $r>0$ 的陈述。** 对于任意满足 $h_n \to h$ 的序列 $h_n \in B_0(r)$，请注意，(A2) 蕴含

$$\begin{aligned}\log(s_n(h_n)) &= \alpha_n\Big[\log f_n\big(X^n\big|\theta^* + \frac{h_n}{\sqrt{\alpha_n n}}\big) - \log f_n(X^n|\theta^*)\Big] \\ &= \sqrt{\alpha_n} h_n^\top V_{\theta^*}\Delta_{n,\theta^*} - \frac{1}{2}h_n^\top V_{\theta^*} h_n + o_{f_{0,n}}(1) = -\frac{1}{2}h_n^\top V_{\theta^*} h_n + o_{f_{0,n}}(1), \end{aligned} \tag{104}$$

其中，我们根据 (A0) 使用了 $\alpha_n \to 0$ 和 $V_{\theta^*}\Delta_{n,\theta^*} = O_{f_{0,n}}(1)$，因为根据 (A2)，$V_{\theta^*}$ 是正定的。根据密度 $\phi_n(h_n)$ 的定义，我们有 $\log(\phi_n(h_n)) = -\frac{p}{2}\log(2\pi) + \frac{1}{2}\log(\det(V_{\theta^*})) - \frac{1}{2}h_n^\top V_{\theta^*} h_n$. 因此，

$$\log\left[\frac{s_n(h_n)}{\phi_n(h_n)}\right] = o_{f_{0,n}}(1) + \frac{p}{2}\log(2\pi) - \frac{1}{2}\log(\det(V_{\theta^*})). \tag{105}$$

在下文的 (106)–(108) 中，我们将隐式地以 $\mathcal{E}$ 为条件，并取 $\alpha_n n > N_1$，使得 (106)–(107) 中的每个对象都有良定义。由 (104)–(105) 可知，

$$\log\left[\frac{\phi_n(h_n)s_n(g_n)}{\phi_n(g_n)s_n(h_n)}\right] = o_{f_{0,n}}(1). \tag{106}$$

由 (100) 以及 $\pi$ 的连续性可得 $\pi(\theta^* + g_n/\sqrt{\alpha_n n}) \,\pi(\theta^* + h_n/\sqrt{\alpha_n n}) \to \pi(\theta^*)$，

$$\log\left[\frac{\pi_n(g_n)}{\pi_n(h_n)}\right] = \log\left[\frac{\pi(\theta^* + g_n/\sqrt{\alpha_n n})}{\pi(\theta^* + h_n/\sqrt{\alpha_n n})}\right] = o_{f_{0,n}}(1). \tag{107}$$

由 (106)–(107)，我们得到

$$b_n(g_n, h_n) \equiv \log\Big[\frac{\phi_n(h_n)s_n(g_n)\pi_n(g_n)}{\phi_n(g_n)s_n(h_n)\pi_n(h_n)}\Big] = o_{f_{0,n}}(1). \tag{108}$$

由于 $h_n$ 和 $g_n$ 是 $B_0(r)$ 中任意的序列，结果 (108) 等价于：对于任意固定的 $r$，存在一个 $N_2 \equiv N_2(r,\epsilon,\eta)$，使得当 $\alpha_n n > \max(N_1, N_2)$ 时，$\mathbb{P}_{f_0,n}(|b_n(g_n,h_n)| > \eta|\mathcal{E}) < \epsilon/2$。注意，我们还令 $\alpha_n n$ 大于 $N_1$，以便在给定 $\mathcal{E}$ 的条件下，$f_n(g,h)$ 是良定义的；此外，当 $\alpha_n n > N_0$ 时，我们有 $P(\mathcal{E}) \geq 1-\epsilon/2$。由于 $f_n(g_n,h_n)$ 是 $|b_n(g_n,h_n)|$ 的连续函数，并且假设 $g_n$ 和 $h_n$ 是 $B_0(r)$ 中收敛的连续序列，根据连续映射定理，存在一个整数 $N_2 \equiv N_2(r,\epsilon,\eta)$，使得当 $\alpha_n n > \max(N_1, N_2)$ 时，

$$\mathbb{P}_{f_{0,n}}\Big(\sup_{g,h\in\bar{B}_0(r)} f_n(g,h) > \eta \,\Big|\, \mathcal{E}\Big) < \frac{\epsilon}{2}. \tag{109}$$

于是根据 (103) 和 (109)，对于 $\alpha_n n > \max(N_0,\ N_1,\ N_2)$，我们有

$$\mathbb{P}_{f_{0,n}}\Big(\sup_{g,h\in\bar{B}_0(r)} f_n(g,h) > \eta\Big) \leq \mathbb{P}_{f_{0,n}}\Big(\sup_{g,h\in\bar{B}_0(r)} f_n(g,h) > \eta\Big|\mathcal{E}\Big) + \mathbb{P}_{f_{0,n}}(\mathcal{E}^{\mathsf{c}}) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \tag{110}$$

现在我们已经证明了一个与引理第二个结论类似的结果，不过这里的上确界是在 $B_0(r)$ 上取的，其中半径在 $r$ 中固定。步骤 3 中，我们证明 (110) 对构造出的 $r_n$ 成立。

步骤 3：证明引理 3 中关于 $r_n \to \infty$ 的陈述。回顾步骤 1 中 $N_0(\epsilon, t(\epsilon), r)$ 的定义、同样来自步骤 1 的 $N_1(r, \delta) = \lceil \frac{4r^2}{\delta^2} \rceil$，以及步骤 2 中 $\tilde{N}_2(r, \epsilon, \eta)$ 的定义。令 $N^*(\epsilon, \eta) = \max\{N_0(\epsilon, t(\epsilon), 1), N_1(1, \delta), N_2(1, \epsilon, \eta)\}$，并且对所有 $\alpha_n n > N^*(\epsilon, \eta)$，令

$$r_n = \max\left\{ r \in \mathbb{R} \,\middle|\, r \le \delta\sqrt{\alpha_n n}/2 \quad \text{and} \quad \alpha_n n > \max\{N_0(\epsilon, t(\epsilon), r), \tilde{N}_2(r, \epsilon, \eta))\right\}.$$

对于 $\alpha_n n \le N^*(\epsilon, \eta)$，我们可以任意定义 $r_n$，例如 $r_n = 1$。我们需要检查：**(A) 检查 $r_n$ 是否良定义（即，存在一个 $r$ 满足 $r_n$ 的定义）：** 事实上，$r = 1$ 满足 $\alpha_n n > N^*(\epsilon, \eta)$ 的 $r_n$ 定义：

$$\frac{\delta\sqrt{\alpha_n n}}{2} > \frac{\delta\sqrt{N^*(\epsilon,\eta)}}{2} \ge \frac{\delta\sqrt{4/\delta^2}}{2} = 1 = r,$$

其中我们用到了 $N^*(\epsilon, \eta) \ge N_1(1, \delta) \ge 4/\delta^2$。接下来，注意到 $\alpha_n n > N^*(\epsilon, \eta) \ge \max\{N_0(\epsilon, t(\epsilon), 1), N_2(1, \epsilon, \eta)\} = \max\{N_0(\epsilon, t(\epsilon), r), N_2(r, \epsilon, \eta)\}$，这由 $N^*(\epsilon, \eta)$ 的定义可得。(B) 验证实际上 $r_n \to \infty$：事实上，

$$\begin{aligned} r_n &= \max\left\{ r \in \mathbb{R} \,\middle|\, r \le \frac{\delta\sqrt{\alpha_n n}}{2} \quad \text{and} \quad \alpha_n n > \max\{N_0(\epsilon, t(\epsilon), r), \tilde{N}_2(r, \epsilon, \eta))\right\} \\ &= \min\left\{ \max\left\{ r \in \mathbb{R} \middle| r \le \frac{\delta\sqrt{\alpha_n n}}{2} \right\}, \max\left\{ r \in \mathbb{R} \middle| \alpha_n n > \max\{N_0(\epsilon, t(\epsilon), r), \tilde{N}_2(r, \epsilon, \eta))\right\}\right\}, \end{aligned}$$

显然 $\delta\sqrt{\alpha_n n}/2 \to \infty$，因此 $r_n \to \infty$ if $\max\left\{ r \in \mathbb{R} \middle| \alpha_n n > \max\{N_0(\epsilon, t(\epsilon), r), \tilde{N}_2(r, \epsilon, \eta))\right\} \to \infty$。注意，若要使 $\max\left\{ r \in \mathbb{R} \middle| \alpha_n n > \max\{N_0(\epsilon, t(\epsilon), r), \tilde{N}_2(r, \epsilon, \eta))\right\} \to \infty$ 随着 $n$ 增大，我们需要 $N_0(\epsilon, t(\epsilon), r)$ 和 $N_2(r, \epsilon, \eta)$ 都随着 $r$ 增大而保持有限。事实上，若结论不成立，则存在一个 $r_{\max}$，使得要么 $N_0(\epsilon, t(\epsilon), r) = \infty$，要么 $N_2(r, \epsilon, \eta) = \infty$，其中 $r > r_{\max}$。然而，步骤 2 对任意有限的 $r$ 都成立，例如 $r = r_{\max} + 1$；并且，特别是，步骤 2 的结果意味着 $N_0$ 和 $N_2$ 都是有限的，因此矛盾。(C) 证明对于所有 $\alpha_n n > N^*(\epsilon, \eta)$，都有 $\mathbb{P}_{f_0, n}(\sup_{g, h \in \bar{B}_0(r_n)} f_n(g, h) > \eta) < \epsilon.$。为此，需要针对 $g, h \in B_0(r_n)$ 重复步骤 1 和步骤 2：

步骤 1：首先，我们检查 $f_n(g, h)$ 在 $B_0(r_n)$ 上是良定义的。我们验证 $B_0(r_n)$ 的 (103)，这意味着 $s_n(\cdot)$ 在 $B_0(r_n)$ 上为正。为此，我们注意到，对于每个 $\alpha_n n$，集合 $B_0(r_n)$ 是紧的。也就是说，对于某个 $r' > 0$，$B_0(r_n) = B_0(r')$。因此，根据 (A2)，对于所有 $\epsilon > 0$，存在 $t > 0$ 和 $N_0' \equiv N_0'(\epsilon, t, r')$，使得 (103) 对 $\alpha_n n > N_0'$ 在 $B_0(r')$ 上成立。其次，我们检查 $\pi_n(\cdot)$ 对所有 $g, h \in B_0(r_n)$ 都为正。我们注意到，当 $h \in B_0(r_n)$ 时，$\theta^* + h/\sqrt{\alpha_n n} \in B_{\theta^*}(\delta)$，因为 $\|h\|_2/\sqrt{\alpha_n n} \le r_n/\sqrt{\alpha_n n} \le \delta/2$，这由 $r_n$ 的定义可知。这保证了 $f_n(g, h)$ 是良定义的，正如步骤 1 中所述。

步骤 2：根据 $r_n$ 定义，$\alpha_n n > \max\{N_0(\epsilon, t(\epsilon) r_n), N_2(r_n, \epsilon, \eta)\}$ 对所有 $\alpha_n n > N^*(\epsilon, \eta)$ 都成立。此外，(110) 中的条件 $\alpha_n n > N_1(r_n, \delta) = \lceil \frac{4r_n^2}{\delta^2} \rceil$ 旨在保证

¯当 $h \in B_0(r_n)$ 时， $\theta^* + h/\sqrt{\alpha_n n} \in B_{\theta^*}(\delta)$ 成立，这一点我们已在上文论证。因此，(110) 中的界成立，证明至此完成。

□

**引理 4.** [69, 引理 1] 设密度 $\varphi$ 和 $\psi$ 在给定的紧集 $K$上为正。则对于任意在整个 $\mathbb{R}^p$上非负的函数 $s(h)$,

$$\int_{\mathbb{R}^p} s(h)|\varphi(h) - \psi(h)|dh \leq \Big[\sup_{g,g' \in K} \tilde{f}(g,g')\Big] \int_{\mathbb{R}^p} s(h)(\psi(h) + \varphi(h))dh + \int_{\mathbb{R}^p \setminus K} s(h)(\psi(h) + \varphi(h))dh,$$

其中

$$\tilde{f}(g,g') \equiv \left\{1 - \frac{\varphi(g')\psi(g)}{\psi(g')\varphi(g)}\right\}^+.$$

˜**注 2.** 与 [69, 引理 1], 不同，引理 4 不使用量 $f^-(g,h)$，因为根据 ˜ ˙[69, 式 (21)-(22)], ，我们可知 $\sup_{g,h\in K} f^+(g,h) = \sup_{g,h\in K} f^-(g,h)$。

**引理 5.** 假设 (A0) 成立。则对于序列 $\alpha_n \to 0$ 和 $\alpha_n n \to \infty$，存在常数 $M$ ，使得

$$\begin{aligned}&\lim_{n\to\infty} \mathbb{P}_{f_{0,n}}\Big(\int_{\mathbb{R}^p} \|h\|_2^k \phi_n(h)dh > M\Big) \\ &\quad = \lim_{n\to\infty} \mathbb{P}_{f_{0,n}}\Big(\int_{\mathbb{R}^p} (\alpha_n n)^{-\frac{p}{2}} \|h\|_2^k \phi\Big(h\Big|\sqrt{\alpha_n n}(\hat{\theta} - \theta^*), V_{\theta^*}^{-1}\Big)dh > M\Big) = 0.\end{aligned} \tag{111}$$

ˆ证明。 注意到，通过变量替换 $h = \sqrt{\alpha_n n}(\theta + \theta - \theta^*)$ ，

$$\begin{aligned}\int_{\mathbb{R}^p} \|h\|_2^k \phi_n(h)dh &= \int_{\mathbb{R}^p} (\alpha_n n)^{-\frac{p}{2}} \|h\|_2^k \phi\Big(h \,\Big|\, \sqrt{\alpha_n n}(\hat{\theta} - \theta^*), V_{\theta^*}^{-1}\Big)dh \\ &= \int_{\mathbb{R}^p} (\alpha_n n)^{\frac{k}{2}} \|\theta - \hat{\theta} + \theta^*\|_2^k \phi\Big(\theta\Big|0, \frac{V_{\theta^*}^{-1}}{\alpha_n n}\Big)d\theta \\ &= (\alpha_n n)^{\frac{k}{2}} \mathbb{E}_{\mathcal{N}(0, \frac{V_{\theta^*}^{-1}}{\alpha_n n})}[\|\theta + \hat{\theta} - \theta^*\|_2^k].\end{aligned} \tag{112}$$

在证明的余下部分，我们将用 $\mathbb{E}$ 而非 $\mathbb{E}_{\mathcal{N}(0, V_{\theta^*}^{-1}/(\alpha_n n))}$ 表示相对于 $\theta \sim \mathcal{N}(0, V_{\theta^*}^{-1}/(\alpha_n n))$ 取的期望。不难证明，对于 ., $(a+b)^p \leq 2^p a^p + b^p \qquad p \geq 0 \qquad 112 \qquad \mathbb{E}[\|\theta + \theta - \theta^*\|_2^k] \leq )$ ( )。
$\mathbb{E}[\ \|\theta\|_2 + \|\theta - \theta^*\|_2\ ^k] \leq 2^k \mathbb{E}[\|\theta\|_2^k] + 2^k \|\theta - \theta^*\|_2^k.$
考虑右侧的 ( )，注意到 ˆ ( ˆ ) ˆ 我们已经证明了

$$\int_{\mathbb{R}^p} \|h\|_2^k \phi_n(h)dh \leq 2^k (\alpha_n n)^{\frac{k}{2}} \mathbb{E}[\|\theta\|_2^k] + (2\sqrt{\alpha_n})^k \|\sqrt{n}(\hat{\theta} - \theta^*)\|_2^k. \tag{113}$$

注意，根据 H¨older 不等式， $\|\theta\|_2^k \leq p^{\frac{k}{2}-1} \sum_{i=1}^p |\theta_i|^k$。接下来，由于 $\theta_i \sim \mathcal{N}(0, \ \frac{1}{\alpha_n n}[V_{\theta^*}^{-1}]_{ii})$,

$$\mathbb{E}[|\theta_i|^k] = \frac{2^{\frac{k}{2}}}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right) \left[\frac{[V_{\theta^*}^{-1}]_{ii}}{\alpha_n n}\right]^{\frac{k}{2}},$$

对于每个 $i \in [p]$。因此，(113) 右侧的第一项有一个常数上界：

$$2^k(\alpha_n n)^{\frac{k}{2}}\mathbb{E}[\|\theta\|_2^k] \le 2^k(\alpha_n n)^{\frac{k}{2}}p^{\frac{k}{2}-1}\frac{2^{\frac{k}{2}}}{\sqrt{\pi}}\Gamma\Big(\frac{k+1}{2}\Big)\sum_{i=1}^{p}\Big[\frac{[V_{\theta^*}^{-1}]_{ii}}{\alpha_n n}\Big]^{\frac{k}{2}}$$
$$= \frac{2^{\frac{3k}{2}}p^{\frac{k}{2}-1}}{\sqrt{\pi}}\Gamma\Big(\frac{k+1}{2}\Big)\sum_{i=1}^{p}\big[[V_{\theta^*}^{-1}]_{ii}\big]^{\frac{k}{2}}. \tag{114}$$

因此，要证明 (111) 中的结果，只需证明 (113) 在 $f_{0,n}$-概率下有界。我们如下进行证明。根据假设 (A0)，$\sqrt{n}(\theta-\theta^*)$ 依分布收敛到 $\mathcal{N}(0,V)$，其中 $V$ 是某个正定矩阵。令 $Q\Lambda Q^\top$ 表示 $V$ 的谱分解。于是，$\Lambda^{-\frac{1}{2}}\sqrt{n}(\hat\theta-\theta^*)$ 依分布收敛到 $\mathcal{N}(0,I)$。根据连续映射定理，$\|\Lambda^{-\frac{1}{2}}\sqrt{n}(\hat\theta-\theta^*)\|_2^k = (\|\Lambda^{-\frac{1}{2}}\sqrt{n}(\hat\theta-\theta^*)\|_2^2)^{\frac{k}{2}}$ 依分布收敛到 $(\chi_p^2)^{\frac{k}{2}}$，因此在 $f_{0,n}$-概率下有界。记 $\Lambda$ 的对角元素为 $\lambda_1,\dots,\lambda_p$（即 $V$ 的特征值）。注意到

$$\|\Lambda^{-\frac{1}{2}}\sqrt{n}(\hat\theta-\theta^*)\|_2^k = \Big(\sum_{j=1}^{p}\frac{[\sqrt{n}(\hat\theta_j-\theta^*)_j]^2}{\lambda_j}\Big)^{\frac{k}{2}} \ge \frac{1}{\lambda_{\max}^{k/2}}\Big(\sum_{j=1}^{p}[\sqrt{n}(\hat\theta_j-\theta^*)_j]^2\Big)^{\frac{k}{2}}$$
$$= \frac{\|\sqrt{n}(\hat\theta-\theta^*)\|_2^k}{\lambda_{\max}^{k/2}},$$

由此可得 $\|\sqrt{n}(\hat\theta-\theta^*)\|_2^k \lesssim (\sqrt{\lambda_{\max}})^k\|\Lambda^{-\frac{1}{2}}\sqrt{n}(\hat\theta-\theta^*)\|_2^k$。根据 $\|\Lambda^{-\frac{1}{2}}\sqrt{n}(\hat\theta-\theta^*)\|_2^k$ 在 $f_{0,n}$-概率意义下有界这一事实，以及假设 $\alpha_n\to 0$，我们得到如下结论：对于任意 $\epsilon>0$ 以及对于任意 $M_0>0$，都存在一个整数 $N_1\equiv N(M_0,\epsilon,k)$，使得只要 $n>N_1$，

$$\mathbb{P}_{f_{0,n}}\Big((2\sqrt{\alpha_n})^k\|\sqrt{n}(\hat\theta-\theta^*)\|_2^k > M_0\Big) \le \mathbb{P}_{f_{0,n}}\Big((2\sqrt{\lambda_{\max}\alpha_n})^k\|\Lambda^{-\frac{1}{2}}\sqrt{n}(\hat\theta-\theta^*)\|_2^k > M_0\Big) \le \epsilon. \tag{115}$$

利用这一点以及式 (114) 中的界，令 $M$ 为一个大于式 (114) 中常数的常数，则只要 $n>N_1$，

$$\begin{aligned}
&\mathbb{P}_{f_{0,n}}\Big(\int_{\mathbb{R}^p}\|h\|_2^k\phi_n(h)dh > M\Big)\\
&\le \mathbb{P}_{f_{0,n}}\Big(\frac{2^{\frac{3k}{2}}p^{\frac{k}{2}-1}}{\sqrt{\pi}}\Gamma\Big(\frac{k+1}{2}\Big)\sum_{i=1}^{p}\big[[V_{\theta^*}^{-1}]_{ii}\big]^{\frac{k}{2}} + (2\sqrt{\alpha_n})^k\|\sqrt{n}(\hat\theta-\theta^*)\|_2^k > M\Big)\\
&= \mathbb{P}_{f_{0,n}}\Big((2\sqrt{\alpha_n})^k\|\sqrt{n}(\hat\theta-\theta^*)\|_2^k > M - \frac{2^{\frac{3k}{2}}p^{\frac{k}{2}-1}}{\sqrt{\pi}}\Gamma\Big(\frac{k+1}{2}\Big)\sum_{i=1}^{p}\big[[V_{\theta^*}^{-1}]_{ii}\big]^{\frac{k}{2}}\Big) < \epsilon.
\end{aligned} \tag{116}$$

□

**引理 6.** 设随机变量 $Z\overset{d}{\equiv}Y|X^n$ 在 $\mathbb{R}^p$ 上具有密度 $f_Z(\cdot)$。对于某个 $k\ge 0$，假设存在一个 $\gamma>0$，使得 $\mathbb{E}[\|Z\|_2^{k(1+\gamma)}] = O_{f_{0^n}}(1)$。那么，对于任意 $\epsilon>0$、任意 $M>0$ 以及任意 $r_n\to\infty$，存在一个整数 $N>0$，使得 $\mathbb{P}_{f_{0,n}}(\int_{\bar B_0(r_n)^c}\|z\|_2^k f_Z(z)dz > M) < \epsilon$ 对所有 $\alpha_n n > N$ 均成立。

证明。引理 6 的证明几乎可以逐字照搬 [69, 引理 2] 的证明。令 $f_{\|Z\|_2^k}(\cdot)$ 表示 $\|Z\|_2^k$ 的密度函数。定义随机变量

$$V \equiv \|Z\|_2^k \mathbb{1}\{\|Z\|_2 > r_n\} = \begin{cases} 0 & \text{if } \|Z\|_2 \le r_n, \\ \|Z\|_2^k & \text{if } \|Z\|_2 > r_n. \end{cases}$$

在证明的其余部分，我们取 $\alpha_n n > N_0$，其中 $N_0$ 是一个足够大的整数，使得对所有 $\alpha_n n > N_0$，$r_n > 0$。因此，$V$ 的密度为

$$f_V(v) = \begin{cases} \mathbb{P}(\|Z\|_2 \le r_n) & \text{if } v = 0, \\ f_{\|Z\|_2^k}(v) & \text{if } v > r_n^k. \end{cases}$$

因此，我们可以将 $V$ 的期望表示为

$$\mathbb{E}[V] = \int_0^{r_n^k} \mathbb{P}(V > t)dt + \int_{r_n^k}^{\infty} \mathbb{P}(V > t)dt. \tag{117}$$

利用 $V$ 的定义及其密度函数，我们研究（117）。考虑右侧的第一项，

$$\begin{aligned} \int_0^{r_n^k} \mathbb{P}(V > t)dt = \int_0^{r_n^k} \int_t^{\infty} f_V(v)dvdt &\overset{(a)}{=} \int_0^{r_n^k} \mathbb{P}(\|Z\|_2^k > r_n^k)dt \\ &= r_n^k \mathbb{P}(\|Z\|_2^k > r_n^k) \overset{(b)}{\le} \frac{\mathbb{E}[\|Z\|_2^{k(1+\gamma)}]}{r_n^{k\gamma}}, \end{aligned} \tag{118}$$

其中，$\gamma$ 满足 $\mathbb{E}[\|Z\|_2^{k(1+\gamma)}] = O_{f_0,\, n}(1)$。式（118）中的步骤（$a$）成立，是因为对于 $0 < t < r_n^k$，我们有 $\int_t^{\infty} f_V(v)dv = \int_{r_n^k}^{\infty} f_{\|Z\|_2^k}(v)dv = \mathbb{P}(\|Z\|_2^k > r_n^k)$，而步骤（$b$）则由马尔可夫不等式得出。对于（117）右侧的第二项，我们再次使用马尔可夫不等式：

$$\int_{r_n^k}^{\infty} \mathbb{P}(V > t)dt \le \int_{r_n^k}^{\infty} \frac{\mathbb{E}[V^{1+\gamma}]}{t^{1+\gamma}}dt \overset{(c)}{\le} \int_{r_n^k}^{\infty} \frac{\mathbb{E}[\|Z\|_2^{k(1+\gamma)}]}{t^{1+\gamma}}dt = \frac{\mathbb{E}[\|Z\|_2^{k(1+\gamma)}]}{\gamma r_n^{k\gamma}}. \tag{119}$$

（$c$）这一步由 (119) 得出，因为 $V^{1+\gamma} = (\|Z\|_2^k \mathbb{1}\{\|Z\|_2 > r_n\})^{1+\gamma} \le \|Z\|_2^{k(1+\gamma)}$。

将 (118) 和 (119) 代入 (117)，可得

$$\int_{\bar{B}_0(r_n)^c} \|z\|_2^k f_Z(z)dz = \mathbb{E}[V] \le \frac{\gamma+1}{\gamma r_n^{k\gamma}} \mathbb{E}\left[\|Z\|_2^{k(1+\gamma)}\right]. \tag{120}$$

根据式（120），我们有 $\mathbb{P}_{f_0,\, n}(\int_{\bar{B}_0\,(r_n)^{\,c}} \|z\|_2^k f_Z(z)dz > M) \le \mathbb{P}_{f_0,\, n}(\frac{\gamma+1}{\gamma r_n^{k\gamma}} \mathbb{E}[\|Z\|_2^{k(1+\gamma)}] > M)$ $\le \mathbb{P}_{f_0,\, n}(\frac{\gamma+1}{\gamma r_n^{k\gamma}} \mathbb{E}[\|Z\|_2^{k(1+\gamma)}] > M)$。由于根据假设，$r_n \to \infty$ and $\mathbb{E}[\|Z\|_2^{k(1+\gamma)}] = O_{f_0,\, n}(1)$，对于所有 $\epsilon > 0$，存在 $N \equiv N(\epsilon,\ k,\ \gamma) > N_0$，使得只要 $\alpha_n n > N$，上述右侧就被 $\epsilon$ 所上界。因此，我们得到了所需结果。 □

## B.1 辅助引理

**引理 7.** 给定向量 $v \in \mathbb{R}^p$，对于式（5）中定义的 $\|v^{\otimes k}\|_1$，有 $\|v^{\otimes k}\|_1 \le p^{\frac{k}{2}} \|v\|_2^k$。

证明。首先，回顾 $\|v\|_1 \le \sqrt{p}\|v\|_2$，这是由柯西-施瓦茨不等式得出的。现在，回顾式 (5) 中 $\|v^{\otimes k}\|_1$ 的定义，可得 $\|v^{\otimes k}\|_1 = \sum_{i_1=1}^p \cdots \sum_{i_k=1}^p |v_{i_1} \times \cdots \times v_{i_k}| \le \sum_{i_1=1}^p |v_{i_1}| \cdots \sum_{i_k=1}^p |v_{i_k}| = \|v\|_1^k \le p^{\frac{k}{2}} \|v\|_2^k$. □

众所周知，[75] 的高斯集中不等式可推出以下结果（另见 [13]）。

**引理 8。** 设 $X \sim \mathcal{N}(0, \Sigma)$。则，对所有 $s > 0$，$\mathbb{P}(\|X\|_2 > s) \le \exp\left(\frac{-(s-\sqrt{\mathrm{tr}(\Sigma)})^2}{2\|\Sigma\|_{\mathrm{op}}}\right)$。

以下凸性引理直接源于 $W_p^p(\mu, \nu)$ 的对偶形式[19,定理 1。20]

**引理 9.** 设 $q \in [0, 1]$。对于概率测度 $\mu_q = q\mu_1 + (1-q)\mu_2$ 和 $\nu_q = q\nu_1 + (1-q)\nu_2$，有 $d_{\mathrm{W}_p}(\mu_q, \nu_q)^p \le q d_{\mathrm{W}_p}(\mu_1, \nu_1)^p + (1-q) d_{\mathrm{W}_p}(\mu_2, \nu_2)^p$.

