arXiv:2602.06931v1 [math.ST] 6 Feb 2026

贝叶斯后验分布中的微模态及其对马尔可夫链蒙特卡罗的影响

BY SANKET AGRAWAL[1,a], SEBASTIANO GRAZZI[2,b] 和 GARETH O. ROBERTS[3,c]

[1]哥本哈根大学，a*sanket.agrawal@math.ku.dk*

[2]博科尼大学，b*sebastiano.grazzi@unibocconi.it*

[3]华威大学，c*gareth.o.roberts@warwick.ac.uk*

我们研究贝叶斯分析所得后验分布中局部众数的存在性及其严重程度。已知重尾误差模型（例如稳健回归中使用的模型）会导致后验尾部出现这类局部众数。为清楚理解这一现象，我们详细考察样本量为 n、维度为 d 且误差服从 Student-t 分布的位置模型。对于尾部足够重的数据生成分布，随着 $n \to \infty$，极端观测值会变得越来越孤立。我们证明，每个这样的观测值都会以趋近于 1 的概率诱发一个唯一的局部后验模态。我们将这类局部众数称为微模态。这些微模态通常高度较低，但其吸引域很大，并且随 n 呈多项式增长。随后，我们将这种后验几何与计算联系起来。我们为一维分段确定性蒙特卡罗算法离开这些微模态所需的时间确立了阿伦尼乌斯定律。我们的分析识别出一个相变：模型设定错误且过度欠分散时，退出时间会急剧增加，导致采样性能显著恶化。

1. 引言。 现代贝叶斯方法越来越多地应用于大样本情形，并采用重尾模型以适应非典型观测值。计算通常通过通用算法完成，这些算法主要通过评估对数密度及其导数来与后验分布交互。在此类情形下，后验分布的全局性质（例如尾部行为和集中性）可能与由个别孤立观测值诱导的局部几何特征共存。理解这些局部特征如何产生、它们在重尾数据生成机制下出现的频率，以及它们如何影响蒙特卡罗算法的性能，是本文的主要动机。

为在一个易于处理的情形下对这些现象进行解析探索和研究，设 $y_1, \ldots\ldots y_n \in \mathbb{R}^d$ 为来自某个未知分布 P的大小为 n 的独立同分布样本。考虑关于 P的简单贝叶斯位置模型：

$$y_j = x + \varepsilon_j, \qquad \varepsilon_1, \ldots, \varepsilon_n \overset{\text{i.i.d.}}{\sim} t_\nu \tag{1.1}$$

其中， $x \in \mathbb{R}^d$ 是未知的位置参数， t_ν 是具有 $\nu > 0$ 个自由度的标准 d维 Student-t 分布。此类重尾模型因其针对离群值的稳健性而常见于贝叶斯回归中 [（例如参见11, 18, 20, 23, 24, 25, 26, 30, 31, 32, 36]）。

令 $\pi^{(n)}$ 表示模型（1.1）下参数 x 的贝叶斯后验分布，并假设采用非信息先验。在一维情形下，对于 $P = t_1$ 和 $\nu = 1$， [38] 研究了 $\pi^{(n)}$的局部最大值分布。他们证明，当 $n \to \infty$时，每个观测

MSC2020 学科分类： 主要分类 62E20、62-08；次要分类 62F15、65C05。关键词和短语： 稳健贝叶斯推断、离群值、极端次序统计量、多峰性、重尾数据、大样本渐近、马尔可夫链蒙特卡洛。

大于 $2n-\sqrt{n}$ 时,¯以概率1，它在单位距离内恰好对应一个局部最大值。事实证明，这类局部众数的存在性可以在更一般的条件下论证，例如在维度大于一时，以及只要 P 具有足够重的尾部。为说明这一思路，令 $d=1$，并考察负对数后验分布的一阶和二阶导数

$$\frac{-\partial \log \pi^{(n)}(x)}{\nu+1}=\sum_{j=1}^{n}\frac{x-y_j}{\nu+(x-y_j)^2};\qquad \frac{-\partial^2 \log \pi^{(n)}(x)}{\nu+1}=\sum_{j=1}^{n}\frac{(\nu-(x-y_j)^2)}{(\nu+(x-y_j)^2)^2}. \tag{1.2}$$

假设观测值的配置满足以下条件：其中一个观测值（记作 y_n 不失一般性地，远离数据集中的其他观测值且与之孤立，即 $|y_n-y_j|\approx|y_n|\gg 1$ 对所有$j\le n-1$。那么，当 x 接近 y_n时，我们可以用 1.3 近似式（1.2）中的各项

$$\frac{-\partial \log \pi^{(n)}(x)}{\nu+1}\approx\frac{n-1}{x}+\frac{(x-y_n)}{\nu+(x-y_n)^2};\qquad \frac{-\partial^2 \log \pi^{(n)}(x)}{\nu+1}\approx-\frac{n-1}{x^2}+\frac{(\nu-(x-y_n)^2)}{(\nu+(x-y_n)^2)^2}.$$

若 $y_n\gg n$，则 $n/|x|\approx n/|y_n|\ll 1$，上述两个表达式中的第一项都可以忽略。因此，$y_n$ 附近后验分布 $\pi^{(n)}$ 的梯度和曲率仅受 y_n影响。这会使后验密度出现一个异常的凸起，并表现为一个局部最大值。

对于尾部足够厚的真实分布 P，上述这类配置并不少见。随着数据集大小 n或维度 d 的增加，这类情况会变得更加常见。无论是哪种情况，离群观测值都会变得更加偏远，并在其附近形成一个局部最大值。我们将这些局部最大值称为微模态，因为它们是后验分布的局部众数，尽管在各自的吸引域内很高，但绝对高度却很小。图1的上方各图展示了 $d=2$中的一个示例。

微模态给用于拟合贝叶斯模型的大多数计算算法带来了困扰。它们显然会成为优化方法的瓶颈，使其可能无法收敛到全局众数。它们也给马尔可夫链蒙特卡罗（MCMC；概述见[40] ）方法带来挑战，因为这类方法通常会陷入局部众数 [（例如，见 15, 29]）。本文聚焦于 [7], 提出的 Zig-Zag 过程（ZZP），这是一种连续时间、非可逆、分段确定性的蒙特卡罗算法。ZZP 作为一种有望加速蒙特卡罗估计量 [3, 6, 8, 21, 27, 35, 41] 收敛的方法，近期受到理论界和实践界的关注，并得到在多峰情形 [34, 42]中使用的支持。该算法旨在减轻扩散型随机游走行为，转而促进弹道动力学和快速混合。此外，ZZP 具有有原则的子采样特性 [7], ：在算法的每次迭代中，都可以随机选取一部分数据来估计似然，而不会使马尔可夫链蒙特卡罗输出产生偏差。从计算角度看，这使得算法的总体计算成本随数据集规模[1], 呈次线性增长，因此，带子采样的 Zig-Zag 是一种可扩展且比其他传统方法更高效的替代方案。

尽管文献中对重尾模型的贝叶斯分析已有深入研究，但在稳健贝叶斯统计和贝叶斯渐近理论文献中，此类微模态的出现却在很大程度上被忽视了。本文尝试迈出一步，开展整合性分析，将这一情形下的后验几何与计算方法的行为联系起来。本文的主要贡献如下：

1. 我们对后验分布 $\pi^{(n)}$ 中微模态的出现与识别进行了严格分析。假设 P 具有阶数为 $\beta>0$的多项式尾部，极端次序统计量会以 $n^{1/\beta}$的尺度增长并远离中心。我们证明，对于每个这样的极端

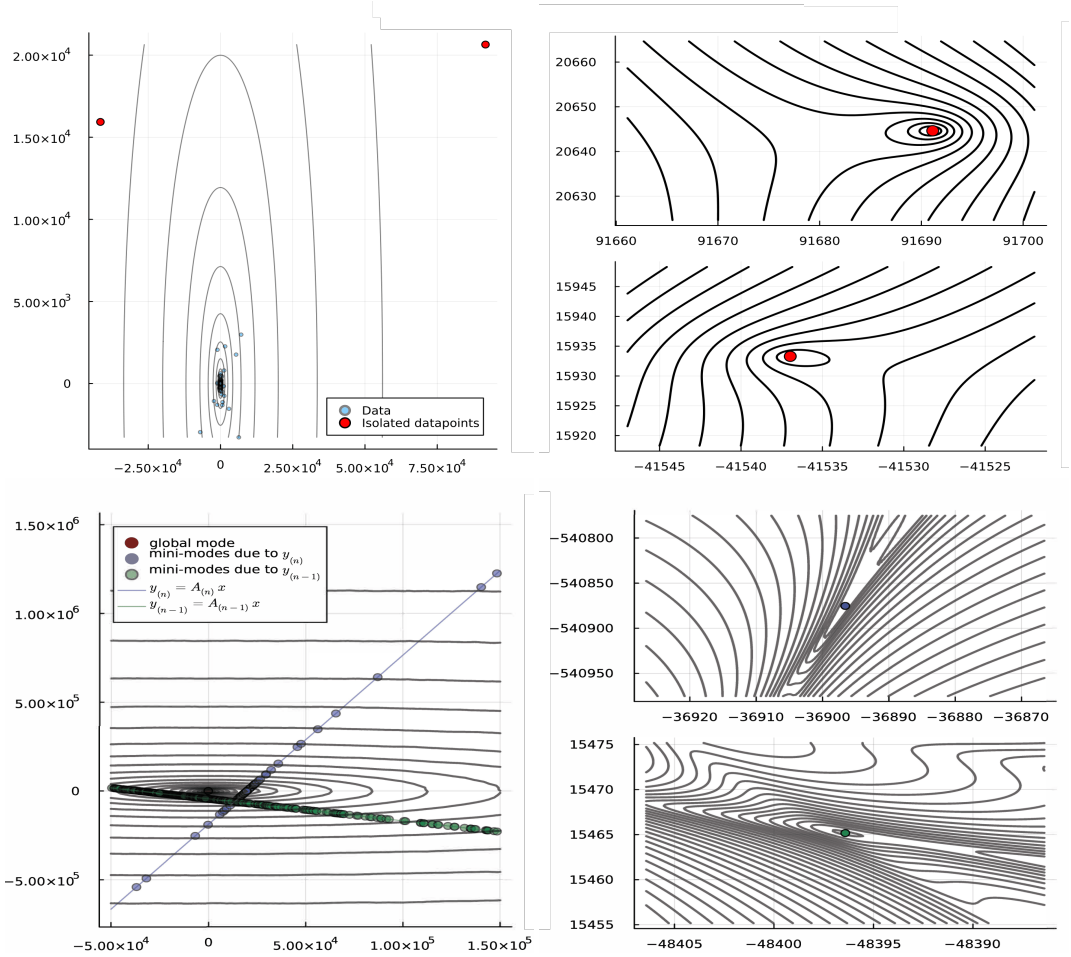


图 1：具有 $n = 10^5$ 个数据点的二维后验分布的等密度等高线。上排图展示模型 (1.1)，其中 $d = 2$。蓝点和红点表示数据点。下排图展示线性模型 $y_j = A_j x + \varepsilon_j$，$d = 2$ 个协变量和 $n = 10^5$ 个数据点。蓝点和绿点表示沿着 $y_i = A_i x$ 所在直线找到的迷你众数，其中 y_i 是最大和第二大的观测值。对于这两个模型，$\varepsilon_j \sim t_1$ 且 $\nu = 1$。右侧图：微模态附近的放大图。

