arXiv:2602.10256v1 [math.ST] 10 Feb 2026

对数凹后验分布的伯恩斯坦-冯·米塞斯定理

Victor-Emmanuel Brunel *

摘要： 我们证明了伯恩斯坦-冯·米塞斯定理的新型一般形式，适用于正确设定和设定错误的模型；在这些模型中，对数似然关于参数是凹函数，且先验分布是对数凹的。不同于伯恩斯坦-冯·米塞斯定理的经典形式，我们的结果不需要技术性光滑假设，而仅依赖凸分析。

关键词与短语： 贝叶斯估计、伯恩斯坦-冯·米塞斯定理、凸分析、对数凹性、模型设定错误。

1. 引言

1.1 框架

设 (E, E) 为一个可测空间，并考虑由 $(P_\theta)_{\theta\in\Theta_0}$ 构成的概率分布族，它们定义在 (E, E) 上并满足以下条件：

- 集合 Θ_0 是 $\mathbb{R}^d$ 的一个非空开凸子集，其中 $d \geq 1$；
- 族 $(P_\theta)_{\theta\in\Theta_0}$ 是支配的——也就是说，存在一个 σ-有限测度 μ，定义在 (E, E) 上，使得对于所有 $\theta \in \Theta_0$，P_θ 相对于 μ 具有密度 f_θ；
- 对于所有 $\theta \in \Theta_0$，且对于 μ-几乎所有的 $x \in E$，$f_\theta(x) > 0$；
- 对于 μ-几乎所有的 $x \in E$，映射 $\theta \in \Theta_0 \mapsto \phi(x, \theta) := -\log f_\theta(x)$ 是凸的。

考虑一列 $X_1, X_2, \ldots\ldots$定义在 E 上的独立同分布（i.i.d.）随机变量[-]，其分布由某个 $\theta \in \Theta_0$ 给定为 $P_{\bar\theta}$。令 π 为 $\mathbb{R}^d$ 上的一个对数凹分布，其支撑集是一个（闭）凸集 $\Theta \subseteq \Theta_0$。在本文中，我们假设 Θ 具有非空内部。根据 [4, 定理3.2], ，这意味着 π 相对于勒贝格测度具有如下形式的密度： e^{-V}，其中某个凸函数 $V : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ 满足：对于所有 $\theta \in \Theta$， $V(\theta) < \infty$；对于所有 $\theta \notin \Theta$， $V(\theta) = \infty$。

在贝叶斯框架中，如果将 π 视为先验分布，那么对于每个 $n \geq 1$，后验分布记为 $\tilde{Q}_n$，它是在 $\mathbb{R}^d$ 上的（随机）对数凹概率测度，其关于勒贝格测度的密度为：

$$\tilde{q}_n(\theta) = \tilde{Z}_n e^{-n\Phi_n(\theta)-V(\theta)}\mathbb{1}_{\theta\in\Theta}, \quad \forall\theta \in \Theta_0 \tag{1}$$

其中 $\Phi_n(\theta) = n^{-1}\sum_{i=1}^n \phi(X_i, \theta)$， $\theta \in \Theta_0$ 是负对数似然，而 $\tilde{Z}_n$ 是一个（随机）归一化常数（见下文定理 1）。

本文研究在给定X_1、……、 X_n的条件下，对 θ 的重新缩放版本的渐近后验分布，其中适当的重新缩放取决于自然几何条件。我们假设映射 V 性质良好，即 $V(\theta^*) \neq \infty$，并且 V 在 θ^* 处可微（沿切锥 T）。只有当 θ^* 位于 Θ 的边界上时，这才是一项严格条件，因为

*CREST-ENSAE, victor.emmanuel.brunel@ensae.fr

这排除了先验密度 e^{-V} 在 θ 趋近 θ^*时趋于零的情形。尽管这是一个有趣的情形，但需要进一步的技术论证，我们将留待未来工作。在本文其余部分，为简单起见，我们假设 $V = \delta_\Theta$是 Θ的指示函数，也就是说，（可能是不适当的）先验密度为 $\mathbb{1}_{\cdot\in\Theta}$。读者可以自行验证，鉴于上述假设，这只是表述上的简化，而非技术上的简化。

下文中，我们用

$$\Phi(\theta) = \mathbb{E}[\phi(X_1,\theta)] = \int_E \phi(x,\theta)e^{-\phi(x,\theta^*)}\,\mathrm{d}\mu(x) \in \mathbb{R}\cup\{\infty\}, \quad \theta\in\Theta_0.$$

函数 Φ 是凸函数。为简便起见，我们假设对所有 $\theta\in\Theta_0$, $\Phi(\theta)<\infty$ 。注意，$\Phi(\theta)-\Phi(\theta) = \mathrm{KL}(P_{\bar\theta}\|P_\theta)$，其中 KL 表示 Kullback-Leibler 散度。在全文中，我们用 $\theta^* = \mathrm{argmin}_{\theta\in\Theta}\Phi(\theta)$ 表示，并假设 Φ 在 θ^* 的某个邻域内二阶连续可微，且 $\nabla^2\Phi(\theta^*)$ 是正定的。当 $\theta^*\in\mathrm{int}(\Theta)$ 时，$\bar\theta=\theta^*$，我们称贝叶斯模型是设定正确的。当 θ^* 位于 Θ 的边界上，且 $\nabla\Phi(\theta)=0$时，¯，此时同样有 $\theta=\theta^*$。在这种情况下，我们称贝叶斯模型是近似误设的。最后，当 θ^* 位于 Θ 的边界上，且 $\nabla\Phi(\theta^*)\neq 0$时，我们称贝叶斯模型是误设的：¯先验分布的支撑集 π 未能包含 θ。在这种情况下，θ^* 只是参数 $\theta\in\Theta$ 的取值，使得 P_θ 尽可能接近 $P_{\bar\theta}$ ——即 X_i的分布。换言之，P_{θ^*} 是在模型 $(P_\theta)_{\theta\in\Theta}$中对数据分布作出的最佳 KL 近似。在这种情况下，关于 θ^* 的一阶最优性条件指出，$-\nabla\Phi(\theta^*)$ 必须属于 Θ在 θ^*处的法锥。

注意，$\nabla^2\Phi(\theta^*)$ 的正定性意味着 θ^* 是可识别的，即 $P_\theta = P_{\theta^*}$ 仅当 $\theta=\theta^*$时才成立，因此 θ^* 是 Φ 在 Θ上的唯一极小点。在正确设定和近似设定错误情形下，矩阵 $\nabla^2\Phi(\theta^*)$ 就是 Fisher 信息矩阵，因而是正定的。此外，ϕ关于 Θ 的凸性，以及 Φ 在 θ^* 的某个邻域内二阶连续可微这一事实，表明后验分布是适当的，从而验证了˜(Z_n 中的定义（1），如下列定理所示。

THEOREM 1. 在上述假设下，以概率1，映射 $\theta\in\Theta_0\mapsto e^{-n\Phi_n(\theta)}\mathbb{1}_{\theta\in\Theta}$关于勒贝格测度的积分有限。

该定理的证明也对我们的整体工作进行了合理性验证，见附录A节。

以下结果将在本文中用到。根据 [5,定理 2]，，可以定义 $\mathbb{R}^d$中的独立同分布随机变量$U_1,\dots,U_n$，它们具有如下性质：对于每个 $i=1,\dots,n$，U_i几乎必然是$\phi(X_i,\cdot)$在 θ^*处的一个次梯度。此外，[5, 定理 3] 表明这些 U_i是可积的，且 $\mathbb{E}[U_i]=\nabla\Phi(\theta^*)$。我们还对这些 U_i作出如下假设；若 $\phi(X_1,\theta)$在 θ^*的任意小邻域内对所有 θ 都平方可积，则根据 [5, 定理3和推论 2]，，该假设自动成立。

ASSUMPTION 1. $\mathbb{E}[\|U_1\|^2]<\infty$。

本文结构如下。在第2节，我们讨论良设定情形。在第3节，我们处理近似设定错误情形，第 4 节则专门讨论模型设定错误情形。模型设定错误情形与前两种情形有本质区别，并且需要更多几何工具。我们将看到，θ 的适当缩放因情形而异。仅在最后一种情形中，我们假定 Θ 可以写成由二阶可微凸函数所定义的有限多个约束的集合。在结论（5）中，我们简要讨论为何在覆盖前两种情形时对 Θ 无需任何假设，而对 Θ的边界、至少在 θ^*附近的某些假设，在模型设定错误情形中似乎是必要的。

1.2 相关工作与贡献

在贝叶斯框架中，伯恩斯坦-冯·米塞斯定理给出了后验分布在样本量趋于无穷时的极限形状——通常为高斯分布—— [8, 14, 19]。常见假设偏于技术性，包括对数似然函数的光滑性以及一致支配条件。借助似然几何，伯恩斯坦-冯·米塞斯定理向非正则模型的推广也已针对指数族得到探讨 [11]。

伯恩斯坦-冯·米塞斯定理也已扩展到非参数和高维设定 [7] （近期综述见 [6] ）。与我们的工作密切相关的一系列研究关注正则的错设参数模型，即先验分布的支撑集不包含数据分布的真实参数的情形 [1, 2, 13]。然而，这些研究也依赖于对数似然函数的技术条件，并且通常假设Θ具有非常简单的形式，例如 $\Theta = (\mathbb{R}_+)^d$。近期另一系列研究也探讨了用于误差量化的伯恩斯坦-冯·米塞斯定理的拉普拉斯近似 [12, 16]。

在本文中，我们关注经典伯恩斯坦-冯·米塞斯定理的推广，只假设后验分布具有对数凹性。我们的目标是证明，在这样的凸性条件下——并额外要求 Kullback-Leibler 散度在 θ^* 附近二阶可微——无需任何光滑性或支配性假设。我们同时考虑正确设定和设定错误的模型；在这些模型中，对数凹先验分布的支撑集可能不包含真实参数。不同于近期的研究 [2]，我们考虑了范围广泛的支撑集，包括光滑凸集和多面体。特别地，我们证明，后验分布的极限形状支撑在 Θ 沿着 $-\nabla\Phi(\theta^*)$ 方向、位于 θ^* 处的二阶切集上。据我们所知，即使在正确设定情形下，我们的结果也是新的。