**引理 10.** [23, 推论 2.51] 若 $f$ 是定义在 $\mathbb{R}^p$ 上的非负或可积勒贝格可测函数，且满足 $f(x) = g(|x|)$，其中 $g$是定义在 $(0, \infty)$ 上的函数，则 $\int_{\mathbb{R}^p} f(x)dx = \mathrm{vol}(\mathbb{S}^{p-1}) \int_0^\infty \rho^{p-1} g(\rho) d\rho$.

# C 验证主要定理的条件

## C.1 假设 (A2) 的充分条件

命题 4 中的条件是通常的正则性条件，也蕴含 $\sqrt{n}$- 随机局部渐近正态性 [76, 章 7]，并且对于独立同分布数据可以放宽：参见 [50,引理 2.1-2.2],其中给出了确保存在满足 (121) 的 $\Delta_{n,\theta^*}$ 和 $V_{\theta^*}$ 的充分条件。

**命题 4.** 假设存在一个 $r_0 > 0$，使得在 $B_{\theta^*}(r_0)$ 上满足以下条件：

1. 对数似然 $\log f_n(X^n|\theta)$ 关于 $\theta$ 三次连续可微。

2. 令 $\frac{1}{n}\nabla^3 \log f_n(X^n|\theta) \in \mathbb{R}^{p\times p\times p}$ 为 $\log f_n(X^n|\theta)$ 的三阶偏导数张量。存在一个函数 $\tilde{M}: \mathcal{X}^n \to \mathbb{R}$，其中 $\mathbb{E}_{f_{0,n}} |\frac{1}{n}\tilde{M}(X^n)| < \infty$，使得 $\nabla^3 \log f_n(X^n|\theta)$ 的各项满足 $|[\nabla^3 \log f_n(X^n|\theta)]_{i,j,k}| = |\frac{\partial^3}{\partial\theta_i \partial\theta_j \partial\theta_k} \log f_n(X^n|\theta)| < \tilde{M}(X^n)$。

3. 令 $\nabla \log f_n(X^n|\theta)$ 和 $\nabla^2 \log f_n(X^n|\theta)$ 分别表示 $\log f_n(X^n|\theta)$ 的梯度和海森矩阵。假设在 $f_{0,n}$-概率下，

$$\frac{1}{\sqrt{n}} \nabla \log f_n(X^n|\theta^*) - V_{\theta^*}\Delta_{n,\theta^*} \to 0, \quad \text{and} \quad \frac{1}{n}\nabla^2 \log f_n(X^n|\theta^*) + V_{\theta^*} \to 0, \quad (121)$$

对于某个向量 $\Delta_{n,\theta^*} \in \mathbb{R}^p$ 以及一个正定矩阵 $V_{\theta^*} \in \mathbb{R}^{p\times p}$。

在上述条件下，对于式 (16) 中定义的 $R_n(h)$，对于所有 $M>0$、$r>0$ 和 $\epsilon>0$，存在一个整数 $N(M,r,\epsilon)$，使得对于所有 $\alpha_n n>N$，

$$\mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\bar{B}_0(r)}|R_n(h)|>M\Big)<\epsilon. \tag{122}$$

证明。证明分为三个步骤。首先，我们利用条件 3，对 $\frac{1}{n}\log f_n(X^n|\theta^*+h/\sqrt{\alpha_n n})$ 在 $\frac{1}{n}\log f_n(X^n|\theta^*)$ 处进行三阶泰勒展开。其次，我们研究泰勒展开的余项在形如 $B_0(r)$ 的集合上的收敛性，其中任意 $r>0$ 均适用。最后，我们结合这些结果，得到 (122) 中所述的 (**A2**) 条件。

**步骤 1：对数似然的泰勒展开。**设 $g(t)\equiv\frac{1}{n}\log f_n(X^n|u(t))$，其中 $u(t)\equiv\theta^*+th/\sqrt{\alpha_n n}$。对 $g(1)$ 在 $g(0)$ 处进行二阶泰勒展开可得

$$\begin{aligned}g(1)&=g(0)+g'(0)(1-0)+\frac{1}{2}g''(0)(1-0)^2+\frac{1}{2}\int_0^1(1-t)^2g'''(t)dt\\&=g(0)+g'(0)+\frac{1}{2}g''(0)+\frac{1}{2}\int_0^1(1-t)^2g'''(t)dt.\end{aligned}\tag{123}$$

注意，在 $g(1)$ 处对 $h/\sqrt{\alpha_n n}$ 展开 $\frac{1}{n}f_n(X^n|\theta^*)$ 与在 $g(0)$ 处对 $\theta^*$ 展开 $\frac{1}{n}f_n(X^n|\theta^*)$，等价于在 $\log f_n(X^n|\theta^*+h/\sqrt{\alpha_n n})$ 处对 $\log f_n(X^n|\theta^*)$ 进行展开，因为

$$g(1)=\frac{1}{n}\log f_n\Big(X^n\Big|\theta^*+\frac{h}{\sqrt{\alpha_n n}}\Big)\quad\text{and}\quad g(0)=\frac{1}{n}\log f_n(X^n|\theta^*).\tag{124}$$

此外，

$$\begin{aligned}&g'(0)=\frac{h^\top}{\sqrt{\alpha_n n}}\Big[\frac{1}{n}\nabla\log f_n(X^n|\theta^*)\Big],\qquad g''(0)=\frac{h^\top}{\sqrt{\alpha_n n}}\Big[\frac{1}{n}\nabla^2\log f_n(X^n|\theta^*)\Big]\frac{h}{\sqrt{\alpha_n n}},\\&g'''(t)=\Big\langle\frac{1}{n}\nabla^3\log f_n\Big(X^n\Big|\theta^*+\frac{th}{\sqrt{\alpha_n n}}\Big),\Big(\frac{h}{\sqrt{\alpha_n n}}\Big)^{\otimes 3}\Big\rangle,\end{aligned}\tag{125}$$

其中，我们用式 (4) 中定义的张量积记法表示 $g'''(t)$。将 (124) 和 (125) 代入 (123)，得到

$$\begin{aligned}\frac{1}{n}\log f_n\Big(X^n\Big|\theta^*+\frac{h}{\sqrt{\alpha_n n}}\Big)=&\frac{1}{n}\log f_n(X^n|\theta^*)+\frac{h^\top}{\sqrt{\alpha_n n}}\Big[\frac{1}{n}\nabla\log f_n(X^n|\theta^*)\Big]\\&+\frac{1}{2}\frac{h^\top}{\sqrt{\alpha_n n}}\Big[\frac{1}{n}\nabla^2\log f_n(X^n|\theta^*)\Big]\frac{h}{\sqrt{\alpha_n n}}+\mathrm{rem}_n(h),\end{aligned}\tag{126}$$

其中，$\mathrm{rem}_n(h)$ 是泰勒近似的余项，其表达式为

$$\mathrm{rem}_n(h)=\frac{1}{2}\int_0^1(1-t)^2g'''(t)dt=\frac{1}{2(\alpha_n n)^{\frac{3}{2}}}\int_0^1(1-t)^2\Big\langle\frac{1}{n}\nabla^3\log f_n\Big(X^n\Big|\theta^*+\frac{th}{\sqrt{\alpha_n n}}\Big),h^{\otimes 3}\Big\rangle dt.$$

于是，根据假设 (A2) 中 $R_n(h)$ 的定义以及 (126) 中的泰勒近似，

$$
\begin{aligned}
|R_n(h)| &= \Big|\alpha_n\Big[\log f_n\big(X^n\big|\theta^* + \frac{h}{\sqrt{\alpha_n n}}\big) - \log f_n(X^n|\theta^*)\Big] - \sqrt{\alpha_n}h^\top V_{\theta^*}\Delta_{n,\theta^*} + \frac{1}{2}h^\top V_{\theta^*}h\Big| \\
&= \Big|\alpha_n n\Big(\frac{h^\top}{\sqrt{\alpha_n n}}\Big[\frac{1}{n}\nabla \log f_n(X^n|\theta^*)\Big] + \frac{1}{2}\frac{h^\top}{\sqrt{\alpha_n n}}\Big[\frac{1}{n}\nabla^2 \log f_n(X^n|\theta^*)\Big]\frac{h}{\sqrt{\alpha_n n}} + \mathrm{rem}_n(h)\Big) \\
&\quad - \sqrt{\alpha_n}h^\top V_{\theta^*}\Delta_{n,\theta^*} + \frac{1}{2}h^\top V_{\theta^*}h\Big| \\
&\le \Big|\sqrt{\alpha_n}h^\top\Big[\frac{1}{\sqrt{n}}\nabla \log f_n(X^n|\theta^*) - V_{\theta^*}\Delta_{n,\theta^*}\Big] + \frac{h^\top}{2}\Big[\frac{1}{n}\nabla^2 \log f_n(X^n|\theta^*) + V_{\theta^*}\Big]h\Big| \\
&\quad + |\alpha_n n \mathrm{rem}_n(h)|.
\end{aligned}
\tag{127}
$$

在继续进行步骤 2 之前，我们先研究 (127) 的第一项。特别地，我们将证明，对于所有 $\epsilon > 0$、$M > 0$ 以及任意 $r > 0$，存在 $N_0 \equiv N_0(\epsilon, r, M)$，使得对于 $\alpha_n n > N_0$，

$$
\mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\bar{B}_0(r)}\Big|\sqrt{\alpha_n}h^\top\Big[\frac{\nabla \log f_n(X^n|\theta^*)}{\sqrt{n}} - V_{\theta^*}\Delta_{n,\theta^*}\Big] + \frac{h^\top}{2}\Big[\frac{\nabla^2 \log f_n(X^n|\theta^*)}{n} + V_{\theta^*}\Big]h\Big| > \frac{M}{2}\Big) < \frac{\epsilon}{2}.
\tag{128}
$$

令 $T_1(h) := \sqrt{\alpha_n}h^\top[\frac{1}{\sqrt{n}}\nabla \log f_n(X^n|\theta^*) - V_{\theta^*}\Delta_{n,\theta^*}]$，且 $T_2(h) := \frac{1}{2}h^\top[\frac{1}{n}\nabla^2 \log f_n(X^n|\theta^*) + V_{\theta^*}]h$。我们首先注意到，

$$
\sup_h |T_1(h) + T_2(h)| \le \sup_h |T_1(h)| + \sup_h |T_2(h)| \le 2\max\Big\{\sup_h |T_1(h)|, \sup_h |T_2(h)|\Big\}.
\tag{129}
$$

现在，为了证明 (128)，{ $\sup_h |T_1(h)| > \frac{M}{4}$ } ∪ { $\sup_h |T_2(h)| > \frac{M}{4}$ }. 129 注意到 max sup (), sup () if sup () sup () 因此，根据 () 和并集界，我们有

$$
\begin{aligned}
&\mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\bar{B}_0(r)}|T_1(h) + T_2(h)| > \frac{M}{2}\Big) \\
&\qquad \le \mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\bar{B}_0(r)}|T_1(h)| > \frac{M}{4}\Big) + \mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\bar{B}_0(r)}|T_2(h)| > \frac{M}{4}\Big).
\end{aligned}
\tag{130}
$$

我们使用柯西-施瓦茨不等式对（130）中的第一项进行上界估计如下：

$$
\begin{aligned}
\mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\bar{B}_0(r)}|T_1(h)| > \frac{M}{4}\Big) &= \mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\bar{B}_0(r)}\Big|\sqrt{\alpha_n}h^\top\Big[\frac{1}{\sqrt{n}}\nabla \log f_n(X^n|\theta^*) - V_{\theta^*}\Delta_{n,\theta^*}\Big]\Big| > \frac{M}{4}\Big) \\
&\le \mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\bar{B}_0(r)}\sqrt{\alpha_n}\|h\|_2\Big\|\frac{1}{\sqrt{n}}\nabla \log f_n(X^n|\theta^*) - V_{\theta^*}\Delta_{n,\theta^*}\Big\|_2 > \frac{M}{4}\Big) \\
&= \mathbb{P}_{f_{0,n}}\Big(\sqrt{\alpha_n}\Big\|\frac{1}{\sqrt{n}}\nabla \log f_n(X^n|\theta^*) - V_{\theta^*}\Delta_{n,\theta^*}\Big\|_2 > \frac{M}{4r}\Big).
\end{aligned}
\tag{131}
$$

式（131）中的等式源于对 $B_0(r)$ 取上确界。根据条件 3 以及 $\alpha_n \to 0$ 这一假设，对于所有 $\epsilon > 0$、$M > 0$ 和 $r > 0$，存在

一个整数 $N_{01} \equiv N_{01}(\epsilon,\ r,\ M)$，使得对于 $\alpha_n n > N_{01}$，式（131）中的概率上界为 $\epsilon/4$。

现在我们对式 (130) 的第二项作上界估计。首先，根据引理 7，我们有

$$\begin{aligned}
2 \sup_{h\in\bar{B}_0(r)} |T_2(h)| &= \sup_{h\in\bar{B}_0(r)} \Big| h^\top \Big[\frac{1}{n}\nabla^2 \log f_n(X^n|\theta^* + V_{\theta^*}\Big] h \Big| \\
&= \sup_{h\in\bar{B}_0(r)} \Big| \sum_{i,j} \Big[\frac{1}{n}\nabla^2 \log f_n(X^n|\theta^*) + V_{\theta^*}\Big]_{ij} h_i h_j \Big| \\
&\le \max_{1\le i,j\le p} \Big| \Big[\frac{1}{n}\nabla^2 \log f_n(X^n|\theta^*) + V_{\theta^*}\Big]_{ij} \Big| \sup_{h\in\bar{B}_0(r)} \sum_{i,j} |h_i h_j| \\
&\le \max_{1\le i,j\le p} \Big| \Big[\frac{1}{n}\nabla^2 \log f_n(X^n|\theta^*) + V_{\theta^*}\Big]_{ij} \Big| \sup_{h\in\bar{B}_0(r)} p\|h\|_2^2 \\
&= pr^2 \max_{1\le i,j\le p} \Big| \Big[\frac{1}{n}\nabla^2 \log f_n(X^n|\theta^*) + V_{\theta^*}\Big]_{ij} \Big|.
\end{aligned}$$

因此，我们有

$$\mathbb{P}_{f_{0,n}}\Big( \sup_{h\in\bar{B}_0(r)} |T_2(h)| > \frac{M}{4} \Big) \le \mathbb{P}_{f_{0,n}}\Big( \max_{1\le i,j\le p} \Big| \Big[\frac{1}{n}\nabla^2 \log f_n(X^n|\theta^*) + V_{\theta^*}\Big]_{ij} \Big| > \frac{M}{2pr^2} \Big). \quad (132)$$

根据条件 3，对于所有 $\epsilon > 0$、$M > 0$ 和 $r > 0$，存在一个整数 $N_{02} \equiv N_{02}(\epsilon, p, r, M)$，使得对于 $\alpha_n n > N_{02}$，式 (132) 中的概率有上界 $\epsilon/4$。结合式 (131) 和式 (132) 的结果，我们得到当 $\alpha_n n > N_0 \equiv \max(N_{01}, N_{02})$ 时，式 (128) 中的结论成立。

**步骤 2：对 $\alpha_n \mathrm{rem}_n(h)$ 在 $h \in B_0(r)$ 上的分析。** 在此步骤中，我们验证对于所有 $\epsilon > 0$、$r > 0$ 和 $M > 0$，且当 $\alpha_n n$ 足够大时，

$$\mathbb{P}_{f_{0,n}}\Big( \sup_{h\in\bar{B}_0(r)} |\alpha_n n \mathrm{rem}_n(h)| > \frac{M}{2} \Big) < \frac{\epsilon}{2}. \quad (133)$$

首先，我们说明当 $\alpha_n n$ 足够大时，$\mathrm{rem}_n(h)$ 在 $B_0(r)$ 上是良定义的。也就是说，我们将证明，对于任意 $h \in B_0(r)$，向量 $\theta^* + th/\sqrt{\alpha_n n}$ 属于 $\theta^*$ 的邻域，其中 $\nabla^3 \log f_n(X^n|\theta^* + th/\sqrt{\alpha_n n})$ 是连续的。根据假设，存在 $r_0 > 0$，使得 $\log f_n(X^n|\theta)$ 在 $B_{\theta^*}(r_0)$ 上具有连续的三阶导数。对于任意 $r > 0$，存在 $N_1 \equiv N_1(r_0, r) = \lceil r^2/r_0^2 \rceil$，使得只要 $h \in \bar{B}_0(r)$ 且 $\alpha_n n \ge N_1$，就有 $\theta^* + th/\sqrt{\alpha_n n} \in \bar{B}_{\theta^*}(r_0)$。为说明这一点，注意到如果 $\|h\|_2 \le r$ 且 $\alpha_n n \ge r^2/r_0^2$，那么对于 $0 \le t \le 1$，

$$\Big\| \theta^* + \frac{th}{\sqrt{\alpha_n n}} - \theta^* \Big\|_2 \le \frac{tr}{\sqrt{r^2/r_0^2}} < t r_0 \le r_0. \quad (134)$$

因此，对于任意 $h \in B_0(r)$，当 $\alpha_n n$ 足够大时，余项 $\mathrm{rem}_n(h)$ 是良定义的。

接下来，我们可以利用式 (4) 中张量内积的定义来界定 $\alpha_n n\mathrm{rem}_n(h)$:

$$
\begin{aligned}
|\alpha_n n\mathrm{rem}_n(h)| &= \Big|\frac{1}{2\sqrt{\alpha_n n}}\int_0^1 (1-t)^2\Big\langle \frac{1}{n}\nabla^3 \log f_n\big(X^n\big|\theta^* + \frac{th}{\sqrt{\alpha_n n}}\big), h^{\otimes 3}\Big\rangle dt\Big| \\
&\le \frac{\|h^{\otimes 3}\|_1}{2\sqrt{\alpha_n n}}\int_0^1 (1-t)^2 \max_{1\le i,j,k\le p}\Big|\Big[\frac{1}{n}\nabla^3 \log f_n\big(X^n\big|\theta^* + \frac{th}{\sqrt{\alpha_n n}}\big)\Big]_{i,j,k}\Big| dt \\
&\le \frac{\|h^{\otimes 3}\|_1}{2\sqrt{\alpha_n n}} \sup_{t'\in[0,1]} \max_{1\le i,j,k\le p}\Big|\Big[\frac{1}{n}\nabla^3 \log f_n\big(X^n\big|\theta^* + \frac{t'h}{\sqrt{\alpha_n n}}\big)\Big]_{i,j,k}\Big| \int_0^1 (1-t)^2 dt \\
&\le \frac{p^{\frac{3}{2}}\|h\|_2^3}{6\sqrt{\alpha_n n}} \sup_{t\in[0,1]} \max_{1\le i,j,k\le p}\Big|\Big[\frac{1}{n}\nabla^3 \log f_n\big(X^n\big|\theta^* + \frac{th}{\sqrt{\alpha_n n}}\big)\Big]_{i,j,k}\Big|. \qquad (135)
\end{aligned}
$$

最后一个不等式由引理 7 并通过计算积分得到。利用上述结果,

$$
\begin{aligned}
\sup_{h\in\bar{B}_0(r)} |\alpha_n n\mathrm{rem}_n(h)| &\le \sup_{h\in\bar{B}_0(r)} \frac{p^{\frac{3}{2}}\|h\|_2^3}{6\sqrt{\alpha_n n}} \sup_{t\in[0,1]} \max_{1\le i,j,k\le p}\Big|\Big[\frac{1}{n}\nabla^3 \log f_n\big(X^n\big|\theta^* + \frac{th}{\sqrt{\alpha_n n}}\big)\Big]_{i,j,k}\Big| \\
&\le \frac{p^{\frac{3}{2}}r^3}{6\sqrt{\alpha_n n}} \sup_{\substack{h\in\bar{B}_0(r)\\ t\in[0,1]}} \max_{1\le i,j,k\le p}\Big|\Big[\frac{1}{n}\nabla^3 \log f_n\big(X^n\big|\theta^* + \frac{th}{\sqrt{\alpha_n n}}\big)\Big]_{i,j,k}\Big| \\
&\le \frac{p^{\frac{3}{2}}r^3}{6\sqrt{\alpha_n n}} \max_{\theta\in\bar{B}_{\theta^*}(r_0)} \max_{1\le i,j,k\le p}\Big|\Big[\frac{1}{n}\nabla^3 \log f_n(X^n|\theta)\Big]_{i,j,k}\Big|. \qquad (136)
\end{aligned}
$$

(136) 中的不等式是通过注意到：根据 (134) 中的论证，当 $\alpha_n n \ge N_1$ 时， $h \in B_0(r)$ 蕴含 $\theta^* + th/\sqrt{\alpha_n n} \in B_{\theta^*}(r_0)$；此外， $\nabla^3 \log f_n(X^n|\theta)$ 在紧集 $B_{\theta^*}(r_0)$ 上的连续性使我们可以将 $\sup_{\theta\in\bar{B}_{\theta^*}(r_0)}$ 改为 $\max_{\theta\in\bar{B}_{\theta^*}(r_0)}$。

取 $\alpha_n n \ge N_1$，我们对 (133) 中的概率应用 (136) 和马尔可夫不等式:

$$
\begin{aligned}
&\mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\bar{B}_0(r)} |\alpha_n n\mathrm{rem}_n(h)| > \frac{M}{2}\Big) \\
&\le \mathbb{P}_{f_{0,n}}\Big(\frac{p^{\frac{3}{2}}r^3}{6\sqrt{\alpha_n n}} \max_{\theta\in\bar{B}_{\theta^*}(r_0)} \max_{1\le i,j,k\le p}\Big|\Big[\frac{1}{n}\nabla^3 \log f_n(X^n|\theta)\Big]_{i,j,k}\Big| > \frac{M}{2}\Big) \\
&\le \frac{p^{\frac{3}{2}}r^3}{3M\sqrt{\alpha_n n}}\mathbb{E}_{f_{0,n}}\Big[\max_{\theta\in\bar{B}_{\theta^*}(r_0)} \max_{i,j,k}\Big|\Big[\frac{1}{n}\nabla^3 \log f_n(X^n\big|\theta)\Big]_{i,j,k}\Big|\Big] \\
&\le \frac{p^{\frac{3}{2}}r^3}{3M\sqrt{\alpha_n n}}\mathbb{E}_{f_{0,n}}\Big[\Big|\frac{1}{n}\tilde{M}(X^n)\Big|\Big] < \frac{\epsilon}{2}. \qquad (137)
\end{aligned}
$$

上述内容中，(137) 中的第一个不等式由条件2的假设得出，第二个不等式则通过进一步取 $\alpha_n n > \max(N_1, N_2)$ 得出，其中

$$
N_2 \equiv N_2(M, r, \epsilon) = \Big\lceil \frac{4p^3 r^6}{9M^2\epsilon^2}\mathbb{E}_{f_{0,n}}\Big[\Big|\frac{1}{n}\tilde{M}(X^n)\Big|\Big]^2\Big\rceil.
$$

注意到 $N_2$ 存在（即 $N_2$ 是有限的），因为根据假设，$\mathbb{E}_{f_{0,n}}[|\frac{1}{n}\tilde{M}(X^n)|]$ 是有限的。

**步骤 3：最终界。** 我们结合 (127)、(128) 和 (133) 中的结果，得到 (122)，如下所示。考虑 (128)，令

$$f(h) \equiv \Big|\sqrt{\alpha_n} h^\top \Big[\frac{\nabla \log f_n(X^n|\theta^*)}{\sqrt{n}} - V_{\theta^*}\Delta_{n,\theta^*}\Big] + \frac{h^\top}{2}\Big[\frac{\nabla^2 \log f_n(X^n|\theta^*)}{n} + V_{\theta^*}\Big]h\Big|.$$

然后，对于任意 $M > 0$、 $r > 0$ 和 $\epsilon > 0$，我们取 $\alpha_n n > \max(N_0, N_1, N_2)$，应用（127）中的界，得到如下结果：

$$\begin{aligned}
&\mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\bar{B}_0(r)} |R_n(h)| > M\Big) \le \mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\bar{B}_0(r)} f(h) + \sup_{h\in\bar{B}_0(r)} |\alpha_n n \mathrm{rem}_n(h)| > M\Big)\\
&\le \mathbb{P}_{f_{0,n}}\Big(\max\Big\{\sup_{h\in\bar{B}_0(r)} f(h), \sup_{h\in\bar{B}_0(r)} |\alpha_n n \mathrm{rem}_n(h)|\Big\} > \frac{M}{2}\Big)\\
&= \mathbb{P}_{f_{0,n}}\Big(\Big\{\sup_{h\in\bar{B}_0(r)} f(h) > \frac{M}{2}\Big\} \cup \Big\{\sup_{h\in\bar{B}_0(r)} |\alpha_n n \mathrm{rem}_n(h)| > \frac{M}{2}\Big\}\Big)\\
&\le \mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\bar{B}_0(r)} f(h) > \frac{M}{2}\Big) + \mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\bar{B}_0(r)} |\alpha_n n \mathrm{rem}_n(h)| > \frac{M}{2}\Big) \le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \qquad (138)
\end{aligned}$$

第二个不等式 （138）中的不等式由（128）和（133）推出。至此，证明完毕 f.

□

## C.2 假设（A3）的充分条件

下一个命题 表明，条件 1 和 2 蕴含（**A3**），即使先验分布不适定 s.

**命题 5。** 假设 (**A0**)–(**A1**) 成立，并假设 $\pi(\theta) \le \kappa$ 对某个 $\kappa > 0$成立。此外，假设条件 1 和 2 成立。则假设 (**A3**) 成立；也就是说，对于所有 $\epsilon > 0$ 和所有 $k_0 \ge 0$，存在常数 $M < \infty$ 和整数 $N$，使得对于 $\alpha_n n \ge N$，有 $\mathbb{P}_{f_{0,n}}(\mathbb{E}_{\underline{\pi_{n,\alpha_n}}(\theta|X^n)}[\|\sqrt{\alpha_n n}(\theta-\theta^*)\|_2^{k_0}] > M) < \epsilon$.