观测，例如 Y，后验分布以趋近于 1 的概率存在一个 唯一的 微模态，位于 $\sqrt{\nu}$-邻域内，中心为 Y（定理 2.3）。我们给出了微模态位置的显式界，并通过其宽度刻画其几何特征 W_n，表明 W_n按多项式尺度变化（定理 $n^{1/\beta-1}$（定理 2.4）。证明结合了极值理论的结果，以及对空间远处区域中经验得分函数的高精度近似（引理 3.1 和定理 3.2）。这将微众数形成转化为一个易于处理的局部求根问题，其替代势函数具有强凸性。

2. 我们利用上述结果研究 ZZP 在一维中逃离微模态所需的时间。通过标准的更新论证，我们证明，对于典范 ZZP 和子采样 ZZP，期望退出时间的阶为 $n^{(1/\beta-1)(\nu+1)}$（定理 2.5）。这些结果表明，采用子采样的 ZZP 可以达到相当的退出时间，同时比其典范对应方法计算成本更低。

3. 我们的分析阐明了假定模型（1.1）与数据生成分布 P之间的相互作用，及其对马尔可夫链蒙特卡罗算法的启示。具体而言，我们

表明，尾部衰减的不匹配会不利地改变后验景观。这会导致一种相变：过于轻尾（欠分散）的假定模型会使 ZZP 的退出时间急剧增加，从而导致采样性能显著恶化。

我们的分析聚焦于（1.1）中的简单模型，但微众数形成背后的定性机制并不局限于这一设定。事实上，只要满足以下条件，类似现象就会更普遍地出现：（i）各个观测值彼此充分孤立；（ii）似然具有重尾特性。图1 底部面板展示了回归模型 $y_i = A_i x + \varepsilon_i$的这一现象，其中误差为单变量重尾噪声 $\varepsilon_i \sim t_1$，协变量 A_i则是一般形式。对于两个最极端的观测值 y_i，在超平面$\{x\colon\ y_i = A_i x\}$上数值发现了数百个微模态。

尽管我们的工作属于理论研究，但我们认为，使用重尾模型时存在重要的方法论启示。首先，微模态很可能会出现。虽然这些众数所包含的后验质量相对于 n呈指数级微小，但其宽广的范围可能会给马尔可夫链蒙特卡罗（以及其他）算法带来严重问题。实践中，从后验分布的全局众数出发启动马尔可夫链蒙特卡罗算法往往非常困难。不过，选择的模型若能确保误差分布至少与实际数据分布一样重尾，就可以减轻微模态带来的最严重影响。

1.1. 相关文献。 我们的研究部分受到 [38],的启发，他们考察了当假定模型和真实分布均为 t_1 （即柯西分布）时，一维情况下微模态的分布。定理2.3通过考虑 d 维中更一般的重尾数据，并允许模型设定错误，在一定程度上呼应并推广了这些结果。一般而言，求解模型（1.1）的似然方程的根，是文献中长期存在的一个问题 [例如，参见 5, 22, 24]。我们的分析推动了识别并刻画这些根的研究。

现有的马尔可夫链蒙特卡罗分析主要关注单峰和对数凹分布[16, 28]。我们考虑的后验分布不属于这一类。历史上，对多峰目标分布上马尔可夫过程的分析一直是在亚稳态框架下进行的 [13]。第2.2 节中的马尔可夫链蒙特卡罗分析拓展了 [34, 37],的研究；他们在双阱势的例子中证明了 ZZP 穿越能量屏障所需时间满足阿伦尼乌斯定律。具体而言，他们表明，当两个势阱之间的距离保持固定时，期望退出时间的对数尺度增长速率与模态的高度（或深度）ε^{-1} 相同，且 $\varepsilon \to \infty$。相比之下，在我们的设定中，能量势垒不仅取决于微模态的高度，也取决于其宽度。最后，与我们的研究思路类似， [4]证明了冷启动下高维贝叶斯逆问题中基于梯度的马尔可夫链蒙特卡罗存在能量势垒。

1.2. 论文结构。 本文其余部分的组织如下。第 2 节给出本文的主要结果。第 2.1 节讨论依赖于 β （见假设 2.1）。第2.2 节给出一维情形下 ZZP 从微模态退出时间的阿伦尼乌斯定律。第 2.3 节讨论上述结果的意义，以及 ν 与 β不匹配所导致的算法性能下降。

第3 节在一般设定下形式化了（1.3）中提出的启发式方法，并利用大数定律，以高精度近似空间远处位置处的经验得分。第4 节给出了定理2.3–2.4的证明，从而确立了微模态的存在性及其宽度。第5 节给出了定理2.5的证明。第6节讨论了这项工作的可能扩展。更多技术性陈述及其证明见附录。

2. 主要结果。

2.1. 微模态的存在性与刻画。令(Ω $\mathcal{F}$， $\mathbb{P}$）为一个任意的概率空间，且足够丰富，能够定义以下所有随机量。对于所有 n， $\{Y_1,\dots,Y_n\}$ 是一个大小为 n 、取值于 $\mathbb{R}^d$ 的随机数据集，其元素独立同分布，服从概率分布 P。对于固定的 k，在 Y_1、……、Y_n 中半径第 k 小的随机向量记为 $Y_{(k)}$。 $Y_{(k)}$ 的相应概率分布记为 $P_{(k)}$。分布 P 满足以下重尾假设。

假设 2.1。令 $|\cdot|$ 表示 $\mathbb{R}^d$ 中的欧几里得范数。分布 P 是 $\mathbb{R}^d$ 上的一个概率分布，其密度为 $p(y) = h(|y|)$，其中 $h\colon (0,\infty) \to (0,\infty)$ 局部有界，且 $\lim_{t\downarrow 0} h(t)$ 为正且有限。函数 h 是正则变差的[1]，其指数为 $-(\beta+d)$，其中 $\beta>0$，并满足 $\lim_{t\uparrow\infty} t^{\beta+d}h(t) = K(\beta,d)$，其中某个常数 $K(\beta,d)>0$。最后，存在一个 t_1，使得 h 在 $[t_1,\infty)$ 上非递增。

粗略地说， P 是一个 d–维各向同性概率分布，其径向密度呈现阶数为 β的多项式尾部。尽管假设2.1 看起来可能有所限制，但这是分析重尾数据时常见的假设，参见例如 [39]。还请注意，标准的多元 Student-t 分布在具有 $\beta>0$ 自由度时满足假设2.1。

在参数贝叶斯模型（1.1）下，对于 P，并假设采用非信息先验，参数 x的后验分布（记作 $\pi^{(n)}$）为：

$$\pi^{(n)}(x) \propto \prod_{j=1}^{n} \frac{1}{(\nu+|x-Y_j|^2)^{(\nu+d)/2}}, \quad x\in\mathbb{R}^d. \tag{2.1}$$

令 $S(x,\ y)\colon = (x-y)/(\nu+|x-y|^2)$。我们用 S_n 表示经验得分

$$S_n(x) := \frac{-1}{\nu+d}\nabla_x \log \pi^{(n)}(x) = \sum_{j=1}^{n} S(x,Y_j), \quad x\in\mathbb{R}^d. \tag{2.2}$$

同理

$$S_n'(x) := \frac{-1}{\nu+d}\nabla_x^2 \log \pi^{(n)}(x) = \sum_{j=1}^{n} S'(x,Y_j), \quad x\in\mathbb{R}^d, \tag{2.3}$$

表示经验信息矩阵。

寻找式（ $\pi^{(n)}$ 2.1）中定义的后验分布的局部众数 $S_n(x)=0$，等价于寻找满足二阶条件$S_n'(x)\succ 0$的$S_n'(x)$，即 [38], 为正定的$\pi^{(n)}$ 的根。重复 $\sqrt{\nu}-$ 中的论证可知，的任意局部最大值都必须位于某个数据点周围半径为 $n^{1/\beta}$的球内。然而，并非所有点的邻域内都会出现局部最大值。如引言所述，数据点能否产生局部最大值，取决于它与数据集其余部分的隔离程度。假设2.1 的一个直接推论是，数据集中的极值在径向分量上的尺度以 $n^{1/\beta}$的多项式速率增长。

命题 2.1. 假设 P 对某个 $\beta>0$满足假设 *2.1* 。那么，对于每个固定的 $k\geq 0$，存在一个 G_k 服从伽马分布，其形状参数为$k+1$，尺度为单位尺度，满足

$$\frac{|Y_{(n-k)}|}{n^{1/\beta}} \xrightarrow{d} \left(\frac{G_k}{A}\right)^{-1/\beta},$$

[1]可测函数 $h\colon (0,\infty)\to(0,\infty)$ 具有指标为 $\rho\in\mathbb{R}$ 的正则变差，如果对所有 $r>0$，$\lim_{t\to\infty} h(rt)/h(t) = r^{\rho}$。在这种情况下，我们记作 $h\in RV_{\rho}$。

当 $n \to \infty$ 时，其中 $A = \frac{2\pi^{d/2} K(\beta,\ d)\ \beta\ \Gamma(d/2)}{}$ 表示伽马函数。

此外，对于所有 $0 < \epsilon < 1/\beta$，当 $n \to \infty$ 时，

$$\mathbb{P}\left(\min_{j \neq n-k} |Y_{(j)} - Y_{(n-k)}| > |Y_{(n-k)}|/2n^{\epsilon} \right) \longrightarrow 1.$$

为使本文余下部分的构造明确，我们假设命题 2.1 中的分布收敛几乎必然成立。注意，由于斯科罗霍德表示定理，这一假设是可行的。

命题 2.1 意味着，对于每个固定的 k，半径排名第 $k-$ 大的点 $Y_{(n-k)}$ 会以速率 $n^{1/\beta}$ 变得离其余数据集越来越远且孤立。由（1.3）可知，邻近 $Y_{(n-k)}$ 的其余数据集的得分贡献此时可近似为 $n^{1-1/\beta}$。当 $\beta > 1$ 时，$n^{1-1/\beta}$ 趋于无穷。因此，正如下一个定理所示，$Y_{(n-k)}$ 的孤立程度不足以使 $Y_{(n-k)}$ 的 $\sqrt{\nu}-$ 邻域内出现微模态。令 $B_r(y)$ 表示以 y 为中心、半径为 r 的球。