这与最大似然估计量（MLE）的渐近分析有相似之处：在设定正确的模型中，经典结果在对数似然满足技术性的光滑性／支配性假设时，推导出经过缩放的 MLE 的渐近分布[15, 17, 18] ；对于约束最大似然估计量的研究也扩展到了设定错误的模型 [10]。这些研究所作的假设与经典伯恩斯坦-冯·米塞斯定理中的假设相似。近期的一项研究 [5] 在对损失函数和约束集作出最小凸性假设的条件下，推导出经过缩放的约束最大似然估计量（更一般地说，约束 M 估计量）的渐近分布，从而能够考虑非光滑模型。我们的工作借鉴了 [5]，的相似思路，这些思路源自凸分析和凸优化。

1.3 记号

集合 $A \subseteq \mathbb{R}^d$ 的内部记为 $\mathrm{int}(A)$。若 A 是凸集，其相对内部记为 $\mathrm{relint}(A)$：这是相对于 A 的仿射张成空间的拓扑而言， A 的内部。

给定两个随机变量 U 和 V，我们将 U 在给定V条件下的条件分布记为 $P_{U|V}$。

两个定义在 $\mathbb{R}^d$ 上的 Borel概率测度 P 和 Q之间的全变差距离定义为

$$\mathrm{TV}(P,Q) = \sup_{B\in\mathcal{B}(\mathbb{R}^d)} |P(B) - Q(B)| = \int_{\mathbb{R}^d} |p(x) - q(x)|\,\mathrm{d}\mu(x)$$

其中，$\mathrm{B}(\mathbb{R}^d)$ 是 $\mathbb{R}^d$ 和 p 的 Borel σ-代数，而 q 和 P 分别是 Q 和 μ相对于共同控制测度 μ 的密度（本文中μ 将取为勒贝格测度）。

在下文中，我们将 Θ 在 θ^*处的切锥记为 T，即 $T = \bigcup p(\Theta - \theta^*)$，并将$p\geq 1$ C 的闭包称为 Θ 在 θ^*处的支撑锥。 Θ 在 θ^* 处的法锥记为 N，它是 C的极锥： $N = \{t \in \mathbb{R}^d : t^\top(\theta - \theta^*) \leq 0, \forall \theta \in \Theta\} = \{t \in \mathbb{R}^d = t^\top u \leq 0, \forall u \in C\}$.

2. 良设定情形

在本节中，我们假设 θ^* 是 Θ的一个内点。因此， $C = \mathbb{R}^d$ 成立，且 θ^* 的一阶条件可简化为 $\nabla\Phi(\theta^*) = 0$。特别地，这意味着 Φ 的约束极小化子与无约束极小化子重合。在这种情况下，我们考虑对 θ进行如下缩放：$t = \sqrt{n}(\theta - \theta^*)$，我们的目标是推导在给定 $X_1,\ldots,X_n$的条件下， t的后验分布Q_n的渐近行为。本节的主要结果是后验分布的如下渐近正态性，它推广了光滑情形下经典的伯恩斯坦-冯·米塞斯定理。记为 $Y_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n U_i$。

—=

THEOREM 2. 设假设 *1* 成立，并假设 $\theta^* \in \mathrm{int}(\Theta)$。那么，

$$TV(Q_n, \mathcal{N}_d(-\nabla^2\Phi(\theta^*)^{-1}Y_n, \nabla^2\Phi(\theta^*)^{-1})) \xrightarrow[n\to\infty]{} 0$$

依概率。

注意到 $\sqrt{n}(\hat{\theta}_n - \theta^*) + \nabla^2\Phi(\theta^*)Y_n \xrightarrow[n\to\infty]{} 0$ 依概率，正如 [5]，中定理 7 的证明所示，其中 $\hat{\theta}_n$ 是受约束的[1] 最大似然估计量。因此，很容易验证，上述陈述也可以等价地写作：

$$\mathrm{TV}(Q_n, \mathcal{N}_d(\sqrt{n}(\hat{\theta}_n - \theta^*), \nabla^2\Phi(\theta^*)^{-1})) \xrightarrow[n\to\infty]{} 0$$

依概率，或者等价地写作：

$$\mathrm{TV}(P_{\sqrt{n}(\theta-\hat{\theta}_n)|X_1,\ldots,X_n}, \mathcal{N}_d(0, \nabla^2\Phi(\theta^*)^{-1})) \xrightarrow[n\to\infty]{} 0$$

$\hat{}$ 依概率（注意 θ_n是 X_1，……， X_n 的可测函数）。

特别地，当 $\Theta_0 = \Theta = \mathbb{R}^d$ 时，我们在极弱的假设下恢复了伯恩斯坦-冯·米塞斯定理最经典的版本：除其期望的二阶导数存在之外，我们不对负对数似然函数 ϕ 作任何光滑性假设。我们认为这一结果是新的。

PROOF OF THEOREM 2. 首先，不失一般性——也为了简便——我们假设 $\nabla^2\Phi(\theta^*) = I_d$。事实上，我们可以考虑用重新缩放 $\tilde{t} := \nabla^2\Phi(\theta^*)^{-1/2}t$ 代替t，这将需要类似的分析。

利用（1），注意 Q_n 关于勒贝格测度具有密度，该密度正比于 $t \in \mathbb{R}^d \mapsto e^{-n\left(\Phi_n(\theta^*+t/\sqrt{n})-\Phi(\theta^*)\right)}\mathbb{1}_{t\in\sqrt{n}(\Theta-\theta^*)}$。

对于 $t \in \mathbb{R}^d$，记

$$G_n(t) = n\left(\Phi_n(\theta^* + t/\sqrt{n}) - \Phi(\theta^*)\right) - t^\top Y_n.$$

那么， q_n 的 Q_n 密度可写为

$$q_n(t) = \frac{Z_n}{(2\pi)^{d/2}}\exp\left(-G_n(t) - t^\top Y_n - \|Y_n\|^2/2\right)\mathbb{1}_{t\in\sqrt{n}(\Theta-\theta^*)}, \quad \forall t \in \mathbb{R}^d, \tag{2}$$

对于某个（随机的）归一化常数 $Z_n > 0$。我们的目标是证明 $Z_n \xrightarrow[n\to\infty]{} 1$ 依概率。

[1]由于 $\theta^* \in \mathrm{int}(\Theta)$，根据 [5, 定理 4]，当 n 足够大时，约束最大似然估计量与无约束最大似然估计量以概率 1 重合。

任取一个 $\varepsilon > 0$。在假设 1下，中心极限定理表明序列 $(Y_n)_{n\geq 1}$是紧的。因此，可以找到 $R > 0$ 使得对所有 $n \geq 1$,

$$P(\|Y_n\| \leq R) \geq 1 - \varepsilon. \tag{3}$$

下面的引理在 [5,定理 7] 的证明中得到了证明，是我们分析的基础。

LEMMA 1. 对于所有紧子集 $K \subseteq \mathbb{R}^d$,

$$\sup_{t\in K} \left| G_n(t) - \frac{\|t\|^2}{2} \right| \xrightarrow[n\to\infty]{} 0$$

依概率。

因此，取 $K = B(0, 4R)$，可得对于所有充分大的 $n \geq 1$，以至少 $1-\varepsilon$的概率,

$$\sup_{\|t\|\leq 4R} |G_n(t) - \|t\|^2/2| \leq \varepsilon. \tag{4}$$

暂且假设 $\|Y_n\| \leq R$ 且 $\sup_{\|t\|\leq 4R}|G_n(t) - \|t\|^2/2| \leq \varepsilon$ – 结合（3）和（4），由并集界可知，该事件发生的概率至少为 $1-2\varepsilon$，对所有充分大的 n。

G_n 的凸性意味着，对于所有满足 $\|t\| > 4R$ 的 $t \in \mathbb{R}^d$ （将 $4Rt/\|t\|$ 表示为 t 与 0 的凸组合），

$$\begin{aligned} G_n(t) &\geq \frac{\|t\|}{4R} G_n(4Rt/\|t\|) - \frac{\|t\| - 4R}{4R} G_n(0) \geq \frac{\|t\|}{4R}(8R^2 - \varepsilon) - \frac{\|t\| - 4R}{4R}\varepsilon \\ &= (2R - \varepsilon/(2R))\|t\| + \varepsilon \geq (2R - \varepsilon/(2R))\|t\|. \end{aligned} \tag{5}$$

因此，对于所有 $t \in \mathbb{R}^d$,

$$\begin{aligned} q_n(t) &\leq \frac{Z_n}{(2\pi)^{d/2}} \left(e^{\varepsilon - \frac{1}{2}\|t+Y_n\|^2} \mathbb{1}_{\|t\|\leq 4R} + e^{-\varepsilon - (2R-\varepsilon/(2R))\|t\| - t^\top Y_n - \|Y_n\|^2/2} \mathbb{1}_{\|t\|>4R} \right) \\ &= \frac{Z_n}{(2\pi)^{d/2}} (A + B) \end{aligned} \tag{6}$$

其中，对于第二项（对应于 $\|t\| > 2R$），我们使用了（5）。

首先，将指示函数简单地上界为 1，以控制 A 项，从而得到 $A \leq e^{\varepsilon} e^{\frac{1}{2}\|t+Y_n\|^2}$。对于 B 项，我们使用柯西-施瓦茨不等式，得到 $-t^\top Y_n \leq R\|t\|$，因此

$$B \leq e^{-\varepsilon - (R-\varepsilon/(2R))\|t\|} \mathbb{1}_{\|t\|>4R}.$$

F最后，对 $\mathbb{R}^d$ 上的勒贝格测度对（6）积分，得到 n

$$1 = \int_{\mathbb{R}^d} q_n(t)\,\mathrm{d}t \leq \frac{Z_n}{(2\pi)^{d/2}} \left(e^{\varepsilon}(2\pi)^{d/2} + e^{-\varepsilon} \int_{\mathbb{R}^d \setminus B(0,4R)} e^{-(R-\varepsilon/(2R))\|t\|}\,\mathrm{d}t \right).$$

由于通过选择足够大的 R，右侧出现的积分可以变得任意小（回想一下，R 是由（3）定义的，并且可以取任意大），因此我们可以选择 R，使得第二个积分不大于 $(2\pi)^{d/2}\varepsilon$，从而得到 $1 \leq Z_n(e^{\varepsilon} + \varepsilon)$，也就是说，$Z_n \geq (e^{\varepsilon} + \varepsilon)^{-1}$。

另一方面，由（2），我们先假设 n 足够大，从而将 q_n 从下方估计为 $B(0,\ 4R) \subseteq \sqrt{n}(\Theta - \theta^*)$（这是可能的，因为 θ^* 位于 Θ的内部）。由此得到$\mathbb{1}_{t\in\sqrt{n}(\Theta-\theta^*)} \geq \mathbb{1}_{\|t\|\leq 4R}$，因此对于所有 $t \in \mathbb{R}^d$，

$$q_n(t) \geq \frac{Z_n}{(2\pi)^{d/2}} e^{-\varepsilon - \frac{1}{2}\|t+Y_n\|^2} \mathbb{1}_{\|t\|\leq 4R} \geq \frac{Z_n}{(2\pi)^{d/2}} e^{-\varepsilon - \frac{1}{2}\|t+Y_n\|^2} \mathbb{1}_{\|t+Y_n\|\leq 3R}$$