证明。 首先，对于条件 1 和 2 中的常数 $L, r, c_3, c_4$ 以及函数 $\gamma$，我们定义以下事件。我们任取一个常数 $M_0 > 0$，其定义将在下文给出。

$$\begin{aligned}
&\mathcal{A} = \{\log f_n(X^n|\theta) - \log f_n(X^n|\theta^*) \le -Ln\gamma(\|\theta-\theta^*\|_2) \text{ for all } \theta \in B_{\theta^*}(r)^c\},\\
&\mathcal{B} = \big\{-nc_3\|\theta-\hat{\theta}\|_2^2 \le \log f_n(X^n|\theta) - \log f_n(X^n|\hat{\theta}) \le -nc_4\|\theta-\hat{\theta}\|_2^2 \text{ for all } \theta \in B_{\hat{\theta}}(2r)\big\},\\
&\mathcal{C} = \{\|\hat{\theta}-\theta^*\|_2 < r\}, \qquad \mathcal{D} = \{\|\sqrt{n}(\hat{\theta}-\theta^*)\|_2 < M_0\}.
\end{aligned} \qquad (139)$$

令 $\epsilon > 0$ 为任意值。根据条件 1，存在 $N_0 \equiv N_0(L, c_0, c_1, c_2, r, \epsilon)$，使得只要 $\alpha_n n > N_0$，就有 $\mathbb{P}_{f_{0,n}}(\mathcal{A}^c) < \epsilon/4$。根据条件2，存在 $N_1 \equiv N_1(c_0, c_1, c_2, c_3, c_4, r, \epsilon)$，使得只要 $\alpha_n n > N_1$，就有 $\mathbb{P}_{f_{0,n}}(\mathcal{B}^c) < \epsilon/4$。根据假设（A0），存在 $N_2 \equiv N_2(r, \epsilon)$，使得只要 $\alpha_n n > N_2$，就有 $\mathbb{P}_{f_{0,n}}(\mathcal{C}^c) < \epsilon/4$；并且存在 $M_0$ 和 $N_3 \equiv N_3(M_0, \epsilon)$，使得只要 $\alpha_n n > N_3$，就有 $\mathbb{P}_{f_{0,n}}(\mathcal{D}^c) < \epsilon/4$。

令 $\mathcal{E}=\{\mathcal{A}\cap\mathcal{B}\cap\mathcal{C}\cap\mathcal{D}\}$。根据并集界，只要 $\alpha_n n>\max(N_0,N_1,N_2,N_3)$，就有 $\mathbb{P}_{f_0,n}(\mathcal{E}^c)\le\mathbb{P}_{f_0,n}(\mathcal{A}^c)+\mathbb{P}_{f_0,n}(\mathcal{B}^c)+\mathbb{P}_{f_0,n}(\mathcal{C}^c)+\mathbb{P}_{f_0,n}(\mathcal{D}^c)<\epsilon$. 因此，当 $\alpha_n n>\max(N_0,N_1,N_2,N_3)$ 时，利用针对 $k_0\ge 0$ 的界 $(a+b)^{k_0}\le 2^{k_0}(a^{k_0}+b^{k_0})$，

$$\begin{aligned}
&\mathbb{P}_{f_{0,n}}\big(\mathbb{E}_{\pi_{n,\alpha_n}(\theta|X^n)}[\|\sqrt{\alpha_n n}(\theta-\theta^*)\|_2^{k_0}]>M\big)\\
&\quad\le\mathbb{P}_{f_{0,n}}\big(\mathbb{E}_{\pi_{n,\alpha_n}(\theta|X^n)}[\|\sqrt{\alpha_n n}(\theta-\theta^*)\|_2^{k_0}]>M\cap\mathcal{E}\big)+\epsilon\\
&\quad\le\mathbb{P}_{f_{0,n}}\big(2^{k_0}\mathbb{E}_{\pi_{n,\alpha_n}(\theta|X^n)}[\|\sqrt{\alpha_n n}(\theta-\hat\theta)\|_2^{k_0}]+2^{k_0}\|\sqrt{\alpha_n n}(\hat\theta-\theta^*)\|_2^{k_0}>M\cap\mathcal{E}\big)+\epsilon\\
&\quad\le\mathbb{P}_{f_{0,n}}\big(2^{k_0}\mathbb{E}_{\pi_{n,\alpha_n}(\theta|X^n)}[\|\sqrt{\alpha_n n}(\theta-\hat\theta)\|_2^{k_0}]>M-(2\sqrt{\alpha_n}M_0)^{k_0}\cap\mathcal{E}\big)+\epsilon,
\end{aligned}\tag{140}$$

其中，最后一个不等式是由与事件 $\mathcal{D}$取交集得到的。

现在，对于任意 $\tau>0$，定义 $M'\equiv 2^{-k_0}(M-(2\tau M_0)^{k_0})$，并注意到，由于 $\alpha_n\to 0$，存在 $N_4\equiv N_4(\tau)$，使得当 $\alpha_n n>N_4$时，$\sqrt{\alpha_n}<\tau$。特别地，取足够小的 $\tau$，使得 $M'>0$。于是，当 $\alpha_n n>N_4$时，$M-(2\sqrt{\alpha_n}M_0)^{k_0}\ge M-(2\tau M_0)^{k_0}$。因此，由（140）可得，

$$\mathbb{P}_{f_{0,n}}(\mathbb{E}_{\pi_{n,\alpha_n}(\theta|X^n)}[\|\sqrt{\alpha_n n}(\theta-\theta^*)\|_2^{k_0}]>M)\le\mathbb{P}_{f_{0,n}}(\mathbb{E}_{\pi_{n,\alpha_n}(\theta|X^n)}[\|\sqrt{\alpha_n n}(\theta-\hat\theta)\|_2^{k_0}]>M'\cap\mathcal{E})+\epsilon.\tag{141}$$

接下来，我们的目标是证明存在一个足够大的 $M'$，使得

$$\mathbb{P}_{f_{0,n}}\big(\mathbb{E}_{\pi_{n,\alpha_n}(\theta|X^n)}[\|\sqrt{\alpha_n n}(\theta-\hat\theta)\|_2^{k_0}]>M'\cap\mathcal{E}\big)=0.\tag{142}$$

也就是说，在证明的剩余部分，我们将研究事件 $\mathcal{E}$ 上的 $\mathbb{E}_{\pi_{n,\alpha_n}(\theta|X^n)}[\|\sqrt{\alpha_n n}(\theta-\hat\theta)\|_2^{k_0}]$，并证明（141）右侧的概率等于 0。根据期望的尾部公式，我们有如下结果：

$$\begin{aligned}
k_0^{-1}\mathbb{E}_{\pi_{n,\alpha_n}(\theta|X^n)}[\|\sqrt{\alpha_n n}(\theta-\hat\theta)\|_2^{k_0}]&=k_0^{-1}\int_0^\infty\mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}(\|\sqrt{\alpha_n n}(\theta-\hat\theta)\|_2^{k_0}>s)ds\\
&=\int_0^\infty t^{k_0-1}\mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}(\|\sqrt{\alpha_n n}(\theta-\hat\theta)\|_2>t)dt,
\end{aligned}\tag{143}$$

其中在式 (143) 中，我们使用了变量替换 $s=t^{k_0}\implies ds=k_0t^{k_0-1}dt$。现在由式 (143) 可得，

$$\begin{aligned}
k_0^{-1}\mathbb{E}_{\pi_{n,\alpha_n}(\theta|X^n)}[\|\sqrt{\alpha_n n}(\theta-\hat\theta)\|_2^{k_0}]&=\int_0^{2r\sqrt{\alpha_n n}}t^{k_0-1}\mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}(\|\sqrt{\alpha_n n}(\theta-\hat\theta)\|_2>t)dt\\
&\quad+\int_{2r\sqrt{\alpha_n n}}^\infty t^{k_0-1}\mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}(\|\sqrt{\alpha_n n}(\theta-\hat\theta)\|_2>t)dt.
\end{aligned}\tag{144}$$

将式 (144) 右侧的项分别记为 $I_1$ 和 $I_2$。令 $M''=M'/(2k_0)$，我们证明了

$$\mathbb{P}_{f_{0,n}}\big(\mathbb{E}_{\pi_{n,\alpha_n}(\theta|X^n)}[\|\sqrt{\alpha_n n}(\theta-\hat\theta)\|_2^{k_0}]>M'\cap\mathcal{E}\big)=\mathbb{P}_{f_{0,n}}\big(I_1+I_2>M'/k_0\cap\mathcal{E}\big).\tag{145}$$

接下来，我们注意到，

$$\begin{aligned}
\mathbb{P}_{f_{0,n}}\big(I_1+I_2>M'/k_0\cap\mathcal{E}\big)&\le\mathbb{P}_{f_{0,n}}\big(\max(I_1,I_2)>M''\cap\mathcal{E}\big)\\
&=\mathbb{P}_{f_{0,n}}\big(\{I_1>M''\cup I_2>M''\}\cap\mathcal{E}\big)\\
&=\mathbb{P}_{f_{0,n}}\big(\{I_1>M''\cap\mathcal{E}\}\cup\{I_2>M''\cap\mathcal{E}\}\big)\\
&\le\mathbb{P}_{f_{0,n}}\big(I_1>M''\cap\mathcal{E}\big)+\mathbb{P}_{f_{0,n}}\big(I_2>M''\cap\mathcal{E}\big).
\end{aligned}$$

因此，由上述内容及（145）可得

$$\mathbb{P}_{f_{0,n}}\big(\mathbb{E}_{\pi_{n,\alpha_n}(\theta|X^n)}[\|\sqrt{\alpha_n n}(\theta-\hat{\theta})\|_2^{k_0}] > M' \cap \mathcal{E}\big) \le \mathbb{P}_{f_{0,n}}(I_1 > M'' \cap \mathcal{E}) + \mathbb{P}_{f_{0,n}}(I_2 > M'' \cap \mathcal{E}). \tag{146}$$

我们将证明存在某个 $M''$（更准确地说，存在某个 $M$，因为 $2k_0M'' = M' = 2^{-k_0}(M - (2\tau M_0)^{k_0})$），使得 (146) 右侧的两个概率均为 0。

**(146) 的第一项：**我们将证明，在给定 $\mathcal{E}$ 的条件下，当 $\alpha_n n$ 足够大时，(144) 中定义的 $I_1$ 有一个常数上界。因此，我们可以令 $M''$ 大于该常数，使得 $\mathbb{P}_{f_{0,n}}(I_1 > M'' \cap \mathcal{E}) = 0$。

首先，考虑 (144) 中定义的 $I_1$，我们的目标是为

$$\begin{aligned}\mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}(\|\sqrt{\alpha_n n}(\theta-\hat{\theta})\|_2 > t) &= \mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}\Big(\|\theta-\hat{\theta}\|_2 > \frac{t}{\sqrt{\alpha_n n}}\Big) \\ &= \mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}\Big(2r > \|\theta-\hat{\theta}\|_2 > \frac{t}{\sqrt{\alpha_n n}}\Big) + \mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}\Big(\|\theta-\hat{\theta}\|_2 > 2r\Big).\end{aligned}$$

因此，对于 (144) 中的 $I_1$

$$\begin{aligned}I_1 = &\int_0^{2r\sqrt{\alpha_n n}} t^{k_0-1}\mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}\Big(2r > \|\theta-\hat{\theta}\|_2 > \frac{t}{\sqrt{\alpha_n n}}\Big)dt \\ &+ \int_0^{2r\sqrt{\alpha_n n}} t^{k_0-1}\mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}\Big(\|\theta-\hat{\theta}\|_2 > 2r\Big)dt.\end{aligned} \tag{147}$$

我们将 (147) 中的各项记为 $I_{11}$ 和 $I_{12}$。接下来我们将证明，在给定 $\mathcal{E}$ 的条件下，当 $\alpha_n n$ 足够大时，$I_{11}$ 和 $I_{12}$ 有一个常数上界。

**项 $I_{11}$（(147)）：**令 $\ell_n(\theta) = \frac{1}{n}\log f_n(X^n|\theta)$。则

$$\mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}\Big(2r > \|\theta-\hat{\theta}\|_2 > \frac{t}{\sqrt{\alpha_n n}}\Big) = \frac{\int_{2r>\|\theta-\hat{\theta}\|_2>t/\sqrt{\alpha_n n}} e^{\alpha_n n(\ell_n(\theta)-\ell_n(\hat{\theta}))}\pi(\theta)d\theta}{\int e^{\alpha_n n(\ell_n(\theta)-\ell_n(\hat{\theta}))}\pi(\theta)d\theta} \equiv \frac{\text{num}}{\text{den}}. \tag{148}$$

我们将通过对分子取上界、对分母取下界，来为上述概率给出上界。在事件 $\mathcal{B}$ 上，我们有

$$\text{num} \le \int_{2r>\|\theta-\hat{\theta}\|_2>\frac{t}{\sqrt{\alpha_n n}}} e^{-\alpha_n n c_4\|\theta-\hat{\theta}\|_2^2}\pi(\theta)d\theta \le \kappa\int_{2r>\|\theta-\hat{\theta}\|_2>\frac{t}{\sqrt{\alpha_n n}}} e^{-\alpha_n n c_4\|\theta-\hat{\theta}\|_2^2}d\theta,$$

其中，第二个不等式由先验分布 $\pi(\theta) \le \kappa$ 的界得出。对上式应用变量替换 $h = \sqrt{\alpha_n n}(\theta - \hat{\theta}) \implies d\theta = (\alpha_n n)^{-p/2}dh$，可得

$$\text{num} \le \kappa(\alpha_n n)^{-\frac{p}{2}}\int_{2r\sqrt{\alpha_n n}>\|h\|_2>t} e^{-c_4\|h\|_2^2}dh \le \kappa(\alpha_n n)^{-\frac{p}{2}}\int_{\|h\|_2>t} e^{-c_4\|h\|_2^2}dh.$$

特别地，由于对于标准高斯向量 $Z \sim \mathcal{N}(0, \mathbb{I})$，有 $\int_{\|h\|_2>t}(2\pi)^{-\frac{p}{2}}(2c_4)^{\frac{p}{2}}e^{-\frac{2c_4}{2}\|h\|_2^2}dh = \mathbb{P}(\|Z\|_2 > t\sqrt{2c_4})$，根据引理 8 可得 $\int_{\|h\|_2>t}(2\pi)^{-\frac{p}{2}}(2c_4)^{\frac{p}{2}}\exp\{-\frac{2c_4}{2}\|h\|_2^2\}dh \le \exp\{-c_4(t - \sqrt{p/(2c_4)})^2\}$。因此，令 $\tilde{C} \equiv \kappa(2\pi)^{p/2}(2c_4)^{-p/2}$，上述界蕴含

$$\text{num} \le \kappa(\alpha_n n)^{-\frac{p}{2}}(2\pi)^{\frac{p}{2}}(2c_4)^{-\frac{p}{2}}e^{-c_4(t-\sqrt{p/(2c_4)})^2} \equiv (\alpha_n n)^{-\frac{p}{2}}\tilde{C}e^{-c_4(t-\sqrt{p/(2c_4)})^2}. \tag{149}$$

类似地，我们可以对分母给出下界。在此之前，注意到根据假设(**A1**)，存在 $r'>0$，使得在 $B_{\theta^*}(r')$ 上有 $\pi(\theta)>0$。此外，在事件 $\hat{}\,\hat{}\mathcal{C}$ 上，$B_{\theta^*}(r)\subseteq B_{\hat{\theta}}(2r)$。这是由三角不等式得出的，因为 $\|\theta-\theta\|_2\le\|\theta-\theta^*\|_2+\|\theta^*-\theta\|_2<r+r=2r$。因此，$\theta\in B_{\theta^*}(r)\implies\theta\in B_{\hat{\theta}}(2r)$。令 $\tilde{r}=\min(r',r)$。那么，在事件 $\mathcal{C}\cap\mathcal{B}$ 上，

$$\begin{aligned}\text{den}=\int e^{\alpha_n n(\ell_n(\theta)-\ell_n(\hat{\theta}))}\pi(\theta)d\theta&\ge\int_{B_{\theta^*}(\tilde{r})}e^{\alpha_n n(\ell_n(\theta)-\ell_n(\hat{\theta}))}\pi(\theta)d\theta\\&\ge\inf_{\theta'\in B_{\theta^*}(\tilde{r})}\pi(\theta')\int_{B_{\theta^*}(\tilde{r})}e^{\alpha_n n(\ell_n(\theta)-\ell_n(\hat{\theta}))}d\theta\\&\ge L(\theta^*,\tilde{r})\int_{B_{\theta^*}(\tilde{r})}e^{-\alpha_n n c_3\|\theta-\hat{\theta}\|_2^2}d\theta,\end{aligned}$$

其中 $L(\theta^*,\tilde{r})\equiv\inf_{\theta\in B_{\theta^*}(\tilde{r})}\pi(\theta)$，根据假设 (**A1**) 且由于 $\tilde{r}\le r'$，该值严格为正。由于积分也为正，num/den 的比值是良定义的（即分母不为 0）。应用变量替换 $h=\sqrt{\alpha_n n}(\theta-\theta^*)$，可得 $\theta=(\alpha_n n)^{-1/2}h+\theta^*$。因此，$\theta\in B_{\theta^*}(\tilde{r})\implies(\alpha_n n)^{-1/2}h+\theta^*\in B_{\theta^*}(\tilde{r})\implies h\in B_0(\tilde{r}\sqrt{\alpha_n n})$。由此，通过变量替换，我们得到

$$\begin{aligned}\text{den}\ge L(\theta^*,\tilde{r})\int_{B_{\theta^*}(\tilde{r})}e^{-\alpha_n n c_3\|\theta-\hat{\theta}\|_2^2}d\theta&=L(\theta^*,\tilde{r})(\alpha_n n)^{-\frac{p}{2}}\int_{B_0(\tilde{r}\sqrt{\alpha_n n})}e^{-\alpha_n n c_3\left\|\theta^*+\frac{h}{\sqrt{\alpha_n n}}-\hat{\theta}\right\|_2^2}dh\\&=L(\theta^*,\tilde{r})(\alpha_n n)^{-\frac{p}{2}}\int_{B_0(\tilde{r}\sqrt{\alpha_n n})}e^{-c_3\left\|h+\sqrt{\alpha_n n}(\theta^*-\hat{\theta})\right\|_2^2}dh.\end{aligned}$$

注意，由于对于 $p\ge0$，$(a+b)^p\le2^p(a^p+b^p)$，我们有 $\|h+\sqrt{\alpha_n n}(\theta^*-\hat{\theta})\|_2^2\le4\|h\|_2^2+4\|\sqrt{\alpha_n n}(\theta^*-\hat{\theta})\|_2^2$。接下来，注意在事件 $\mathcal{D}$ 上，我们有 $\|\sqrt{n}(\theta^*-\hat{\theta})\|_2^2<M_0^2$，而对于 $\alpha_n n>N_4$，我们有 $\alpha_n<\tau^2$。因此，我们可以进一步对上式进行界定

$$\left\|h+\sqrt{\alpha_n n}(\theta^*-\hat{\theta})\right\|_2^2\le4\|h\|_2^2+4\tau^2M_0^2.\tag{150}$$

接下来，利用（150）中的界可知，当 $\mathcal{D}$ 且 $\alpha_n n$ 足够大时，

$$\begin{aligned}\text{den}&\ge L(\theta^*,\tilde{r})(\alpha_n n)^{-\frac{p}{2}}\int_{B_0(\tilde{r}\sqrt{\alpha_n n})}e^{-c_3\left\|h+\sqrt{\alpha_n n}(\theta^*-\hat{\theta})\right\|_2^2}dh\\&\ge L(\theta^*,\tilde{r})(\alpha_n n)^{-\frac{p}{2}}\int_{B_0(\tilde{r}\sqrt{\alpha_n n})}e^{-c_3(4\|h\|_2^2+4\tau^2M_0^2)}dh\\&=L(\theta^*,\tilde{r})e^{-4c_3\tau^2M_0^2}(\alpha_n n)^{-\frac{p}{2}}\int_{B_0(\tilde{r}\sqrt{\alpha_n n})}e^{-4c_3\|h\|_2^2}dh\ge C_0(\alpha_n n)^{-\frac{p}{2}},\end{aligned}\tag{151}$$

其中 $C_0=L(\theta^*,\tilde{r})e^{-4c_3\tau^2M_0^2}\int_{B_0(\tilde{r})}e^{-4c_3\|h\|_2^2}dh$，且最后一个不等式在 $\alpha_n n>1$ 时成立。将 (149) 和 (151) 中的界代入 (148)，得到

$$\mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}\Big(2r>\|\theta-\hat{\theta}\|_2>\frac{t}{\sqrt{\alpha_n n}}\Big)=\frac{\text{num}}{\text{den}}\le\frac{(\alpha_n n)^{-\frac{p}{2}}\tilde{C}e^{-c_4(t-\sqrt{p/(2c_4)})^2}}{(\alpha_n n)^{-\frac{p}{2}}C_0}=Ce^{-c_4(t-\sqrt{\frac{p}{2c_4}})^2},$$

˜其中 $C \equiv C/C_0$。现在，我们利用上述结果对 (147) 的 $I_{11}$ 进行如下界定：

$$
\begin{aligned}
I_{11} &= \int_0^{2r\sqrt{\alpha_n n}} t^{k_0-1} \mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}\Big(2r > \|\theta-\hat{\theta}\|_2 > \frac{t}{\sqrt{\alpha_n n}}\Big) dt \\
&\le C \int_0^{2r\sqrt{\alpha_n n}} t^{k_0-1} e^{-c_4(t-\sqrt{p/(2c_4)})^2} dt \le C \int_0^{\infty} t^{k_0-1} e^{-c_4(t-\sqrt{p/(2c_4)})^2} dt, \qquad (152)
\end{aligned}
$$

这是一个与 $n$ 无关的有限常数。

**项 $I_{12}$ 的 (147)：** 首先注意到

$$
\begin{aligned}
I_{12} &= \mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}\big(\|\theta-\hat{\theta}\|_2 > 2r\big) \int_0^{2r\sqrt{\alpha_n n}} t^{k_0-1} dt \\
&= \frac{(2r)^{k_0}(\alpha_n n)^{\frac{k_0}{2}}}{k_0} \cdot \frac{e^{\alpha_n n \ell_n(\theta^*)} \int_{\|\theta-\hat{\theta}\|_2>2r} e^{\alpha_n n(\ell_n(\theta)-\ell_n(\theta^*))}\pi(\theta)d\theta}{e^{\alpha_n n \ell_n(\hat{\theta})} \int e^{\alpha_n n(\ell_n(\theta)-\ell_n(\hat{\theta}))}\pi(\theta)d\theta}. \qquad (153)
\end{aligned}
$$

然后，由于 $e^{\alpha_n n \ell_n(\theta^*)} \le e^{\alpha_n n \ell_n(\hat{\theta})}$ 因为 $\hat{\theta}$ 是最大似然估计，我们有

$$
\begin{aligned}
I_{12} &\le \frac{(2r)^{k_0}(\alpha_n n)^{\frac{k_0}{2}} \int_{\|\theta-\hat{\theta}\|_2>2r} e^{\alpha_n n(\ell_n(\theta)-\ell_n(\theta^*))}\pi(\theta)d\theta}{k_0 \int e^{\alpha_n n(\ell_n(\theta)-\ell_n(\hat{\theta}))}\pi(\theta)d\theta} \qquad (154) \\
&\le \frac{\kappa(2r)^{k_0}(\alpha_n n)^{\frac{k_0+p}{2}}}{C_0 k_0} \int_{\|\theta-\hat{\theta}\|_2>2r} e^{\alpha_n n(\ell_n(\theta)-\ell_n(\theta^*))} d\theta, \qquad (155)
\end{aligned}
$$

其中，上面的第二个不等式使用了 (151) 中的界以及假设 $\pi(\theta) \le \kappa$。^ ^ 接下来，利用三角不等式 $\|\theta-\theta\|_2 \le \|\theta-\theta^*\|_2 + \|\theta-\theta^*\|_2$，并注意到我们处于 ^ 事件 $\mathcal{C}$ 上，可得 $\{\theta : \|\theta-\theta\|_2 > 2r\} \subseteq \{\theta : \|\theta-\theta^*\|_2 > r\}$，因此，

$$
\int_{\|\theta-\hat{\theta}\|_2>2r} e^{\alpha_n n(\ell_n(\theta)-\ell_n(\theta^*))} d\theta \le \int_{\|\theta-\theta^*\|_2>r} e^{\alpha_n n(\ell_n(\theta)-\ell_n(\theta^*))} d\theta \le \int_{\|\theta-\theta^*\|_2>r} e^{-L\alpha_n n\gamma(\|\theta-\theta^*\|_2)} d\theta. \qquad (156)
$$

上述最后一个不等式是在条件于式（139）中定义的事件 $\mathcal{A}$ 的情况下得出的。现在，转到球坐标（其中 vol($\mathbb{S}^{p-1}$) 表示单位球面的体积——见推论10），

$$
\begin{aligned}
\int_{\|\theta-\theta^*\|_2>r} e^{-L\alpha_n n\gamma(\|\theta-\theta^*\|_2)} d\theta &= \mathrm{vol}(\mathbb{S}^{p-1}) \int_r^{\infty} \rho^{p-1} e^{-L\alpha_n n\gamma(\rho)} d\rho \\
&\le \mathrm{vol}(\mathbb{S}^{p-1}) e^{Lc_2\alpha_n n} \int_r^{\infty} \rho^{p-1} e^{-Lc_0\alpha_n n \log(1+c_1\rho)} d\rho \qquad (157) \\
&\le \mathrm{vol}(\mathbb{S}^{p-1}) e^{Lc_2\alpha_n n} c_1^{-Lc_0\alpha_n n} \int_r^{\infty} \rho^{p-1-Lc_0\alpha_n n} d\rho \qquad (158) \\
&= \mathrm{vol}(\mathbb{S}^{p-1}) e^{Lc_2\alpha_n n} (rc_1)^{-Lc_0\alpha_n n} \frac{r^p}{Lc_0\alpha_n n - p}. \qquad (159)
\end{aligned}
$$

在式（157）中，我们用到了条件 1 假设，$\gamma(v) \geq c_0\log(1+c_1 v) - c_2$ for$v \geq r$；在式（158）中，我们使用了 $(1+c_1\rho)^{-Lc_0\alpha_n n} \leq c_1^{-Lc_0\alpha_n n}\rho^{-Lc_0\alpha_n n}$（因为 $c_1 > 0$）。最后，在式（159）中，我们利用了对于 $\alpha_n n > (p+1)/(Lc_0)$，有 $p - Lc_0\alpha_n n < -1$，因此

$$\int_r^\infty \rho^{p-1-Lc_0\alpha_n n}d\rho = \frac{1}{p-Lc_0\alpha_n n}\rho^{p-Lc_0\alpha_n n}\Big|_r^\infty = \frac{1}{Lc_0\alpha_n n - p}r^{p-Lc_0\alpha_n n}.$$

将 (156)–(159) 中的界代入 (155),

$$I_{12} \leq \frac{\mathrm{vol}(\mathbb{S}^{p-1})\kappa(2r)^{k_0}(\alpha_n n)^{\frac{k_0+p}{2}}}{C_0 k_0} \times \frac{r^p}{Lc_0\alpha_n n - p} \times \Big(\frac{e^{c_2}}{(c_1 r)^{c_0}}\Big)^{L\alpha_n n}. \tag{160}$$

(160) 中的第三项随 $\alpha_n n$ 指数级衰减，因为我们已取 $r > \frac{e^{c_2/c_0}}{c_1}$。因此，对于某个 $\tau_1 > 0$, 存在 $N_5 \equiv N_5(\tau_1)$，使得当 $\alpha_n n > \max((p+1)/(Lc_0), N_5)$ 时，$I_{12} < \tau_1$。

现在回看 (146) 右侧的第一项。结合 (152) 中的界以及上文 (160) 中的论证，取足够大的 $\alpha_n n$，

$$\begin{aligned}\mathbb{P}_{f_{0,n}}(I_1 > M'' \cap \mathcal{E}) &= \mathbb{P}_{f_{0,n}}(I_{11} + I_{12} > M'' \cap \mathcal{E}) \\ &\leq \mathbb{P}_{f_{0,n}}\Big(C\int_0^\infty t^{k_0-1}e^{-c_4(-t-\sqrt{p/(2c_4)})^2}dt + \tau_1 > M'' \cap \mathcal{E}\Big),\end{aligned} \tag{161}$$

并且我们可以取 $M'' = C\int_0^\infty t^{k_0-1}e^{c_4\left(-t-\sqrt{p/(2c_4)}\right)^2}dt + \tau_1 + 1 < \infty$，使得右侧的概率如预期为 0。

**(146) 的第二项(146)：** 我们证明，在给定 $\mathcal{E}$ 的条件下，当 $\alpha_n n$ 足够大时，$I_2$ 有一个常数上界。因此，我们可以选择大于该常数的 $M''$，使得$\mathbb{P}_{f_{0,n}}(I_2 > M'' \cap \mathcal{E}) = 0$。首先，通过变量替换 $u = \frac{t}{\sqrt{\alpha_n n}} \implies dt = \sqrt{\alpha_n n}du$,

$$\begin{aligned}I_2 &= \int_{2r\sqrt{\alpha_n n}}^\infty t^{k_0-1}\mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}\Big(\|\theta-\hat\theta\|_2 > \frac{t}{\sqrt{\alpha_n n}}\Big)dt \\ &= (\alpha_n n)^{\frac{k_0}{2}}\int_{2r}^\infty u^{k_0-1}\mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}(\|\theta-\hat\theta\|_2 > u)du.\end{aligned} \tag{162}$$

我们将概率表示如下:

$$\begin{aligned}\mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}(\|\theta-\hat\theta\|_2 > u) &= \frac{e^{\alpha_n n\ell_n(\theta^*)}\int_{\|\theta-\hat\theta\|_2>u} e^{\alpha_n n(\ell_n(\theta)-\ell_n(\theta^*))}\pi(\theta)d\theta}{e^{\alpha_n n\ell_n(\hat\theta)}\int e^{\alpha_n n(\ell_n(\theta)-\ell_n(\hat\theta))}\pi(\theta)d\theta} \\ &\leq (\alpha_n n)^{\frac{p}{2}}\kappa C_0^{-1}\int_{\|\theta-\hat\theta\|_2>u} e^{\alpha_n n(\ell_n(\theta)-\ell_n(\theta^*))}d\theta,\end{aligned}$$

ˆ 其中我们用到了 $\ell_n(\theta^*) \leq \ell_n(\theta)$、(151) 中的界以及对先验分布所作的界假设。我们的目标是为上述量得到一个对 $u > 2r$ 有效的界。与 (156)–(160) 中一样，在事件 $\mathcal{A}\cap\mathcal{C}$ 上，对于足够大的 $\alpha_n n$，我们可以将上述量界定如下:

$$\begin{aligned}\mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}(\|\theta-\hat\theta\|_2 > u) &\leq (\alpha_n n)^{p/2}\kappa C_0^{-1}\int_{\|\theta-\hat\theta\|_2>u}\exp\left(\alpha_n n(\ell_n(\theta)-\ell_n(\theta^*))\right)d\theta \\ &\leq \frac{\mathrm{vol}(\mathbb{S}^{p-1})e^{Lc_2\alpha_n n}c_1^{-Lc_0\alpha_n n}(\alpha_n n)^{\frac{p}{2}}\kappa}{C_0} \times \frac{(u/2)^{p-Lc_0\alpha_n n}}{Lc_0\alpha_n n - p}.\end{aligned} \tag{163}$$