定理 2.2. 假设2.1 在 $\beta > 1$ 下成立。令 $k \geq 0$ 固定。随着 $n \to \infty$，以趋近于1 的概率，对于所有 $x \in B_{\sqrt{\nu}}(Y_{(n-k)})$，存在一个 $v \in \mathbb{S}^{d-1}$，使得$v^T S_n(x) < 0$。

另一方面，当 $\beta \leq 1$时，(1.3) 中的第一项随着 n 的增加而变得可以忽略。特别是，只要 $|Y_{(n-k)}| > 2n\sqrt{\nu}$，$Y_{(n-k)}$ 在其 $\sqrt{\nu}-$邻域内的似然贡献就会超过数据集的其余部分，从而产生一个微模态。

定理 2.3. 假设 2.1 在 $0 < \beta \leq 1$下成立。令 $k \geq 0$ 固定。对于$n > k$，令 E_n 为(4.1)中定义的事件。存在一个 $N_0 \geq 4$，使得对于所有 $n > N_0$，在事件 E_n上，存在 x_n^+ 在 $\pi^{(n)}$ 的 $\sqrt{\nu}-$ 邻域内的唯一局部最大值 $Y_{(n-k)}$。此外，x_n^+ 满足：

$$|Y_{(n-k)} - x_n^+| \leq \left(\frac{4\nu}{G_k + c_\beta} \right) n^{1-1/\beta}; \qquad c_\beta = \begin{cases} 0, & 0 < \beta < 1, \\ 2\sqrt{\nu}, & \beta = 1. \end{cases} \tag{2.4}$$

其中 G_k $2n\sqrt{\nu}\}$ $E_n \subset A_n$ n $n \to \infty$ 2.1 $A_n = \{|Y_{(n-k)}| >$ 如命题 . 的陈述所示。此外，定义 :。那么，当 足够大时，且当 时，

$$\mathbb{P}(E_n \mid A_n) \to 1, \quad \mathbb{P}(A_n) \longrightarrow \begin{cases} 1, & 0 < \beta < 1, \\ \mathbb{P}(G_k > 2\sqrt{\nu}), & \beta = 1. \end{cases}$$

定理 2.3 是命题 2.1 给出的 $Y_{(n-k)}$ 缩放关系，以及引理 3.1 中通过大数定律对经验得分 S_n 进行近似的结果。鉴于 $|Y_{(n-k)}| > 2n\sqrt{\nu}$，这意味着以趋于 1 的概率，在距离 $Y_{(n-k)}$ 为 $\sqrt{\nu}-$的范围内恰有一个局部最大值。这一结论可轻易推广至任意有限个 K极值 $\{Y_{(n-k_1)} \cdots, Y_{(n-k_K)}\}$：只需应用简单的并集界以及极端次序统计量的联合弱收敛 [，参见 19，定理 2.1。1]。定理 2.3 将 [38] 推广至维度大于一的情形、更一般的数据分布，以及可能错设的模型（此处为 ν 与 β），而此前该结论仅针对设定正确的模型中的一维柯西分布。不过，这并非完整的推广，因为定理 2.3 仅适用于有限多个点，因此并未涵盖所有大于 $2n\sqrt{\nu}$ 的观测值（另见引理 3.1 后的讨论）。

令 k 固定。对于较大的 n，令 x_n^+ 为定理 2.3 保证的、在 $Y_{(n-k)}$ 邻域内 $\pi^{(n)}$ 的唯一局部最大值。于是，$S_n(x_n^+) = 0$ 和 $S_n'(x_n^+)$ 为正

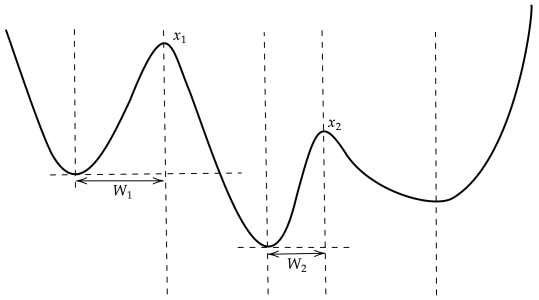


图 2：对于某个任意函数 f， x_1 和 x_2 处局部极大值的宽度。

定的。 S_n 的连续性意味着存在一个 $t>0$，使得对于所有 $v\in\mathbb{S}^{d-1}$ 和 $0<u<t$，$v^T S_n(x+vu)>0$。这意味着从 x_n^+ 出发，后验密度在所有方向上、距离为 t 的范围内单调递减。我们将与 $Y_{(n-k)}$ 关联的微模态的宽度定义为最大的此类 t，并用 W_n 表示。更准确地说，我们定义：

$$W_n := \sup\left\{t\geq 0 : v^T S_n(x_n^+ + vt) > 0 \;\forall\, v\in\mathbb{S}^{d-1}\right\}. \tag{2.5}$$

请注意，局部最大值宽度的概念 x 对于任意函数 f 也可以作更一般的定义。粗略地说，它是从 x开始， f 在所有方向上单调递减所能达到的最大距离。图2 展示了一维情形。下一个定理表明， W_n 按 $n^{1/\beta-1}$ 的尺度变化，其中 $n\to\infty$。

定理 2.4。假设条件2.1 在 $0<\beta\leq 1$下成立。令 $k\geq 0$ 为固定值，并令n 满足定理2.3 的条件。令 W_n 为按(2.5)定义的微模态宽度。那么，在事件 E_n上，

$$\left(\frac{G_k + c_\beta}{2}\right) n^{1/\beta-1} - 2\sqrt{\nu} < W_n < 4G_k n^{1/\beta-1} + \sqrt{\nu},$$

其中c_β 定义见(2.4)。

定理2.4与定理2.3共同刻画了由极端观测值引起的微模态。特别是，定理2.4 给出了微模态存在时其大小的估计。从计算角度来看，这些结果意味着后验分布中存在一个能量势垒，马尔可夫链蒙特卡罗算法必须克服该势垒才能离开微模态。在下一节中，我们将研究 ZZP 从某个微模态初始化时，克服这一能量势垒所需的时间。

2.2. *ZZP* 离开微模态的退出时间。本节余下部分，我们将注意力限制在一维情形，并特别关注与 $Y_{(n)}$ （即最远数据点）相关的微模态。读者可能会发现，即便是这个看似简单的设定，也需要进行非平凡的分析。

一维 ZZP $(X_t,V_t)_{t\geq 0}$ 是定义在位置与速度增广空间 $\mathcal{Z}=\mathbb{R}\times\{-1,+1\}$ 上的连续时间马尔可夫过程。给定切换率 $\lambda\colon \mathcal{Z}\mapsto\mathbb{R}^+$，其动力学由扩展马尔可夫生成元确定

$$\mathcal{L}f(x,v) = v\cdot\partial_x f(x,v) + \lambda(x,v)\left\{f(x,-v)-f(x,v)\right\},\quad (x,v)\in\mathcal{Z}, \tag{2.6}$$

对于生成元 [核心中的函数 f，参见例如 17]。粗略地说， X_t 以恒定速度 V_t 运动，并在速率为 $\lambda(X_t,V_t)$的随机时刻，在速度值 $+1$ 和 -1之间切换。对于每个 n，令 $Z_t^n=(X_t^n,V_t^n)$ 为一维 ZZP，其

切换率 λ_n。若 $\lambda_n(x,1) - \lambda_n(x,-1) = -\partial \log \pi^{(n)}(x) = (\nu+1)S_n(x)$ 对所有 x成立，则 X_t^n的边缘分布对 $\pi^{(n)}$ [是不变的，详见 7，了解详情]。我们考虑以下两种切换率：

$$\lambda_n^{\text{can}}(x,v) = \left(v\partial \log \pi^{(n)}(x)\right)_+ = (\nu+1)\left(v\sum_{j=1}^n S(x,Y_j)\right)_+ ,$$
$$\lambda_n^{\text{ss}}(x,v) = (\nu+1)\sum_{j=1}^n \left(vS(x,Y_j)\right)_+ . \tag{2.7}$$

选择 λ_n^{can}和 λ_n^{ss} 分别对应于 [7]的典范 ZZP 和子采样 ZZP。虽然典范 ZZP 的混合效果可能优于子采样 ZZP [2], ，但子采样 ZZP 的优势在于能将每次迭代的实现成本降低 n倍。这是因为 $\lambda_n^{\text{ss}}(x,v)$ 可以通过随机选择单个观测值（例如 Y_i），再计算 $n(\nu+1)(vS(x,Y_i))_+$来进行无偏估计，因此与典范 ZZP 相比，通常能提高整体计算效率 [; 详情请参见1]。当上下文能够明确 λ_n 选择的是哪一种时，我们将省略上标。

当 n 足够大时，令 x_n^+ 为定理 2.3 在事件 E_n 上给出的、位于 $Y_{(n)}$ 附近的局部最大值点。令 W_n 表示由 (2.5) 定义的相应微模态宽度。假设以 $\pi^{(n)}$ 为目标的 ZZP 的初始条件为 $X_0^n = x_n^+$。该过程退出微模态所需的时间，是满足 $|X_t^n - x_n^+| > W_n$ 的第一个时刻 $t \geq 0$。不失一般性，假设 $Y_{(n)} > 0$ （关于 $Y_{(n)} < 0$ 的论证类似）。由 (2.2) 以及 $\sqrt{\nu}$-邻域内 $Y_{(n)}$ 处 S_n' 的正定性（引理 4.2），可得 $S_n(x) > 0$ 对所有 $x \geq x_n^+$ 均成立，即 $\pi^{(n)}$ 在半直线 $[x_n^+, \infty)$ 上单调递减。在这种情况下，向右退出并不重要，因为该过程总会以概率 1 返回微模态。因此，我们将过程首次越过左侧宽度 W_n 时视为退出微模态。更准确地说，对于 ZZP $Z_t^n = (X_t^n, V_t^n)$，我们将退出时间 τ_n 定义为

$$\tau_n = \inf\left\{t \geq 0 : x_n^+ - X_t^n \geq W_n \mid (X_0^n, V_0^n) = (x_n^+, -1)\right\}. \tag{2.8}$$

要研究 τ_n 的精确情况，点 x_n^+和宽度 W_n 都必须确切已知。即使对于较小的 n，这也是一项艰巨的任务，参见例如 [22]。然而，事实证明，定理 2.4 中对宽度的估计足以将 τ_n 估计到一个数量级。

定理 2.5. 设假设 *2.1* 在 $0 < \beta \leq 1$下成立。对所有 $n \geq 3$，令 Z_t^n为切换率由 (2.7)给出的 *Zig-Zag* 过程。令 τ_n 如 (2.8)所定义。则对 λ_n^{can}和 λ_n^{ss}两者，当 $n \to \infty$

$$\mathbb{E}(\tau_n \mid E_n) = O_p\left(n^{(1/\beta-1)(\nu+1)}\right),$$

其中 E_n 如(4.1)所示。

因此，子采样 ZZP 与典范 ZZP 离开微模态所需的时间大致相同，鉴于其有利的计算成本，子采样 ZZP 是一个有吸引力的选择。

2.3. 本质微模态与模型设定错误。 众所周知，对重尾数据集使用尾部更轻的模型，即 $\nu > \beta$，会导致分配给离群值的似然权重更大。因此，马尔可夫链蒙特卡罗方法会在后验分布相应微模态附近花费更长时间。定理2.5 以严格方式刻画了（1.1）中的贝叶斯模型与ZZP的这一关系。