因此，通过对勒贝格测度进行积分，

$$1 = \int_{\mathbb{R}^d} q_n(t)\,\mathrm{d}t \geq \frac{Z_n}{(2\pi)^{d/2}} e^{-\varepsilon} \int_{B(0,3R)} e^{-\|u\|^2/2}\,\mathrm{d}u.$$

由于右端的积分在 $R \to \infty$时趋于$(2\pi)^{d/2}$，我们选取足够大的 R，使得该积分不小于$(2\pi)^{d/2}e^{-\varepsilon}$。由此可得 $Z_n \leq e^{2\varepsilon}$。

最后，回顾一下，我们已经证明：对任意固定的 $\varepsilon > 0$，当 n足够大时，$P((e^{\varepsilon}+\varepsilon)^{-1} \leq Z_n \leq e^{2\varepsilon})$ $\geq 1 - 2\varepsilon$。因此，$Z_n \xrightarrow[n\to\infty]{} 1$ 依概率成立，这正是我们想要证明的。

现在，对于证明的最后一步，我们定义 $p_n : t \in \mathbb{R}^d \mapsto (2\pi)^{-d/2} e^{-\|t+Y_n\|^2/2}$，它是$\mathrm{N}_d(-Y_n, I_d)$的密度。我们现在的目标是证明

$$\int_{\mathbb{R}^d} |p_n(t) - q_n(t)|\,\mathrm{d}t \xrightarrow[n\to\infty]{} 0$$

依概率。首先，注意对所有 $x \in \mathbb{R}$, $|x| = 2x_+ - x$，其中 $x_+ = \max(x,0)$。将此应用于上面积分中所有 $t \in \mathbb{R}^d$ 对应的 $x = p_n(t) - q_n(t)$，并回想 p_n 和 q_n 都是密度，因此它们的积分均为 1，我们要证明

$$\int_{\mathbb{R}^d} \left(1 - Z_n e^{-(G_n(t) - \|t\|^2/2)} \mathbb{1}_{t\in\sqrt{n}(\Theta-\theta^*)}\right)_+ p_n(t)\,\mathrm{d}t \xrightarrow[n\to\infty]{} 0$$

依概率。此外，由于 $x \in \mathbb{R}_+ \mapsto x_+$ 的次可加性，上述积分可按如下方式界定

$$\begin{aligned}
&\int_{\mathbb{R}^d} \left(1 - Z_n e^{-(G_n(t) - \|t\|^2/2)} \mathbb{1}_{t\in\sqrt{n}(\Theta-\theta^*)}\right)_+ p_n(t)\,\mathrm{d}t \\
&\qquad = \int_{\mathbb{R}^d} \left(1 - Z_n + Z_n - Z_n e^{-(G_n(t) - \|t\|^2/2)} \mathbb{1}_{t\in\sqrt{n}(\Theta-\theta^*)}\right)_+ p_n(t)\,\mathrm{d}t \\
&\qquad \leq (1 - Z_n)_+ \int_{\mathbb{R}^d} p_n(t)\,\mathrm{d}t + Z_n \int_{\mathbb{R}^d} \left(1 - e^{-(G_n(t) - \|t\|^2/2)} \mathbb{1}_{t\in\sqrt{n}(\Theta-\theta^*)}\right)_+ p_n(t)\,\mathrm{d}t \\
&\qquad = (1 - Z_n)_+ + Z_n \int_{\mathbb{R}^d} \left(1 - e^{-(G_n(t) - \|t\|^2/2)} \mathbb{1}_{t\in\sqrt{n}(\Theta-\theta^*)}\right)_+ p_n(t)\,\mathrm{d}t.
\end{aligned}$$

由于 $Z_n \xrightarrow[n\to\infty]{} 1$ 依概率成立，因此只需证明

$$\int_{\mathbb{R}^d} \left(1 - e^{-(G_n(t) - \|t\|^2/2)} \mathbb{1}_{t\in\sqrt{n}(\Theta-\theta^*)}\right)_+ p_n(t)\,\mathrm{d}t \xrightarrow[n\to\infty]{} 0 \tag{7}$$

依概率。固定 $\varepsilon > 0$ 并且，如上所述，令 R 满足（3）。再次假设 $\|Y_n\| \leq R$ 且 $\sup_{\|t\|\leq 4R} |G_n(t)$ $- \|t\|^2/2| \leq \varepsilon$。

首先，将（7）中的积分分解为在 $B(0,\ 4R)$ 上的一个积分，以及在$\mathbb{R}^d \setminus B(0,\ 4R)$ 上的另一个积分。对于第一个积分，利用映射 $x \in \mathbb{R} \mapsto x_+$ 是单调不减的这一事实，可得

$$\begin{aligned}
\int_{B(0,4R)} \left(1 - e^{-(G_n(t) - \|t\|^2/2)} \mathbb{1}_{t\in\sqrt{n}(\Theta-\theta^*)}\right)_+ p_n(t)\,\mathrm{d}t &\leq \int_{\mathbb{R}^d} (1 - e^{-\varepsilon} \mathbb{1}_{t\in\sqrt{n}(\Theta-\theta^*)})_+ p_n(t)\,\mathrm{d}t \\
&= \int_{\mathbb{R}^d} (1 - e^{-\varepsilon})_+ p_n(t)\,\mathrm{d}t = 1 - e^{-\varepsilon}
\end{aligned}$$

只要 n 足够大，使得在最后一个等号中 $B(0,\ 4R) \subseteq \sqrt{n}(\Theta - \theta^*)$。对于第二个积分，我们只需对所有 $x \geq 0$ 使用不等式 $(1-x)_+ \leq 1$，即可得到

$$\begin{aligned}
&\int_{\mathbb{R}^d \setminus B(0,4R)} \left(1 - e^{-(G_n(t) - \|t\|^2/2)} \mathbb{1}_{t \in \sqrt{n}(\Theta - \theta^*)}\right)_+ p_n(t)\,\mathrm{d}t \\
&\qquad \leq \int_{\mathbb{R}^d \setminus B(0,4R)} p_n(t)\,\mathrm{d}t \\
&\qquad \leq (2\pi)^{-d/2} \int_{\mathbb{R}^d \setminus B(0,4R)} e^{-\frac{1}{2}(\|t\|^2 - 2R\|t\|)}\,\mathrm{d}t \\
&\qquad \leq (2\pi)^{-d/2} \int_{\mathbb{R}^d \setminus B(0,4R)} e^{-\frac{\|t\|^2}{4}}\,\mathrm{d}t
\end{aligned}$$

根据柯西-施瓦茨不等式，并结合 $\|Y_n\| \leq R$这一事实。由于最后一个积分在 $R \to \infty$时趋于 0，因此我们可以选择 R 足够大，使得最后一个公式右侧不大于 ε。最后，假设 n 足够大（因此 $B(0,4R) \subseteq \sqrt{n}(\Theta - \theta^*)$），我们证明了：根据并集界，以至少 $1-2\varepsilon$ 的概率成立的事件上，式（7）中的积分不超过 $1 - e^{-\varepsilon} + \varepsilon$。证明至此结束。

□

3. 近似设定错误情形

这里，我们假设 θ^* 位于 Θ 的边界上，且 $\nabla\Phi(\theta^*) = 0$。在这种情况下，Φ 的约束极小化子与无约束极小化子仍然重合，但分析后验分布还需要进一步论证。我们仍然考虑以下变量替换：$t = \sqrt{n}(\theta - \theta^*)$，其后验分布 Q_n 相对于勒贝格测度具有密度 q_n，可写为

$$q_n(t) = Z_n A_n e^{-G_n(t) - t^\top Y_n - \|Y_n\|^2/2} \mathbb{1}_{t \in \sqrt{n}(\Theta - \theta^*)}, \quad \forall t \in \mathbb{R}^d$$

其中

- A_n 是满足 $A_n \int_T e^{-\frac{\|t+Y_n\|^2}{2}}\,\mathrm{d}t = 1$ 的（随机）归一化常数——请回忆，T是 Θ 在 θ^* 处的切锥；并且由于假设 Θ 具有非空内部，T 也具有非空内部；
- Z_n 是确保 $\int_{\mathbb{R}^d} q_n(t)\,\mathrm{d}t = 1$ 的（随机）归一化常数。

下文中，我们记 $T_n = \sqrt{n}(\Theta - \theta^*)$，因此序列 $(T_n)_{n \geq 1}$ 单调不减，并满足 $\bigcup_{n \geq 1} T_n = T$。

本节的主要结果如下定理。给定 $a \in \mathbb{R}^d$ 和一个正定对称矩阵 $A \in \mathbb{R}^{d \times d}$，我们用 $\mathbf{N}_d^T(a,\ A)$ 表示均值为 a、协方差矩阵为 A，且条件限制在 T中的高斯分布。也就是说，其相对于勒贝格测度的密度正比于 $t \in \mathbb{R}^d \mapsto e^{-\frac{\|\Sigma^{-1/2}(t-a)\|^2}{2}} \mathbb{1}_{t \in T}$。

Theorem 3. 设假设 *1* 成立。假设 θ^* 位于 Θ 的边界上，且 $\nabla\Phi(\theta^*) = 0$。那么，

$$TV(Q_n, \mathcal{N}_d^T(-\nabla^2\Phi(\theta^*)^{-1} Y_n, \nabla^2\Phi(\theta^*)^{-1})) \xrightarrow[n \to \infty]{} 0$$

依概率。

Remark 1. 请注意，在这种情况下，$\sqrt{n}(\hat{\theta}_n - \theta^*) + \nabla^2\Phi(\theta^*)^{-1} Y_n \xrightarrow[n \to \infty]{} 0$不再依概率成立，其中 θ_n 是约束最大似然估计量；而在

上文第 2 节中，对于足够大的 n，约束最大似然估计量与无约束最大似然估计量几乎必然相等。事实上，由 [5]中定理 7 的证明可知，$\sqrt{n}(\hat{\theta}_n - \theta^*) - \pi^{\nabla^2\Phi(\theta^*)T_n}(-Y_n) \longrightarrow_{n\to\infty} 0$依概率成立，其中 $\pi^{\nabla^2\Phi(\theta^*)T_n}$是相对于由 $\nabla^2\Phi(\theta^*)$定义的欧几里得结构，向 T_n作出的度量投影。