将 (163) 中的最终界代入 (162) 中关于 $I_2$ 的表达式，得到

$$\begin{aligned} I_2 &= (\alpha_n n)^{\frac{k_0}{2}} \int_{2r}^{\infty} u^{k_0-1} \mathbb{P}_{\pi_{n,\alpha_n}(\theta|X^n)}(\|\theta - \hat{\theta}\|_2 > u) du \\ &\leq \frac{\mathrm{vol}(\mathbb{S}^{p-1}) e^{Lc_2\alpha_n n} c_1^{-Lc_0\alpha_n n} \kappa(\alpha_n n)^{\frac{k_0+p}{2}}}{2^{p-Lc_0\alpha_n n} C_0 (Lc_0\alpha_n n - p)} \int_{2r}^{\infty} u^{k_0-1+p-Lc_0\alpha_n n} du \\ &= \frac{\mathrm{vol}(\mathbb{S}^{p-1}) r^{k_0+p} \kappa(\alpha_n n)^{\frac{k_0+p}{2}}}{2^p C_0 (Lc_0\alpha_n n - p)(Lc_0\alpha_n n - (k_0+p))} \times \Big(\frac{e^{c_2}}{(c_1 r)^{c_0}}\Big)^{L\alpha_n n}. \end{aligned}$$

我们看到，由于取了 $r > \frac{e^{c_2/c_0}}{c_1}$，上述第二项关于 $\alpha_n n$ 以指数速度衰减至 0。因此，对于某个 $\tau_2 > 0$，存在 $N_6 \equiv N_6(\tau_2)$，使得只要 $\alpha_n n > N_6$，就有 $I_2 < \tau_2$。然后取 $\alpha_n n > N_6$，我们得到

$$\mathbb{P}_{f_{0,n}}(I_2 > M'' \cap \mathcal{E}) \leq \mathbb{P}_{f_{0,n}}(\tau_2 > M'' \cap \mathcal{E}), \tag{164}$$

并且我们可以取 取 $M'' = \tau_2 + 1 < \infty$，使得右侧的概率等于 0。

**最终界：** 取 $\alpha_n n$ 足够大，并且

$$M'' = \max\Big\{C \int_0^{\infty} t^{k_0-1} e^{-c_4(-t-\sqrt{p/(2c_4)})^2} dt + \tau_1 + 1,\ \tau_2 + 1\Big\} < \infty,$$

根据 (146)、(161) 和 (164) 中的界，我们有

$$\begin{aligned} \mathbb{P}_{f_{0,n}}(\mathbb{E}_{\pi_{n,\alpha_n}(\theta|X^n)}[\|\sqrt{\alpha_n n}(\theta - \theta^*)\|_2^{k_0}] > M) &\leq \mathbb{P}_{f_{0,n}}(I_1 > M'' \cap \mathcal{E}) + \mathbb{P}_{f_{0,n}}(I_2 > M'' \cap \mathcal{E}) + \epsilon \\ &= \epsilon. \end{aligned}$$

最后，我们如下恢复原常数 $M$。回忆 $M'' = M'/(2k_0) \implies M' = 2k_0 M''$，并且 $M = 2^{k_0} M' + (2\tau M_0)^{k_0} = 2^{k_0+1} k_0 M'' + (2\tau M_0)^{k_0}$。因此，

$$M = 2^{k_0+1} k_0 \max\Big\{C \int_0^{\infty} t^{k_0-1} e^{-c_4(t-\sqrt{p/(2c_4)})^2} dt + \tau_1 + 1,\ \tau_2 + 1\Big\} + (2\tau M_0)^{k_0} < \infty.$$

□

## C.3 假设 (A2') 的充分条件

假设 (**A2'**) 可以被视为一种高阶 LAN 条件。(**A2**) 中的标准 LAN 条件，是以伪真参数为中心对数似然比进行二次近似。假设 (**A2'**) 则是以最大似然估计为中心对数似然比进行三次近似。命题 4 表明，通过对数似然的三阶导数施加正则性条件，我们可以检验 (**A2**)。下面的命题表明，通过对数似然的四阶导数施加正则性条件，我们可以检验 (**A2'**)。

**命题 6。** 假设(**A0**) 成立。此外，假设存在某个 $\gamma > 0$，使得以下条件在 $B_{\theta^*}(\gamma)$ 上成立：

1. 对数似然函数 $\log f_n(X^n|\theta)$ 是四阶连续可微的。

2. 令 $\frac{1}{n}\nabla^4 \log f_n(X^n|\theta) \in \mathbb{R}^{p\times p\times p\times p}$ 为一个包含 $\log f_n(X^n|\theta)$ 偏导数的四阶张量。对于开球闭包中的所有 $\theta$，存在一个函数$\tilde{M}$：$\mathcal{X}^n \to \mathbb{R}$，其中 $\mathbb{E}_{f_{0,n}}[\frac{1}{n}\tilde{M}(X^n)] < \infty$，使得 $\nabla^4 \log f_n(X^n|\theta)$ 的各项满足 $|[\nabla^4 \log f_n(X^n|\theta)]_{i,j,k,l}| = |\frac{\partial^4}{\partial\theta_i\partial\theta_j\partial\theta_k\partial\theta_l} \log f_n(X^n|\theta)| < \tilde{M}(X^n)$。

3. $-\log f_n(X^n|\theta)$ 的海森矩阵以趋近于 1 的概率为正定。也就是说，$-\log f_n(X^n|\theta)$ 的海森矩阵的最小特征值以趋近于 1 的概率为正；即，$\mathbb{P}_{f_{0,n}}(\lambda_{\min}(-\nabla^2 \frac{1}{n}\log f_n(X^n|\theta)) > 0) \to 1$.

那么，式 (23) 中关于 $\tilde{R}_n(h)$ 的条件在 $\delta = \gamma/2$, $H_n = -\frac{1}{n}\nabla^2 \log f_n(X^n|\hat{\theta})$（它以趋于 1 的概率为对称正定矩阵）、$S_n = \nabla^3 \frac{1}{n}\log f_n(X^n|\hat{\theta})$ 和 $M = \frac{p^2}{12\epsilon}\mathbb{E}_{f_{0,n}}[\frac{1}{n}\tilde{M}(X^n)] + 1$ 下成立。

证明。我们分两步证明命题 6。首先，对 $\frac{1}{n}\log f_n(X^n|\hat{\theta} + h/\sqrt{\alpha_n n})$ 在 $\frac{1}{n}\log f_n(X^n|\hat{\theta})$ 处进行三阶泰勒展开。其次，我们研究泰勒展开余项在集合 $\mathcal{H}_n$ 上的收敛性，并说明 $M = \frac{p^2}{12\epsilon}\mathbb{E}_{f_{0,n}}[\frac{1}{n}\tilde{M}(X^n)] + 1$ 的选择如何导出式 (23) 中所述的条件 (**A2'**)。

**步骤 1：对数似然的泰勒展开**。令 $g(t) \equiv \frac{1}{n}\log f_n(X^n|u(t))$，其中$u(t) \equiv \theta + th/\sqrt{\alpha_n n}$。与式 (123) 类似，将 $g(1)$ 在 $g(0)$ 处进行三阶泰勒展开可得

$$g(1) = g(0) + g'(0) + \frac{1}{2}g''(0) + \frac{1}{6}g'''(0) + \frac{1}{6}\int_0^1 (1-t)^3 g''''(t)dt. \tag{165}$$

请注意，将 $g(1)$ 在 $g(0)$ 处展开，等价于将 $\frac{1}{n}\log f_n(X^n|\hat{\theta} + h/\sqrt{\alpha_n n})$ 在 $\frac{1}{n}\log f_n(X^n|\hat{\theta})$ 处展开，因为 $g(1)$ 和 $g(0)$ 如式 (124) 所示，其中以 $\hat{\theta}$ 替换 $\theta^*$。此外，$g'(0)$、$g''(0)$ 和 $g'''(0)$ 如式 (125) 所示，其中以 $\theta$ 替换 $\theta^*$，并且

$$g''''(t) = \left\langle \frac{1}{n}\nabla^4 \log f_n\big(X^n\big|\hat{\theta} + \frac{th}{\sqrt{\alpha_n n}}\big), \Big(\frac{h}{\sqrt{\alpha_n n}}\Big)^{\otimes 4}\right\rangle. \tag{166}$$

将这些值代入式 (165) 得

$$\begin{aligned}\frac{1}{n}\log f_n\Big(X^n\Big|\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\Big) = \frac{1}{n}\log f_n(X^n|\hat{\theta}) + \frac{h^\top}{\sqrt{\alpha_n n}}\left[\frac{1}{n}\nabla \log f_n(X^n|\hat{\theta})\right] \\ + \frac{1}{2}\frac{h^\top}{\sqrt{\alpha_n n}}\left[\frac{1}{n}\nabla^2 \log f_n(X^n|\hat{\theta})\right]\frac{h}{\sqrt{\alpha_n n}} \\ + \frac{1}{6}\left\langle \frac{1}{n}\nabla^3 \log f_n(X^n|\hat{\theta}), \Big(\frac{h}{\sqrt{\alpha_n n}}\Big)^{\otimes 3}\right\rangle + \mathrm{rem}_n(h),\end{aligned} \tag{167}$$

其中 $\mathrm{rem}_n(h)$ 是泰勒近似的余项，其表达式为

$$\mathrm{rem}_n(h) = \frac{1}{6}\int_0^1 (1-t)^3 g''''(t)dt = \frac{1}{6(\alpha_n n)^2}\int_0^1 (1-t)^3 \Big\langle \frac{1}{n}\nabla^4 \log f_n\big(X^n\big|\hat{\theta} + \frac{th}{\sqrt{\alpha_n n}}\big), h^{\otimes 4}\Big\rangle dt. \tag{168}$$

˜因此，根据假设 （**A2'**） 中 $R_n(h)$ 的定义以及式（167）中的泰勒近似，

$$\tilde{R}_n(h) = \alpha_n\Big[\log f_n\big(X^n\big|\hat{\theta}+\frac{h}{\sqrt{\alpha_n n}}\big)-\log f_n(X^n|\hat{\theta})\Big]+\frac{1}{2}h^\top H_n h-\frac{1}{6\sqrt{\alpha_n n}}\langle S_n, h^{\otimes 3}\rangle$$
$$=\alpha_n n\Big(\frac{h^\top}{\sqrt{\alpha_n n}}\Big[\frac{1}{n}\nabla\log f_n(X^n|\hat{\theta})\Big]+\frac{1}{2}\frac{h^\top}{\sqrt{\alpha_n n}}\Big[\frac{1}{n}\nabla^2\log f_n(X^n|\hat{\theta})\Big]\frac{h}{\sqrt{\alpha_n n}}$$
$$+\frac{1}{6}\Big\langle\frac{1}{n}\nabla^3\log f_n(X^n|\hat{\theta}),\Big(\frac{h}{\sqrt{\alpha_n n}}\Big)^{\otimes 3}\Big\rangle+\mathrm{rem}_n(h)\Big)+\frac{1}{2}h^\top H_n h-\frac{1}{6\sqrt{\alpha_n n}}\langle S_n, h^{\otimes 3}\rangle$$
$$=\frac{1}{2}h^\top\Big[\frac{1}{n}\nabla^2\log f_n(X^n|\hat{\theta})+H_n\Big]h+\frac{1}{6\sqrt{\alpha_n n}}\Big\langle\frac{1}{n}\nabla^3\log f_n(X^n|\hat{\theta})-S_n, h^{\otimes 3}\Big\rangle+\alpha_n n\mathrm{rem}_n(h) \tag{169}$$
$$=\alpha_n n\mathrm{rem}_n(h). \tag{170}$$

ˆ (169) 中的等式由 (**A0**) 推出，该条件意味着 $\nabla\log f_n(X^n|\theta)=0$，再结合公共项即可得到。结合 $H_n=-\frac{1}{n}\nabla^2\log f_n(X^n|\hat{\theta})$ 和 $S_n=\frac{1}{n}\nabla^3\log f_n(X^n|\hat{\theta})$，可得到 (170) 中的等式。

–

在此步骤中，还需证明这一 $H_n$ 是对称且正定的，并且 $f_{0,n}$ 概率趋于 1。回顾可知，我们令 $\gamma$ 表示满足条件 1–3 的开球半径。给定该 $\gamma$，定义以下事件：

$$\mathcal{E}_1=\Big\{\lambda_{\min}\Big(-\nabla^2\frac{1}{n}\log f_n(X^n|\hat{\theta})\Big)>0\Big\}\quad\text{and}\quad\mathcal{E}_2=\{\hat{\theta}\in B_{\theta^*}(\gamma/2)\}. \tag{171}$$

首先，我们证明 $H_n$ 是正定的，并且 $f_{0,n}$ 概率趋于 1。为此，请注意， $\mathcal{E}_1$ 是我们希望控制的事件，但只有当对数似然在 $\theta$ 处二阶可微时，该事件才有意义。事件 $\mathcal{E}_2$ 对应于 $\theta$ 落在 $\theta^*$ 的某个邻域内，在该邻域中， $\nabla^2\log f_n(X^n|\theta)$ 是连续的。事实上，根据条件 1，假设 $\nabla^2\log f_n(X^n|\theta)$ 在更大的球 $B_{\theta^*}(\gamma)$ 上连续，但较为受限的事件 $\mathcal{E}_2$ 将在证明的步骤 2 中发挥作用。

我们将证明，对于任意 $\epsilon>0$，存在一个整数 $N_0$，使得对于 $\alpha_n n>N_0$，

$$\mathbb{P}_{f_{0,n}}(\mathcal{E}_1\cap\mathcal{E}_2)\geq 1-\epsilon. \tag{172}$$

根据假设 (**A0**)，存在一个整数 $N_{00}\equiv N_{00}(\epsilon,\gamma)$，使得对于 $\alpha_n n>N_{00}$，

$$\mathbb{P}_{f_{0,n}}(\mathcal{E}_2^{\mathsf{c}})=\mathbb{P}_{f_{0,n}}\Big(\|\hat{\theta}-\theta^*\|_2\geq\frac{\gamma}{2}\Big)<\frac{\epsilon}{4}<\frac{\epsilon}{2}. \tag{173}$$

此外，根据条件 3 的假设，存在某个 $N_{01}\equiv N_{01}(\epsilon,\gamma)$，使得对于$\alpha_n n>N_{01}$，我们有 $\mathbb{P}_{f_{0,n}}(\mathcal{E}_1^{\mathsf{c}}\cap\mathcal{E}_2)<\frac{\epsilon}{4}$.事实上，我们假设 $\lambda_{\min}^-(-\nabla^2\frac{1}{n}\log f_n(X^n|\theta))>0$以 $f_{0,n}$-概率趋于 1，对于 $\theta\in B_{\theta^*}(\gamma)$（条件 3）。在 $\mathcal{E}_1$ 中，我们在 $\theta=\hat{\theta}$ 处计算$\lambda_{\min}(-\nabla^2\frac{1}{n}\log f_n(X^n|\theta))$，因此根据 ˆ$\theta$是否属于 $B_{\theta^*}(\gamma)$，将概率 $\mathbb{P}(\mathcal{E}_1^{\mathsf{c}})$ 分情况讨论。在事件 $\mathcal{E}_1^{\mathsf{c}}\cap\{\theta\in B_{\theta^*}(\gamma)\}$ 上，我们运用条件 3。因此，根据 (173)，只要 $\alpha_n n>N_0\equiv\max(N_{00},N_{01})$，

$$\mathbb{P}_{f_{0,n}}(\mathcal{E}_1^{\mathsf{c}})\leq\mathbb{P}_{f_{0,n}}(\mathcal{E}_1^{\mathsf{c}}\cap\mathcal{E}_2)+\mathbb{P}_{f_{0,n}}(\mathcal{E}_2^{\mathsf{c}})<\frac{\epsilon}{4}+\frac{\epsilon}{4}=\frac{\epsilon}{2}. \tag{174}$$

由 (173) 和 (174) 以及并集界可知，对于 $\alpha_n n > N_0$，我们已证明 (172)。因此，我们得出结论，$-\frac{1}{n}\nabla^2 \log f_n(X^n|\hat{\theta})$ 以趋于 1 的概率为正定。现在，我们证明 $H_n$ 以 $f_{0,n}$-概率趋于 1 时是对称的。事实上，在 (171) 中定义的事件 $\mathcal{E}_2$ 上，情况确实如此。在 $\mathcal{E}_2$ 上，我们有 $\hat{\theta} \in B_{\theta^*}(\gamma)$。于是，由于 $-\frac{1}{n}\nabla^2 \log f_n(X^n|\theta)$ 在 $B_{\theta^*}(\gamma)$ 上连续，

$$[H_n]_{ij} = -\frac{\partial^2}{\partial\theta_i \partial\theta_j}\frac{1}{n}\nabla^2 \log f_n(X^n|\theta)\Big|_{\theta=\hat{\theta}} = -\frac{\partial^2}{\partial\theta_j \partial\theta_i}\frac{1}{n}\nabla^2 \log f_n(X^n|\theta)\Big|_{\theta=\hat{\theta}} = [H_n]_{ji}$$

关于 $\mathcal{E}_2$，由Clairaut定理可知。结合该结论与（173），我们证明了 $H_n$ 关于 $f_{0,n}$-概率趋于 1 时是对称的。

**步骤 2：分析 $\boldsymbol{\alpha_n n \mathrm{rem}_n(h)}$。**回顾一下，$\gamma$ 是满足条件 1-3 的开球半径。在此步骤中，我们旨在证明，对于所有 $\epsilon > 0$ 以及充分大的 $\alpha_n n$，

$$\mathbb{P}_{f_{0,n}}(\mathcal{E}') < \epsilon, \tag{175}$$

其中 $\mathcal{E}' = \{\sup_{h\in\mathcal{H}_n}[\alpha_n n \mathrm{rem}_n(h) - \frac{M}{\alpha_n n}\|h\|_2^4] > 0\}$，且 $\mathcal{H}_n = \{h \in \mathbb{R}^p | \frac{\|h\|_2}{\sqrt{\alpha_n n}} \le \delta\}$ 其中 $\delta = \gamma/2$ 和 $M = \frac{p^2}{12\epsilon}\mathbb{E}_{f_{0,n}}[\frac{1}{n}\tilde{M}(X^n)] + 1$。回顾在（171）中定义的事件 $\mathcal{E}_2$，并定义 $\mathcal{E}_3 = \{\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}} \in B_{\theta^*}(\gamma)\}$。注意，$\mathcal{E}_{3^c} \subset \mathcal{E}_2^c$，因为在事件 $\mathcal{E}_{3^c}$ 上，对于所有 $h \in \mathcal{H}_n$，我们同样有

$$\gamma \le \left\|\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}} - \theta^*\right\|_2 \le \|\hat{\theta} - \theta^*\|_2 + \frac{\gamma}{2} \iff \hat{\theta} \notin B_{\theta^*}(\gamma/2).$$

利用这一事实和 (173)，可知取 $\alpha_n n > N_{00}$ 后，我们得到 $\mathbb{P}_{f_{0,n}}(\mathcal{E}_{3^c}) \le \mathbb{P}_{f_{0,n}}(\mathcal{E}_2^c) < \frac{\epsilon}{4}$。现在定义 $\mathcal{E} \equiv \mathcal{E}_2 \cap \mathcal{E}_3$，并注意到，由并集界及 (173) 中得到的界可得 $\mathbb{P}_{f_{0,n}}(\mathcal{E}^c) = \mathbb{P}_{f_{0,n}}((\mathcal{E}_2 \cap \mathcal{E}_3)^c) \le \mathbb{P}_{f_{0,n}}(\mathcal{E}_{2^c}) + \mathbb{P}_{f_{0,n}}(\mathcal{E}_{3^c}) \le \frac{\epsilon}{4} + \frac{\epsilon}{4} = \frac{\epsilon}{2}$。取 $\alpha_n n > N_{00}$ 后，我们可以如下界定 (175) 中的概率：

$$\mathbb{P}_{f_{0,n}}(\mathcal{E}') \le \mathbb{P}_{f_{0,n}}(\mathcal{E}' \cap \mathcal{E}) + \mathbb{P}_{f_{0,n}}(\mathcal{E}^c) \le \mathbb{P}_{f_{0,n}}(\mathcal{E}' \cap \mathcal{E}) + \frac{\epsilon}{2}. \tag{176}$$

因此，接下来只需证明，对于我们选择的 $M$，$\mathbb{P}_{f_{0,n}}(\mathcal{E}' \cap \mathcal{E}) < \frac{\epsilon}{2}$。

首先，我们说明在事件 $\mathcal{E}$ 上，$\mathrm{rem}_n(h)$ 在 $\mathcal{H}_n$ 上是良定义的。也就是说，我们将证明，对于所有 $0 \le t \le 1$ 以及所有满足 $\|h\|_2/\sqrt{\alpha_n n} \le \delta$ 的 $h \in \mathbb{R}^p$，向量 $\theta + th/\sqrt{\alpha_n n}$ 属于 $B_{\theta^*}(\gamma)$，即 $\theta^*$ 的邻域，其中 $\log f_n(X^n|\theta)$ 是四阶连续可微的。首先注意，对所有 $h \in \mathcal{H}_n$，都有 $\frac{t\|h\|_2}{\sqrt{\alpha_n n}} \le \delta = \frac{\gamma}{2}$。因此，

$$\left\|\hat{\theta} + \frac{th}{\sqrt{\alpha_n n}} - \theta^*\right\|_2 \le \|\hat{\theta} - \theta^*\|_2 + \frac{t\|h\|_2}{\sqrt{\alpha_n n}} \le \|\hat{\theta} - \theta^*\|_2 + \frac{\gamma}{2} < \frac{\gamma}{2} + \frac{\gamma}{2} = \gamma, \tag{177}$$

这里我们还用到了事件 $\mathcal{E}_2 = \{\theta \in B_{\theta^*}(\gamma/2)\}$ 成立这一事实。我们已经证明了 $\hat{\theta} + th/\sqrt{\alpha_n n} \in B_{\theta^*}(\gamma)$。此外，根据条件 1，$\nabla^4 \log f_n(X^n|\theta)$ 在 $B_{\theta^*}(\gamma)$ 上连续；因此，式 (168) 中定义的 $\mathrm{rem}_n(h)$ 在 $\mathcal{E}$ 上是良定义的。此外，在 $\mathcal{E}$ 上，我们可以利用引理 7，像式 (135) 中那样对 $|\alpha_n n \mathrm{rem}_n(h)|$ 关于 $h \in \mathcal{H}_n$ 给出上界：

$$|\alpha_n n \mathrm{rem}_n(h)| \le \frac{p^2\|h\|_2^4}{24\alpha_n n} \sup_{t\in[0,1]} \max_{1\le i,j,k,l\le p} \left|\left[\frac{1}{n}\nabla^4 \log f_n\big(X^n\big|\hat{\theta} + \frac{th}{\sqrt{\alpha_n n}}\big)\right]_{i,j,k,l}\right|. \tag{178}$$

接下来，回顾式 ( ) 中 $\mathcal{E}'$ 175 $\mathcal{E}' = \{$ $_{h\in\mathcal{H}_n}[\alpha_n n$ $_n$ $h$ $-\frac{M}{\alpha_n n}\|h\|_2^4] >$ 的定义；即，
0} 178
sup rem ( )。注意，利用式 ( ) 中的界，我们有

$$\begin{aligned}
&\sup_{h\in\mathcal{H}_n}\Big[\alpha_n n\mathrm{rem}_n(h) - \frac{M\|h\|_2^4}{\alpha_n n}\Big] \\
&\le \sup_{h\in\mathcal{H}_n}\Big[\frac{\|h\|_2^4}{\alpha_n n}\Big(\frac{p^2}{24}\sup_{t\in[0,1]}\max_{1\le i,j,k,l\le p}\Big|\Big[\frac{1}{n}\nabla^4\log f_n\big(X^n\big|\hat\theta+\frac{th}{\sqrt{\alpha_n n}}\big)\Big]_{i,j,k,l}\Big| - M\Big)\Big] \\
&\le \delta^4\alpha_n n\Big(\frac{p^2}{24}\sup_{\substack{t\in[0,1]\\ h\in\mathcal{H}_n}}\max_{1\le i,j,k,l\le p}\Big|\Big[\frac{1}{n}\nabla^4\log f_n\big(X^n\big|\hat\theta+\frac{th}{\sqrt{\alpha_n n}}\big)\Big]_{i,j,k,l}\Big| - M\Big), \qquad (179)
\end{aligned}$$

其中，(179) 中的不等式是由应用界 $\|h\|_2 \le \delta\sqrt{\alpha_n n}$ 得到的，而该界由 $\mathcal{H}_n$ 的定义推出。现在我们使用 (179) 中的界来对 $\mathbb{P}_{f_0,\ n}(\mathcal{E}'\cap\mathcal{E})$ 进行上界估计：

$$\begin{aligned}
&\mathbb{P}_{f_{0,n}}(\mathcal{E}'\cap\mathcal{E}) = \mathbb{P}_{f_{0,n}}\Big(\Big\{\sup_{h\in\mathcal{H}_n}\Big[\alpha_n n\mathrm{rem}_n(h) - \frac{M}{\alpha_n n}\|h\|_2^4\Big] > 0\Big\}\cap\mathcal{E}\Big) \\
&\le \mathbb{P}_{f_{0,n}}\Big(\Big\{\delta^4\alpha_n n\Big(\frac{p^2}{24}\sup_{\substack{t\in[0,1]\\ h\in\mathcal{H}_n}}\max_{1\le i,j,k,l\le p}\Big|\Big[\frac{1}{n}\nabla^4\log f_n\big(X^n\big|\hat\theta+\frac{th}{\sqrt{\alpha_n n}}\big)\Big]_{i,j,k,l}\Big| - M\Big) > 0\Big\}\cap\mathcal{E}\Big) \\
&= \mathbb{P}_{f_{0,n}}\Big(\Big\{\frac{p^2}{24}\sup_{\substack{t\in[0,1]\\ h\in\mathcal{H}_n}}\max_{1\le i,j,k,l\le p}\Big|\Big[\frac{1}{n}\nabla^4\log f_n\big(X^n\big|\hat\theta+\frac{th}{\sqrt{\alpha_n n}}\big)\Big]_{i,j,k,l}\Big| > M\Big\}\cap\mathcal{E}\Big). \qquad (180)
\end{aligned}$$

接下来, $\hat\theta$ 根据 (177) 中的论证，对所有 $h\in\mathcal{H}_n$，均有 $+th/\sqrt{\alpha_n n}\in B_{\theta^*}(\gamma)$。因此，

$$\begin{aligned}
&\sup_{\substack{t\in[0,1]\\ h\in\mathcal{H}_n}}\max_{1\le i,j,k,l\le p}\Big|\Big[\frac{1}{n}\nabla^4\log f_n\big(X^n\big|\hat\theta+\frac{th}{\sqrt{\alpha_n n}}\big)\Big]_{i,j,k,l}\Big| \\
&\qquad\le \sup_{\theta\in B_{\theta^*}(\gamma)}\max_{1\le i,j,k,l\le p}\Big|\Big[\frac{1}{n}\nabla^4\log f_n\big(X^n\big|\theta\big)\Big]_{i,j,k,l}\Big|.
\end{aligned}$$

我们将上 述结果应用于 (180) 中关于 $\mathbb{P}_{f_0,\ n}(\mathcal{E}'\cap\mathcal{E})$ 的界，并使用马尔可夫不等式 ：

$$\begin{aligned}
\mathbb{P}_{f_{0,n}}(\mathcal{E}'\cap\mathcal{E}) &\le \mathbb{P}_{f_{0,n}}\Big(\frac{p^2}{24}\sup_{\theta\in B_{\theta^*}(\gamma)}\max_{1\le i,j,k,l\le p}\Big|\Big[\frac{1}{n}\nabla^4\log f_n\big(X^n\big|\theta\big)\Big]_{i,j,k,l}\Big| > M\Big) \\
&\le \frac{p^2}{24M}\mathbb{E}_{f_{0,n}}\Big[\sup_{\theta\in B_{\theta^*}(\gamma)}\max_{1\le i,j,k,l\le p}\Big|\Big[\frac{1}{n}\nabla^4\log f_n\big(X^n\big|\theta\big)\Big]_{i,j,k,l}\Big|\Big] \\
&\le \frac{p^2}{24M}\mathbb{E}_{f_{0,n}}\Big[\frac{1}{n}\tilde M(X^n)\Big] < \frac{\epsilon}{2}. \qquad (181)
\end{aligned}$$