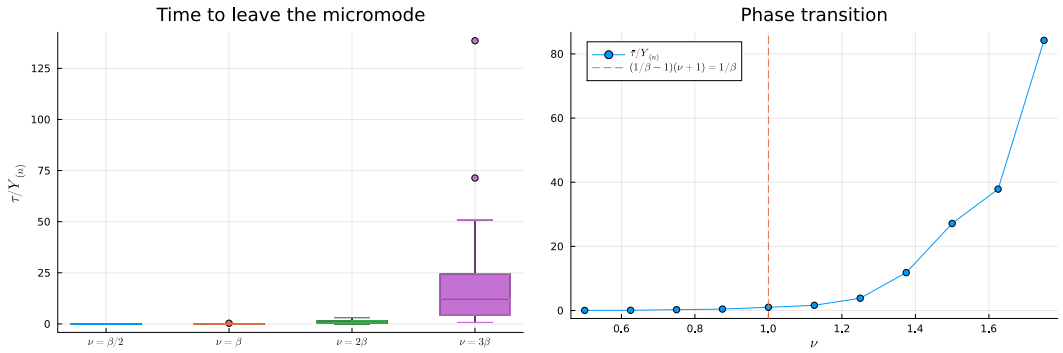


图 3：对于以 π_ν 为目标的典范 ZZP，从 $Y_{(n)}$ 附近的微模态开始，在不同 ν 值（x 轴）下，离开 $Y_{(n)}$ 附近微模态所需的时间，并按 $|Y_{(n)}|$ 进行归一化。此处，$n = 3000$ 和 $\beta = 1/2$。左：每个箱线图由 40 条独立轨迹计算得出。右：对于以 π_ν 为目标的 40 条独立轨迹，在不同 ν（x 轴）下计算得到的离开微模态的平均时间。相变 $\beta/(1-\beta) = 1$ 处以虚线表示。

现考虑一个在其位置分量上初始化于 $Y_{(n)}$ 的ZZP。若它未陷入某个微模态，仍需要 $O_p(|Y_{(n)}|)$ 的时间才能到达后验集中，因为该过程以单位速度 ± 1移动。由命题2.1，$Y_{(n)} = O_p(n^{1/\beta})$，故该行进时间具有相同的阶，并随 $n \to \infty$发散。因此，我们将该过程的时间按因子 $n^{1/\beta}$重新标度。若原过程的退出时间为 $\tau_n \ll n^{1/\beta}$，则重新标度后，该时间相对于到达分布主体所需的总时间可忽略不计。反之，若 $\tau_n \gg n^{1/\beta}$，则即使在重新标度之后，该过程仍会在 $Y_{(n)}$ 附近滞留不合理的时间。自然，后一种情形是不可取的。此时，我们称该微模态为本质微模态。更精确地说，一个微模态是本质微模态，如果

$$\frac{\mathbb{E}\left(\tau_n \mid E_n\right)}{|Y_{(n)}|} \to \infty \tag{2.9}$$

如 $n \to \infty$所示。根据定理 2.5 和命题 2.1，当满足以下条件时会出现本质微模态：

$$(1/\beta - 1)(\nu + 1) > 1/\beta \iff \nu > \frac{\beta}{1-\beta},$$

以及 t 因此，这取决于假定模型与真实数据分布的组合 P .

鉴于上述结果，请注意，设定正确的模型，即 $\beta = \nu$，永远不会产生本质微模态，因为 $0 < \beta \leq 1$；对于以下两种情况，结论同样成立：**(i)** 过度离散模型，即 $\nu < \beta$；以及 **(ii)** 当 $\beta = 1$时。然而，当 $0 < \beta < 1$时，在阈值 $\beta/(1-\beta)$ 之外的轻尾模型会出现相变，并产生会使过程长时间陷入其中的本质微模态。这一相变在图 3 中针对$\beta = 1/2$进行了说明。如图所示，$\tau_n/|Y_{(n)}|$ 在 $\nu > 1$时开始发散。

3. 经验得分近似。 在本节中，我们将第1节所阐述的论证形式化。设 Y_1，……，$Y_n \sim P$为 $\mathbb{R}^d$中独立分布的随机向量，并假设 P满足假设2.1 （在 $0 < \beta \leq 1$下）。

本节的主要技术结果是引理 3.1，该引理表明，在空间的远处位置，经验得分可以通过与数据无关的函数以很高的精度近似。

引理 3.1。 对每个 $n \geq 4$，定义

$$\mathbb{S}_n(z) = \sup_{x:\, |x-z| \leq 2|z|/n} \frac{2|z|}{n} \left| S_n(x) - \frac{nx}{|x|^2} \right|. \tag{3.1}$$

令 $0 < \epsilon < 1$ 任意。存在一个 $T_0(\epsilon)$，使得对所有 $n \geq T_0(\epsilon)$ 以及 $|z| > 2n\sqrt{\nu}$，

$$\mathbb{S}_n(z) \leq 48\nu n^{2\epsilon-1} + \frac{|z|^{-\beta}}{n^{\epsilon(d-1)}} L'(|z|) + g(|z|) + R_n(|z|) \tag{3.2}$$

其中 $L'(|z|) \in RV_0$、$g(|z|) \in RV_{-\beta}$，而 R_n 是一个依赖于 $Y_1, \ldots, Y_n$ 和 $|z|$ 的随机变量，并且对所有 $t > 0$，

$$\mathbb{P}\left(R_n(|z|) > t\right) \leq 5^d \exp\left(-\frac{n^{1-2\epsilon}t^2}{512}\right) + C_d \frac{5^d |z|^{-\beta}}{2^d n^{d\epsilon-1}} L(|z|). \tag{3.3}$$

这里 C_d 是一个依赖于维数的常数，而 $L(t) = t^{\beta+d}h(t)$ 对于 h 在假设 2.1 中。

引理 3.1 的结果本质上是经验得分 S_n 的大数定律误差界。对于 $|z| > 2n\sqrt{\nu}$ 和任意的 $\epsilon > 0$，以高概率，数据集中的大部分数据点与球 $|z|/n^{2\epsilon}$ 相距至少 $B_{2|z|/n}(z)$。对于所有这类数据点，得分函数一致有界于 $n^{1+2\epsilon}/|z|$，因此适用标准的集中不等式。另一方面，数据集中距离该球较近或位于球内的部分所产生的总得分贡献，可由这类数据点的数量直接控制。我们通过结合这些观察结果以及对期望得分的一致逼近（即 $\mathbb{E}(S_n(x))$ 由函数 $x \mapsto nx/|x|^2$ 逼近），证明引理 3.1 。详情见附录 C 。

注记3.1。设 $(z_n)_{n=1}^{\infty}$ 是一个满足 $|z_n| > 2n\sqrt{\nu}$ 的序列。引理3.1 给出了在 z_n 的邻域内对 S_n 进行近似的误差界。随着 $|z_n|$ 趋于 ∞，近似会变得更加精确。如果 $\mathbb{S}_n(z_n)$ 随着 $n \to \infty$ 趋于0 ，这意味着在 $(|z_n|/n)-$ z_n 邻域内，经验得分 S_n 可由 $nx/|x|^2$ 很好地近似。假设 $|z_n| \sim n^{\alpha}$ 对某个 $\alpha > 0$ 成立，且满足 $|z_n| \succ 2n\sqrt{\nu}$。通过选择 $\epsilon < 1/2$，由(3.3)可知，对于任意满足 $\alpha\beta > (1-d/2)$ 的 α，(3.2)中的右端趋于0 。因此，若将一个数据点人为放置在 z_n 上，根据数据集其余部分的构形，它可能成为一个潜在的微模态。若该数据点与数据集其余部分充分隔离，这个潜在的微模态就会成为实际的微模态。

备注 3.2. 上述讨论可推广到任何独立于数据集 $\{Y_1, Y_2, \cdots, Y_n\}$ 定义的随机序列 Z_n，其中 $_{n=1}^{\infty}$。假设 $|Z_n| \sim n^{\alpha}$ 对于 $\alpha > (1-d/2)/\beta$。那么，结合引理 3.1 与一个简单的条件化论证可得，对所有 $\delta < 1/2$，

$$\mathbb{P}\left(\mathbb{S}_n(Z_n) > n^{1-\alpha-\delta} \mid |Z_n| > 2n\sqrt{\nu}\right) \to 0$$

当 $n \to \infty$ 时。注意， α 必须满足 $\alpha \geq 1$，否则 $\mathbb{P}(|Z_n| > 2n\sqrt{\nu}) \to 0$，结论便是空泛的。

固定 $k \in \mathbb{N}$。根据命题 2.1， $|Y_{(n-k)}| \sim n^{1/\beta}$ 且对所有 β 和 d， $1/\beta > (1-d/2)/\beta$。因此，引理 3.1 表明，在 $Y_{(n-k)}$ 附近，其余 $(n-1)$ 个观测值的得分贡献可用 $(n-1)x/|x|^2$ 近似。受此启发，我们在 $Y_{(n-k)}$（即第 $(n-k)-$ 个次序统计量）附近定义一个近似得分函数。定义 $\mathcal{F}_n = \{x : \ |x - Y_{(n-k)}| < 2|Y_{(n-k)}|/n\}$ 并

$$\hat{S}_n(x) = S_n(x)\mathbf{1}_{\mathcal{F}_n^c} + \left(\frac{(n-1)x}{|x|^2} + S(x, Y_{(n-k)})\right)\mathbf{1}_{\mathcal{F}_n}, \quad x \in \mathbb{R}^d. \tag{3.4}$$

我们现在陈述本节的主要结果。

定理 3.2. 假设 $0 < \beta \leq 1$ 并固定 $k \geq 0$。对于 $n \geq k$，令事件 $A_n = \{|Y_{(n-k)}| > 2n\sqrt{\nu}\}$。则对所有 $\delta < 1/2$，当 $n \to \infty$ 时，

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}^d}\left|S_n(x) - \hat{S}_n(x)\right| > n^{1-1/\beta-\delta} \mid A_n\right) \longrightarrow 0. \tag{3.5}$$

注释 3.3. 注意，$Y_{(n-k)}$ 与数据集中的其余部分并非相互独立。因此，定理3.2并不能像注释3.1 中所述，直接由引理3.2推出。不过，可以通过定义3.2 中所述的过程，使 Z_n 与 $Y_{(n-k)}$ 具有相同的分布，并以 Z_n 确实是数据集$(n-k)$–次次序统计量 $\{Y_1, \ldots, Y_{n-1}, Z_n\}$这一事件为条件。详情请见第A.1 节。