PROOF. 为简单起见，我们再次假设 $\nabla^2\Phi(\theta^*) = I_d$。与定理 2 的证明一样，第一步是证明 $Z_n \longrightarrow 1$ 依概率成立。为达到这一目的，必要且充分的条件是证明$_{n\to\infty}$

$$A_n \int_{\mathbb{R}^d} e^{-G_n(t) - t^\top Y_n - \|Y_n\|^2/2} \mathbb{1}_{t\in T_n}\, \mathrm{d}t \xrightarrow[n\to\infty]{} 1$$

依概率成立，因为 Z_n 是该量的逆。等价地，我们来证明

$$A_n \left(\int_{\mathbb{R}^d} e^{-G_n(t) - t^\top Y_n - \|Y_n\|^2/2} \mathbb{1}_{t\in T_n}\, \mathrm{d}t - \int_{\mathbb{R}^d} e^{-\frac{\|t+Y_n\|^2}{2}} \mathbb{1}_{t\in T}\, \mathrm{d}t \right) \xrightarrow[n\to\infty]{} 0$$

依概率成立。我们分两步进行：首先，证明序列 $(A_n)_{n\geq 1}$ 是紧的；然后，证明前一个公式中方括号内的量依概率收敛到 0。

固定 $\varepsilon > 0$ 和 $R > 0$，使得对于所有 $n \geq 1$，都有 $P(\|Y_n\| \leq R) \geq 1 - \varepsilon$，如上一节所示。于是，只要 $\|Y_n\| \leq R$成立，就可以写出以下不等式：

$$1 = A_n \int_T e^{-\frac{\|t+Y_n\|^2}{2}}\, \mathrm{d}t \geq A_n \int_T e^{-\|t\|^2 - \|Y_n\|^2}\, \mathrm{d}t \geq e^{-R^2} \int_T e^{-\|t\|^2}\, \mathrm{d}t,$$

由此得到 $A_n \leq e^{R^2}\left(\int_T e^{-\|t\|^2}\mathrm{d}t\right)^{-1}$。这表明 $(A_n)_{n\geq 1}$是紧的。

现在，让我们证明

$$\int_{\mathbb{R}^d} e^{-G_n(t) - t^\top Y_n - \|Y_n\|^2/2} \mathbb{1}_{t\in T_n}\, \mathrm{d}t - \int_{\mathbb{R}^d} e^{-\frac{\|t+Y_n\|^2}{2}} \mathbb{1}_{t\in T}\, \mathrm{d}t \xrightarrow[n\to\infty]{} 0 \tag{8}$$

依概率。

再次，假设 $\|Y_n\| \leq R$。由于对所有 $n \geq 1$，$T_n \subseteq T$

$$\begin{aligned}\left| \int_{\mathbb{R}^d} e^{-\frac{\|t+Y_n\|^2}{2}} \mathbb{1}_{t\in T}\, \mathrm{d}t - \int_{\mathbb{R}^d} e^{-\frac{\|t+Y_n\|^2}{2}} \mathbb{1}_{t\in T_n}\, \mathrm{d}t \right| &= \int_{\mathbb{R}^d} e^{-\frac{\|t+Y_n\|^2}{2}} (\mathbb{1}_{t\in T} - \mathbb{1}_{t\in T_n})\, \mathrm{d}t \\ &\leq \int_{\mathbb{R}^d} e^{-\frac{\|t\|^2 - 2R\|t\|^2}{2}} (\mathbb{1}_{t\in T} - \mathbb{1}_{t\in T_n})\, \mathrm{d}t\end{aligned}$$

其中，最后一个积分趋于 0 $n \to \infty$，这是由控制收敛定理得出的。这表明，前一个公式中的两个积分之差依概率收敛于 0，而证明8)归结为证明

$$\int_{T_n} \left(e^{-G_n(t) - t^\top Y_n - \|Y_n\|^2/2} - e^{-\frac{\|t+Y_n\|^2}{2}} \right) \mathrm{d}t \xrightarrow[n\to\infty]{} 0 \tag{9}$$

依概率。

固定 $R' > 0$，并假设 $\sup_{\|t\|\leq R'} |G_n(t) - \|t\|^2/2| \leq \varepsilon$。根据引理1，当 n 取足够大时，这种情况发生的概率至少为 $1 - \varepsilon$。此外，假设 $\|Y_n\| \leq R$。将上述积分拆分为两部分：一个是在 $T_n \cap B(0, R')$ 上的积分 I_n，另一个是在 $T_n \setminus B(0, R')$ 上的积分 J_n。我们分别对 I_n 和 J_n 进行界定。首先，将 I_n 写作

$$I_n = \int_{B(0,R')} e^{-\|t+Y_n\|^2/2} \left(e^{-(G_n(t) - \|t\|^2/2)} - 1 \right) \mathbb{1}_{t\in T_n}\, \mathrm{d}t$$

因此，根据三角不等式，

$$\begin{aligned}|I_n| &\le \int_{B(0,R')} e^{-\|t+Y_n\|^2/2}\left|e^{-(G_n(t)-\|t\|^2/2)}-1\right|\mathbb{1}_{t\in T_n}\,\mathrm{d}t\\ &\le (e^{\varepsilon}-1)\int_{B(0,R')} e^{-\|t+Y_n\|^2/2}\mathbb{1}_{t\in T_n}\,\mathrm{d}t \le (e^{\varepsilon}-1)\int_{\mathbb{R}^d} e^{-\|t+Y_n\|^2/2}\,\mathrm{d}t\\ &= (e^{\varepsilon}-1)\int_{\mathbb{R}^d} e^{-\|t\|^2/2}\,\mathrm{d}t = (2\pi)^{d/2}(e^{\varepsilon}-1)\end{aligned}$$

其中我们在第三个不等式中简单地将 $\mathbb{1}_{t\in T_n}$ 用 1 界定，并在第一个等式中进行了变量替换。其次，对于 J_n，使用（5）（以 R' 代替 $4R$），并使用三角不等式来写出

$$\begin{aligned}|J_n| &\le \int_{\mathbb{R}^d\setminus B(0,R')} e^{-\|t+Y_n\|^2/2}\left(e^{-(G_n(t)-\|t\|^2/2)}+1\right)\mathrm{d}t\\ &= \int_{\mathbb{R}^d\setminus B(0,R')} e^{-\|t+Y_n\|^2/2}\,\mathrm{d}t + \int_{\mathbb{R}^d\setminus B(0,R')} e^{-G_n(t)-t^\top Y_n-\|Y_n\|^2/2}\,\mathrm{d}t\\ &\le \int_{\mathbb{R}^d\setminus B(0,R')} e^{-\|t+Y_n\|^2/2}\,\mathrm{d}t + \int_{\mathbb{R}^d\setminus B(0,R')} e^{-(R'/2-2\varepsilon/R')\|t\|-\varepsilon-t^\top Y_n-\|Y_n\|^2/2}\,\mathrm{d}t\\ &\le \int_{\mathbb{R}^d\setminus B(0,R'-R)} e^{-\|u\|^2/2}\,\mathrm{d}u + \int_{\mathbb{R}^d\setminus B(0,R')} e^{-(R'/2-2\varepsilon/R'-R)\|t\|-\varepsilon}\,\mathrm{d}t\end{aligned}$$

其中，对于第一个积分，我们进行了变量替换 $u = t + Y_n$，其范数必须至少为 $R' - R$，而对于第二个积分，我们使用了柯西-施瓦茨不等式。由于这两个积分在 $R' \to \infty$ 时都趋于 0，可以选择 R' 使得 $|J_n| \le \varepsilon$。

总结来说，我们证明了：若 n 足够大，则以至少 $1-2\varepsilon$ 的概率（即当$\|Y_n\| \le R$ 且 $\sup_{t\in B(0,R')} |G_n(t) - \|t\|^2/2| \le \varepsilon$），（9）中的积分的绝对值被 $(2\pi)^{d/2}(e^{\varepsilon}-1)+\varepsilon$所界定。这证明了（9）。因此，我们得到 $Z_n \xrightarrow[n\to\infty]{} 1$ 依概率。

现在，我们可以证明定理 2，也就是说。既然我们已经证明 $Z_n \longrightarrow 1_{n\to\infty}$依概率成立，这就意味着需要证明

$$\int_{\mathbb{R}^d}\left|A_n e^{-G_n(t)-t^\top Y_n-\|Y_n\|^2/2}\mathbb{1}_{t\in T_n} - A_n e^{-\frac{\|t+Y_n\|^2}{2}}\mathbb{1}_{t\in T}\right|\mathrm{d}t \xrightarrow[n\to\infty]{} 0$$

依概率成立。此外，既然我们还证明了序列 $(A_n)_{n\ge1}$ 是紧的，我们只需证明

$$\int_{\mathbb{R}^d}\left|e^{-G_n(t)-t^\top Y_n-\|Y_n\|^2/2}\mathbb{1}_{t\in T_n} - e^{-\frac{\|t+Y_n\|^2}{2}}\mathbb{1}_{t\in T}\right|\mathrm{d}t \xrightarrow[n\to\infty]{} 0$$

依概率成立，这可直接由前面的计算推出。 □

4. 模型设定错误情形

现在，假设 θ^* 是一个边界点，且 $u := \nabla\Phi(\theta^*) \neq 0$。为简单起见，我们假设 $\theta^* = 0$ 且 $\nabla^2\Phi(\theta^*) = I_d$。这并不失一般性，因为只需对 θ进行线性重新参数化即可。本节的主要定理在一般情形下陈述（其中 $\nabla^2\Phi(\theta^*)$ 不一定是单位矩阵），见附录中的B 节（参见定理 5）。

在本节中 n，我们假设 Θ 可以表示为一组光滑凸约束：

$$\Theta = \{\theta \in \mathbb{R}^d : g_j(\theta) \le 0, \forall j = 1,\dots,p\}$$

其中， $p \geq 1$ 是一个整数， g_1、……、 $g_p : \mathbb{R}^d \to \mathbb{R}$ 是二阶可微的凸函数。我们令 $J = \{j = 1$、……、 $p : g_j(0) = 0\}$ 为 0 处的有效约束集合。在下文中，对于所有 $j \in J$，我们将 $u_j = \nabla g_j(0)$ 记为。