(181) 中的第一个不等式由条件2推出，第二个不等式则是通过选择 $M$ 使得 $M > \frac{p^2}{12\epsilon}\mathbb{E}_{f_{0,\ n}}[\frac{1}{n}\tilde M(X^n)$ ] 得到的。例如，选择 $M = \frac{p^2}{12\epsilon}\mathbb{E}_{f_{0,\ n}}[\frac{1}{n}\tilde M(X^n)]+1$，根据条件2，该值有限。(176) 和 (181) 中的界给出了 (175) 中所需的结果。

˜**结论**。(175) 中的结果以及第 ˜1 步（见 (170)）所示的等式 $R_n(h) = \alpha_n n\mathrm{rem}_n(h)$ 表明， $R_n(h)$ 满足所需的表示形式，其中 $\delta = \gamma/2$,
$H_n = -\frac{1}{n}\nabla^2$ $f_n$ $X^n|\theta$ log ( ˆ)（如
(172) 所示，其为对称正定矩阵的概率趋近于 1， $S_n = \nabla^3\frac{1}{n}$log $f_n(X^n|\hat\theta)$，且 $M = \frac{p^2}{12\epsilon}\mathbb{E}_{f_{0,\ n}}[\frac{1}{n}\tilde M(X^n)] +$□。

## C.4 假设（ALap）的充分条件（命题 2 的证明）

ˆ证明。令 $g(t) \equiv g(u(t))$，其中 $u(t) \equiv \theta + th/\sqrt{\alpha_n n}$。对$g(1) = b(\hat{}\theta + h/\sqrt{\alpha_n n})$在 $g(0) = b(\hat{}\theta)$ 处作一阶泰勒展开，得到 $g(1) = g(0) + g'(0) + \int_0^1 (1-t)g''(t)dt$.利用

$$g'(0) = \frac{h^\top}{\sqrt{\alpha_n n}} \nabla b(\hat{\theta}), \quad \text{and} \quad g''(t) = \frac{h^\top}{\sqrt{\alpha_n n}} \Big[\nabla^2 b\Big(\hat{\theta} + \frac{th}{\sqrt{\alpha_n n}}\Big)\Big] \frac{h}{\sqrt{\alpha_n n}},$$

我们得到

$$\begin{aligned} b\Big(\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\Big) &= b(\hat{\theta}) + \frac{h^\top \nabla b(\hat{\theta})}{\sqrt{\alpha_n n}} + \mathrm{rem}_n(h); \\ \mathrm{rem}_n(h) &:= \int_0^1 \frac{(1-t)}{\alpha_n n} h^\top \Big[\nabla^2 b\Big(\hat{\theta} + \frac{th}{\sqrt{\alpha_n n}}\Big)\Big] h dt. \end{aligned} \tag{182}$$

根据定义　　假设（**ALap**）中 $R_b(h)$ 的定义以及式 (182) 中的泰勒近似,

$$R_b(h) = \Big[b\Big(\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\Big) - b(\hat{\theta})\Big] - \frac{v^\top h}{\sqrt{\alpha_n n}} = \frac{h^\top \nabla b(\hat{\theta})}{\sqrt{\alpha_n n}} + \mathrm{rem}_n(h) - \frac{v^\top h}{\sqrt{\alpha_n n}} = \mathrm{rem}_n(h), \tag{183}$$

ˆ ˆ ˆ ˆ ˆ 其中，最后一个不等式是通过令 $v = \nabla b(\theta) = \pi(\theta)\nabla q(\theta) + q(\theta)\nabla \pi(\theta)$ 得到的。由于如上所示，$R_b(h)$ =rem$_n(h)$，为完成证明，我们证明对所有 $\epsilon > 0$,

$$\mathbb{P}_{f_{0,n}}\Big(\sup_{h \in \mathcal{H}_{n,b}} \mathrm{rem}_n(h) - \frac{M}{\alpha_n n}\|h\|_2^2 > 0\Big) < \epsilon, \tag{184}$$

当 $M = p\,$sup$_{\theta \in B_{\theta^*}(\gamma)}$ max$_{1 \le i,\ j \le p} |[\nabla^2 b(\theta)]_{i,\ j}|$ 充分大时成立。下文中，我们记 $\mathcal{E}' = \{$ sup$_{h \in \mathcal{H}_{n,\ b}}$rem$_n(h) - \frac{M}{\alpha_n n}\|h\|_2^2 > 0\}$, 使得式 (184) 可写为 $\mathbb{P}_{f_{0,\ n}}(\mathcal{E}') < \epsilon$。

ˆ ˆ 定义事件 $\mathcal{E}_1 \equiv \{\theta \in B_{\theta^*}(\gamma/2)\}$ 和 $\mathcal{E}_2 \equiv \{\theta + h/\sqrt{\alpha_n n} \in \overline{B_{\theta^*}(\gamma)}\}$. 。根据假设（A0），存在一个 $N_0 \equiv N_0(\epsilon, \gamma)$，使得只要 $\alpha_n n > N_0$，就有 $\mathbb{P}_{f_0,n}(\mathcal{E}_1^{\mathsf{c}}) = \hat{} \hat{} \mathbb{P}_{f_0,n}(\|\theta - \theta^*\|_2 \ge \frac{\gamma}{2}) < \frac{\epsilon}{2}$.。接下来，注意到 $\mathcal{E}_{\mathsf{c}2} \subset \mathcal{E}_{\mathsf{c}1}$, 因为 $\gamma \le \|\theta + h/\sqrt{\alpha_n n} - \theta^*\|_2 \le \|\theta - \theta^*\|_2 + \frac{\gamma}{2}$ˆ 为真当且仅当 $\theta \notin B_{\theta^*}(\gamma/2)$。因此，取 $\alpha_n n > N_0$，我们有 $\mathbb{P}_{f_0,n}(\mathcal{E}_2^{\mathsf{c}}) \le \mathbb{P}_{f_0,n}(\mathcal{E}_1^{\mathsf{c}}) < \frac{\epsilon}{2}$。现在定义 $\mathcal{E} \equiv \mathcal{E}_1 \cap \mathcal{E}_2$，由并集界可得，当 $\alpha_n n > N_0$ 时, $\mathbb{P}_{f_0,n}(\mathcal{E}\mathsf{c}) \le \epsilon$。因此，取 $\alpha_n n > N_0$，我们可以对（184）中的概率进行界定:

$$\mathbb{P}_{f_{0,n}}\Big(\sup_{h \in \mathcal{H}_{n,b}} \mathrm{rem}_n(h) - \frac{M\|h\|_2^2}{\alpha_n n} > 0\Big) = \mathbb{P}_{f_{0,n}}(\mathcal{E}') \le \mathbb{P}_{f_{0,n}}(\mathcal{E}' \cap \mathcal{E}) + \mathbb{P}_{f_{0,n}}(\mathcal{E}^{\mathsf{c}}) \le \mathbb{P}_{f_{0,n}}(\mathcal{E}' \cap \mathcal{E}) + \epsilon.$$

因此，为完成证明，我们将证明 $\mathbb{P}_{f_{0,n}}(\mathcal{E}' \cap \mathcal{E}) = 0$对于我们选择的 $M$ 成立。

首先，我们注意到，依据类似于（177）附近的论证，在事件 $\mathcal{E}$ 上，我们有 ˆthat $\theta + th/\sqrt{\alpha_n n} \in \overline{B_{\theta^*}(\gamma)}$；因此，rem$_n(h)$ 在 $\mathcal{H}_{n,\ b}$ 上是良定义的。接下来，利用引理 7

并与事件 $\mathcal{E}$ 取交集，我们可以用类似于（135）中的方法对 $\mathcal{H}_{n,b}$ 上的 $\mathrm{rem}_n(h)$ 进行界定，从而得到

$$\mathrm{rem}_n(h) \le \frac{p\|h\|_2^2}{2\alpha_n n} \sup_{t\in[0,1]} \max_{1\le i,j\le p} \Big| \Big[\nabla^2 b\Big(\hat{\theta} + \frac{th}{\sqrt{\alpha_n n}}\Big)\Big]_{i,j}\Big|.$$

接下来，由于 $\theta \mp th/\sqrt{\alpha_n n} \in B_{\theta^*}(\gamma)$，通过类似于 (136)（或紧接在 (180) 之后）的论证，

$$\sup_{h\in\mathcal{H}_{n,b}} \mathrm{rem}_n(h) \le \sup_{h\in\mathcal{H}_{n,b}} \frac{p\|h\|_2^2}{2\alpha_n n} \sup_{\theta\in B_{\theta^*}(\gamma)} \max_{1\le i,j\le p} \Big|[\nabla^2 b(\theta)]_{i,j}\Big|. \tag{185}$$

最后，我们应用 (185) 来界定 $\mathbb{P}_{f_{0,n}}(\mathcal{E}' \cap \mathcal{E})$。

$$\begin{aligned}
\mathbb{P}_{f_{0,n}}(\mathcal{E}' \cap \mathcal{E}) &= \mathbb{P}_{f_{0,n}}\Big( \sup_{h\in\mathcal{H}_{n,b}} \mathrm{rem}_n(h) - \frac{M\|h\|_2^2}{\alpha_n n} > 0 \cap \mathcal{E}\Big) \\
&\le \mathbb{P}_{f_{0,n}}\Big( \sup_{h\in\mathcal{H}_{n,b}} \frac{\|h\|_2^2}{\alpha_n n}\Big(\frac{p}{2} \sup_{\theta\in B_{\theta^*}(\gamma)} \max_{1\le i,j\le p} \big|[\nabla^2 b(\theta)]_{i,j}\big| - M\Big) > 0 \cap \mathcal{E}\Big) \\
&\le \mathbb{P}_{f_{0,n}}\Big( \frac{\gamma^2}{4}\Big(\frac{p}{2} \sup_{\theta\in B_{\theta^*}(\gamma)} \max_{1\le i,j\le p} \big|[\nabla^2 b(\theta)]_{i,j}\big| - M\Big) > 0 \cap \mathcal{E}\Big) = 0.
\end{aligned} \tag{186}$$

最后一个不等式由 $\mathcal{H}_{n,b} = \{h \in \mathbb{R}^p | \frac{\|h\|_2}{\sqrt{\alpha_n n}} \le \delta\}$ 结合 $\delta = \gamma/2$ 得出；最后一个等式则由代入 $M = p\sup_{\theta\in B_{\theta^*}(\gamma)} \max_{1\le i,j\le p} |[\nabla^2 b(\theta)]_{i,j}|$ 得出。 □

## C.5 验证我们的主要结果对 $f_n(X^n|\theta$ 的条件 )

本节将证明：如果定理 1 和定理 2 的条件对 $f_n(X^n|\theta)$ 成立，那么倾斜似然 $\tilde{f}_n(X^n|\theta) = f_n(X^n|\theta)^{\frac{\alpha_n}{\alpha_n}}$ ——其中 $\alpha_n$ 满足 $\frac{1}{n} \ll \alpha_n \ll 1$，且 $\frac{\alpha_n}{\alpha_n} \to 1$ 以 $f_{0,n}$ 概率成立——也满足这些假设，且取值为相同的 $\theta$、$\theta^*$ 和 $V_{\theta^*}$。因此，只需将 $f_n(X^n|\theta)$ 替换为 $f_n(X^n|\theta)$，即可证明定理 1 和定理 2 的结果对 $\alpha_n$ 后验分布成立。需要指出的是，假设 (**A1**) 和 (**A1'**)仅关于先验分布，因此无需对其进行检验。

第一个结果检验定理 1 的假设 (**A0**) 和 (**A2**)，并表明对于 $f_n(X^n|\theta)$，两者均满足 ，且使用这些假设中为 $f_n(X^n|\theta)$ 指定的相同 $\theta$、$\theta^*$ 和 $V_{\theta^*}$。

**命题 7.** 假设 $f_n(X^n|\theta)$ 满足假设(**A0**) 和 (**A2**)。此外，假设对于一个依赖于数据的序列 $\alpha_n$，存在一个序列 $\frac{1}{n} \ll \alpha_n \ll 1$，使得 $\alpha_n/\alpha_n \to 1$ 以 $f_{0,n}$-概率成立。那么倾斜似然 $\tilde{f}_n(X^n|\theta) = f_n(X^n|\theta)^{\frac{\alpha_n}{\alpha_n}}$ 也满足 (**A0**) 和 (**A2**)，且具有相同的 $\theta$、$\theta^*$ 和 $V_{\theta^*}$。

证明。我们首先验证 $f_n(X^n|\theta)$ 满足假设 (**A0**)。具体而言，我们将证明 关于 $f_n(X^n|\theta)$ 的最大似然估计（记为 $\theta$）等于关于 $f_n(X^n|\theta)$ 的最大似然估计（记为 $\theta$）。事实上，由于 $\alpha_n/\alpha_n$ 不依赖于 $\theta$，因此

$$\tilde{\theta} = \arg\max_\theta \log \tilde{f}_n(X^n|\theta) = \arg\max_\theta \frac{\hat{\alpha}_n}{\alpha_n} \log f_n(X^n|\theta) = \arg\max_\theta \log f_n(X^n|\theta) = \hat{\theta}.$$

˜接下来，我们验证 $f_n(X^n|\theta)$ 满足假设 (**A2**)。具体而言，对于最初在 (**A2**) 中为 $f_n(X^n|\theta)$ 指定的 $V_{\theta^*}$，我们将式 (16) 中以 $f_n(X^n|\theta)$ 替换 $f_n(X^n|\theta)$ 后的值定义为 $R_n(h)$。根据假设，对于任何 $r>0$，以 $f_0$ 下的 $n$-概率，$\sup_{h\in\bar{B}_0(r)}|R_n(h)|\to 0$；现在我们证明，对于任何 $r>0$，以 $f_0$ 下的 $n$-概率，$\sup_{h\in\bar{B}_0(r)}|R_n(h)|\to 0$。注意到

$$
\begin{aligned}
|\tilde{R}_n(h)| &= \Big|\alpha_n\Big[\log\tilde{f}_n\big(X^n\big|\theta^*+\frac{h}{\sqrt{\alpha_n n}}\big)-\log\tilde{f}_n(X^n|\theta^*)\Big]-\sqrt{\alpha_n}h^\top V_{\theta^*}\Delta_{n,\theta^*}+\frac{1}{2}h^\top V_{\theta^*}h\Big| \\
&= \Big|\frac{\hat{\alpha}_n}{\alpha_n}\Big(\alpha_n\Big[\log f_n\big(X^n\big|\theta^*+\frac{h}{\sqrt{\alpha_n n}}\big)-\log f_n(X^n|\theta^*)\Big]\Big)-\sqrt{\alpha_n}h^\top V_{\theta^*}\Delta_{n,\theta^*}+\frac{1}{2}h^\top V_{\theta^*}h\Big| \\
&= \Big|\frac{\hat{\alpha}_n}{\alpha_n}R_n(h)-\Big(1-\frac{\hat{\alpha}_n}{\alpha_n}\Big)\sqrt{\alpha_n}h^\top V_{\theta^*}\Delta_{n,\theta^*}+\Big(1-\frac{\hat{\alpha}_n}{\alpha_n}\Big)\frac{1}{2}h^\top V_{\theta^*}h\Big| \\
&\le \frac{\hat{\alpha}_n}{\alpha_n}|R_n(h)|+\Big|1-\frac{\hat{\alpha}_n}{\alpha_n}\Big|\,\big|\sqrt{\alpha_n}h^\top V_{\theta^*}\Delta_{n,\theta^*}\big|+\Big|1-\frac{\hat{\alpha}_n}{\alpha_n}\Big|\,\Big|\frac{1}{2}h^\top V_{\theta^*}h\Big|.
\end{aligned}
$$

现在，由于$\frac{\hat{\alpha}_n}{\alpha_n}=1+o_{f_0,n}(1)$ 且 $\sup_{h\in\bar{B}_0(r)}|R_n(h)|=o_{f_0,n}(1)$，根据上述结果可得

$$
\sup_{h\in\bar{B}_0(r)}|\tilde{R}_n(h)|\le o_{f_{0,n}}(1)+o_{f_{0,n}}(1)\Big(\sup_{h\in\bar{B}_0(r)}\Big\{\big|\sqrt{\alpha_n}h^\top V_{\theta^*}\Delta_{n,\theta^*}\big|+\big|\frac{1}{2}h^\top V_{\theta^*}h\big|\Big\}\Big)=o_{f_{0,n}}(1).
$$

在最后一个等式中，我们利用了括号中的项为 $O_{f_{0,n}}(1)$ 这一事实。事实上，根据引理 7，

$$
\big|\frac{1}{2}h^\top V_{\theta^*}h\big|\le\frac{1}{2}\sum_{1\le i,j\le p}[V_{\theta^*}]_{ij}|h_ih_j|\le\max_{1\le i,j\le p}[V_{\theta^*}]_{ij}\frac{1}{2}\sum_{1\le i,j\le p}|h_ih_j|\le\frac{p}{2}\|h\|_2^2\max_{1\le i,j\le p}[V_{\theta^*}]_{ij}.
$$

利用上述结果和柯西-施瓦茨不等式，我们得到

$$
\begin{aligned}
\sup_{h\in\bar{B}_0(r)}\Big\{\big|\sqrt{\alpha_n}h^\top V_{\theta^*}\Delta_{n,\theta^*}\big|+\big|\frac{h^\top V_{\theta^*}h}{2}\big|\Big\} &\le \sup_{h\in\bar{B}_0(r)}\Big\{\sqrt{\alpha_n}\|h\|_2\|V_{\theta^*}\Delta_{n,\theta^*}\|_2+\frac{p\|h\|_2^2}{2}\max_{1\le i,j\le p}[V_{\theta^*}]_{ij}\Big\} \\
&\le \sqrt{\alpha_n}r\|V_{\theta^*}\Delta_{n,\theta^*}\|_2+\frac{pr^2}{2}\max_{1\le i,j\le p}[V_{\theta^*}]_{ij}. \qquad (187)
\end{aligned}
$$

由于 $\alpha_n\to 0$ 且 $\Delta_{n,\theta^*}=O_{f_{0,n}}(1)$，我们有 (187) 是 $O_{f_{0,n}}(1)$。 □

第二个结果检验了当 $f_n(X^n|\theta)$ 满足命题 5 的条件时，假设 (**A3**) 是否成立；命题 5 给出了检验似然是否满足 (**A3**) 的条件。在下文的命题中˜，我们证明：若 $f_n(X^n|\theta)$ 满足这些条件，则 $f_n(X^n|\theta)$ 也满足命题 5 的条件。进而，这保证了假设 (**A3**) 对 $f_n(X^n|\theta)$ 成立。

**命题 8**。 假设 $f_n(X^n|\theta)$ 和 $\pi(\theta)$ 满足命题 5 的条件。此外，假设对于一个依赖于数据的序列 ˆ$\alpha_n$，存在一个序列 $\frac{1}{n}\ll\alpha_n\ll 1$，使得 ˙$\frac{\alpha_n}{\alpha_n}\to 1$ 以 $f_{0,n}$-概率成立。则假设 (**A3**)˝ 对 ˆ$\alpha_n$-后验分布成立。¯

˙证明。 设 $\tilde{f}_n(X^n|\theta)=f_n(X^n|\theta)^{\frac{\alpha_n}{\hat{\alpha}_n}}$。我们将证明 $\tilde{f}_n(X^n|\theta)$ 满足命题 5 的条件，由此可知 (**A3**) 对 ˆ$\alpha_n$-后验分布成立。注意，˜我们已经证明 (**A0**) 对 $f_n(X^n|\theta)$ 成立，并且已假设先验分布

满足 (A1) 以及界 $\pi(\theta) \le \kappa$，其中某个 $\kappa > 0$。接下来需要证明~条件 1 和 2 对 $f_n(X^n|\theta)$ 成立。我们将在下文给出证明。

**检验条件 1**。为检验条件 1，我们将证明，对于任意 $\epsilon > 0$，存在~一个常数 $L > 0$ 和一个函数 $\tilde{\gamma}$：$[0,\infty) \to [0,\infty)$，使得 $\mathbb{P}_{f_{0,n}}(\mathcal{A}) \le \epsilon$，其中

$$\mathcal{A} = \Big\{ \sup_{\theta \in B_{\theta^*}(r)^c} \log \tilde{f}_n(X^n|\theta) - \log \tilde{f}_n(X^n|\theta^*) > -\tilde{L} n \tilde{\gamma}(\|\theta - \theta^*\|_2) \Big\},$$

当 $n$ 足够大，且对任意满足条件陈述中界限的 $r > 0$。为此，定义事件 $\mathcal{E} = \{\hat{\alpha}_n/\alpha_n > 1 - t\}$，其中某个 $t \in (0, 1)$。根据假设，存在$N_0 \equiv N_0(t, \epsilon)$，使得只要 $n > N_0$，就有 $\mathbb{P}_{f_{0,n}}(\mathcal{E}^c) < \epsilon/2$。于是，取 $n > N_0$，可得

$$\begin{aligned}
\mathbb{P}_{f_{0,n}}(\mathcal{A}) &\le \mathbb{P}_{f_{0,n}}(\mathcal{A} \cap \mathcal{E}) + \frac{\epsilon}{2} \\
&\le \mathbb{P}_{f_{0,n}}\Big( \Big\{ \sup_{\theta \in B_{\theta^*}(r)^c} \log f_n(X^n|\theta) - \log f_n(X^n|\theta^*) > -\frac{\tilde{L} n \tilde{\gamma}(\|\theta - \theta^*\|_2)}{1-t} \Big\} \cap \mathcal{E} \Big) + \frac{\epsilon}{2} \\
&\le \mathbb{P}_{f_{0,n}}\Big( \sup_{\theta \in B_{\theta^*}(r)^c} \log f_n(X^n|\theta) - \log f_n(X^n|\theta^*) > -\frac{\tilde{L} n \tilde{\gamma}(\|\theta - \theta^*\|_2)}{1-t} \Big) + \frac{\epsilon}{2}.
\end{aligned}$$

~由于我们假设条件 1 对 $f_n(X^n|\theta)$ 成立，令 $L := (1-t)L$（由于 $t < 1$，该值为正），并令 $\tilde{\gamma}(\cdot) = \gamma(\cdot)$，则存在一个 $N_1 \equiv N_1(t, \epsilon)$，使得当 $n > N_1$ 时，上述概率有上界 $\frac{1}{2}\epsilon$。因此，当 $n > \max(N_0, N_1)$ 时，所需结果成立。~注意，我们证明了条件 1 对 $f_n(X^n|\theta)$ 成立，且其 $\gamma(\cdot)$ 与对 $f_n(X^n|\theta)$ 成立时相同。

检验条件 2。首先注意，$r$ 的值正是使条件 1 成立的值（如上所示，对 $f_n(X^n|\theta)$ 和 $f_n(X^n|\theta)$ 均成立）。接下来，令 $g_n \equiv (\theta, X^n)$ ~ ~ ^ $\log f_n(X^n|\theta) - \log f_n(X^n|\theta)$。那么，对某些常数 $\tilde{c}_3, \tilde{c}_4 > 0$,

$$\begin{aligned}
&\Big\{ -n\tilde{c}_3 \|\theta - \hat{\theta}\|_2^2 \le g_n(\theta, X^n) \le -n\tilde{c}_4 \|\theta - \hat{\theta}\|_2^2 \quad \text{for all } \theta \in B_{\hat{\theta}}(2r) \Big\}^c \\
&\quad = \Big\{ g_n(\theta, X^n) < -n\tilde{c}_3 \|\theta - \hat{\theta}\|_2^2 \ \cup \ g_n(\theta, X^n) > -n\tilde{c}_4 \|\theta - \hat{\theta}\|_2^2 \quad \text{for some } \theta \in B_{\hat{\theta}}(2r) \Big\}.
\end{aligned}$$

因此，验证条件2等价于证明：对于所有 $\epsilon > 0$，都存在 $N$ 和常数 $\tilde{c}_3 > 0$ 及 $\tilde{c}_4 > 0$，使得只要 $n > N$,

$$\mathbb{P}_{f_{0,n}}\Big( g_n(\theta, X^n) < -n\tilde{c}_3 \|\theta - \hat{\theta}\|_2^2 \ \cup \ g_n(\theta, X^n) > -n\tilde{c}_4 \|\theta - \hat{\theta}\|_2^2 \quad \text{for some } \theta \in B_{\hat{\theta}}(2r) \Big) < \epsilon.$$

根据并集界，我们可以对上述概率作如下上界估计：

$$\begin{aligned}
&\mathbb{P}_{f_{0,n}}\Big( g_n(\theta, X^n) < -n\tilde{c}_3 \|\theta - \hat{\theta}\|_2^2 \ \cup \ g_n(\theta, X^n) > -n\tilde{c}_4 \|\theta - \hat{\theta}\|_2^2 \quad \text{for some } \theta \in B_{\hat{\theta}}(2r) \Big) \\
&\quad \le \mathbb{P}_{f_{0,n}}\Big( g_n(\theta, X^n) < -n\tilde{c}_3 \|\theta - \hat{\theta}\|_2^2 \quad \text{for some } \theta \in B_{\hat{\theta}}(2r) \Big) \\
&\qquad + \mathbb{P}_{f_{0,n}}\Big( g_n(\theta, X^n) > -n\tilde{c}_4 \|\theta - \hat{\theta}\|_2^2 \quad \text{for some } \theta \in B_{\hat{\theta}}(2r) \Big) \\
&\quad = \mathbb{P}_{f_{0,n}}\Big( \frac{\hat{\alpha}_n}{\alpha_n} \big( \log f_n(X^n|\theta) - \log f_n(X^n|\hat{\theta}) \big) < -n\tilde{c}_3 \|\theta - \hat{\theta}\|_2^2 \quad \text{for some } \theta \in B_{\hat{\theta}}(2r) \Big) \\
&\qquad + \mathbb{P}_{f_{0,n}}\Big( \frac{\hat{\alpha}_n}{\alpha_n} \big( \log f_n(X^n|\theta) - \log f_n(X^n|\hat{\theta}) \big) > -n\tilde{c}_4 \|\theta - \hat{\theta}\|_2^2 \quad \text{for some } \theta \in B_{\hat{\theta}}(2r) \Big) \\
&\quad \equiv P_1 + P_2.
\end{aligned}$$

现在，定义 $\mathcal{E}=\{1-t\le\hat{}\alpha_n/\alpha_n\le 1+t\}$，其中 $t\in\in(0,1)$。根据假设，存在$N_0\equiv N_0(t,\epsilon)$，使得 $\mathbb{P}_{f_{0,n}}(\mathcal{E}^c)<\frac{1}{4}\epsilon$，当 $n>\bar{}N_0$ 时。我们用此来界定 $P_1$和 $P_2$。

ˆ首先，考虑项 $P_1$。注意，根据 $\theta$ 的定义，对于所有 $\theta\in B_{\hat{\theta}}(2r)$，都有 $\log f_n(X^n|\theta)-\log f_n(X^n|\theta)\le 0$ˆ。此外，若 $f_n(X^n|\theta)$ 满足条件 2，则存在一个 $N_1(\epsilon,c_3,c_4,t)$，使得只要 $n>N_1$，条件2中的事件就以至少 $1-\epsilon/4$ 的概率成立。因此，取 $n>\max\{N_0,N_1\}$ 和 $\tilde{c}_3=(1+t)c_3$，其中 $c_3>0$ 是使条件2对 $f_n(X^n|\theta)$ 成立的常数，我们有

$$
\begin{aligned}
P_1&\le\mathbb{P}_{f_{0,n}}\Big(\Big\{\frac{\hat{\alpha}_n}{\alpha_n}\big(\log f_n(X^n|\theta)-\log f_n(X^n|\hat{\theta})\big)<-n\tilde{c}_3\|\theta-\hat{\theta}\|_2^2 \text{ for some } \theta\in B_{\hat{\theta}}(2r)\Big\}\cap\mathcal{E}\Big)+\frac{\epsilon}{4}\\
&\le\mathbb{P}_{f_{0,n}}\Big(\log f_n(X^n|\theta)-\log f_n(X^n|\hat{\theta})<-\frac{n\tilde{c}_3\|\theta-\hat{\theta}\|_2^2}{1+t} \text{ for some } \theta\in B_{\hat{\theta}}(2r)\Big)+\frac{\epsilon}{4}\\
&=1-\mathbb{P}_{f_{0,n}}\Big(\log f_n(X^n|\theta)-\log f_n(X^n|\hat{\theta})\ge -nc_3\|\theta-\hat{\theta}\|_2^2 \text{ for all } \theta\in B_{\hat{\theta}}(2r)\Big)+\frac{\epsilon}{4}\\
&\le\frac{\epsilon}{4}+\frac{\epsilon}{4}=\frac{\epsilon}{2}.
\end{aligned}
$$