注释 3.4. 由于（3.5）本质上是大数定律的结果，因此还可以进一步改进误差率，证明 $|S_n - \hat{S}_n|_\infty \sim n^{1/2-1/\beta}$。不过，本文不再探讨这一推广。

定理 3.2 的用处在于，它使人们能够将经验得分 S_n替换为某种简单得多、也更易于分析的量。我们在下一节中使用定理 3.2 来证明在 $Y_{(n-k)}$ 的邻域中存在一个局部最大值（定理 2.3），并度量其宽度（定理 2.4）。这是用于推导定理 2.5 中 micromodeforZZP 的退出时间的两个主要要素，其证明在第 5 节中给出。

4. 微模态的存在性与宽度。 在此，我们将严格证明定理 2.3 和定理 2.4。注意，由于 $\pi^{(n)}$ 是连续可微的，因此在闭球 $B_{\sqrt{\nu}}(Y_{(n-k)})$ 上，对任意 $k \geq 0$ 都存在一个局部最大值。我们将证明，当 $n \to \infty$ 时，这样的点是唯一的，并且处于阶为 $n^{1-1/\beta}$、自 $Y_{(n-k)}$ 算起的距离处。此外，我们将证明相应微模态的宽度（2.5）为 $n^{1/\beta-1}$ 的阶。我们首先陈述一个一般性结果，用以证明 $\pi^{(n)}$ 在给定邻域中恰好存在一个局部最大值。

引理 4.1. 设 $U : \mathbb{R}^d \to \mathbb{R}$ 在某个闭球 $B_r(y)$ 的邻域内二阶连续可微，其中 $r > 0$ 和 $y \in \mathbb{R}^d$任意。假设
1. 对所有 $x \in \partial B_r(y)$，$\nabla U(x)^T(x-y) > 0$.2. 对所有 $x \in B_r(y)$，$\nabla^2 U(x) \succ 0$.
那么，存在唯一的 $x^* \in B_r(y)$ 使得

$$U(x^*) = \min_{x\in\overline{B_r(y)}} U(x)$$

此外，$\nabla U(x^*) = 0$。

证明。由于 U在闭球 $B_r(y)$的邻域内连续，因此它在 $B_r(y)$上取得最小值。设 $x^* \in B_r(y)$ 是 U 的一个极小点。为导出矛盾，假设 $x^* \in \partial B_r(y)$。令 $t \in (0,1)$。那么，根据假设，

$$U(x^* + t(y - x^*)) \geq U(x^*)$$

因为 $x^* + t(y-x^*) = (1-t)x^* + ty \in \overline{B_r(y)}$。除以 t 并取 $t \downarrow 0$，得到 ，

$$\nabla U(x^*)^T(y-x^*) \geq 0 \iff \nabla U(x^*)^T(x^*-y) \leq 0$$

这就产生了矛盾。因此， x^* 必定位于球 $B_r(y)$内。由于 x^* 是开球 $B_r(y)$内 U的无约束极小值点，因此必须满足一阶条件 $\nabla U(x^*)=0$。唯一性由凸性假设得出，即对所有 $x\in B_r(y)$，都有 $\nabla^2 U(x)\succ 0$。

□

设 G_k 如命题 2.1 的陈述所述。任取 $0<\epsilon,\delta<1/2$。对于 $n\geq 3$，定义事件：

$$A:=\{\omega\in\Omega: G_k(\omega)>c_\beta\},\quad A_n':=\left\{\min_{j\neq n-k}|Y_{(j)}-Y_{(n-k)}|>|Y_{(n-k)}|/2n^\epsilon\right\},$$
$$B_n:=\left\{\sup_x|S_n(x)-\hat{S}_n(x)|<n^{1-1/\beta-\delta}\right\},\quad E_n:=A_n'\cap B_n\cap A, \tag{4.1}$$

其中 S_n 定义见（3.4），c_β 由（2.4）给出。为证明定理 2.3，我们将验证引理 4.1 中关于 $U(x)=(-\log\pi^{(n)})(x)$的两个条件。首先是强凸性条件。回顾一下，$S_n'(x)$ 是对应于 $S_n(x)$ 的雅可比矩阵，即 $S_n'(x)=\sum_{j=1}^n S'(x,Y_j)$，其中

$$S'(x,y):=\frac{1}{\nu+|x-y|^2}I_d-\frac{2}{(\nu+|x-y|^2)^2}(x-y)(x-y)^T.$$

引理 4.2. 任取 $0\leq r<\sqrt{\nu}$。存在一个依赖于 r 的 N，且足够大，使得对所有 $n>N$，在事件 E_n 上，$S_n'(x)\succ 0\ \forall\, x\in B_r(Y_{(n-k)})$。

证明。对于任意满足 $|a|=1$ 的 $a\in\mathbb{R}^d$，考虑

$$a^TS'(x,y)a=\frac{1}{\nu+|x-y|^2}-\frac{2|a\cdot(x-y)|^2}{(\nu+|x-y|^2)^2}\geq\frac{(\nu-|x-y|^2)}{(\nu+|x-y|^2)^2}.$$

在 $E_n\subseteq A_n'$ 上，最小值$_{j\neq n-k}|Y_{(j)}-Y_{(n-k)}|>|Y_{(n-k)}|/2n^\epsilon$。因此，对于所有 $x\in\overline{B_r(Y_{(n-k)})}$，可知在事件 A_n' 上，最小值$_{j\neq n-k}|x-Y_{(j)}|\geq|Y_{(n-k)}|/2n^\epsilon-\sqrt{\nu}=:\delta_n$。由于$|Y_{(n-k)}|/2n^\epsilon$ 会随着 n 变大而趋于无穷大，因此当 $n\to\infty$时，$\delta_n\to\infty$。注意，函数 $s\mapsto(\nu-s^2)/(\nu+s^2)^2$ 在 $(\sqrt{3\nu},\infty)$ 上递增。因此，对于足够大的 n，只要 $\delta_n>\sqrt{3\nu}$，就有对所有 $j\neq(n-k)$，

$$a^TS'(x,Y_{(j)})a\geq\frac{(\nu-\delta_n^2)}{(\nu+\delta_n^2)^2},\quad x\in\overline{B_r(Y_{(n-k)})}.$$

另一方面，对于 $j=(n-k)$，

$$a^TS'(x,Y_{(n-k)})a\geq\frac{\nu-r^2}{(\nu+r^2)^2}.$$

综合两者可得，在事件 A_n' 上，且对所有 $x\in\overline{B_r(Y_{(n-k)})}$，

$$a^TS_n'(x)a\geq\frac{(n-1)(\nu-\delta_n^2)}{(\nu+\delta_n^2)^2}+\frac{\nu-r^2}{(\nu+r^2)^2}\to\frac{\nu-r^2}{(\nu+r^2)^2}$$

如 $n\to\infty$。为此，注意第一项按 $-n/\delta_n^2$ 的尺度变化，其中 $\delta_n\sim|Y_{(n-k)}|/n^\epsilon\sim n^{1/\beta-\epsilon}$，由命题 2.1 可知。因此，$n/\delta_n^2\to 0$，因为 $0<\beta\leq 1$ 且 $0<\epsilon<1/2$。由于 $r<\sqrt{\nu}$，上述结论意味着存在一个 N，使得对于所有 $n\geq N$，在事件 A_n' 上，我们有$a^TS_n'(x)a>0$；对于所有 $x\in B_r(Y_{(n-k)})$ 以及所有满足 $|a|=1$ 的 a，这一结论均成立。证明至此完成。□

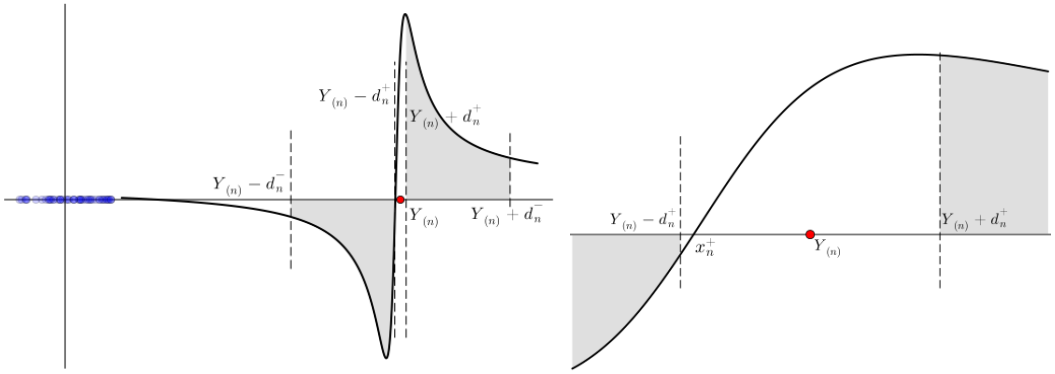


图 4：示意图展示了一维中 $S_n(x)$ 以及 d_n^+ 和 d_n^- 的构造。长度 $d_n^\pm$ 的选取满足：对于所有 $t \in [d_n^+, d_n^-]$ 和 $v = \pm 1$，$vS_n(Y_{(n)} + vt) > 0$。右侧面板放大显示了 S_n 穿过 x–轴的区域。

上述引理意味着，如果 $\pi^{(n)}$ 在球 $B_{\sqrt{\nu}}(Y_{(n-k)})$ 中存在局部最大值，那么该局部最大值唯一，并且满足 $S_n(x) = 0$。剩下要证明的是，该局部最大值在 $Y_{(n-k)}$ 附近存在。在事件 A 上，令 m_1, m_2 满足 $0 \le c_\beta < m_1(\omega) < m_2(\omega) < G_k(\omega)$，对每个 $\omega \in A$ 均成立。定义，

$$d_n := \frac{n^{1-1/\beta}}{m_1}; \quad d_n^\pm := \frac{1 \mp \sqrt{1 - 4d_n^2\nu}}{2d_n}; \quad n \ge 3. \tag{4.2}$$

请注意，$d_n^\pm$ 并非常数，而是通过 m_1依赖于 G_k 的取值。由（4.2）可知，只有当 $4d_n^2\nu < 1$ 时，$d_n^\pm$才为实数。对于 $\beta = 1$，根据 m_1 的选取和c_β的定义，$d_n = 1/m_1 < 1/c_\beta = 1/(2\sqrt{\nu})$。对于 $\beta < 1$，当 $n \to \infty$时，$d_n \to 0$。因此，存在 N_1，使得对于所有 $n > N_1$，$d_n < 1/(2\sqrt{\nu})$ 且 $d_n^\pm$为实数。此外，在事件 A上，$n^{-1/\beta}|Y_{(n-k)}| \to G_k > m_2$。因此，存在 N_2，使得对于所有 $n > N_2$，$n^{-1/\beta}|Y_{(n-k)}| > m_2$。这意味着对于所有 $n > \max\{N_1, N_2\}$，

$$d_n^+ \le 2d_n\nu < \sqrt{\nu} < \frac{\sqrt{\nu}}{2d_n} < d_n^- < \frac{1}{d_n} < \frac{m_2}{n^{1-1/\beta}} < \frac{|Y_{(n-k)}|}{n}.$$