那么，0 处的 N 相对于 Θ 的法锥是由 u_j 的锥包生成的，其中 $j \in J$，即 $N = \{\sum_{j\in J} \lambda_j u_j : \lambda_j \geq 0, \forall j \in J\}$。0 处 Θ 的支撑锥 C 是 N 的极锥，即，

$$C = \{\theta \in \mathbb{R}^d : u_j^\top \theta \leq 0, \forall j \in J\}.$$

这是一个凸多面体，因为它由有限个线性约束定义。注意，这些约束中有些˜可能是冗余的。下文中，我们用 J 表示所有满足以下条件的 $j \in J$ 的集合： $C \cap u_j^\perp$ 是 C 的一个面（即，一个（$d-1$）维面）。

根据一阶条件，由于 0 是 Θ 上 Φ 的一个极小点， $-u \in N$。因此， $C \cap u^\perp$ 必定是 C 的一个面（维数介于 0 和 $d-1$ 之间）。也就是说，它是 C 的有限个面˜的交集。我们用 J^* 表示所有满足以下条件的 $j \in J$ 的集合：面 $C \cap u_j^\perp$ 包含面 $C \cap u^\perp$。也就是说， $-u$ 位于由 u_j 张成的锥的相对内部，其中 $j \in J^*$，即，对于某些正系数 $\lambda_j > 0, j \in J^*$， $-u = \sum_{j\in J^*} \lambda_j u_j$。

现在，我们令 $L = \text{span}(C \cap u^\top) = \{t \in \mathbb{R}^d : u_j^\top t = 0, \ \forall j \in J^*\}$ 为面 $C \cap u^\top$ 的线性包 C。其正交空间为 $L^\perp = \text{span}(\{u_j : j \in J^*\})$。

于是，我们可以将 $C \cap u^\top$ 改写为

$$C \cap u^\top = \{t \in L : u_j^\top t \leq 0, \forall j \in \tilde{J}, u_j^\top t = 0, \forall j \in J^*\}. \tag{10}$$

特别地，其相对内部为

$$\begin{aligned} \text{relint}(C \cap u^\top) &= \{t \in L : u_j^\top t < 0, \forall j \in \tilde{J} \setminus J^*, u_j^\top t = 0, \forall j \in J^*\} \\ &= \{t \in C \cap u^\perp : J(t) = J^*\} \end{aligned} \tag{11}$$

˜其中，对于所有 $t \in L$，我们将 $J(t) = \{j \in J : u_j^\top t = 0\}$ 记为。

对于 $n \geq 1$，我们令 $C_2^{(n)} = \{(t, \ s) \in L \times L^\perp : t/\sqrt{n} + s/n \in \Theta\}$，并将 Θ 处于 $\theta^* = 0$ （相对于 u）的二阶切集定义为

$$\begin{aligned} C_2 &= \{(t,s) \in L \times L^\perp : u_j^\top t \leq 0, \forall j \in J, \frac{1}{2} t^\top \nabla^2 g_j(0) t + u_j^\top s \leq 0, \forall j \in J(t)\} \\ &= \{(t,s) \in (C \cap u^\perp) \times L^\perp : s \in C_2(t)\} \end{aligned}$$

其中 $C_2(t) = \{s \in L^\perp : \frac{1}{2} t^\top \nabla^2 g_j(0) t + u_j^\top s \leq 0, \ \forall j \in J(t)\}$。相关的切锥集概念可参见 [3, 第 3] 章。重要的是，集合 C_2 可视为集合序列的极限集 $C_2^{(n)}$， $n \geq 1$，见下文引理 6 （关于 ´到 $C_2^{(n)}$ 的 Painlev C_2 é–Kuratowski 收敛可以给出精确表述，但引理 6 已足以满足我们的目的）。

为方便起见，我们也令 $C_{2\,0}^{(n)} = \{(t, \ s) \in L \times L^\perp : t/\sqrt{n} + s/n \in \Theta_0\}$。对于所有 $(t, \ s) \in C_{2\,0}^{(n)}$，记 $G_n(t, \ s)$ 为 $= n(\Phi_n(t/\sqrt{n} + s/n) - \Phi_n(0))$（回想一下，我们已假设 $\theta^* = 0$）。由于 Θ_0 是开集且包含 0，对于任意固定的 $R > 0$，当 n 充分大时， $B_R := \{(t, \ s) \in L \times L^\perp : \|t\| + \|s\| \leq R\} \subseteq C_{2\,0}^{(n)}$。

现在，我们将每个 $\theta \in \Theta$ 分解为 $\theta = \frac{t}{\sqrt{n}} + \frac{s}{n}$，其中 $(t, \ s) \in C_2^{(n)}$ 是唯一确定的，并将 $(t, \ s)$ 的联合后验密度写作

$$q_n(t,s) = \tilde{Z}_n e^{-n\left(\Phi_n(t/\sqrt{n} + s/n) - \Phi_n(0)\right)} \mathbb{1}_{(t,s) \in C_2^{(n)}},$$

˜对所有 $(t,\ s) \in L \times L^{\perp}$ 均成立，其中 Z_n 是一个随机的正归一化常数。我们将 Q_n 记作 $L \times L^{\perp}$ 上相对于 $L \times L^{\perp}$ 的勒贝格测度具有密度 q_n 的（随机）分布。[2]

将 R_n 记为定义在 $L \times L^{\perp}$ 上的概率测度，其相对于勒贝格测度的密度为 $r_n : (t,s) \in L \times L^{\perp} \mapsto A_n e^{-t^{\top}Y_n - \frac{\|t\|^2}{2} - s^{\top}u} \mathbb{1}_{(t,s)\in C_2}$，其中 $A_n > 0$ 是一个（随机的）正归一化常数，满足

$$A_n \int_{L\times L^{\perp}} e^{-t^{\top}Y_n - \frac{\|t\|^2}{2} - s^{\top}u} \mathbb{1}_{(t,s)\in C_2} \,\mathrm{d}t\,\mathrm{d}s = 1. \tag{12}$$

引理 4 和 5 分别表明，(12) 中的积分几乎必然有限且非零，因此 A_n 定义良好。于是，我们有以下定理。

Theorem 4. 设假设 *1* 成立。假设 θ^* 位于 Θ 的边界上，且 $\nabla\Phi(\theta^*) \neq 0$。则，$TV(Q_n, R_n) \xrightarrow[n\to\infty]{} 0$ 依概率收敛。

Proof. 首先，写作

$$q_n(t,s) = Z_n e^{-\left(G_n(t,s) - t^{\top}Y_n - \frac{\|t\|^2}{2} - s^{\top}u\right)} A_n e^{-t^{\top}Y_n - \frac{1}{2}\|t\|^2 - s^{\top}u} \mathbb{1}_{(t,s)\in C_2^{(n)}},$$

˜对所有 $(t,s) \in L \times L^{\perp}$，其中 Z_n 仅仅是 Z_n / A_n。

证明策略与前几节类似。首先，我们将证明 $Z_n \xrightarrow[n\to\infty]{} 1$依概率成立，方法是证明

$$A_n \int_{L\times L^{\perp}} e^{-\left(G_n(t,s) - t^{\top}Y_n - \frac{\|t\|^2}{2} - s^{\top}u\right)} e^{-t^{\top}Y_n - \frac{\|t\|^2}{2} - s^{\top}u} \mathbb{1}_{(t,s)\in C_2^{(n)}} \,\mathrm{d}t\,\mathrm{d}s \xrightarrow[n\to\infty]{} 1 \tag{13}$$

依概率。通过与定理证明中类似的论证2 和3，序列 $(A_n)_{n\geq 1}$是紧的，因此要证明（13），只需证明

$$\int_{L\times L^{\perp}} e^{-t^{\top}Y_n - \frac{\|t\|^2}{2} - s^{\top}u} \left(e^{-\left(G_n(t,s) - t^{\top}Y_n - \frac{\|t\|^2}{2} - s^{\top}u\right)} \mathbb{1}_{(t,s)\in C_2^{(n)}} - \mathbb{1}_{t\in C\cap u^{\perp}, s\in C_2(t)} \right) \mathrm{d}t\,\mathrm{d}s \xrightarrow[n\to\infty]{} 0 \tag{14}$$

依概率成立。

固定 $\varepsilon > 0$。由于 $(Y_n)_{n\geq 1}$ 根据中心极限定理是紧的，因此可以选取 $r > 0$，使得$P(\|Y_n\| \leq r) \geq 1 - \varepsilon$ 对所有 $n \geq 1$均成立。现在，固定 $R > 0$。根据 [5, 引理 4],，我们有

$$\sup_{(t,s)\in B_R} \left| G_n(t,s) - t^{\top}Y_n - \frac{\|t\|^2}{2} - s^{\top}u \right| \xrightarrow[n\to\infty]{} 0 \tag{15}$$

依概率。假设 $\|Y_n\| \leq r$ 且 $\sup_{(t,\ s)\in B_R} \left| G_n(t,\ s) - t^{\top}Y_n - \frac{\|t\|^2}{2} - s^{\top}u \right| \leq \varepsilon$。如果 n 足够大，则这两个事件同时发生的概率至少为 $1 - 2\varepsilon$。注意，由于 G_n 是凸的，我们必须（根据之前证明中使用过的相同论证）对所有 $(t,\ s) \notin B_R$ 都有：若将 $z = (t,\ s)$ 和 $\|z\| = \|t\| + \|s\|$ 分别记为，

$$\begin{aligned} G_n(z) &\geq t^{\top}Y_n + \frac{R\|t\|^2}{2\|z\|} + s^{\top}u - \frac{\|z\| - R - 1}{R}\varepsilon \\ &\geq -r\|t\| + \frac{R\|t\|^2}{2\|z\|} + s^{\top}u - \frac{\|z\|\varepsilon}{R} \end{aligned} \tag{16}$$

[2] $L \times L^{\perp}$ 的勒贝格测度等于 L 和 $L^{\perp}$的勒贝格测度的乘积。

（此不等式可通过将 $Rz/\|z\|$ 表示为 0 和 z的凸组合得到）。此外，根据引理 2 和 3，若 $(t,\ s) \in C_2^{(n)}$，或者若 $t \in C \cap u^\perp$ 且 $s \in C_2(t)$，则有 $u^\top s \geq \alpha > 0$，其中 α 是一个正常数。因此，对所有此类 $z = (t,\ s) \notin B_R$，我们得到

$$\begin{aligned} G_n(z) &\geq -r\|t\| + \frac{R\|t\|^2}{2(\|t\|+\|s\|)} + \alpha\|s\| - \frac{\varepsilon}{R}(\|t\|+\|s\|) \\ &= \frac{1}{\|t\|+\|s\|}\left((R/2 - r - \varepsilon/R)\|t\|^2 + (\alpha - r - 2\varepsilon/R)\|t\|\|s\| + (\alpha - \varepsilon/R)\|s\|^2\right) \end{aligned}$$