$P_2<\epsilon/2$ 的证明类似，由此得到所需结果。 □

˜本节的最后一个结果检验 $f_n(X^n|\theta)$ 是否满足假设 **(A2')** 和**(A3')**；应用引理 1 和定理 2 至 $f_n(X^n|\theta)$ 需要这些假设。

**命题 9.** 假设 $f_n(X^n|\theta)$ 满足假设**(A0)**、**(A2')**和**(A3')**，并且对于一个依赖于数据的序列 ˆ$\alpha_n$，存在一个序列 $\frac{1}{n}\ll\alpha_n\ll 1$，使得 ˆ$\alpha_n/\alpha_n\to 1$ 以 $f_{0,n}$-概率收敛。那么 $\tilde{f}_n(X^n|\theta)=f_n(X^n|\theta)^{\frac{\alpha_n}{\alpha_n}}$ 也满足**(A2')** 和**(A3')**。

˜证明。我们验证 $f_n(X^n|\theta)$ 满足假设**(A2')**。只需证明存在一个矩阵 $H_n\in\mathbb{R}^{p\times p}$（它以高概率是对称正定的）、一个张量 $S_n\in\mathbb{R}^{p\times p\times p}$和一个常数 $M<\infty$，使得对于所有 $\epsilon>0$ 以及某个 $\delta>0$,

$$
\mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\mathcal{H}_n}\tilde{R}_n(h)-\frac{\tilde{M}}{\alpha_n n}\|h\|_2^4>0\Big)<\epsilon, \tag{188}
$$

其中 $\mathcal{H}_n=\{h\in\mathbb{R}^p|\,\frac{\|h\|_2}{\sqrt{\alpha_n n}}<\delta\}$ 和

$$
\begin{aligned}
\tilde{R}_n(h)&\equiv\alpha_n\Big[\log\tilde{f}_n\big(X^n\big|\hat{\theta}+\frac{h}{\sqrt{\alpha_n n}}\big)-\log\tilde{f}_n(X^n|\hat{\theta})\Big]+\frac{1}{2}h^\top\tilde{H}_n h-\frac{1}{6\sqrt{\alpha_n n}}\langle\tilde{S}_n,h^{\otimes 3}\rangle\\
&=\alpha_n\frac{\hat{\alpha}_n}{\alpha_n}\Big[\log f_n\big(X^n\big|\hat{\theta}+\frac{h}{\sqrt{\alpha_n n}}\big)-\log f_n(X^n|\hat{\theta})\Big]+\frac{1}{2}h^\top\tilde{H}_n h-\frac{1}{6\sqrt{\alpha_n n}}\langle\tilde{S}_n,h^{\otimes 3}\rangle.
\end{aligned}
$$

定义事件 $\mathcal{A}=\{\hat{}\alpha_n/\alpha_n<1+t\}$，其中某个 $t>0$。根据假设，存在$N_0\equiv N_0(t,\ \epsilon)$，使得 $\mathbb{P}_{f_{0,\ n}}(\mathcal{A}^c)<\epsilon/2$ 只要 $n>N_0$。于是，取 $n>N_0$，我们可以将 **(188)** 中的概率界定如下:

$$
\mathbb{P}_{f_{0,n}}\Big(\sup_{h\in\mathcal{H}_n}\tilde{R}_n(h)-\frac{\tilde{M}}{\alpha_n n}\|h\|_2^4>0\Big)\le\mathbb{P}_{f_{0,n}}\Big(\Big\{\sup_{h\in\mathcal{H}_n}\tilde{R}_n(h)-\frac{\tilde{M}}{\alpha_n n}\|h\|_2^4>0\Big\}\cap\mathcal{A}\Big)+\frac{\epsilon}{2}.
$$

ˆ还需要界定上式右侧的第一个概率。根据 $\theta$ 的定义，我们知道，对于任意 $h \in \mathcal{H}_n$，$\log f_n(X^n|\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}) - \log f_n(X^n|\hat{\theta}) \le 0$。于是，在事件 $\mathcal{A}$ 上，

$$\tilde{R}_n(h) \le (1+t)\alpha_n \Big[ \log f_n\big(X^n\big|\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\big) - \log f_n(X^n|\hat{\theta}) \Big] + \frac{1}{2} h^\top \tilde{H}_n h - \frac{1}{6\sqrt{\alpha_n n}} \langle \tilde{S}_n, h^{\otimes 3} \rangle,$$

因此取决于 $\mathcal{A}$

$$\begin{aligned}
&\frac{1}{1+t}\Big( \tilde{R}_n(h) - \frac{\tilde{M}\|h\|_2^4}{\alpha_n n} \Big) \\
&\le \alpha_n \Big[ \log f_n\big(X^n\big|\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\big) - \log f_n(X^n|\hat{\theta}) \Big] + \frac{h^\top \tilde{H}_n h}{2(1+t)} - \frac{\langle \frac{1}{1+t}\tilde{S}_n, h^{\otimes 3} \rangle}{6\sqrt{\alpha_n n}} - \frac{\tilde{M}\|h\|_2^4}{\alpha_n n (1+t)}.
\end{aligned}$$

˜令 $H_n := (1+t)H_n$（通过选择˜ ˜t < 1，该矩阵以高概率为正定），$S_n := (1+t)S_n$，以及 $M = (1+t)M < \infty$，则有

$$\begin{aligned}
&\mathbb{P}_{f_{0,n}}\Big( \Big\{ \sup_{h \in \mathcal{H}_n} \tilde{R}_n(h) - \frac{\tilde{M}}{\alpha_n n}\|h\|_2^4 > 0 \Big\} \cap \mathcal{A} \Big) \\
&\le \mathbb{P}_{f_{0,n}}\Big( \sup_{h \in \mathcal{H}_n} \alpha_n \Big[ \log f_n\big(X^n\big|\hat{\theta} + \frac{h}{\sqrt{\alpha_n n}}\big) - \log f_n(X^n|\hat{\theta}) \Big] + \frac{h^\top H_n h}{2} - \frac{\langle S_n, h^{\otimes 3} \rangle}{6\sqrt{\alpha_n n}} - \frac{M\|h\|_2^4}{\alpha_n n} > 0 \Big) \\
&\le \epsilon/2,
\end{aligned}$$

其中，最后一行对于足够大的 $n$ 成立，因为我们假设假设 **（A2'）** ˆ˜ ˜ ˜ ˜对 $f_n(X^n|\theta)$成立。因此，对于 $f_n(X^n|\theta)$，取上述 $H_n$、$S_n$和 $M$时，**（A2'）** 成立。

˜**假设（A3'）：** 我们将证明，将$f_n(X^n|\theta)$替换为 $f_n(X^n|\theta)$后，条件（24）仍然成立。更具体地说，我们证明，对于所有 $\epsilon > 0$ 和 $\delta' > 0$，存在常数 $\tilde{c}$，使得当 $n$ 充分大时

$$\mathbb{P}_{f_{0,n}}\Big( \sup_{\theta \in B_{\hat{\theta}}(\delta')^{\mathsf{c}}} \log \tilde{f}_n(X^n|\theta) - \log \tilde{f}_n(X^n|\hat{\theta}) > -\tilde{c} \Big).$$

为此，定义事件 $\mathcal{E} = \{\hat{\alpha}_n/\alpha_n > 1 - t\}$ 对于 $t \in (0, 1)$，并注意到存在 $N_0 \equiv N_0(\epsilon, t)$，使得只要 $n > N_0$，就有 $\mathbb{P}_{f_{0,n}}(\mathcal{E}^{\mathsf{c}}) < \epsilon/2$。于是取 $n > N_0$，我们得到

$$\begin{aligned}
&\mathbb{P}_{f_{0,n}}\Big( \sup_{\theta \in B_{\hat{\theta}}(\delta')^{\mathsf{c}}} \log \tilde{f}_n(X^n|\theta) - \log \tilde{f}_n(X^n|\hat{\theta}) > -\tilde{c} \Big) \\
&\quad \le \mathbb{P}_{f_{0,n}}\Big( \Big\{ \sup_{\theta \in B_{\hat{\theta}}(\delta')^{\mathsf{c}}} \log \tilde{f}_n(X^n|\theta) - \log \tilde{f}_n(X^n|\hat{\theta}) > -\tilde{c} \Big\} \cap \mathcal{E} \Big) + \frac{\epsilon}{2} \\
&\quad = \mathbb{P}_{f_{0,n}}\Big( \Big\{ \frac{\hat{\alpha}_n}{\alpha_n} \Big[ \sup_{\theta \in B_{\hat{\theta}}(\delta')^{\mathsf{c}}} \log f_n(X^n|\theta) - \log f_n(X^n|\hat{\theta}) \Big] > -\tilde{c} \Big\} \cap \mathcal{E} \Big) + \frac{\epsilon}{2} \\
&\quad \le \mathbb{P}_{f_{0,n}}\Big( \Big\{ \sup_{\theta \in B_{\hat{\theta}}(\delta')^{\mathsf{c}}} \log f_n(X^n|\theta) - \log f_n(X^n|\hat{\theta}) > -\frac{\tilde{c}}{1-t} \Big\} \cap \mathcal{E} \Big) + \frac{\epsilon}{2} \\
&\quad \le \mathbb{P}_{f_{0,n}}\Big( \sup_{\theta \in B_{\hat{\theta}}(\delta')^{\mathsf{c}}} \log f_n(X^n|\theta) - \log f_n(X^n|\hat{\theta}) > -\frac{\tilde{c}}{1-t} \Big) + \frac{\epsilon}{2}.
\end{aligned}$$

由于我们假设 (**A3'**) 对 $f_n(X^n|\theta)$ 成立，若令 $\tilde{c}=(1-t)c$（由于 $t<1$，该值为正），则存在一个 $N_1\equiv N_1(\epsilon,\ t,\ c)$，使得只要 $n>N_1$，上述概率就不超过 $\tilde{}\ \epsilon/2$。取 $n>\max(N_0,\ N_1)$，即可得到带有 $f_n(X^n|\theta)$ 的 (**A3'**)。 □

# D 补充示例

## D.1 命题1的证明

证明。不难看出,式 (21) 中先验分布的特征函数 s

$$\varphi_\pi(t)=\int_H \exp(it^\top\eta)\exp\left(\eta^\top\xi-\nu A(\eta)-\psi(\xi,\nu)\right)d\eta=\exp\left(\psi(it+\xi,\nu)-\psi(\xi,\nu)\right). \quad (189)$$

由式 (22) 可知，后验分布与先验分布 (21) 属于同一族，而 $\alpha_n$-后验分布的特征函数为 $\varphi_{\pi_n,\alpha_n}(t)=\exp(\psi(it+\xi_n,\nu_n)-\psi(\xi_n,\nu_n))$。现在，假设对于某个 $\alpha_0\geq 0$，$\frac{1}{n}T(X^n)$ 以 $f_{0,n}$-概率收敛到某个极限 $g(\eta^*)$，其中 $\eta^*$ 表示自然参数的真实值。那么，在 $f_{0,n}$-概率下，$\xi_n=\alpha_n n(\frac{1}{n}T(X^n))+\xi\to\alpha_0 g(\eta^*)+\xi\equiv\xi'$，并且 $\nu_n=\alpha_n n+\nu\to\alpha_0+\nu\equiv\nu'$.由于 $\varphi_{\pi_n,\alpha_n}(t)$ 关于 $\xi_n$ 和 $\nu_n$ 连续，$\pi_{n,\alpha_n}$ 的特征函数在 $f_{0,n}$-概率下的极限为 $\varphi_{\pi_n,\alpha_n}(t)\to\exp(\psi(it+\xi',\nu')-\psi(\xi',\nu'))$，其形式与先验分布 (189) 的特征函数相同。根据 L´évy 连续性定理，$\alpha_n$-后验分布弱收敛到属于式 (21) 中分布族的一个分布。 □

## D.2 当 $\alpha_n\asymp 1/n$ 时贝叶斯—冯·米泽斯结果的显式反例

我们详细推导一些满足命题1的似然模型和先验分布的例子。我们明确计算这些例子的极限特征函数，以此说明伯恩斯坦-冯·米塞斯结果不可能成立。

指数-伽马：令 $X^n=(X_1,\ldots,X_n)$ 为独立同分布的 $\sim\text{Exp}(\lambda)$，且 $\lambda\sim\Gamma(a,b)$。记真实参数值为 $\lambda^*$。我们可以写成 $f(X^n|\lambda)=\exp(-\lambda\sum_{i=1}^n X_i-n[-\log(\lambda)]-0)$。因此，按式 (20) 的形式，我们取了 $\eta=-\lambda$, $A(\eta)=-\log(-\eta)$、$T(X^n)=\sum_{i=1}^n X_i$，以及$B(X^n)=0$。我们可以写成 $\pi(\lambda)=\exp(-\lambda b-(1-a)(-\log(\lambda))-\log(\frac{\Gamma(a)}{b^a}))$。因此，将$\pi(\eta)$写成式 (21) 的形式可得 $\xi=b$、$\nu=1-a$，以及 $\psi(\xi,\nu)=\log(\Gamma(1-\nu)/\xi^{1-\nu})$。$\pi(\lambda)$的特征函数为 $\varphi_\pi(t)=\exp(\log(\Gamma(\xi+it)^{1-\nu})-\log(\Gamma(1-\nu)/\xi^{1-\nu}))$。根据式 (22)，$\alpha_n$-后验分布的密度为 $\pi_{n,\alpha_n}(\lambda|X^n)=\exp(-\lambda\xi_n-\log(\Gamma(1-\nu_n))-\psi(\xi_n,\nu_n))$。其中我们取 $\xi_n\equiv\alpha_n\sum_{i=1}^n X_i+b$ 和 $\nu_n\equiv\alpha_n n+1-a.$。我们可以将 $\pi_{n,\alpha_n}(\lambda|X^n)$的特征函数写成

$$\begin{aligned}\varphi_{\pi_{n,\alpha_n}}(t)&=\exp\left(\log\left(\frac{(\xi_n+it)^{1-\nu}}{\Gamma(1-\nu_n)}\right)-\log\left(\frac{\xi_n^{1-\nu_n}}{\Gamma(1-\nu_n)}\right)\right)\\&\to\exp\left(\log\left(\frac{(\xi'+it)^{1-\nu'}}{\Gamma(1-\nu')}\right)-\log\left(\frac{\xi'^{1-\nu}}{\Gamma(1-\nu')}\right)\right),\end{aligned}$$

其中，极限是以 $f_{0,n}$-概率收敛的，并且我们取 $g(\eta^*)=-\frac{1}{\eta^*}=\frac{1}{\lambda^*}$, $\xi'=\alpha_0 g(\eta^*)+\xi=\frac{\alpha_0}{\lambda^*}+b$,以及 $\nu'=\alpha_0+\nu=\alpha_0+1-a.$

帕累托-伽马：设 $X^n=(X_1,\dots,X_n)$ 独立同分布于$\sim$帕累托分布$(x_m,k)$，其中 $x_m$ 已知，且$k\sim\Gamma(a,b)$。用 $k^*$ 表示真实参数值。我们可以写成 $f(X^n|k)=\exp(-(k+1)\sum_{i=1}^n\log(\frac{X_i}{x_m})-n\log(\frac{x_m}{k})-0)$。因此，按照（20）的形式可得 $\eta=-(k+1)$，$A(\eta)=\log(\frac{x_m}{-\eta-1})$，$T(X^n)=\sum_{i=1}^n\log(\frac{X_i}{x_m})$，以及 $B(X^n)=0$。我们可以写成 $\pi(k)=\exp(-(k+1)b-(1-a)\log(\frac{x_m}{k})-(-\log(\frac{x_m^{a-1}b^a}{\Gamma\ a}(\ )\ )-b))$。因此，将 $\pi(\eta)$ 写成（21）的形式可得 $\xi=b$、$\nu=1-a$，以及 $\psi(\xi,\nu)=-\log(\frac{x_m^{-\nu}\xi^{1-\nu}}{\Gamma\ 1-\nu}(\ )\ )-\xi$。$\pi(k)$ 的特征函数为 $\varphi_\pi(t)=\exp(-\log(\frac{x_m^{-\nu}(\ \xi+it\ ^{1-\nu})}{\Gamma\ 1-\nu}(\ )\ )+\log(\frac{x_m^{-\nu}\xi^{1-\nu}}{\Gamma\ 1-\nu}(\ )\ )-it)$。根据（22），$\alpha_n$-后验分布的密度为 $\pi_{n,\alpha_n}(k|X^n)=\exp(-(k+1)\xi_n-\nu_n\log(\frac{x_m}{k})-\psi(\xi_n,\nu_n))$，其中我们取$\xi_n\equiv\alpha_n\sum_{i=1}^n\log(\frac{X_i}{x_m})+b$和 $\nu_n\equiv\alpha_n n+(1-a)$。我们可以将 $\pi_{n,\alpha_n}(k|X^n)$ 的特征函数写为

$$\begin{aligned}\varphi_{\pi_{n,\alpha_n}}(t)&=\exp\Big(-\log\Big(\frac{x_m^{-\nu}(\xi_n+it)^{1-\nu_n}}{\Gamma(1-\nu_n)}\Big)+\log\Big(\frac{x_m^{-\nu}\xi_n^{1-\nu_n}}{\Gamma(1-\nu_n)}\Big)-it\Big)\\&\to\exp\Big(-\log\Big(\frac{x_m^{-\nu'}(\xi'+it)^{1-\nu'}}{\Gamma(1-\nu')}\Big)+\log\Big(\frac{x_m^{-\nu'}\xi'^{1-\nu'}}{\Gamma(1-\nu')}\Big)-it\Big),\end{aligned}$$

其中极限以 $f_{0,\ n}$-概率收敛，且 $g(\eta^*)=-\frac{1}{\eta^*+1}=\frac{1}{k^*}$，$\xi'=\alpha_0 g(\eta^*)+\xi=\frac{\alpha_0}{k^*}+b$，以及 $\nu'=\alpha_0+\nu=\alpha_0+1-a$.

伯努利-贝塔：设 $X^n=(X_1,\dots,X_n)$ 独立同分布于 $\sim$Bernoulli$(p)$，且 $p\sim$服从贝塔分布 $a$, $b$。记真实参数值为 $p^*$。我们可以写成 $f(X^n|p)=\exp(\log(\frac{p}{1-p})\sum_{i=1}^n X_i-n\log(\frac{1}{1-p})-0)$。因此，以 (20) 的形式，我们取 $\eta=\log(\frac{p}{1-p})$，$A(\eta)=\log(1+e^\eta)$，$T(X^n)=\sum_{i=1}^n X_i$，以及$B(X^n)=0$。我们可以写成 $\pi(p)=\exp((a-1)\log(\frac{p}{1-p})-(a+b+2)\log(\frac{1}{1-p})-\log(\frac{\Gamma(a)\Gamma(b)}{\Gamma\ a+b}(\ )))$。因此，以 (21) 的形式，我们取 $\xi=a-1$、$\nu=a+b+2$，以及 $\psi(\xi,\nu)=\log(\frac{\Gamma(1+\xi)\Gamma(\nu-\xi-3)}{\Gamma\ \nu-2}(\ ))$。$\pi(p)$ 的特征函数为

$$\varphi_\pi(t)=\exp\Big(\log\Big(\frac{\Gamma(1+\xi+it)\Gamma(\nu-\xi-it-3)}{\Gamma(\nu-2)}\Big)-\log\Big(\frac{\Gamma(1+\xi)\Gamma(\nu-\xi-3)}{\Gamma(\nu-2)}\Big)\Big).$$

根据式 (22)，$\alpha_n$-后验分布的密度为 $\pi_{n,\ \alpha_n}(p|X^n)=\exp(\xi_n\log(\frac{p}{1-p})-\nu_n\log(\frac{1}{1-p})-\psi(\xi_n,\ \nu_n))$，其中我们使用了 $\xi_n=\alpha_n\sum_{i=1}^n X_i+a-1$ 和 $\nu_n=\alpha_n n+a+b+2$。我们可以将 $\pi_{n,\ \alpha_n}(p|X^n)$ 的特征函数写为

$$\begin{aligned}\varphi_{\pi_{n,\alpha_n}}(t)&=\exp\Big(\log\Big(\frac{\Gamma(1+\xi_n+it)\Gamma(\nu_n-\xi_n-it-3)}{\Gamma(\nu_n-2)}\Big)-\log\Big(\frac{\Gamma(1+\xi_n)\Gamma(\nu_n-\xi_n-3)}{\Gamma(\nu_n-2)}\Big)\Big)\\&\to\exp\Big(\log\Big(\frac{\Gamma(1+\xi'+it)\Gamma(\nu'-\xi'-it-3)}{\Gamma(\nu'-2)}\Big)-\log\Big(\frac{\Gamma(1+\xi')\Gamma(\nu'-\xi'-3)}{\Gamma(\nu'-2)}\Big)\Big),\end{aligned}$$

其中极限是按 $f_{0\,n}$ $g\ \eta^*$ ——— $\xi'=\alpha_0 g\ \eta^*\ +\xi=\alpha_0 p^*+a-,$ -概率收敛的，且 $\frac{1}{(\ )}=\frac{e^{\eta^*}}{1+e^{\eta^*}}=p^*$，$(\ )$，以及 $\nu'=\alpha_0+\nu=\alpha_0+a+b+2.$

**高斯-高斯：** 设 $X^n=(X_1,\dots,X_n)$ 独立同分布 $\sim\mathcal{N}(\mu,\sigma^2)$，其中 $\sigma^2$ 已知，且 $\mu\sim\mathcal{N}(\mu_0,\sigma_0^2)$。用 $\mu^*$ 表示真实参数值。我们可以写成 $f(X^n|\mu)=\exp(\mu\frac{\sum_{i=1}^n X_i}{\sigma^2}-n(\frac{\mu^2}{2\sigma^2})-(\frac{n}{2}\log(2\pi\sigma^2)+\frac{1}{2\sigma^2}\sum_{i=1}^n X_i^2))$。因此，将 $f_n(X^n|\eta)$ 写成 (20) 的形式，可得

$\eta = \mu$, $A(\eta) = \frac{\eta^2}{2\sigma^2}$, $T(X^n) = \frac{1}{\sigma^2}\sum_{i=1}^n X_i$, 且 $B(X^n) = \frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^n X_i^2$。我们可以将 $\pi(\mu)$ 写为 $\pi(\mu) = \exp(\mu\frac{\mu_0}{\sigma_0^2} - \frac{\sigma^2}{\sigma_0^2}(\frac{\mu^2}{2\sigma^2}) - (\frac{1}{2}\log(2\pi\sigma_0^2) + \frac{1}{2\sigma_0^2}\mu_0^2))$。因此，将 $\pi(\eta)$ 写成 (21) 的形式，可得 $\xi = \mu_0/\sigma_0^2$, $\nu = \sigma^2/\sigma_0^2$, 以及 $\psi(\xi,\nu) = \frac{1}{2}\log(\frac{2\pi\sigma^2}{\nu}) + \frac{\xi^2\sigma^2}{2\nu}$。$\pi(\mu)$ 的特征函数为 $\varphi_\pi(t) = \exp(\frac{(\xi+it)^2\sigma^2}{2\nu} - \frac{\xi^2\sigma^2}{2\nu})$。根据 (22), $\alpha_n$-后验分布的密度为 $f(X^n|\mu) = \exp(\mu\xi_n - \nu_n(\frac{\mu^2}{2\sigma^2}) - \psi(\xi_n,\nu_n))$。其中我们使用了 $\xi_n = \frac{\alpha_n\sum_{i=1}^n X_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}$ 和 $\nu_n = \alpha_n n + \frac{\sigma^2}{\sigma_0^2}$。我们将 $f(X^n|\mu)$ 的特征函数写为

$$\varphi_{\pi_{n,\alpha_n}}(t) = \exp\Big(\frac{(\xi_n+it)^2\sigma^2}{2\nu_n} - \frac{\xi_n^2\sigma^2}{2\nu_n}\Big) \to \exp\Big(\frac{(\xi'+it)^2\sigma^2}{2\nu'} - \frac{\xi'^2\sigma^2}{2\nu'}\Big),$$

其中，极限是在 $f_{0,n}$-概率意义下，且 $g(\eta^*) = \frac{\eta^*}{\sigma^2} = \frac{\mu^*}{\sigma^2}$, $\xi' = \alpha_0 g(\eta^*) + \xi = \frac{\alpha_0\mu^*}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}$, 并且 $\nu' = \alpha_0 + \nu = \alpha_0 + \sigma^2/\sigma_0^2$. 按照式 (21) 的形式书写，根据定理 1，$\alpha_n$-后验分布的所谓极限分布为 $\phi(\mu|\hat{\mu}, \frac{\sigma^2}{\alpha_n n}) = \exp(\mu\frac{\alpha_n n\hat{}}{\sigma^2}\mu - \alpha_n n(\frac{\mu^2}{2\sigma^2}) - (\frac{1}{2}\log(\frac{2\pi\sigma^2}{\alpha_n n}) + \alpha_n n\hat{}\frac{\mu^2}{2\sigma^2}))$. 定义 $\xi_n' = \frac{\alpha_n n\hat{}}{\sigma^2}\mu$ 和 $\nu_n' = \alpha_n n$，我们可以将 $\phi(\mu|\hat{\mu}, \frac{\sigma^2}{\alpha_n n})$ 的特征函数写为

$$\varphi_{\pi_{n,\alpha_n}}(t) = \exp\Big(\frac{(\xi_n'+it)^2\sigma^2}{2\nu_n'} - \frac{\xi_n'^2\sigma^2}{2\nu_n'}\Big) \to \exp\Big(\frac{(\xi''+it)^2\sigma^2}{2\nu''} - \frac{\xi''^2\sigma^2}{2\nu''}\Big),$$

其中极限是关于 $f_{0,n}$-概率的，并且 $\xi'' = \frac{\alpha_0\mu^*}{\sigma^2} \neq \xi'$ 且 $\nu'' = \alpha_0 \neq \nu'$.