参见图4，查看一维情况下的示意图。

引理 4.3. 对于足够大的 n，在事件 E_n 上，

$$\inf\left\{v^T S_n(Y_{(n-k)} + vt) : t \in [d_n^+, d_n^-], v \in \mathbb{S}^{d-1}\right\} > 0.$$

证明。假设 $n >$ 最大值$\{N_1, N_2\}$，则 $d_n^- \le |Y_{(n-k)}|/n$，并且对于所有 $v \in \mathbb{S}^{d-1}$ 和 $t \in [d_n^+, d_n^-]$，我们有，

$$\begin{aligned} v^T S_n(Y_{(n-k)} + vt) &> v^T \hat{S}_n(Y_{(n-k)} + vt) - n^{1-1/\beta-\delta} \\ &= \frac{(n-1)}{|Y_{(n-k)} + vt|^2} v^T(Y_{(n-k)} + vt) + \frac{t}{\nu + t^2} - n^{1-1/\beta-\delta}. \end{aligned}$$

在事件 $E_n \subseteq B_n$ 上。令 $t \in [d_n^+, d_n^-]$ 固定。对于任意 $v \in \mathbb{S}^{d-1}$，令 $\phi = v^T Y_{(n-k)}$。于是，$\phi \in [-|Y_{(n-k)}|, |Y_{(n-k)}|]$。考虑函数，

$$f(\phi) = (n-1)\frac{\phi + t}{|Y_{(n-k)}|^2 + t^2 + 2\phi t} \implies f'(\phi) = (n-1)\frac{|Y_{(n-k)}|^2 - t^2}{(|Y_{(n-k)}|^2 + t^2 + 2\phi t)^2}.$$

由于 $t < d_n^- < |Y_{(n-k)}|$，函数 f 关于 ϕ 单调递增。因此，令 $\phi = -|Y_{(n-k)}|$，可得，

$$v^T S_n(Y_{(n-k)} + vt) \geq -\frac{n-1}{|Y_{(n-k)}| - t} + \frac{t}{\nu + t^2} - n^{1-1/\beta-\delta},$$

对所有 $v \in \mathbb{S}^{d-1}$。第一项关于 t 递减。此外，由构造可知，$t/(\nu+t^2) > d_n$对所有 $t \in [d_n^+, d_n^-]$。因此，

$$v^T S_n(Y_{(n-k)} + vt) \geq -\frac{n-1}{|Y_{(n-k)}| - d_n^-} + d_n - n^{1-1/\beta-\delta} \geq \frac{-n}{|Y_{(n-k)}|} + d_n - n^{-\delta} m_1 d_n,$$

因为 $d_n^- < |Y_{(n-k)}|/n$。但同时，$|Y_{(n-k)}|/n > m_2/n^{1-1/\beta}$ 由此可得，

$$v^T S_n(Y_{(n-k)} + vt) \geq n^{1-1/\beta}\left(\frac{-1}{m_2} + \frac{1}{m_1} - n^{-\delta}\right) > 0$$

当 n 足够大时。 □

我们现在已准备好证明定理2.3和定理 2.4。

定理 2.3 的证明。令 $r = \sqrt{\nu}/2$（当 $0 < \beta < 1$ 时）且 $r = 2\nu/m_1$（当 $\beta = 1$ 时）。则对$n > \max\{N_1, N_2\}$，$d_n^+ < r < \sqrt{\nu}$对事件 E_n 上的任意 $\beta \leq 1$。设 $N > \max\{N_1, N_2\}$使得对所有 $n > N$，引理 4.2 和引理 4.3 的陈述均成立。则对 ~~$n > N$，在~~事件 E_n 上，可得 $S_n'(x) \succ 0$ 位于闭球 $B_r(Y_{(n-k)})$ 内，且$v^T(S_n(Y_{(n-k)} + vt)) > 0$对所有 $v \in \mathbb{S}^{d-1}$和 $t \in [d_n^+, d_n^-]$。由引理 4.1，这意味着对所有 $n > N$，在事件 E_n 上，存在唯一的 $x_n^+ \in B_{d_n^+}(Y_{(n-k)})$，使得 $S_n(x_n^+) = 0$且 $S_n'(x_n^+) \succ 0$。此外，由于 $d_n^- > \sqrt{\nu}$，这进一步意味着在环域 $B_{\sqrt{\nu}}(Y_{(n-k)}) \setminus B_{d_n^+}(Y_{(n-k)})$ 上不可能存在 $S_n(x)$ 的任何其他根。因此，在 $Y_{(n-k)}$ 的 $\sqrt{\nu}$–邻域内存在唯一的局部最大值 x_n^+。进一步，按设计，

$$|Y_{(n-k)} - x_n^+| \leq d_n^+ < 2d_n\nu = \frac{2\nu}{m_1} n^{1-1/\beta}.$$

选取 $m_1 = (G_k + c_\beta)/2$ 即得第一部分。

对于该定理的第二部分，考虑序列 $A_n = \{|Y_{(n-k)}| > 2n\sqrt{\nu}\}$。注意到 A 最终包含于 A_n，即 $A \subset \text{liminf}_{n\to\infty} A_n$。因此，当$n$充分大时，$E_n \subseteq A \subset A_n$ 且 $\mathbb{P}(E_n \mid A_n) = \mathbb{P}(E_n)/\mathbb{P}(A_n)$。此外，由于 G_k 连续，$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$。由定理 3.2 可知，当 $n \to \infty$ 时，$\mathbb{P}(B_n|A_n) \to 1$。由此得到，当 $n \to \infty$时，$\mathbb{P}(A \cap B_n) \to \mathbb{P}(A)$。最后，由命题 2.1，$\mathbb{P}(A_n') \to 1$。这意味着$\mathbb{P}(A_n' \cap B_n \cap A) = \mathbb{P}(E_n) \to \mathbb{P}(A)$，从而 $\mathbb{P}(E_n \mid A_n) \to 1$。 □

定理2.4的证明。定义

$$t_0 = \sup\{t > 0 : |x_n^+ + vt - Y_{(n-k)}| < d_n^- \ \forall\, v \in \mathbb{S}^{d-1}\}.$$

设 $t < t_0$。若 $|x_n^+ + vt - Y_{(n-k)}| > d_n^+$，则$v^T S_n(x_n^+ + vt) > 0$由引理4.3可得。假设 $|x_n^+ + vt - Y_{(n-k)}| < d_n^+ < \sqrt{\nu}$ $-$ $0 < r < t <$。根据泰勒展开，对于某个 $\sqrt{\nu}$

$$v^T S_n(x_n^+ + vt) = v^T S_n(x_n^+) + v^T S_n'(x_n^+ + rv)v = v^T S_n'(x_n^+ + rv)v.$$

由于 $|x_n^+ - Y_{(n-k)}| < d_n^+$，因此连接 x_n^+和 $x_n^+ + vt$ 的线段也完全位于d_n^+–半径为 $Y_{(n-k)}$的球内。因此，根据上文引理 4.2，上面的右侧为正。这意味着

$$W_n \geq t_0 \geq d_n^- - |x_n^+ - Y_{(n-k)}| > d_n^- - d_n^+ = \frac{\sqrt{1 - 4d_n^2\nu}}{d_n} > \frac{1}{d_n} - 2\sqrt{\nu}.$$

将（4.2）中的 d_n 代入 $m_1=(G_k+c_\beta)/2$即可得到下界。

对于上界，设 $r=|Y_{(n-k)}|$, $v=-Y_{(n-k)}/r$ 和 $t=2r/n$。考虑，

$$v^T\hat{S}_n(Y_{(n-k)}+vt)=\frac{2r/n}{\nu+(2r/n)^2}-\frac{n-1}{r-2r/n}<\frac{-n}{r}\left(\frac{1}{2}+\frac{1}{n-2}\right).$$

因此，在事件 E_n 上，$v^TS_n(Y_{(n-k)}+vt)\le v^T\hat{}S_n(Y_{(n-k)}+vt)+n^{1-1/\beta-\delta}<0$ 对足够大的 n成立。这意味着，

$$W_n<|Y_{(n-k)}+vt-x_n^+|\le|Y_{(n-k)}-x_n^+|+2|Y_{(n-k)}|/n.$$

由于 $n^{-1/\beta}|Y_{(n-k)}|\to G_k>c_\beta$，存在 N，使得对于 $n>N$，$|Y_{(n-k)}|/n<2G_kn^{1/\beta-1}$。因此，结果得证。 □

5. 从微模态退出所需的时间。 本节研究一维情况下 ZZP 退出微模态所需的时间，并证明定理2.5。

对于 $n\ge4$，令 $Z_t^n=(X_t^n,V_t^n)$ 为典范 ZZP 或采用子采样、切换率为 λ_n 的 ZZP，其中切换率由（2.7）定义。假设 n 足够大，使得在事件 E_n发生时，定理 2.3 和定理 2.4 均成立。给定初始条件 $(X_0^n,V_0^n)=(x_n^+,-1)$，我们关注由（2.8）定义的 Z_t^n的退出时间。定义 $x_n^-:=x_n^+-W_n$。于是，

$$\tau_n=\inf\{t\ge0:X_t^n\le x_n^-\mid(X_0^n,V_0^n)=(x_n^+,-1)\}.$$

按照 [34], 的方法，我们使用标准的更新时刻论证研究 Z_t^n 的退出时间。为此，定义：

$$\begin{aligned}\eta_n&=\inf\{t>0:X_t^n\in\{x_n^-,x_n^+\}\};\quad p_{\tau,n}=\mathbb{P}(X_{\eta_n}^n=x_n^-).\\T_n(y,v)&=\inf\{t>0:(X_t^n,V_t^n)=(x_n^+,-1)\mid(X_0^n,V_0^n)=(y,v)\}\end{aligned}\tag{5.1}$$

从 $(x_n^+,-1)$ 出发，η_n 是 ZZP 离开微模态或回升至局部最大值 x_n^+ 的首次时刻。根据构造，$Z_{\eta_n}^n\in\{(x_n^-,-1),(x_n^+,+1)\}$ 以概率 1 成立。给定初始条件，若 $Z_{\eta_n}^n=(x_n^-,-1)$，则 $\tau_n=\eta_n$，且过程从左侧离开微模态。反之，若 $Z_{\eta_n}^n=(x_n^+,+1)$，ZZP 还需经历一个 $T_n(x_n^+,+1)$ 时段，才能返回初始条件 $(x_n^+,-1)$ 并开始下一次游程。由于 ZZP 是遍历的，[（例如参见 9]），且以 $\mathbb{P}-$ 概率 1 成立，因此 $\mathbb{P}(T_n(x_n^+,+1)<\infty)=1$。由此得到退出时间满足以下递推关系

$$\tau_n=\begin{cases}\eta_n & \text{with prob. } p_{\tau,n}\\ \eta_n+T_n'(x_n^+,+1)+\tau_n' & \text{with prob. } 1-p_{\tau,n}.\end{cases}$$