通过选择足够大的 $\|t\|,\ \|s\|$，无论 R 的取值如何，都可以使其非负。因此，若将式（14）中的积分拆分为在 $I_n^{(1)}$上的积分 B_R 和在$I_n^{(2)}$上的积分 $(L \times L^\perp) \setminus B_R$，便可得到

$$|I_n^{(2)}| \leq 2 \int_{(L\times L^\perp)\setminus B_R} e^{r\|t\| - \frac{\|t\|^2}{2} - \alpha\|s\|}\, \mathrm{d}t\, \mathrm{d}s.$$

由于映射 $(t,\ s) \in L \times L^\perp \mapsto e^{r\|t\| - \frac{\|t\|^2}{2} - \alpha\|s\|}$ 在 $L \times L^\perp$ 上可积，因此通过选择足够大的 R，可以使上方表达式右侧小于 ε。

现在，我们聚焦于 $I_1^{(n)}$。首先，将 $I_1^{(n)}$ 写为

$$\begin{aligned} I_1^{(n)} &= \int_{B_R} e^{-t^\top Y_n - \frac{\|t\|^2}{2} - s^\top u}\left(e^{-\left(G_n(t,s) - t^\top Y_n - \frac{\|t\|^2}{2} - s^\top u\right)}\mathbb{1}_{(t,s)\in C_2^{(n)}} - \mathbb{1}_{t\in C\cap u^\perp, s\in C_2(t)}\right)\mathrm{d}t\,\mathrm{d}s \\ &= \int_{B_R} e^{-t^\top Y_n - \frac{\|t\|^2}{2} - s^\top u} e^{-\left(G_n(t,s) - t^\top Y_n - \frac{\|t\|^2}{2} - s^\top u\right)}\left(\mathbb{1}_{(t,s)\in C_2^{(n)}} - \mathbb{1}_{t\in C\cap u^\perp, s\in C_2(t)}\right)\mathrm{d}t\,\mathrm{d}s \\ &\quad + \int_{B_R} e^{-t^\top Y_n - \frac{\|t\|^2}{2} - s^\top u}\left(e^{-\left(G_n(t,s) - t^\top Y_n - \frac{\|t\|^2}{2} - s^\top u\right)} - 1\right)\mathbb{1}_{t\in C\cap u^\perp, s\in C_2(t)}\,\mathrm{d}t\,\mathrm{d}s \\ &=: I_3^{(n)} + I_4^{(n)}. \end{aligned}$$

首先，注意到根据假设，

$$|I_3^{(n)}| \leq e^{\varepsilon} \int_{B_R} e^{r\|t\| - \frac{\|t\|^2}{2} - s^\top u}\left|\mathbb{1}_{(t,s)\in C_2^{(n)}} - \mathbb{1}_{t\in C\cap u^\perp, s\in C_2(t)}\right|\mathrm{d}t\,\mathrm{d}s.$$

根据引理 6 和控制收敛定理，右侧趋于 0，当 $n \to \infty$时，因此当 n足够大时，它会小于 ε。

$$\begin{aligned} |I_4^{(n)}| &\leq (e^{\varepsilon} - 1)\int_{B_R} e^{r\|t\| - \frac{\|t\|^2}{2} - s^\top u} 1_{t\in C\cap u^\perp, s\in C_2(t)}\,\mathrm{d}t\,\mathrm{d}s \\ &\leq \alpha_2(e^{\varepsilon} - 1) \end{aligned}$$

对于某个正常数 α_2，该常数仅依赖于 R 和 Θ。

综上，我们已经证明了（14）。现在，只需写出4，定理的其余证明便可轻松完成。

$$\mathrm{TV}(Q_n, R_n) = \int_{L\times L^\perp} |q_n - r_n| = |1 - 1/Z_n| + \int_{L\times L^\perp} |q_n/Z_n - r_n|$$

并注意到，如上所示，第一项依概率趋于 0；利用$(A_n)_{n\geq 1}$ 的紧性，第二项也依概率趋于 0，这与证明（14）所用的论证完全相同。

□

5. 结论

在本文中，我们研究了两种情形下对数凹后验分布的渐近行为。第一种情形是设定正确的模型，即真实参数属于先验分布的支撑集。第二种情形是设定错误的模型，即真实参数不属于先验分布的支撑集。在第二种情形下，我们证明，采用非经典缩放的后验分布具有非平凡极限，其支撑集位于先验分布支撑集的二阶切集上。在所有情形下，我们仅假设对数似然函数关于参数是凹的，且 Fisher 信息矩阵正定。在模型设定错误的情形下，我们假设先验分布的支撑集 Θ 可以用有限个二阶可微凸约束表示。事实上，我们认为，只需假设 Θ 的示性函数在 θ^* 沿方向 $-\nabla\Phi(\theta^*)$ 处二次上图可微，[3] 便可得到类似结果；这将恢复文献 [5] 中关于 MLE 渐近分布所作的假设。此问题留待未来研究。值得注意的是，在设定正确和近似设定错误的情形下，我们对 Θ 没有作任何假设——只要求其为具有非空内部的凸集。这与 MLE 相关情形 [5], 相似，在该情形中， Θ的示性函数总是在方向 θ^* 沿$-\nabla\Phi(\theta^*) = 0$处二次上图可微。

参考文献

[1] Natalia A. Bochkina。伯恩斯坦–冯·米泽斯定理与设定错误的模型：综述。现代统计学基础，第355–380页，2019年。[2] Natalia A. Bochkina 和 Peter J. Green。伯恩斯坦-冯·米塞斯定理与非正则模型。《统计年鉴》，42(5):1850–1878，2014年。[3] J. Fr´ed´eric Bonnans 和 Alexander Shapiro。优化问题的扰动分析。Springer Science & Business Media，2013年。[4] Christer Borell。d 维空间中的凸集函数。《匈牙利数学期刊》，6(2):111–136，1975年。[5] Victor-Emmanuel Brunel。凸性约束下 M-估计的渐近性质。*arXiv* 预印本 *arXiv:2511.04612*，2025年。[6]Isma¨el Castillo。贝叶斯非参数统计，圣弗卢尔讲义。*arXiv* 预印本 *arXiv:2402.16422*，2024年。[7] Isma¨el Castillo 和 Richard Nickl。高斯白噪声中的非参数伯恩斯坦–冯·米泽斯定理。《统计年鉴》，41(4):1999–2028，2013年。[8] A Philip Dawid。关于后验分布的极限正态性。载于剑桥哲学学会数学论文集，第67卷，第625–633页。剑桥大学出版社，1970年。[9] Chi Ngoc Do。自反巴拿赫空间中凸函数的广义二阶导数。《美国数学会会刊》，334(1):281–301，1992年。[10] Charles J. Geyer。关于约束 M-估计的渐近性质。《统计年鉴》，第1993–2010页，1994年。[11] Geoff Goehle。指数族分布的贝叶斯中心极限定理：一种几何方法。《信息几何》，7(2):471–488，2024年。[12] Anya Katsevich。通过新的拉普拉斯近似界改进伯恩斯坦–冯·米泽斯定理中的维数依赖性。《信息与推断：*IMA* 期刊》，14(3):iaaf020，2025年。

[3]根据 [9, 推论4.4]，这等价于在 Θ 上的投影在 $-\nabla\Phi(\theta^*)$处具有方向导数。

[13] BJK Kleijn 和 AW van der Vaart。误设情况下的伯恩斯坦-冯·米塞斯定理。《电子统计学杂志》，6:354–381，2012年。[14] Lucien Le Cam。关于最大似然估计及相关贝叶斯估计的一些渐近性质。《加州大学统计学出版物》，1:277–330，1953年。[15] Lucien Le Cam。关于证明最大似然估计渐近正态性所用的假设。《数学统计年刊》，41(3):802–828，1970年。[16] Vladimir Spokoiny。贝叶斯推断中的非精确拉普拉斯近似及后验均值的使用。《贝叶斯分析》，20(1):1–28，2025年。[17] Aad W Van der Vaart。渐近统计学，第3卷。剑桥大学出版社，2000年。[18] Aad W Van Der Vaart 和 Jon A Wellner。弱收敛与经验过程：及其在统计学中的应用。施普林格，1996年。[19] Andrew M Walker。关于后验分布的渐近行为。《英国皇家统计学会杂志*B*辑：统计方法学》，31(1):80–88，1969年。

附录A：定理 1

根据大数定律，$\Phi_n(\theta) \xrightarrow[n\to\infty]{} \Phi(\theta)$ 几乎必然成立，对所有固定的 $\theta \in \Theta_0$ 均成立。因此，根据 [5, 推论 1]，，对所有满足 $B(\theta^*, r) \subseteq \Theta_0$ 的 $r > 0$ 均成立，

$$\sup_{\theta \in B(\theta^*, r)} |\Phi_n(\theta) - \Phi(\theta)| \xrightarrow[n\to\infty]{} 0 \tag{17}$$

几乎必然成立。令 $\alpha > 0$ 为 $\nabla^2\Phi(\theta^*)$ 的最小特征值。由于 Φ 在 θ^* 的邻域内二阶可微，存在 $r > 0$，使得 $B(\theta^*, r) \subseteq \Theta_0$，且对所有 $\theta \in B(\theta^*, r)$，$\nabla^2\Phi(\theta)$ 的所有特征值均大于 $\alpha/2$。因此，对所有 $\theta \in B(\theta^*, r) \cap \Theta$，

$$\Phi(\theta) \geq \Phi(\theta_0) + \nabla\Phi(\theta^*)^\top(\theta - \theta^*) + \frac{\alpha}{4}\|\theta - \theta^*\|^2. \tag{18}$$

此外，θ^* 的一阶最优性条件表明 $u^\top(\theta - \theta^*) \geq 0$ 对所有 $\theta \in \Theta$ 均成立，因此 (19) 可写为

$$\Phi(\theta) \geq \Phi(\theta_0) + \frac{\alpha}{4}\|\theta - \theta^*\|^2, \tag{19}$$