在此例中，特征函数是高斯分布的特征函数，这正是我们在该情形下根据命题1所预期的结果。然而，1中所述的伯恩斯坦-冯·米塞斯并不成立，因为该陈述中出现的高斯分布并不相同。

## D.3 满足 (A3) 的模型

在本节中，我们验证包括线性回归示例和指数族在内的模型的假设(**A3**) 。我们证明，即使使用会导致更厚尾部的非正规先验，后验分布仍将具有任意多个矩，只要 $\alpha_n n$ 足够大。

**线性回归：** 我们考虑一个类似于第 2.1 节模拟中所用的线性回归模型。对于 $x_i \in \mathbb{R}^p$ 和 $y_i \in \mathbb{R}$, 我们指定线性模型 $y_i = x_i^\top\beta^* + \epsilon_i$, 其中 $\epsilon_i \overset{\text{独立同分布}}{\sim} \mathcal{N}(0,\ 1)$, 且 $x_i \overset{\text{独立同分布}}{\sim} \mathcal{N}(0,\ I_p)$。将最大似然估计/普通最小二乘估计量记为 $\hat{}\beta$，并指定先验 $\pi(\beta) \propto 1$。令 $Y = [y_1,\dots,y_n]^\top$ 和 $X = [x_1^\top\cdots x_n^\top]^\top$ 并使用恒等式 $\|Y - X\beta\|_2^2 = \|Y - X\hat{}\beta\|_2^2 + \|X(\hat{}\beta - \beta)\|_2^2$, 不难看出

$$\pi_{n,\alpha_n}(\beta|X,Y) = \phi\Big(\beta\,\Big|\,\hat{\beta}, \frac{1}{\alpha_n n}\Big(\frac{1}{n}X^\top X\Big)^{-1}\Big). \tag{190}$$

为验证假设(**A3**)，我们证明对于所有 $k_0 \geq 0$, $\mathbb{E}_{\pi_{n,\alpha_n}(\beta|X,Y)}[\|\sqrt{\alpha_n n}(\beta - \beta^*)\|_2^{k_0}] = O_{f_{0,n}}(1)$ 均成立。首先，定义 $v_n \equiv \sqrt{\alpha_n n}(\beta - \hat{}\beta)$ 和 $S_n \equiv \frac{1}{n}X^\top X$。由（190）可知

$$v_n\,|\,X,Y \sim \mathcal{N}\left(\mathbf{0}, S_n^{-1}\right) \implies \|S_n^{1/2}v_n\|_2^2\,|\,X,Y \sim \chi_p^2. \tag{191}$$

接下来，根据关于 $k_0 \geq 0$ 的界 $(a+b)^{k_0} \leq 2^{k_0}(a^{k_0}+b^{k_0})$，我们得到

$$\begin{aligned}
2^{-k_0}\mathbb{E}_{\pi_{n,\alpha_n}(\beta|X,Y)}[\|\sqrt{\alpha_n n}(\beta-\beta^*)\|_2^{k_0}] &\leq \mathbb{E}_{\pi_{n,\alpha_n}(\beta|X,Y)}[\|\sqrt{\alpha_n n}(\beta-\hat{\beta})\|_2^{k_0}+\|\sqrt{\alpha_n n}(\hat{\beta}-\beta^*)\|_2^{k_0}] \\
&= \mathbb{E}[\|v_n\|_2^{k_0} \mid X,Y] + \|\sqrt{\alpha_n n}(\hat{\beta}-\beta^*)\|_2^{k_0} \\
&= \mathbb{E}\big[\big(v_n^\top S_n^{1/2} S_n^{-1} S_n^{1/2} v_n\big)^{k_0/2} \mid X,Y\big] + o_{f_{0,n}}(1) \\
&\leq \lambda_{\min}(S_n)^{-k_0/2}\mathbb{E}_{u\sim\chi_p^2}\big[u^{k_0/2}\big] + o_{f_{0,n}}(1) = O_{f_{0,n}}(1),
\end{aligned}$$

其中，第二个等式由假设 **(A0)** 以及 $\alpha_n \to 0$ 得出，最后一个等式则由 $S_n = \frac{1}{n}X^\top X$ 依概率收敛于一个正定矩阵得出。因此，假设**(A3)** 得到满足。

**线性回归（带随机 $\alpha$）：** Cleary (190) 还表明，$\hat{}\alpha_n$-后验分布可以写成

$$\pi_{n,\hat{\alpha}_n}(\beta|X,Y) = \phi\Big(\beta \,\Big|\, \hat{\beta}, \frac{1}{\hat{\alpha}_n n}\Big(\frac{1}{n}X^\top X\Big)^{-1}\Big), \tag{192}$$

假设 $\alpha_n$ 是一个序列，使得 $\frac{\alpha_n}{\alpha_n} \to 1$ 在 $f_0$ 中以 $n$-概率收敛，根据（191），我们有 $\mathbb{E}_{\pi_{n,\hat{\alpha}_n}(\beta|X,Y)}[\|\sqrt{\frac{\alpha_n}{\alpha_n}}S_n^{1/2}v_n\|_2^2] \sim \chi_p^2$，其中 $v_n = \sqrt{\alpha_n n}(\beta-\hat{\beta})$。因此，按照与确定性 $\alpha_n$ 情形类似的计算，我们有

$$\begin{aligned}
2^{-k_0}\mathbb{E}_{\pi_{n,\hat{\alpha}_n}(\beta|X,Y)}[\|\sqrt{\alpha_n n}(\beta-\beta^*)\|_2^{k_0}] &\leq \mathbb{E}_{\pi_{n,\hat{\alpha}_n}(\beta|X,Y)}\Big[\big\|\sqrt{\alpha_n n}(\beta-\hat{\beta})\big\|_2^{k_0}\Big] + o_{f_{0,n}}(1) \\
&= \Big(\frac{\alpha_n}{\hat{\alpha}_n}\Big)^{k_0}\mathbb{E}_{\pi_{n,\hat{\alpha}_n}(\beta|X,Y)}\Big[\Big(\big\|\sqrt{\frac{\hat{\alpha}_n}{\alpha_n}}v_n\big\|_2^2\Big)^{k_0/2}\Big] + o_{f_{0,n}}(1) \\
&\leq (1+o_{f_{0,n}}(1))\lambda_{\min}(S_n)^{-k_0/2}\mathbb{E}_{u\sim\chi_p^2}[u^{k_0/2}] + o_{f_{0,n}}(1) \\
&= O_{f_{0,n}}(1),
\end{aligned}$$

因此，**(A3)** 对$\hat{}\alpha_n$-后验分布成立。

**指数族：** 设 $X^n$ 具有密度 $f_n(X^n|\eta) = \exp(\eta T(X^n) - n(A(\eta) - B(X^n))$，其中 $\eta \in H \subset \mathbb{R}$ 表示自然参数，且 $A$ 在 $H$的内部二阶可微且严格凸。记最大似然估计为 $\hat{}\eta$，真实自然参数值为$\eta^*$。令 $\pi(\eta) \propto 1$。则 $\alpha_n$-后验分布正比于 $\exp(\alpha_n\eta T(X^n) - \alpha_n n A(\eta))$。我们通过验证命题 5 的条件 1–2，来验证 $\alpha_n$-后验分布满足**(A3)** 。根据命题 8，这也等同于检查**(A3)** 对 $\hat{}\alpha_n$-后验分布是否成立。

条件 *1*。我们将证明，常数 $c_0 = c_1 = 1$ 和 $c_2 = 0$ ——其中 $L$ 稍后指定——会导出条件 1 中的式 (17)。

首先考虑 $\eta > \eta^* + r$ 的情况，其中 $r > \frac{e^{c_2/c_0}}{c_1} = 1$。由于 $\eta > \eta^*$，我们有 $\eta - \eta^* = |\eta - \eta^*|$ 且

$$\begin{aligned}
\log f_n(X^n|\eta) - \log f_n(X^n|\eta^*) &= |\eta-\eta^*|T(X^n) - n(A(\eta)-A(\eta^*)) \\
&= n|\eta-\eta^*|\Big[\frac{1}{n}T(X^n) - A'(\eta^*) - \Big(\frac{A(\eta)-A(\eta^*)}{|\eta-\eta^*|} - A'(\eta^*)\Big)\Big].
\end{aligned} \tag{193}$$

假设在 $f_{0,\ n}$-概率下，$\frac{1}{n}T(X^n) \to A'(\eta^*)$（这在独立同分布设定下成立），则随着样本量增大，该概率趋近于 1，上述表达式可进一步界定如下，其中某个 $t > 0$ 满足

$$\begin{aligned}\log f_n(X^n|\eta) - \log f_n(X^n|\eta^*) &\le -n|\eta-\eta^*|\Big[-t + \frac{A(\eta)-A(\eta^*)}{|\eta-\eta^*|} - A'(\eta^*)\Big] \\ &= -n|\eta-\eta^*|\,[-t + A'(\tilde{\eta}) - A'(\eta^*)] \\ &\le -n|\eta-\eta^*|\Big[-t + \inf_{\{\eta|\eta>\eta^*+r\}}\big(A'(\eta)-A'(\eta^*)\big)\Big] \\ &= -n|\eta-\eta^*|\Big[-t + A'(\eta^*+r) - A'(\eta^*)\Big],\end{aligned}$$

其中 $\tilde{\eta} \in (\eta^*,\ \eta)$，最后一个不等式用到了 $\eta > \eta^* + r$，最后一个等式则由 $A$ 的严格凸性得出。将 $t$ 取得足够小，括号中的项便为正，且 $f_{0,\ n}$-概率趋于 1。

在 $\eta \le \eta^* - r$ 的情况下，我们有 $\eta - \eta^* = -|\eta-\eta^*|$，容易验证可得到类似的界，从而有 $\log f_n(X^n|\eta) - \log f_n(X^n|\eta^*) \le -n|\eta-\eta^*|[-t-(A'(\eta^*-r) - A'(\eta^*))]$. 综合这几种情况，令.....即可满足条件 1 的 (17),

$$L = -t + \max\{A'(\eta^*+r) - A'(\eta^*), A'(\eta^*) - A'(\eta^*-r)\}$$

并且 $\gamma(|\eta-\eta^*|) = |\eta-\eta^*|$。特别地，对于 $v \ge r > 1$，有 $\gamma(v) \ge c_0\log(1+c_1 v) - c_2$，其中常数为 $c_0 = c_1 = 1$ 和 $c_2 = 0$。

条件2。根据中值定理,

$$\log f_n(X^n|\eta) - \log f_n(X^n|\hat{\eta}) = -\frac{n}{2}A''(\hat{\eta} + \xi_n(\eta-\hat{\eta}))(\eta-\hat{\eta})^2,$$

其中 $\xi_n \in (0,1)$。注意，由于 $A$ 的严格 ˆ 凸性，对于所有 $\eta$，$A''(\hat{\eta} + \xi_n(\eta - \hat{\eta}))$ 都严格为正。对于 $\eta \in B_{\hat{\eta}}(2r)$（其中 $r > 1$ 由检验条件 1 得到），我们有 $B_{\hat{\eta}}(2r) \subseteq B_{\eta^*}(4r)$，且当 $f_{0,n}$ 时，该结论以趋于 1 的概率成立，因此,

$$\sup_{\eta\in B_{\hat{\eta}}(2r)} A''(\eta) \le \sup_{\eta\in B_{\eta^*}(4r)} A''(\eta) \quad \text{and} \quad \inf_{\eta\in B_{\hat{\eta}}(2r)} A''(\eta) \ge \inf_{\eta\in B_{\eta^*}(4r)} A''(\eta).$$

结合中值定理的结论与上述界，我们得到该结论以趋于 1 的概率成立,

$$-\frac{n}{2}\inf_{\eta\in B_{\eta^*}(4r)} A''(\eta)(\eta-\hat{\eta})^2 < \log f_n(X^n|\eta) - \log f_n(X^n|\hat{\eta}) < -\frac{n}{2}\sup_{\eta\in B_{\eta^*}(4r)} A''(\eta)(\eta-\hat{\eta})^2,$$

在 $B_{\hat{\eta}}(2r)$ 上一致成立。因此，我们可以取 $c_3 = \frac{1}{2}\inf_{\eta\in B_{\eta^*}(4r)}A''(\eta)$ 和 $c_4 = \frac{1}{2}\sup_{\eta\in B_{\eta^*}(4r)}A''(\eta)$，从而完成证明。

# 事件 E 关于相关工作的补充讨论

贝叶斯统计领域已有大量与我们的贡献相关的研究。最直接相关的，或许是关于开发数据驱动后验温度调整方法的研究脉络。 [30, 31] 提出的 SafeBayesian 算法是该主题最早的论文之一，

探讨了这一主题。该算法选择使累积后验期望对数损失最小的 $\alpha$ 。在[36], 中，作者提出选择 $\alpha$ ，使先验分布与 $\alpha$-后验分布之间的信息期望增益，等于先验分布与标准后验分布之间的信息期望增益。在 [73, 81], 的研究中，温度参数经过校准，以使可信区间达到名义频率学派覆盖概率。在理论设定限制更多的情况下， [14]为吉布斯后验中的最优逆温度参数推导了闭式表达式，该参数可使总体风险最小化。近期研究提出选择 $\alpha$ 以优化后验预测，但也指出，这会导致一个定义不明确的优化问题：在大样本下，温度参数不会产生有意义的影响 [64]。然而，类似方法在广义后验推断中的表现优于标准后验分布 [53]。

另一条重要的相关研究脉络是对贝叶斯后验渐近性质的研究。特别是，我们的工作建立在已有的模型设定错误情形下的贝叶斯—冯·米塞斯定理之上，这些定理针对低维参数模型，其中所得后验分布的渐近行为由一个“伪真实”似然模型决定 [50, 5]。此外，还有大量文献研究参数维度随样本量增加的参数模型高维情形下，后验分布的渐近性质。关于参数数量不断增长时贝叶斯线性回归渐近正态性的结果 $p$ ，已在 [26, 12] 中确立；而关于此类后验分布的线性统计量的中心极限定理，最近已在 [54]中确立，适用于 $p \ll n^{2/3}$ 的情形。此外， $p \to \infty$ 时有效的贝叶斯-冯·米泽斯定理／拉普拉斯近似成立所需的充分条件及其适用的渐近情形，见于 [46, 45, 47]。最后，我们指出，半参数情形下针对标准后验分布的贝叶斯-冯·米泽斯定理已在 [17] 中确立，而针对 $\alpha_n$-后验分布的相应结果则见于 [52]。

此外，关于广义后验分布的伯恩斯坦-冯·米塞斯结果还有一个更广泛的类别 [18, 28, 65, 53, 61]。这些后验分布在标准后验构造中用某种由损失函数取指数后定义的吉布斯分布替代似然。值得注意的是，吉布斯后验是若干统计问题的最优解，例如估计量聚合 [78, 20, 70], 随机模型选择 [63], 以及最小复杂度密度估计/信息复杂度最小化 [7, 83, 84]。

$\alpha$后验之所以广受欢迎，很大程度上源于它们在理论和实践中均展现出的稳健性。这些性质最初是在模型设定错误的线性回归模型中得到探讨的；研究表明， $\alpha$后验的表现优于标准后验分布 [31]。在指数族以及校正过度离散的问题中， [36] 也得出了类似结论。在渐近分析中， [5]表明，$\alpha$后验及其变分近似能够稳健地应对似然的一般性设定错误，前提是这种设定错误发生的概率趋于零。 [66]提出的粗化后验以数据受到扰动为条件，而不是像计算标准后验分布那样，假定数据被精确观测到。这种后验构造明确地纳入了一定程度的模型错设稳健性。对扰动幅度设置先验分布后可以发现，渐近地看，粗化后验近似为幂后验，其幂值取决于样本量。与此略有不同， [56]提出了一种方法：通过幂似然方法，将有偏（即模型设定错误的）观察性数据补充到随机对照试验数据中，同时维持名义上的可信区间覆盖率。

此外，也有人提出了一些方法，以增强标准后验构造对离群值的稳健性。其中，[80] 提出的贝叶斯数据重加权方法或许与 $\alpha$后验最为相似。作者提出，为每个观测值分配一个温度参数 $\alpha_i$ （而不是对整个似然应用单个 $\alpha$ ），并通过为 $\alpha_i$设置先验分布，对 $\alpha_i$和观测值进行联合建模。这种构造的思路是，单个权重可以降低个别离群值对似然总体贡献的影响 [80]。其他针对离群值的稳健化方法，则是通过采用稳健损失 [37, 28, 3, 62, 61] 和检验统计量 [6]来构造广义后验。也有一些研究认为，$\alpha$后验能够提高对先验分布选择的稳健性。[67]在贝叶斯因子模型选择的背景下倡导了这一观点。同样，幂先验通过基于历史数据计算 $\alpha$后验，并将其用作当前研究的先验分布，从而构造信息性先验 [40, 41, 16]。

最后，我们的工作也与一些关于变分推断（VI）的文献相关。VI 是一种基于优化的方法，用于近似不可处理的后验分布 [42,11, 79]。VI 的核心思想是用一个更易处理的分布来近似后验分布 $\pi_n(\theta|X^n)$。为此，我们指定一个计算上易于处理的分布族 $\mathcal{Q}$，并选择与后验分布在 KL散度意义下距离最近的分布 $q^* \in \mathcal{Q}$ 。这一思想的一个变体与 $\alpha$-后验直接相关，具体形式如下，即证据下界（ELBO）最大化的正则化版本：$q^* = \arg\max_{q\in\mathcal{Q}} \int \log f_n(X^n|\theta)\pi(\theta)q(\theta)d\theta - \frac{1}{\alpha} d_{\mathrm{KL}}(q\|\pi)$，其中 $d_{\mathrm{KL}}(q\|\pi) = \int_{\mathbb{R}^p} q(\theta) \log(\frac{q(\theta)}{\pi(\theta)})$ $d\theta$。事实上，当 $\mathcal{Q}$ 不受限制 [5]时，$\alpha$-后验分布会使上述问题达到最大值。文献中将正则化 ELBO 及相关量称为 $\alpha$-VI [38, 82]，这一目标也已用于训练变分自编码器 [34, 15]。另一种类似的目标通过提高似然的影响权重，已被证明能够有效地为过参数化概率模型调整超参数 [32]。$\alpha$-VI 目标还可以进一步推广：用另一个损失函数替代对数似然；当 $\mathcal{Q}$ 受限时，这种方法称为广义变分推断 [51] 。当 $\mathcal{Q}$ 不受限制时，广义后验分布 [10] 会使这一目标达到最大值。

# F 补充实验细节

## F.1 贝叶斯交叉验证公式

我们推导出用于计算实验中采用的两种交叉验证损失的显式公式。与积分 (10) 和 (11) 相比，这些公式更便于使用。

### F.1.1 $\overline{\mathrm{lppd}}_{\mathrm{LOO}}(\alpha)$计算

回顾留一法 $\mathrm{lppd}_{\mathrm{LOO}}(\alpha)$ 的定义，

$$\overline{\mathrm{lppd}}_{\mathrm{LOO}}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \log \int p(x_i, y_i|\beta)\pi_{n,\alpha}(\beta|X_{-i}Y_{-i})d\beta,$$

其中， $X_{-i}$、 $Y_{-i}$ 分别表示移除第 i$^{th}$个条目后的设计矩阵和因变量。在我们的数值实验中，先验分布为 $N(0, I_p)$，上述的 $\alpha_n$-后验分布为 $\pi_{n,\alpha}(\beta|X_{-i}, Y_{-i}) = \phi(\beta|\mu_{n,-i}(\alpha)\ \Sigma_{n,-i}(\alpha))$, 其中

$$\mu_{n,-i}(\alpha) \equiv \Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha} I_p\Big)^{-1} X_{-i}^\top Y_{-i} \quad \text{and} \quad \Sigma_{n,-i}(\alpha) \equiv \frac{\sigma^2}{\alpha}\Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha} I_p\Big)^{-1}. \quad (194)$$

为获得式（8）中的后验分布，我们令 $\sigma^2 = 1$ 以获得岭回归估量，$\mu_{n,\ -i}(\alpha) \equiv \hat{}\beta_{-i}^{\text{Ridge}}$ ，作为 $\alpha_n$-后验均值。在数据应用示例中，我们将 $\hat{}\sigma^2 = \frac{1}{n-p}\sum_{i=1}^n (Y_i - \hat{}Y_i)^2$ 代入上式。将 $\alpha_n$-后验分布的表达式代入 $\text{lppd}_{\text{LOO}}(\alpha)$，得到

$$\overline{\text{lppd}}_{\text{LOO}}(\alpha) = \frac{1}{n}\sum_{i=1}^n \log \frac{\int \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma^2}(y_i - x_i^\top\beta)^2 + (\beta - \mu_{n,-i})^\top \Sigma_{n,-i}^{-1}(\beta - \mu_{n,-i})\right)\right) d\beta}{(2\pi)^{\frac{p}{2}+\frac{1}{2}}\sigma|\Sigma_{n,-i}|^{\frac{1}{2}}},$$

其中，为使记号稍微简洁一些，我们省略了 $\mu_{n,\ -i}$ 和 $\Sigma_{n,\ -i}$ 对 $\alpha$ 的依赖性。在接下来的计算中，我们将不写出上式的分母，因为它在上式中是一个加性常数，不会影响 $\text{lppd}_{\text{LOO}}(\alpha)$ 的优化。配方可得

$$\begin{aligned}
&\frac{1}{\sigma^2}(y_i - x_i^\top\beta)^2 + (\beta - \mu_{n,-i})^\top\Sigma_{n,-i}^{-1}(\beta - \mu_{n,-i}) \\
&= (\beta - \tilde{\mu}_{n,-i})\tilde{\Sigma}_{n,-i}^{-1}(\beta - \tilde{\mu}_{n,-i}) + \Big(\frac{1}{\sigma^2}y_i^2 + \mu_{n,-i}^\top\Sigma_{n,-i}^{-1}\mu_{n,-i} - \tilde{\mu}_{n,-i}^\top\tilde{\Sigma}_{n,-i}^{-1}\tilde{\mu}_{n,-i}\Big) \\
&= (\beta - \tilde{\mu}_{n,-i})\tilde{\Sigma}_{n,-i}^{-1}(\beta - \tilde{\mu}_{n,-i}) + C_{n,-i}, \qquad (195)
\end{aligned}$$

其中

$$\tilde{\mu}_{n,-i} = \tilde{\Sigma}_{n,-i}\Big(\frac{1}{\sigma^2}x_i y_i + \Sigma_{n,-i}^{-1}\mu_{n,-i}\Big) \quad \text{and} \quad \tilde{\Sigma}_{n,-i} = \Big(\frac{1}{\sigma^2}x_i x_i^\top + \Sigma_{n,-i}^{-1}\Big)^{-1}, \qquad (196)$$

并且在 $C_{n,\ -i}$ 中，我们汇总了不依赖于 $\beta$ 的各项。因此，

$$\begin{aligned}
&\int \exp\Big(-\frac{1}{2}\Big(\frac{1}{\sigma^2}(y_i - x_i^\top\beta)^2 + (\beta - \mu_{n,-i})^\top\Sigma_{n,-i}^{-1}(\beta - \mu_{n,-i})\Big)\Big)d\beta \\
&= e^{-C_{n,-i}/2}\int \exp\Big(\frac{-1}{2}(\beta - \tilde{\mu}_{n,-i})^\top\tilde{\Sigma}_{n,-i}^{-1}(\beta - \tilde{\mu}_{n,-i})\Big)d\beta = e^{-C_{n,-i}/2}(2\pi)^{\frac{p}{2}}|\tilde{\Sigma}_{n,-i}|^{\frac{1}{2}}. \qquad (197)
\end{aligned}$$

由 (194)、(195)、(196) 和 (197) 可得 $\text{argmin}_\alpha\{\text{lppd}_{\text{LOO}}(\alpha)\} = \text{argmin}_\alpha\{T_1(\alpha) + T_2(\alpha) + T_3(\alpha)\}$，其中

$$\begin{aligned}
T_1(\alpha) = -\frac{1}{2n}\sum_{i=1}^n \log\Big(\frac{|\tilde{\Sigma}_{n,-i}(\alpha)^{-1}|}{|\Sigma_{n,-i}(\alpha)^{-1}|}\Big) &= -\frac{1}{2n}\sum_{i=1}^n \log\Big(\frac{|\frac{1}{\sigma^2}x_i x_i^\top + \frac{\alpha}{\sigma^2}(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha}I_p)|}{|\frac{\alpha}{\sigma^2}(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha}I_p)|}\Big) \\
&= -\frac{1}{2n}\sum_{i=1}^n \log\Big|\frac{1}{\alpha}\Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha}I_p\Big)^{-1}x_i x_i^\top + I_p\Big|,
\end{aligned}$$

最后一个表达式可通过繁琐但直接的运算得到。

$$T_2(\alpha) = -\frac{1}{2n}\sum_{i=1}^{n} \mu_{n,-i}^\top \Sigma_{n,-i}^{-1}\mu_{n,-i} = -\frac{1}{2n}\sum_{i=1}^{n} \frac{\alpha}{\sigma^2} Y_{-i}^\top X_{-i}\Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha} I_p\Big)^{-1} X_{-i}^\top Y_{-i},$$

and

$$\begin{aligned} T_3(\alpha) &= \frac{1}{2n}\sum_{i=1}^{n} \tilde{\mu}_{n,-i}^\top \tilde{\Sigma}_{n,-i}^{-1}\tilde{\mu}_{n,-i} \\ &= \frac{1}{2n}\sum_{i=1}^{n} \Big[\frac{x_i y_i}{\sigma^2} + \frac{\alpha}{\sigma^2} X_{-i}^\top Y_{-i}\Big]^\top \Big(\frac{x_i x_i^\top}{\sigma^2} + \frac{\alpha}{\sigma^2}\Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha} I_p\Big)\Big)^{-1} \Big[\frac{x_i y_i}{\sigma^2} + \frac{\alpha}{\sigma^2} X_{-i}^\top Y_{-i}\Big]. \end{aligned}$$

在数值实验中，我们针对 $\lambda$ 优化 $\overline{\text{lppd}}_{\text{LOO}}$，并利用关系 $\alpha = {}^{1}\!/_{(n-1)\lambda}$ 将选定的 $\lambda$ 值“转换”为选定的 $\alpha$ 值（见 F.2节）。因此，我们按 ${}^{1}\!/_{n-1}$ 对留一法矩阵进行重新缩放（例如，我们使用 ${}^{1}\!/_{n-1} X_{-i}^\top X_{-i}$ 代替 $X_{-i}^\top X_{-i}$）。在上述公式中，我们已将 ${}^{1}\!/_{n-1}$ 提出并约去。下一小节计算的 $\overline{\text{lppd}}_{\text{LOO-VI}}$ 也同样如此。

### F.1.2 $\overline{\text{lppd}}_{\text{VI-LOO}}$计算

我们还考虑 $\overline{\text{lppd}}_{\text{VI-LOO}}$，其计算中将式 (8) 的高斯分布均值场变分近似的密度代入式 (10)。该密度的参数为

$$\mu_{n,-i}(\alpha) \equiv \Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha} I_p\Big)^{-1} X_{-i}^\top Y_{-i} \quad \text{and} \quad \Sigma_{n,-i}(\alpha) \equiv \frac{\sigma^2}{\alpha}\Big[\text{diag}\Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha} I_p\Big)\Big]^{-1}.$$

我们使用与计算 $\overline{\text{lppd}}$（F.1.1节）类似的方法，得到 $\text{argmin}_\alpha\{\overline{\text{lppd}}_{\text{LOO-VI}}(\alpha)\} = \text{argmin}_\alpha\{\overline{T_1(\alpha)} + T_2(\alpha) + T_3(\alpha)\}$，其中

$$\tilde{T}_1(\alpha) = \frac{-1}{2n}\sum_{i=1}^{n} \log\Big(\frac{|\tilde{\Sigma}_{n,-i}(\alpha)^{-1}|}{|\Sigma_{n,-i}(\alpha)^{-1}|}\Big) = -\frac{1}{2n}\sum_{i=1}^{n} \log\Big|\frac{1}{\alpha}\Big[\text{diag}\Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha} I_p\Big)\Big]^{-1} x_i x_i^\top + I_p\Big|,$$

$$\begin{aligned} \tilde{T}_2(\alpha) &= \frac{-1}{2n}\sum_{i=1}^{n} \mu_{n,-i}^\top \Sigma_{n,-i}^{-1}\mu_{n,-i} \\ &= \frac{-1}{2n}\sum_{i=1}^{n} \frac{\alpha}{\sigma^2} Y_{-i}^\top X_{-i}\Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha} I_p\Big)^{-1} \text{diag}\Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha} I_p\Big)\Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha} I_p\Big)^{-1} X_{-i}^\top Y_{-i}, \end{aligned}$$

and

$$\begin{aligned} T_3 &= \frac{1}{2n}\sum_{i=1}^{n} \tilde{\mu}_{n,-i}^\top \tilde{\Sigma}_{n,-i}^{-1}\tilde{\mu}_{n,-i} \\ &= \frac{1}{2n}\sum_{i=1}^{n} \Big[\frac{x_i y_i}{\sigma^2} + \frac{\alpha}{\sigma^2}\text{diag}\Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha} I_p\Big)\Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha} I_p\Big)^{-1} X_{-i}^\top Y_{-i}\Big]^\top \\ &\Big(\frac{x_i x_i^\top}{\sigma^2} + \frac{\alpha}{\sigma^2}\text{diag}\Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha} I_p\Big)\Big)^{-1}\Big[\frac{x_i y_i}{\sigma^2} + \frac{\alpha}{\sigma^2}\text{diag}\Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha} I_p\Big)\Big(X_{-i}^\top X_{-i} + \frac{\sigma^2}{\alpha} I_p\Big)^{-1} X_{-i}^\top Y_{-i}\Big]. \end{aligned}$$

| 方法 | 参数 | 间距 | 区间 | 密度 | 映射到 $\alpha$ |
|---|---|---|---|---|---|
| BCV | $\lambda$ | 对数 | $[10^{-12}, 10]$ | 200 | $((n-1)\lambda)^{-1}$ |
| BCV + 变分推断 | $\lambda$ | 对数 | $[10^{-12}, 10]$ | 200 | $((n-1)\lambda)^{-1}$ |
| 留一交叉验证 | $\lambda$ | Linear | $[10^{-12}, 30]$ | 200 | $\lambda^{-1}$ |
| 训练-测试 | $\lambda$ | Linear | $[10^{-12}, 5]$ | 200 | $(n\lambda)^{-1}$ |
| SafeBayes | $\alpha$ | Linear | $[0, 1]$ | 30 | $\alpha$ |

表4：第 2 节中描述的模拟示例所用网格。

| 方法 | 参数 | 间距 | 区间 | 密度 | 映射到 $\alpha$ |
|---|---|---|---|---|---|
| BCV | $\lambda$ | 对数 | $[10^{-12}, 10]$ | 200 | $((n-1)\lambda)^{-1}$ |
| BCV + 变分推断 | $\lambda$ | 对数 | $[10^{-12}, 10]$ | 200 | $((n-1)\lambda)^{-1}$ |
| 留一交叉验证 | $\lambda$ | 线性 | $[10^{-12}, 0.5]$ | 200 | $\lambda^{-1}$ |
| 训练-测试 | $\lambda$ | 线性 | $[10^{-12}, 0.05]$ | 200 | $(n\lambda)^{-1}$ |
| SafeBayes | $\alpha$ | 线性 | $[0, 1]$ | 30 | $\alpha$ |

Ta表 5：第2.4节所述数据应用示例中使用的网格 .