其中$T_n'(x_n^+,+1)$与 τ_n' 分别是 $T_n(x_n^+,+1)$ 与 τ_n的独立同分布副本。因此，在事件E_n上，

$$\mathbb{E}(\tau_n)=\frac{1}{p_{\tau,n}}\mathbb{E}(\eta_n)+\frac{1-p_{\tau,n}}{p_{\tau,n}}\mathbb{E}(T_n(x_n^+,+1)).\tag{5.2}$$

在（5.2），$1/p_{\tau,\ n}$ 表示ZZP为成功退出左侧微模态而进行的游程的期望次数。接下来，我们将证明这是领头阶项。考虑切换率（2.7）。注意到，速率

$$\lambda_n(x,v)=(\nu+1)(v\cdot S_n(x))_+-\gamma_n(x),\quad(x,v)\in\mathbb{R}\times\{-1,+1\},$$

对应于 $\gamma_n\equiv0$ 的典范 ZZP 速率，以及子采样 ZZP 速率

$$\gamma_n(x)=(\nu+1)\left(\sum_{j=1}^n(S(x,Y_j))_+-(S_n(x))_+\right),\quad x\in\mathbb{R}.\tag{5.3}$$

回顾第2.2节， $S_n(x) \geq 0$ 对所有 $x \geq x_n^+$均成立。此外，由式 (2.2) 可知， $S_n(x) \geq 0$ 对所有 $x \geq Y_{(n)}$均成立。因此，必有 $x_n^+ < Y_{(n)}$。另外，根据式 (2.5) 中 W_n 的定义可得，对所有 $x \in (x_n^-, x_n^+)$， $S_n(x) \leq 0$ 均成立。最后，在集合 E_n上，当 n 足够大时， $|Y_{(n)} - Y_{(j)}| > |Y_n|/2n^\epsilon > W_n$成立。因此，对所有 $j \leq n-1$， $x_n^- > |Y_{(j)}|$成立，从而 $S(x, Y_{(j)}) > 0$。当 n 足够大时，我们得到，

$$\frac{\gamma_n(x)}{\nu+1} = \sum_{j=1}^{n} (S(x,Y_j))_+ - (S_n(x))_+ = \begin{cases} S_n(x) - S(x,Y_{(n)}), & x_n^- \leq x < x_n^+, \\ -S(x,Y_{(n)}), & x_n^+ \leq x < Y_{(n)}, \\ 0, & x \geq Y_{(n)}. \end{cases} \tag{5.4}$$

无论如何， $\gamma_n(x)/(\nu+1) \leq S_n(x) - S(x, Y_{(n)})$ 对于 $x \in (x_n^-, Y_{(n)})$。对于 $x \in (x_n^-, Y_{(n)})$，在事件 E_n 中（4.1），

(5.5)

$$\frac{\gamma_n(x)}{\nu+1} \leq S_n(x) - S(x,Y_{(n)}) \leq \hat{S}_n(x) + n^{1-1/\beta-\delta} - S(x,Y_{(n)}) = \frac{n-1}{x} + n^{1-1/\beta-\delta}.$$

右侧的第一项关于 x 是递减的，并且我们有， $x > x_n^- = Y_{(n)} - (Y_{(n)} - x_n^+)$ $- W_n > Y_{(n)} - 2\sqrt{\nu - 4}G_k n^{1/\beta-1}$。但根据定理 2.3的证明， n 满足 $|Y_{(n)}| > m_1 n^{1/\beta}$。由此可得，

$$\frac{\gamma_n(x)}{\nu+1} \leq n^{1-1/\beta} \left(\frac{1}{m_1 - 2\sqrt{\nu} n^{-1/\beta} - 4G_k n^{-1}} - n^{-\delta} \right) \leq \frac{n^{1-1/\beta}}{2m_1}$$

当 n足够大时。定义 $\gamma_n := n^{1-1/\beta}/(G_k + c_\beta)$。那么当 n足够大时， $\gamma_n(x) \leq (\nu+1)$ γ_n 对所有$x \in (x_n^-, x_n^+)$成立。此外，根据定理 2.4，

$$\frac{1}{2} - \frac{2\sqrt{\nu} n^{1-1/\beta}}{G_k + c_\beta} \leq W_n \overline{\gamma_n} \leq \frac{4G_k}{G_k + c_\beta} + \frac{\sqrt{\nu} n^{1-1/\beta}}{G_k + c_\beta} \leq 4 + \frac{\sqrt{\nu} n^{1-1/\beta}}{G_k + c_\beta}. \tag{5.6}$$

令 $p_n := \pi^{(n)}(x_n^-)/\pi^{(n)}(x_n^+)$。根据 [34], 的命题17

$$\frac{p_n}{1 + \overline{\gamma_n} W_n} \leq p_{\tau,n} \leq p_n.$$

将此结果与（5.6）结合可知， $p_{\tau,\ n}$ 与 p_n 的数量级相同。我们有如下结果。

引理 5.1. 假设 $\gamma_n(x) \equiv 0$， 或由下式定义 (5.3)。由于 $n \to \infty$,

$$p_{\tau,n} \mid E_n = O_p\left(n^{(1-1/\beta)(\nu+1)}\right).$$

证明。根据前面的讨论， $p_n(1 - O_p(1)) \leq p_{\tau,\ n}$ 其中 $p_n = \pi^{(n)}(x_n^-)\, \pi^{(n)}(x_n^+)$。定义 $\hat{}p_n := \exp\left((\nu+1) \int_{x_n^-}^{x_n^+} \hat{}S_n(u)\ du\right)$。在事件 E_n 上，

$$|\ln p_n - \ln \hat{p}_n| \leq (\nu+1) \int_{x_n^-}^{x_n^+} \left| S_n(u) - \hat{S}_n(u) \right| ds \leq (\nu+1) n^{1-1/\beta-\delta} W_n.$$

根据定理 2.4，右侧趋于 0（当 $n \to \infty$ 时）。因此，在 E_n 上可得 $|\ln p_n - \ln \hat{}$ $p_n| = o_p(1)$。于是， $p_{\tau,n} = O_p(p_n) = O_p(\hat{}p_n)$。现在，计算 $\hat{}p_n$ 定义中的积分，得到：

$$\hat{p}_n = \left(\frac{x_n^+}{x_n^-}\right)^{(\nu+1)(n-1)} \cdot \left(\frac{\nu + (x_n^+ - Y_{(n)})^2}{\nu + (x_n^- - Y_{(n)})^2}\right)^{(\nu+1)/2}. \tag{5.7}$$

由于 $Y_{(n)} - x_n^+ < \sqrt{\nu} < Y_{(n)} - x_n^-$，且 $x_n^- > Y_n - 2\sqrt{\nu} = 4G_k n^{1/\beta-1}$，我们有：

$$\hat{p}_n \geq (\nu/2)^{(\nu+1)/2}(Y_{(n)} - x_n^-)^{-(\nu+1)} \geq n^{(1-1/\beta)(\nu+1)} \left(\frac{\sqrt{\nu/2}}{2\sqrt{\nu} + 4G_k} \right)^{(\nu+1)}$$

当 n较大时。类似地，根据定理2.4的证明，

$$\frac{x_n^-}{x_n^+} \geq \frac{Y_{(n)} - 2\sqrt{\nu} - 4G_k n^{1/\beta-1}}{Y_{(n)}} \geq 1 - \frac{2n^{1-1/\beta}\sqrt{\nu} + 4G_k}{m_1 n} \geq 1 - \frac{2\sqrt{\nu} + 4G_k}{m_1 n}$$

当 n 较大时。因此，

$$\hat{p}_n \leq \left(1 - \frac{2\sqrt{\nu} + 4G_k}{m_1 n}\right)^{-(\nu+1)(n-1)} (2\nu)^{(\nu+1)/2} W_n^{-(\nu+1)}.$$

由定理 2.4即可得出该结果。 □

引理 5.1 表明，ZZP 在成功从左侧离开微模态之前所经历的游程次数的期望为 $n^{(1/\beta-1)(\nu+1)}$ 量级。此外，以下结果表明，在事件 E_n 下，随着 $n \to \infty$，每次游程的期望长度以及向右游程的期望长度均保持为 $O_p(1)$。

引理 5.2。假设满足 $\gamma_n(x) \equiv 0$，或由(5.3)定义。存在常数C_1, C_2和 C_3，使得，

$$C_1 + O_p(n^{1-1/\beta}) \leq \mathbb{E}(T_n(x_n^+, +1) \mid E_n) \leq C_2 + O_p(n^{1-1/\beta}),$$
$$\mathbb{E}(\eta_n \mid E_n) \leq C_3 + O_p(n^{1-1/\beta}),$$

如 $n \to \infty$。

我们现在可以证明定理2.5。

证明定理2.5。由（5.2），我们有，

$$\frac{1 - p_{\tau,n}}{p_{\tau,n}} \mathbb{E}(T_n(x_n^+, +1)) \leq \mathbb{E}(\tau_n) \leq \frac{1}{p_{\tau,n}} \mathbb{E}(\eta_n) + \frac{1 - p_{\tau,n}}{p_{\tau,n}} \mathbb{E}(T_n(x_n^+, +1)).$$

于是，该结果由引理5.1和引理5.2得出。 □

6. 推广。在本文中，我们选择极端次序统计量，因为它们提供了一个自然且方便的孤立点序列。所有结果都可以很容易地推广到例如中间次序统计量（见 [19]），这得益于引理3.1。在计算方面，我们仅针对最极端的数据点给出了 ZZP 的退出时间分析，$Y_{(n)}$。同样的分析可以推广到 $Y_{(n-k)}$，对于 $k > 0$，只需再对右侧偏离进行额外（但繁琐）的控制。

我们的分析聚焦于一维 ZZP。其他流行的基于梯度的算法（如过阻尼朗之万扩散和欠阻尼朗之万扩散）的退出时间已由[14, 33] 针对双阱势进行了研究。基于这些结果，我们预期这些算法在本文所考虑的设定中表现类似。特别地，可以将我们的微模态退出时间分析推广到维度大于 1 的这些过程。不过，我们预期这一推广在技术上更具挑战性，因为它通常依赖于势理论和谱分析 [13]。

最后，本文将研究范围限定在式（1.1）中的简单模型。不过，正如引言中的图 1 所示，微模态也可能常见于

更复杂的模型中，例如如下形式的线性模型 $y_j = A_j x + \varepsilon_j$，$\varepsilon_j \sim t_1$。在这种情况下，我们预计极端观测值会导致后验曲面沿超平面 $y_{(n)} = A_{(n)}x$ 形成一条脊，如图 1所示。我们目前基于大数定律的证明技巧仍可用于近似超平面周围管状区域内的经验得分，而非 $y_{(n)}$周围球面区域内的经验得分。随后可以进行类似于第4节的分析。