对所有 $\theta \in B(\theta^*, r) \cap \Theta$。令 $\varepsilon = \alpha r^2/16$。根据（17），以概率 1，对于所有充分大的$n \geq 1$，同时对所有 $\theta \in B(\theta^*, r) \cap \Theta$

$$\Phi_n(\theta) \geq \Phi(\theta) - \varepsilon \geq \Phi(\theta^*) - \varepsilon + \frac{\alpha}{8}\|\theta - \theta^*\|^2. \tag{20}$$

and

$$\Phi_n(\theta^*) \leq \Phi(\theta^*) + \varepsilon. \tag{21}$$

特别地，对于所有 $\theta \in B(\theta^*, r) \cap \Theta$，

$$\Phi_n(\theta) - \Phi_n(\theta^*) \geq \alpha r^2/8. \tag{22}$$

如果不存在这样的 $\theta \in \Theta$ 满足 $\|\theta - \theta^*\| > r$，那么 Θ 是紧致的，定理 1 显然成立。否则，固定 $\theta \in \Theta$ 满足 $\|\theta - \theta^*\| > r$。令 $\theta_1 = \theta^* + r(\theta - \theta^*)/\|\theta - \theta^*\|$，它是 $\mathbb{R}_+(\theta - \theta^*)$ 与 $B(\theta^*, r)$的边界的交集。由 Θ的凸性，$\theta_1 \in \Theta$ 以及 Φ_n 的凸性可得

$$\Phi_n(\theta_1) \leq \Phi_n(\theta^*) + \frac{r}{\|\theta - \theta^*\|}(\Phi_n(\theta) - \Phi_n(\theta^*)),$$

这可以写成

$$\Phi_n(\theta) \geq \frac{\|\theta-\theta^*\|}{r}(\Phi_n(\theta_1)-\Phi_n(\theta^*))+\Phi_n(\theta^*). \tag{23}$$

结合（22），我们得到：以概率 1，对所有充分大的 n，都有

$$\Phi_n(\theta) \geq \frac{\alpha r\|\theta-\theta^*\|}{8}+\Phi_n(\theta^*), \quad \forall \theta \in \Theta \setminus B(\theta^*, r).$$

因此，对所有 $\theta \in \Theta_0 \setminus B(\theta^*, r)$，

$$e^{-n\Phi_n(\theta)}\mathbb{1}_{\theta\in\Theta} \leq e^{-\frac{n\alpha r\|\theta-\theta^*\|}{8}-\Phi_n(\theta^*)}$$

并且，因而，该映射在 $\Theta_0 \setminus B(\theta^*, r)$ 上可积，因此在 Θ_0 上也可积，因为根据 Φ_n 的凸性，它在$B(\theta^*, r)$ 上连续。定理的证明至此结束。

Remark 2. 这里，我们证明了定理 *1* 在 $V=\delta_\Theta$ 的特定情形下成立。事实上，很容易验证，只要 V 有下界，我们的证明就同样成立。这大致意味着先验分布的信息量不宜过大，也就是说，它不会在其支撑集的任何特定区域上赋予过多的质量。

附录B：定理4的一般性陈述

在第4节中，我们只在 Fisher 信息矩阵为单位矩阵的情况下陈述了定理4，即 $\nabla^2\Phi(\theta^*)=I_d$。这种简化并未损失一般性，因为一般情形可以通过简单的仿射变换得到。不过，一般版本需要引入一些额外的记号，我们将在此介绍。

记 $S=\nabla^2\Phi(\theta^*)$ 为 Fisher 信息矩阵。根据假设，它是正定的，因此赋予 $\mathbb{R}^d$ 欧几里得结构。对于所有 $x, y \in \mathbb{R}^d$，我们用 $x, y_S := x^\top S y$ 表示相应的内积。相应的欧几里得范数记为 $\|\cdot\|_S$。也就是说，$\|x\|_S=(x^\top S x)^{1/2}$。对于非空子集 $A \subseteq \mathbb{R}^d$，我们用 $A^{\perp S}$ 表示相对于 $\cdot,\cdot_S$ 的 A 的正交子空间。也就是说，

$$A^{\perp S}=\{y \in \mathbb{R}^d : \langle x, y\rangle_S = 0, \forall x \in A\}.$$

Theorem 5. 令假设*1*成立。令 Θ 表示为$\{\theta \in \mathbb{R}^d = g_j(\theta) \leq 0, \ j=1, \ldots, \ p\}$，其中$p \geq 1$是某个整数，且 $g_1, \ldots, g_p : \mathbb{R}^d \to \mathbb{R}$ 是凸且二阶可微的。假设 Θ具有非空内部。

进一步假设 θ^* 位于 Θ 的边界上，且 $\nabla\Phi(\theta^*) \neq 0$，并令 $J=\{j=1、\ldots、\ p : g_j(\theta^*)=0\}$。令 C 为 Θ 在 θ^*处的支撑锥。

将 $u=\nabla\Phi(\theta^*)$ 和 $S=\nabla^2\Phi(\theta^*)$记为。假设 S 是正定的。

*Le*令~J 为所有满足以下条件的 $j \in J$ 的集合： $C \cap u_j^\top$ 是 C 的一个面，并令 J^* 为所有~的集合，其中 $j \in J$ 对应的面包含面 $C \cap u^\perp$。

令 $L=$张成空间（$C \cap \nabla\Phi$（θ^*）$^\top$）以及 $C_2=\{$（t、 s）$\in L \times L^{\perp S} : t \in C \cap u^\perp$、 $\frac{1}{2}t^\top\nabla^2 g_j$（$\theta^*$）$t+u_j^\top s \leq 0$、 $\forall j \in J$（t）$\}$~，其中 J（t）$=\{j \in J : u_j^\top t=0\}$。

然后，通过将 $\theta=\frac{t}{\sqrt{n}}+\frac{s}{n}$表示为唯一的 （$t$、 s） $\in L \times L^{\perp S}$， Q_n 的后验分布满足 （t、 s）

$$TV(Q_n, R_n) \xrightarrow[n\to\infty]{} 0$$

依概率，其中 R_n 是在 $L \times L^{\perp S}$ 上的概率分布，其相对于勒贝格测度的密度正比于

$$(t,s) \in L \times L^\perp \mapsto e^{-t^\top Y_n-\frac{\|t\|_S^2}{2}-s^\top u}\mathbb{1}_{(t,s)\in C_2}.$$

附录 C：中间引理

LEMMA 2. 存在 $\alpha > 0$，使得对于所有 $t \in C \cap u^\perp$ 以及所有 $s \in C_2(t)$，$u^\top s \geq \alpha \|s\|$.

PROOF. 首先，回顾一下，对于所有 $t \in C \cap u^\perp$，集合 $J(t)$ 包含 J^*。此外，由于 g_j 都是凸的，$t^\top \nabla^2 g_j(0) t$ 总是非负的。因此，对于所有 $t \in C \cap u^\perp$，$C_2(t) \subseteq \{s \in Lperp : s^\top u_j \leq 0, \ \forall j \in J^*\}$。

固定某个 $s \in L^\perp$，使得对于所有 $j \in J^*$，$s^\top u_j \leq 0$。假设 $s^\top u = 0$。那么，

$$0 = -s^\top u = \sum_{j \in J^*} \lambda_j s^\top u_j$$

它是若干非负项之和。因此，这些项都必须为零，从而得到对于所有 $j \in J^*$，$s^\top u_j = 0$。也就是说，$s \in L$，因此，s 必须为 0。我们已经证明，在闭锥 $\{s \in L^\perp : s^\top u_j \leq 0, \ \forall j \in J^*\}$ 中，只要 $s \neq 0$，就有 $u^\top s > 0$。通过紧致性论证，我们得到存在 $\alpha > 0$，使得对于所有满足 $s^\top u_j \leq 0$, $\forall j \in J^*$ 且 $\|s\| = 1$ 的 $s \in L^\perp$，都有 $u^\top s \geq \alpha$。由齐次性即可得到结论。 □

我们还有如下结果。

LEMMA 3. 设 $K = \{s \in L^\perp : \exists t \in C \cap u^\perp, t + s \in \Theta\}$。那么，存在 $\alpha > 0$，使得对所有$s \in K$, $u^\top s \geq \alpha \|s\|$。

PROOF. 由于 u_j's $j \in J^*$ 是 Θ 在 0 处的法向量，因此有 $\Theta \subseteq \{\theta \in \mathbb{R}^d : u_j^\top \theta \leq 0, \forall j \in J^*\} =: \Theta^*$。因此，$\inf\{u^\top s/\|s\| : \exists t \in C \cap u^\perp, t+s \in \Theta\} \geq \inf\{u^\top s/\|s\| : \exists t \in C \cap u^\perp, t+s \in \Theta^*\}$。对所有 $t \in C$，$u_j^\top t = 0$ 对所有 $j \in J^*$ 都成立，因为 N 和 C 互为极锥。此外，若 $t \in u^\perp$，则$\sum_{j \in J^*} \lambda_j u_j^\top t = 0$，因此该和式中的每一项都必须为 0，即，对于每个 $j \in J^*$，$u_j^\top t = 0$。由此可得 $\inf\{u^\top s/\|s\| : \exists t \in C \cap u^\perp, t + c \in \Theta\} \geq \inf\{u^\top s/\|s\| : s \in L^\perp, u_j^\top s \leq 0, \forall j \in J^*\}$。利用引理2证明中的相同论证，对于满足 $u_j^\top s \leq 0$、$\forall j \in J^*$ 的所有 $s \in L^\perp$，有 $u^\top s > 0$。因此，通过紧致性论证，必存在 $\alpha > 0$，使得对于满足$u_j^\top s \leq 0$、$\forall j \in J^*$ 和 $\|s\| = 1$ 的所有 $s \in L^\perp$，均有 $u^\top s \geq \alpha$。这就得到了所需结果。 □

LEMMA 4. 存在 $a_1 > 0$，使得对于所有 $t \in C \cap u^\perp$,

$$\int_{L^\perp} e^{-s^\top u} \mathbb{1}_{s \in C_2(t)} \, \mathrm{d}t \leq a_1.$$

PROOF. 根据引理 2，存在某个 $\alpha > 0$，且与 $t \in C \cap u^\perp$ 无关，使得

$$\begin{aligned} \int_{L^\perp} e^{-s^\top u} \mathbb{1}_{s \in C_2(t)} \, \mathrm{d}s &\leq \int_{L^\perp} e^{-\alpha \|s\|} \mathbb{1}_{s \in C_2(t)} \, \mathrm{d}s \\ &\leq \int_{L^\perp} e^{-\alpha \|s\|} \, \mathrm{d}s < \infty. \end{aligned}$$

□

对于下一个引理，回忆一下，J^* 是满足以下条件的指标集合 $j \in J$：$C \cap u_j^\perp$是包含面 $C \cap u^\top$ 的 C的一个面。

LEMMA 5. 集合 C_2 是凸的。其内部非空，并且它由 y

$$\text{int}(C_2) = \{(t, s) \in L \times L^\perp : t \in \text{relint}(C \cap u^\perp), \frac{1}{2} t^\top \nabla^2 g_j(0) t + u_j^\top s < 0, \forall j \in J^*\}. \tag{24}$$

PROOF.