## F.2 网格搜索中使用的网格

我们列出了模拟示例（第 2.3 节，表4）和数据应用示例（第2.4节，表5）中网格搜索所使用的网格。每个表格包含以下列:

- 方法：选择方法/损失函数。各方法的详细信息请参见第2.2节。
- 参数：用于最小化损失函数的参数。除 SafeBayes（其优化的是温度参数 $\alpha$ ）外，所有方法优化的都是 $\lambda$ （式 (9) 中的正则化参数）。
- 间距：网格是按线性间隔还是对数间隔排列。
- 区间：网格的上下界。
- 密度：网格点的数量。
- 映射到 $\alpha$：所选参数如何映射回 $\alpha_n$ 在 $\alpha_n$-后验分布中的取值。

根据表4和表5计算出 $\hat{\alpha}_n$ 后，我们将大于 $10^6$ 的 $\hat{\alpha}_n$ 重新编码为“实际上”无穷大，并在图1–3和图10中将这些值标记为 $\infty$ 。

## F.3 $\hat{\alpha}_n$ 与 $n$之间的关系

本节说明我们如何估计表2和表3中出现的$\hat{\alpha}_n$ 趋于 0 的收敛速率。为了得到$\hat{\alpha}_n$ 相对于 $n$的最佳拟合曲线，我们假设 $\alpha = Cn^{\gamma}$，我们的

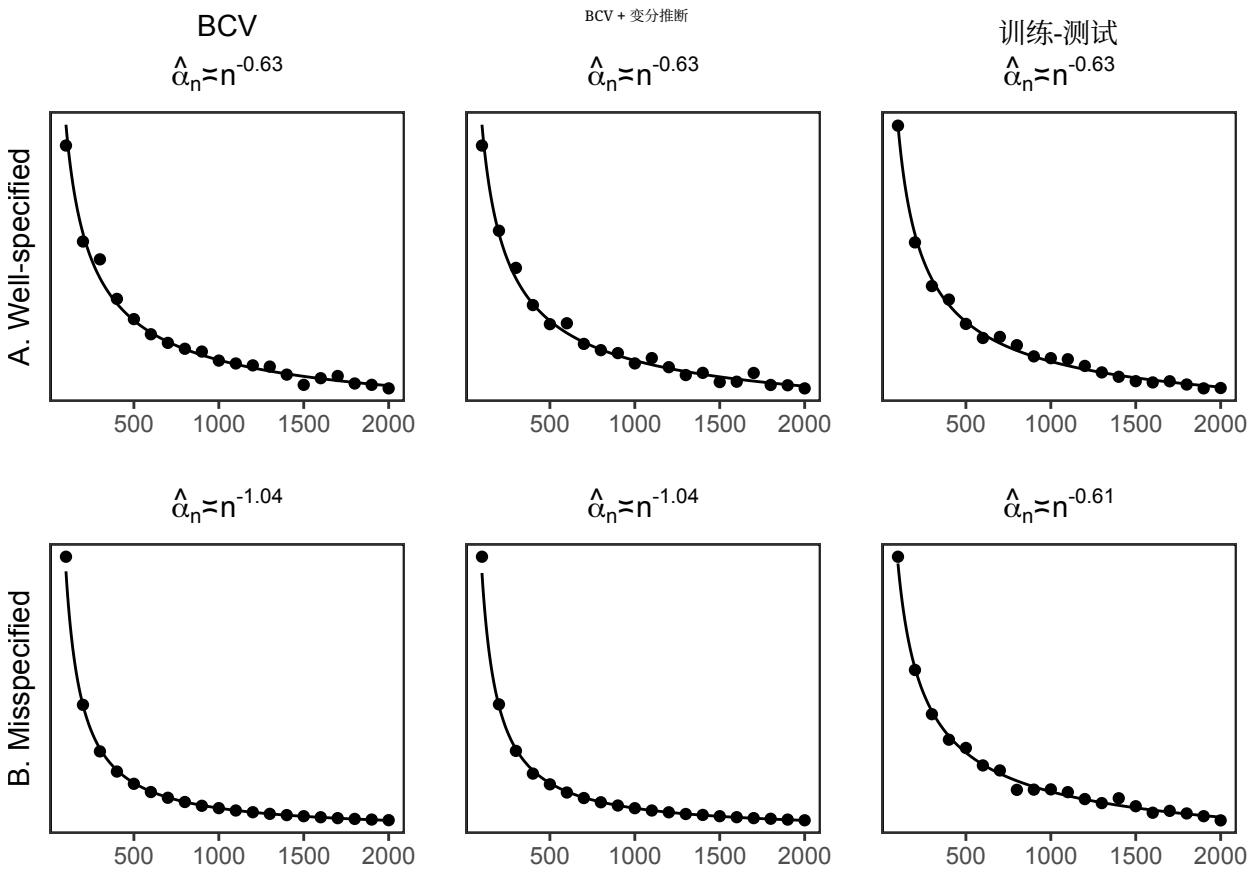



图 6：面板 A 和 B：在模型设定正确的情形（A）和模型设定错误的情形（B）下，$\hat{\alpha}_n$（纵轴）对 $n$（横轴）的依赖性。点表示对 1,000 次重复试验所得的 $\hat{\alpha}_n$ 值取平均后的结果，并叠加了最佳拟合曲线。图标题标示了 $\hat{\alpha}_n$的估计随机阶。在以下情形的绘图和估计中，$\hat{\alpha}_n = \infty$ 处的角点解均被舍弃：BCV/设定正确、BCV+变分推断/设定正确、训练-测试/设定正确以及设定错误。

目标是估计参数 $C$ 和 $\gamma$。我们以 $\log(n)$ 为自变量，对 $\log(\alpha)$ 进行回归，因为 $\log(\alpha) = \log(C) + \gamma \log\, n \equiv \beta_0 + \beta_1 \log\, n$，估计得到的最佳拟合曲线为$\hat{\alpha}_n = e^{\hat{\beta}_0} n^{\hat{\beta}_1}$，且$\hat{\gamma} = \hat{\beta}_1$。我们在表2中报告模拟示例的 $\beta_1$ 估计值及其95%置信区间，并在表3中报告数据应用示例的相应结果。相应的最佳拟合曲线见图 6（模拟示例）和图 7（数据应用示例）。

## F.4 补充图

### F.4.1 在线性尺度上绘制的图2–3

在图2和图3中，我们以对数尺度绘制了使用 BCV、BCV + 变分推断和训练-测试选出的 $\hat{\alpha}_n$ 值。这样做是因为，随着 $n$ 增大，根据这些方法选出的 $\hat{\alpha}_n$ 数量级会发生变化。在统一的对数尺度上绘制 $\hat{\alpha}_n$，便于我们比较随着 $n$ 增大而变化的 $\hat{\alpha}_n$ 分布，同时展示 $\hat{\alpha}_n$ 向 0 收敛的过程；我们还将在表 2 和图 6 中进一步研究这一点。在本附录中，我们还展示了直方图

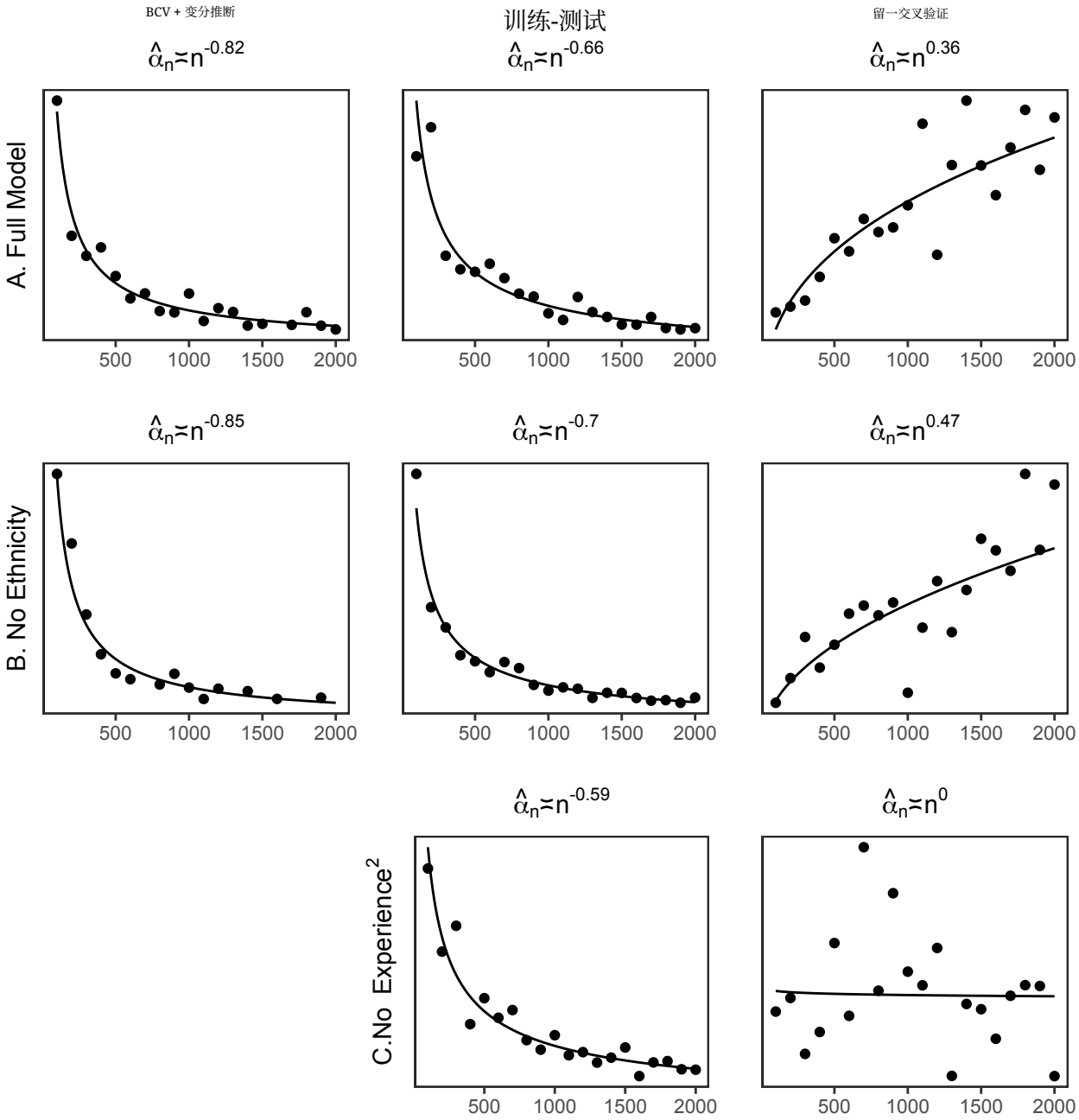



图 7：面板 A、B 和 C：在完整模型设定（A）、剔除族裔的模型（B）以及剔除经验平方项的模型下，$\hat{\alpha}_n$（y 轴）对 $n$（x 轴）的依赖性。各点对应于 $\hat{\alpha}_n$ 值（剔除角点解后），即 100 次重复试验的平均值，并叠加了最佳拟合曲线。图标题表示 $\hat{\alpha}_n$ 的估计随机阶。由于“非角点解”仅在 $n = 100$时出现，因此未绘制剔除经验平方项的模型下 BCV + 变分推断的图。我们还注意到，根据留一交叉验证得到的“非角点”解似乎随着 $n$ 增大，而非减小。为求完整，我们纳入了这一结果。

图2和图3中$\hat{}\alpha_n$ 的$\alpha_n$ ，其$\alpha_n$采用线性尺度绘制（图8和图9）。对于每种方法和样本量，我们使用不同的横轴，以展示随着 $n$ 增长，$\hat{}\alpha_n$的尺度/数量级。

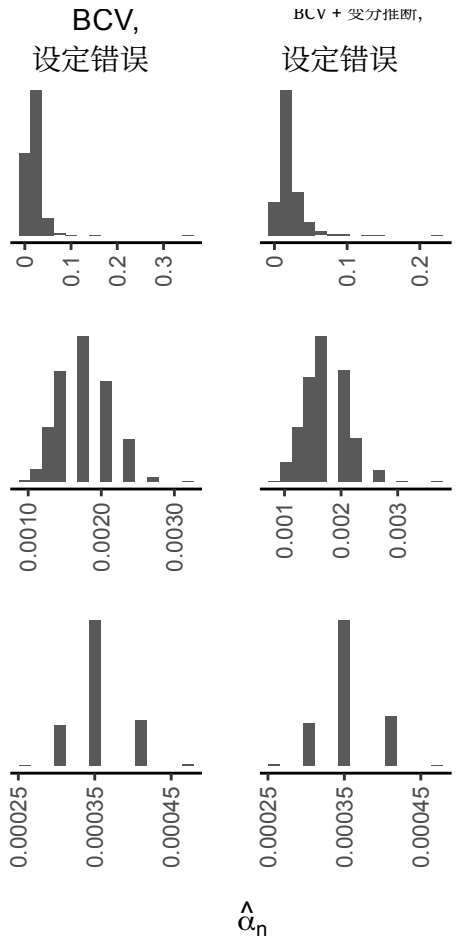



图8：我们复现图2，并在线性尺度上绘制$\hat{}\alpha_n$ 值。对于 $n = 100$、$n = 1$、000和 $n = 5$、000，使用不同的横轴尺度，以展示所选$\hat{}\alpha_n$值随 $n$ 增长而呈现的不同尺度。

### F.4.2 CPS1988示例中所选$\hat{}\alpha_n$ 值的分布

在第2.4节中，我们在 CPS1988 数据集的子样本上重复图1–3中的实验。图10展示了所选$\hat{}\alpha_n$ 的分布。

### F.4.3 子抽样下后验均值估计的稳定性

在第2.4节中，我们考察 CPS1988 数据集，并研究所选$\hat{}\alpha_n$ 以及在模型 (13)–(15) 下得到的$\hat{}\alpha_n$-后验均值估计对子抽样的稳定性。本节在图11a（完整模型）、图11b（不包含族裔的模型）和图11c（不包含工作经验$^2$的模型）中报告$\hat{}\alpha_n$-后验均值估计的结果。

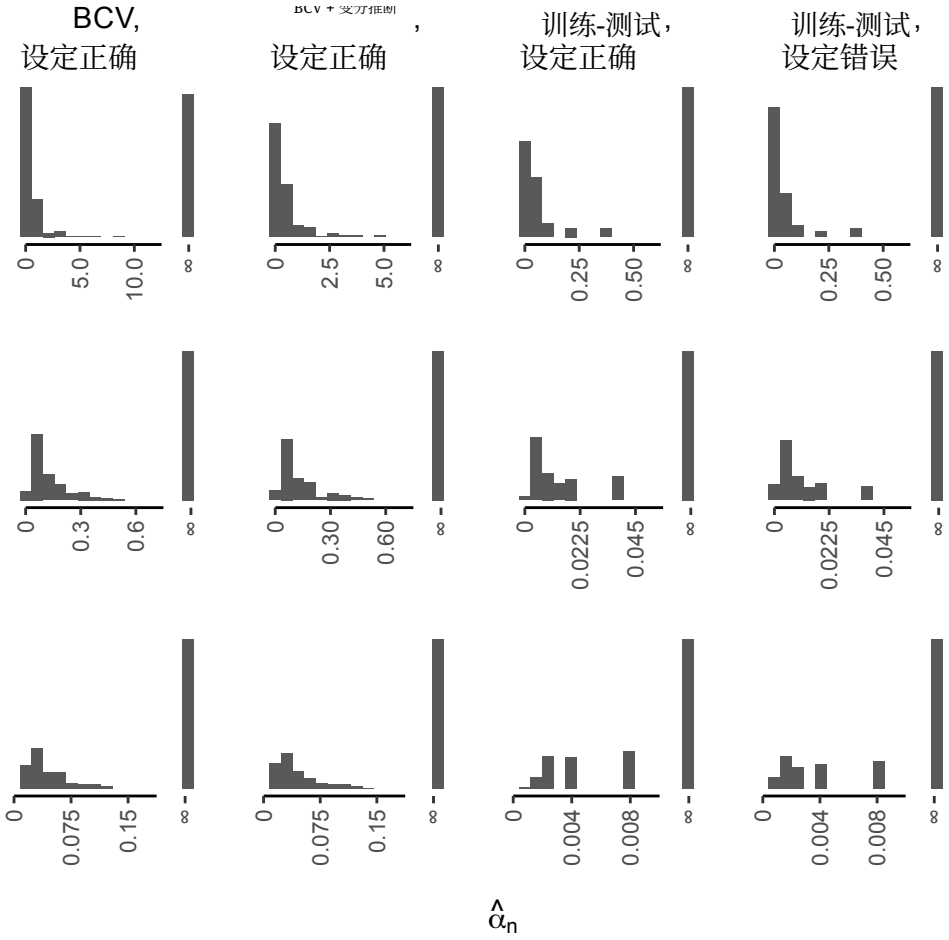



图9：图3中以线性尺度绘制的$\hat{\alpha}_n$ 值。为展示所选$\hat{\alpha}_n$ 值随 $n$ 增长而呈现的不同尺度，$n = 100$、 $n = 1,000$ 和 $n = 5,000$ 使用了不同的横轴尺度。

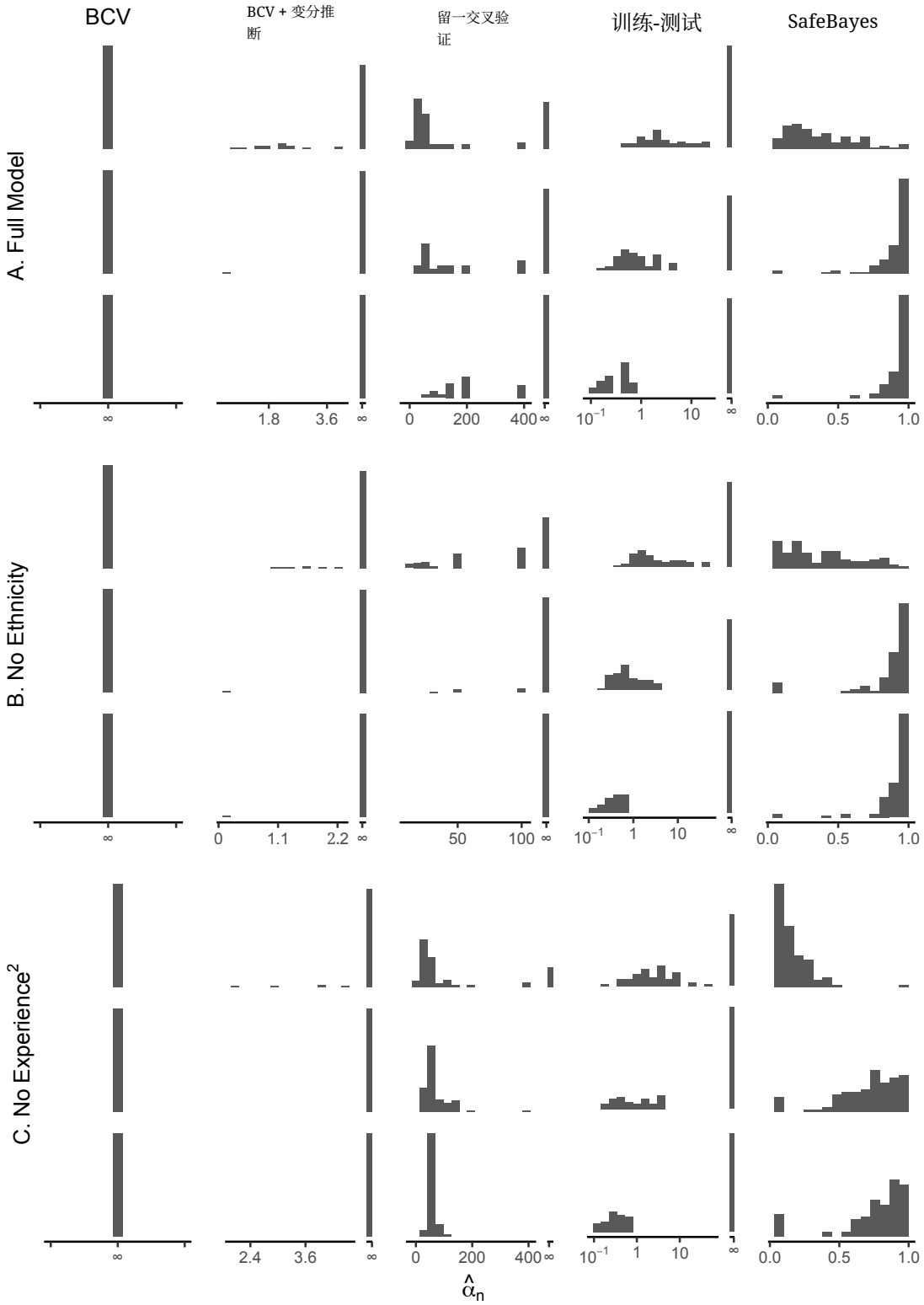
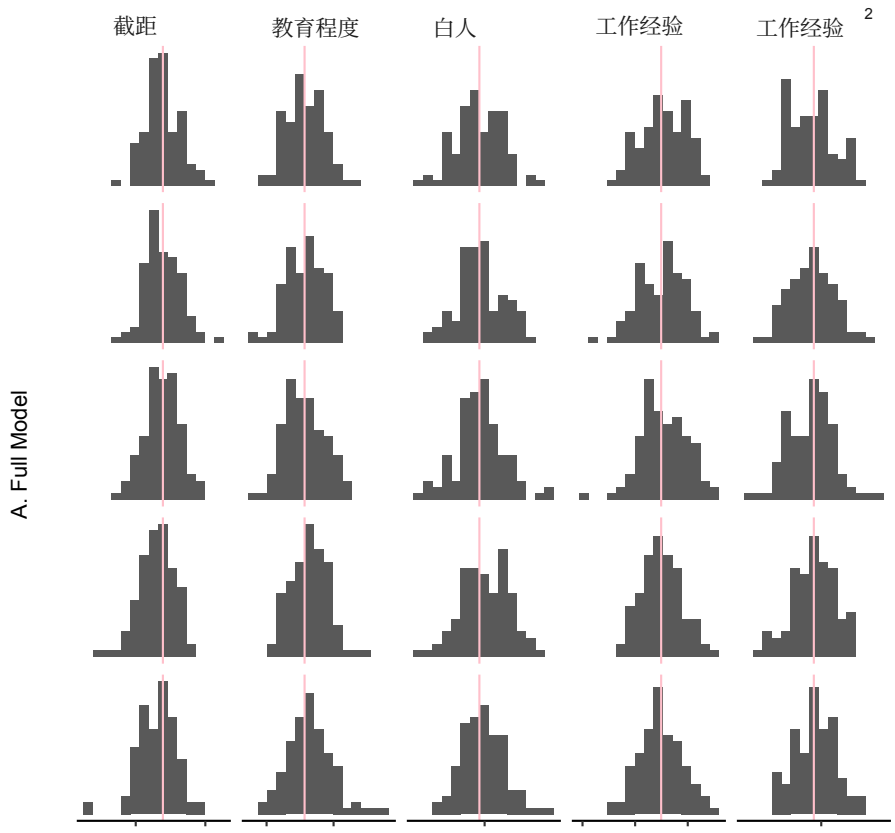
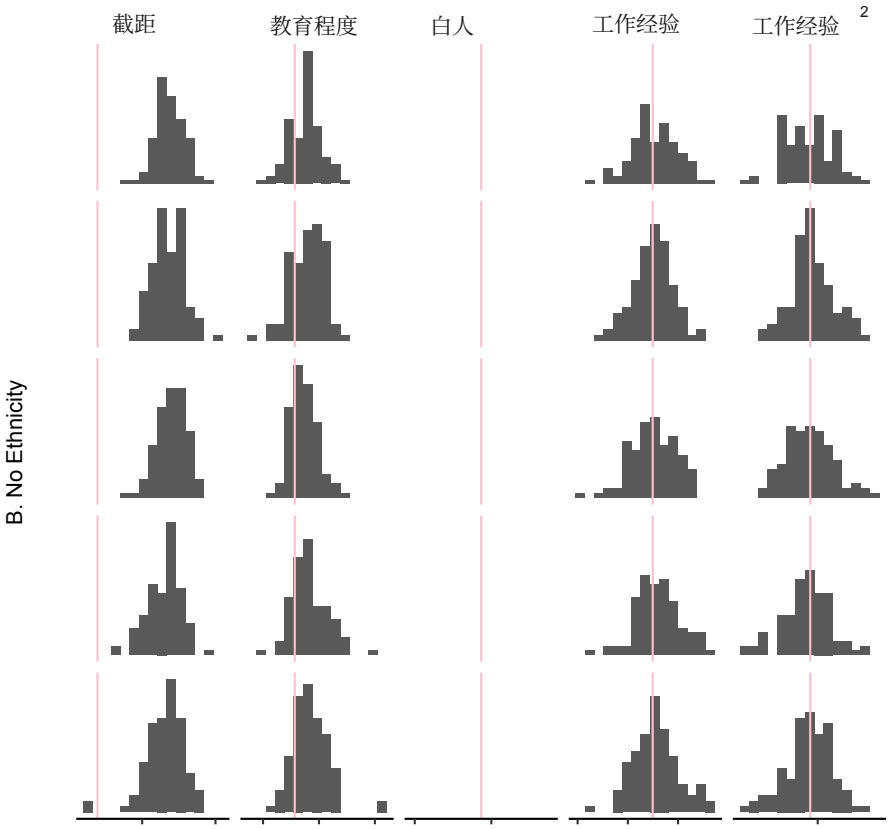


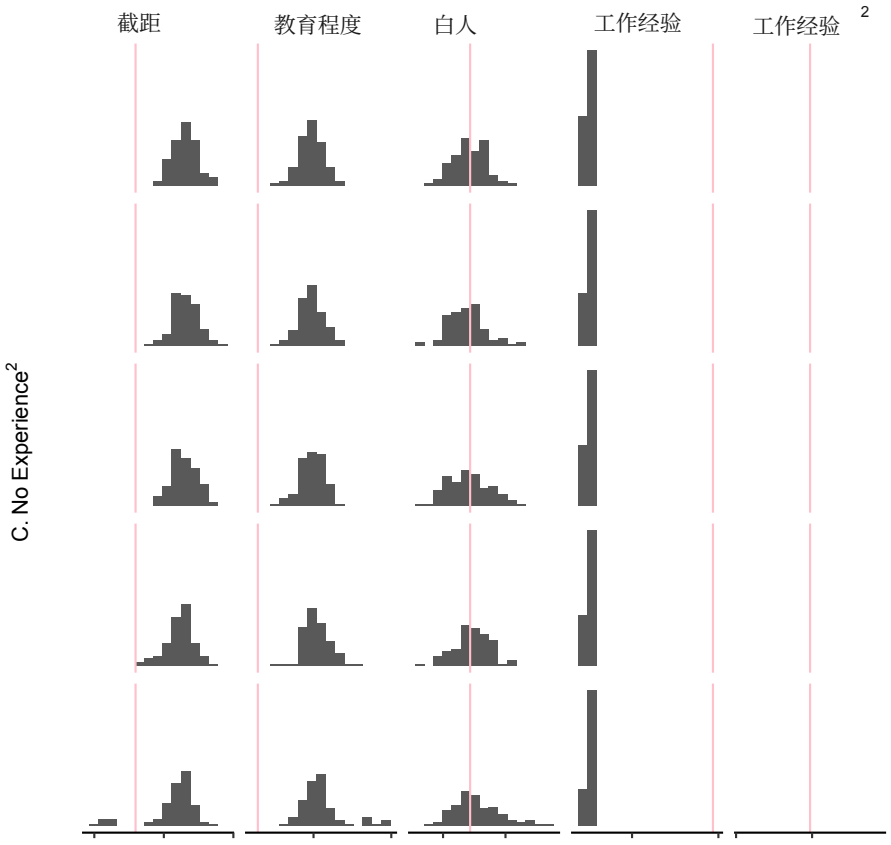

图10：完整模型（面板 A）、不含族裔的模型（面板 B）和不含经验平方项的模型（面板 C）下，CPS1988 数据集的 100 个子样本中 $\hat{\alpha}_n$ 的分布。每个面板中，列对应于 $\alpha$ 的选择方法（贝叶斯交叉验证、BCV $+$ 变分推断、留一交叉验证、训练-测试划分、SafeBayes），行对应于每次重复实验的样本量（$n = 100$, $n = 1,000$, $n = 5,000$）。




（a）完整模型




(b) 不包含族裔的模型




(c) 不包含经验平方项的模型

图11：大小为 $n = 5{,}000$ 的 100 个子样本中各系数所得后验均值的分布图。每一行对应一种选择方法。粉色竖线表示基于完整数据集估计的完整模型系数的最小二乘估计。