的凸性 C_2。我们首先检查 C_2 是一个凸集。设 $(t_0,\ s_0)$，$(t_1,\ s_1)$ 是 C_2 中的两个元素，且 $\lambda \in \in (0,$ 1)。记 $t_\lambda = (1-\lambda)t_0 + \lambda t_1$ 和 $s_\lambda = (1-\lambda)s_0 + \lambda s_1$。我们的目标是证明 $(t_\lambda,\ s_\lambda) \in C_2$。首先，根据集合 $C \cap u^\perp$ 的凸性，$t_\lambda \in C \cap u^\perp$。此外，很容易验证 $J(t_\lambda) \subseteq J(t_0) \cap J(t_1)$。现在，我们只需验证 $s_\lambda \in C_2(t_\lambda)$，也就是，

$$\frac{1}{2} t^\top \nabla g_j(0) t + u_j^\top s \le 0, \quad \forall j \in J(t_\lambda), \tag{25}$$

当 $(t,s) = (t_\lambda, s_\lambda)$ 时。由于 (25) 对 (t_0,s_0) 和 (t_1,s_1) 均成立，且映射 $(t,s) \mapsto t^\top \nabla^2 g_j(0) t + u_j^\top s$ 是凸的，因此 (25) 对 (t_λ, s_λ) 自动成立，由此得到结论。

(24) 的有效性。现在，我们来检验 (24)。右向左包含显然成立，因此只需检验左向右包含。设 $(t,s) \in \mathrm{int}(C_2)$. 必然有 t 属于 $\mathrm{relint}(C \cap u^\top)$。事实上，若非如此，则存在某个 $j \in J \setminus J^*$ 使得 $u_j^\top t = 0$。由于 $j \notin J^*$，必然存在某个 $z \in L$ 满足 $u_j^\top z \neq 0$。不失一般性，假设 $u_j^\top z > 0$。那么，对所有 $\varepsilon > 0$, $t + \varepsilon z \notin C \cap u^\top$，因此 $(t + \varepsilon z, s)$ 不可能属于 C_2，且 (t,s) 不可能是 C_2 的内点。

现在，为了导出矛盾，假设 $\frac{1}{2} t^\top \nabla^2 g_j(0) t + u_j^\top s = 0$ 对某个 $j \in J^*$ 成立。注意，根据 J^* 的定义，$L \subseteq u_j^\perp$。此外，根据 J^* 的定义，$u_j \neq 0$，因为假设 $C \cap u_j^\perp$ 是 C 的一个面——所以它不可能是整个 C。因此，$u_j \notin L$。也就是说，存在 $y \in L^\perp$ 使得 $u_j^\top y \neq 0$。同样，不失一般性地，假设 $u_j^\top y > 0$。因此，对于所有 $\varepsilon > 0$，$\frac{1}{2} t^\top \nabla^2 g_j(0) t + u_j^\top (s + \varepsilon y) > 0$，所以（$t,\ s + \varepsilon y$）不能属于 C_2。

非空内部：C_2。为得出矛盾，假设 $\mathrm{int}(C_2)$ 为空。由于任意非空凸集的相对内部总是非空的，相对内部$(C \cap u^\top) \neq \varnothing$。因此，对所有 $t \in$ 相对内部$(C \cap u^\perp)$ 以及所有 $s \in L^\perp$，必存在 $j \in J^*$ 使得 $\frac{1}{2} t^\top \nabla^2 g_j(0) t + u_j^\top s \ge 0$。也就是说，对所有 $s \in L^\perp$，我们有：

$$\max_{j \in J^*} \left(\frac{1}{2} t^\top \nabla^2 g_j(0) t + u_j^\top s \right), \quad \forall t \in \mathrm{relint}(C \cap u^\perp).$$

令 t 在上一式中趋于 0，可得对所有 $s \in L^\perp$，必存在某个 $j \in J^*$ 使得 $u_j^\top s \ge 0$。换言之，集合 $\{ s \in L^\perp : u_j^\top s < 0,\ \forall j \in J^* \}$ 必为空。这是 $L^\perp$ 中凸多面体 $P := \{ s \in L^\perp : u_j^\top s \le 0,\ \forall j \in J^* \}$ 的相对内部。因此，凸集 $L \times P$ 必具有空内部。然而，该凸集包含 $\{(t,\ s) \in L \times L^\perp : t + s \in \Theta\}$，它与 Θ 等距，且假定 Θ 具有非空内部。由此得出矛盾。

□

Lemma 6. 对于几乎所有[4] $(t,s) \in L \times L^\perp$,

$$\mathbb{1}_{(t,s) \in C_2^{(n)}} \xrightarrow[n \to \infty]{} \mathbb{1}_{(t,s) \in C_2}.$$

Proof. 回忆一下，J 是 0 处有效约束的指标集合，即 $J = \{ j = 1, \dots p : g_j = (0)\ 0 \}$。我们将所有满足 $C \cap u_j^\perp$ 是 C 的一个面 的这类 $j \in J$ 组成的集合定义为 J；将所有满足 $C \cap u_j^\perp$ 是包含面 $C \cap u^\perp$ 的 C 的一个面 的这类 j 组成的集合定义为 $J^* \subseteq J$。因此，可以写成

$$C \cap u^\perp = \{ t \in L : u_j^\top t \le 0, \forall j \in \tilde{J}, u_j^\top t = 0, \forall j \in J^* \}.$$

[4] 关于 $L \times L^\perp$ 的勒贝格测度

因此， $C \cap u^\perp$ 在 L 中的相对内部为

$$\mathrm{relint}(C \cap u^\perp) = \{t \in L : u_j^\top t < 0, \forall j \in \tilde{J} \setminus J^*, u_j^\top t, \forall j \in J^*\},$$

也就是说，$C \cap u^\perp$的相对内部 $= \{t \in C \cap u^\perp : J(t) = J^*\}$，其中我们回顾一下， $J(t) = \{j \in J : u_j^\top t = 0\}$。

固定 $(t, s) \in L \times L^\perp$。回顾一下，$(t, s) \in C_2^{(n)}$ 当且仅当 $t/\sqrt{n} \mp s/n \in \Theta$。由于定义 Θ 的函数g_j, $j = 1, \ldots\ldots, p$均为凸函数，因此它们在 $\mathbb{R}^d$ 上连续。因此，对所有足够大的 n（其具体大小取决于固定的 (t, s)），$t/\sqrt{n} + s/n \in \Theta$ 当且仅当 $g_j(t/\sqrt{n} + s/n) \le 0$对所有 $j \in J$ 均成立。事实上，对 $j \notin J$，根据定义， $g_j(0) < 0$，因此 $g_j(\theta)$ 在原点附近的某个区域内始终为负。所以，对足够大的 n,

$$\mathbb{1}_{(t,s)\in C_2^{(n)}} = \mathbb{1}_{g_j(t/\sqrt{n}+s/n)\le 0, \forall j \in J}.$$

回忆 $J(t) = \{j \in J : u_j^\top t = 0\}$ 的定义。固定 $j \in J$。若 $j \in J(t)$，记为

$$g_j(t/\sqrt{n} + s/n) = \frac{1}{n}\left(\frac{1}{2} t^\top \nabla^2 g_j(0) t + u_j^\top s + \varepsilon_n^{(j)}\right) \tag{26}$$

对某个 $\varepsilon_n^{(j)} \xrightarrow[n\to\infty]{} 0$。若 $j \notin J(t)$，记为

$$g_j(t/\sqrt{n} + s/n) = \frac{1}{\sqrt{n}}\left(u_j^\top t + \eta_n^{(j)}\right) \tag{27}$$

对某个 $\eta_n^{(j)} \longrightarrow 0$。注意， $\varepsilon_n^{(j)}$ 和 $\eta_n^{(j)}$ 依赖于 s 和 t，但这里它们都是固定的。$n\to\infty$

因此，对所有充分大的 n,

$$(t, s) \in C_2^{(n)} \iff \begin{cases} u_j^\top t + \eta_n^{(j)} \le 0, \forall j \in J \setminus J(t) \\ \frac{1}{2} t^\top \nabla^2 g_j(0) t + u_j^\top s + \varepsilon_n^{(j)} \le 0, \forall j \in J(t). \end{cases}$$

假设 $(t, s) \in \mathrm{int}(C_2)$。也就是说，根据引理5， $t \in \mathrm{relint}(C \cap u^\top)$ 且 $s \in \mathrm{relint}(C_2(t))$。也就是说，根据 (11)， $t \in C \cap u^\perp$，其中 $J(t) = J^*$。因此，对所有 $j \in J \setminus J^*$， $u_j^\top t < 0$。由于 $\eta_n^{(j)}$、$j \notin J^*$ 趋于 0，必有 $u_j^\top t + \eta_j^{(n)} \le 0$ 对所有 $j \notin J^*$ 均成立，且对充分大的 n 也成立。此外，由于 $s \in \mathrm{relint}(C_2(t))$，$\frac{1}{2} t^\top \nabla^2 g_j(0) t + u_j^\top s < 0$ 对所有 $j \in J^*$ 均成立。于是，由于对所有 $j \in J(t) = J^*$ 均有 $\varepsilon_n^{(j)} \xrightarrow[n\to\infty]{} 0$，必有 $\frac{1}{2} t^\top \nabla^2 g_j(0) t + u_j^\top s + \varepsilon_n^{(j)} \le 0$ 对所有 $j \in J(t)$ 均成立，且对充分大的 n 也成立。因此我们证明了，若 $(t, s) \in \mathrm{relint}(C_2)$，则对所有充分大的 n， $\mathbb{1}_{(t,s)\in C_2^{(n)}} = 1 = \mathbb{1}_{(t,s)\in C_2}$。

现在，令 $(t, s) \in L \times L^\perp$ 满足 $(t, s) \notin C_2$。也就是说， $t \notin C \cap u^\top$，或者 $t \in C \cap u^\top$ 但 $-\frac{1}{2} t^\top \nabla^2 g_j(0) t + u_j^\top s > 0$ 对某个 $j \in J(t)$ 成立。与上文类似，我们得到对所有充分大的 n， $\mathbb{1}_{(t,s)\in C_2^{(n)}} = 0 = \mathbb{1}_{(t,s)\in C_2}$ 均成立。

最后，由于 C_2 根据引理 5是一个凸集，其边界的勒贝格测度为零，因此我们证明了对几乎所有 $(t, s) \in L \times L^\perp$,

$$\mathbb{1}_{(t,s)\in C_2^{(n)}} \xrightarrow[n\to\infty]{} \mathbb{1}_{(t,s)\in C_2}.$$

□