附录A：最大距离缩放

A.1. 关于次序统计量的一个结果。 回顾一下，Y_1，……，$Y_n \overset{iid}{\sim} P$其中 P满足假设 2.1。令 F_R 表示径向分量的分布函数 P，即$F_R(r) = P(\{|y| \le r\})$。那么，在假设 2.1下，

$$F_R(r) := \int_0^r f_R(t)\,dt = \int_0^r \omega_{d-1} t^{d-1} h(t) dt, \quad r > 0, \tag{A.1}$$

其中，h 与假设 2.1 中的定义相同，$\omega_{d-1} = 2\pi^{d/2}/\Gamma(d/2)$ 是单位球面 $\mathbb{S}^{d-1}$的表面积。我们用 $f_R(\cdot)$表示径向分量的密度。

在任意概率空间 $(\Omega\ \mathcal{F}, \mathbb{Q})$上，定义 $\tilde{Y}_{n,1}, \ldots, \tilde{Y}_{n,n-1} \overset{iid}{\sim} P$。定义 Z_n为˜ ˜一个与 $Y_{n,1}, \ldots, Y_{n,n-1}$独立且分布为 $P_{(n-k)}$的随机向量。在假设2.1下，我们有

$$\mathbb{Q}(Z_n \in A) = P_{(n-k)}(A) = \int_A \frac{f_{R_{(n-k)}}(|z|)}{\omega_{d-1}|z|^{d-1}}\,dz, \quad A \in \mathcal{B}(\mathbb{R}^d),$$

w此处 $f_{R_{(n-k)}}$ 表示 n 个样本中第 $k-$ 大半径的密度，其表达式为 ，

$$f_{R_{(n-k)}}(r) = \frac{n!}{k!(n-k-1)!}[F_R(r)]^{n-k-1}[1 - F_R(r)]^k f_R(r), \quad r > 0.$$

PROPOSITION A.1. 设 $g\colon (\mathbb{R}^d)^n \to \mathbb{R}$ 是关于其前 $n-1$ 个参数对称的任意函数。那么，对于任意 $A \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(g(Y_{-(n-k)}, Y_{(n-k)}) \in A) \le 4\sqrt{k+1} \cdot \mathbb{Q}(g(\tilde{Y}_{1,n-1}, \ldots, \tilde{Y}_{n,n-1}, Z_n) \in A), \tag{A.2}$$

其中，$Y_{-(n-k)}$ 表示移除 $Y_{(n-k)}$ 后的$(n-1)-$元组$(Y_1, \ldots, Y_n)$。

ˆ ˆ证明。设 Y_{n-k} 是服从 $P_{(n-k)}$分布的随机向量。给定 Y_{n-k}，定义ˆ ˆ$Y_1, \ldots Y_{n-k-1}$，使其在条件ˆ径向分量小于 Y_{n-k} 时，独立同分布于 P，即

$$\mathbb{P}(\hat{Y}_j \in B \mid \hat{Y}_{n-k}) = \frac{P(B \cap \{|Y| \le |\hat{Y}_{n-k}|\})}{P(\{|Y| \le |\hat{Y}_{n-k}|\})}, \quad B \in \mathcal{B}(\mathbb{R}^d), j = 1, \ldots, n-k-1.$$

ˆ ˆ类似地，定义 $Y_{n-k+1}, \ldots, Y_n$，使其在径向分量大于 Y_{n-k} 的条件下，独立同分布于ˆP，即

$$\mathbb{P}(\hat{Y}_j \in B \mid \hat{Y}_{n-k}) = \frac{P(B \cap \{|Y| > |\hat{Y}_{n-k}|\})}{P(\{|Y| > |\hat{Y}_{n-k}|\})}, \quad B \in \mathcal{B}(\mathbb{R}^d), j = n-k+1, \ldots, n.$$

ˆ ˆ于是，$\{Y_1, \ldots, Y_n\}$ 的分布与 $\{Y_1, \ldots, Y_n\}$ 的分布相同，至多相差若干置ˆ ˆ换。特别地，$Y_{(n-k)} = Y_{n-k}$ 几乎必然成立，并且我们有，

$$\mathbb{P}(g(Y_{-(n-k)}, Y_{(n-k)}) \in A) = \mathbb{P}(g(\hat{Y}_{-(n-k)}, \hat{Y}_{n-k}) \in A).$$

~ ~ ~ ~设 $\{Y_{(n,1)},\dots,Y_{(n,n-1)}\}$ 为对应于 $\{Y_{n,1},\dots,Y_{n,n-1}\}$的次序统计量。考虑事件，

$$B_k=\{|\tilde{Y}_{(n,n-k-1)}|\le|Z_n|\le|\tilde{Y}_{(n,n-k)}|\}.$$

于是，

$$\begin{aligned}\mathbb{P}(g(\hat{Y}_{-(n-k)},\hat{Y}_{n-k})\in A)&=\mathbb{Q}(g(\tilde{Y}_{n,1},\dots,\tilde{Y}_{n,n-1},Z_n)\in A\mid B_k)\\&\le\frac{\mathbb{Q}(g(\tilde{Y}_{n,1},\dots,\tilde{Y}_{n,n-1},Z_n)\in A)}{\mathbb{Q}(B_k)}.\end{aligned}$$

但同时，

$$\begin{aligned}\mathbb{Q}(B_k)&=\mathbb{Q}(\{|\tilde{Y}_{(n,n-k-1)}|\le|Z_n|\le|\tilde{Y}_{(n,n-k)}|\})\\&=\binom{n-1}{k}\mathbb{E}\left[(F_R(|Z_n|))^{n-k-1}\left(1-F_R(|Z_n|)\right)^k\right]\\&=\binom{n-1}{k}\frac{n!}{(n-k-1)!k!}\int_0^\infty[F_R(z)]^{2n-2k-2}[1-F_R(z)]^{2k}f_R(z)\,dz\\&=\binom{n-1}{k}\frac{n!}{(n-k-1)!k!}B(2n-2k-1,2k+1)\\&\ge\frac{1}{2\cdot4^k}\binom{2k}{k}\ge\frac{1}{4\sqrt{k+1}}.\end{aligned}$$

因此结论成立。 □

在下文中，我们将广泛使用命题A.1，利用 $\{Y_{1,n-1},\dots,Y_{n,n-1},Z_n\}$ 的结果得出关于(~ ~ $n-k$)– 阶次序统计量 $Y_{(n-k)}$ 的结论。

A.2. 命题2.1的证明。 本节将证明命题2.1。

在开始之前，先对记号作一说明。在本节及后续各节中，我们将使用‘$f\sim g$’来表示两个函数 $f,g:(0,\infty)\to(0,\infty)$，用以表示 $\lim_{x\to\infty}f(x)/g(x)=1$。

对于任意函数 f，我们将用‘$f\le RV_\rho$’表示存在 $g\in RV_\rho$ 和 x_0，使得对所有 $x\ge x_0$，$f(x)\le g(x)$。在本节中，我们常常在情形 $|x|\to\infty$（其中 $x\in\mathbb{R}^d$）下讨论。因此，我们将写成‘$\le RV_\rho(|x|)$’，以突出这一情形。

设 h 和 t_1 如假设 2.1 中所述。由于 $h\in RV_{-\beta-d}$，存在一个缓变 L[2]，使得对所有 $t>0$，$h(t)=t^{-\beta-d}L(t)$。由假设，h，即 $t\mapsto t^{-\beta-d}L(t)$ 在 $[t_1,\infty)$ 上非增。此外，根据缓变函数的波特界（见 [10,, 定理 1.5.6]），存在 $t_2>0$，使得对所有 $t>t_2$，$L(t/2)\le2L(t)$。令 $t_0=\max\{t_1,t_2\}$。

我们首先指出，径向分量的分布属于某个极值分布的吸引域。

命题 A.2. 设 P 满足假设2.1，且 F_R 表示径向分量的分布函数。令 $A=\omega_{d-1}K(\beta,d)/\beta$ 为一个常数。则，

$$F_R^n((An)^{1/\beta}r)\to\begin{cases}\exp(-r^{-\beta}), & r>0,\\ 0, & r\le0.\end{cases}\tag{A.3}$$

如 $n\to\infty$。

[2]若对所有 $r>0$，均有 $\lim_{t\to\infty}h(rt)/h(t)=1$，则称可测函数 $h:(0,\infty)\to(0,\infty)$为缓变函数。

证明。由（A.1），注意到，

$$1-F_R(r)=\omega_{d-1}\int_r^\infty t^{d-1}h(t)\,dt,\quad r>0$$

且 $1-F_R(r)=1$ 对所有 $r\le 0$均成立。由于 $h\in RV_{-\beta-d}$，可知 $t^{d-1}h(t)\in RV_{-\beta-1}$。对于$\beta>0$，由 Karamata定理（见 [39, 定理 0.6]）可得，$1-F_R\in RV_{-\beta}$。根据 [39], 的命题1.11，存在一个序列 a_n，使得，

$$F_R^n(a_nr)\to\begin{cases}\exp(-r^{-\beta}), & r>0,\\ 0, & r\le 0.\end{cases}$$

还需证明可以将 a_n 选为$(An)^{1/\beta}$。

同样，由 Karamata定理（[39,定理 0.6]）可得 von-Mises条件，

$$\lim_{r\to\infty}\frac{rF_R'(r)}{1-F_R(r)}=\beta, \tag{A.4}$$

其中请注意，$F_R'(r)=f_R(r)=\omega_{d-1}r^{d-1}h(r)$ 是径向分量的密度函数。根据 [39]的命题1.15，可以将 a_n 选为满足 $a_nF_R'(a_n)\sim\beta/n$。定义，$a_n=(An)^{1/\beta}$。于是，

$$a_nF_R'(a_n)=\omega_{d-1}(a_n)^dh(a_n)=\omega_{d-1}(a_n)^{-\beta}\left(a_n^{\beta+d}h(a_n)\right)=\frac{\omega_{d-1}}{An}\left(a_n^{\beta+d}h(a_n)\right).$$

因此，

$$\lim_{n\to\infty}\frac{a_nF_R'(a_n)}{\beta/n}=\lim_{n\to\infty}\frac{\omega_{d-1}}{A\beta}\left(a_n^{\beta+d}h(a_n)\right)=1,$$

由于 $a_n\to\infty$ 当 $n\to\infty$ 且 $t^{\beta+d}h(t)$ 依假设收敛于 $K(\beta,d)$。结论得证。

□

现在，我们给出空间中远处位置的球概率的一个简单上界。该上界将用于命题 2.1第二部分的证明。

引理 A.3. 设 $Y\sim P$ 其中 P 满足假设 2.1。设 z 为空间中的任意一点，满足 $|z|>t_0$ 并设 $r<|z|/2$。则，

$$\mathbb{P}(|Y-z|\le r)\le\frac{\omega_{d-1}}{d\cdot 2^{-\beta-d-1}}\cdot r^d|z|^{-\beta-d}L(|z|)$$

其中 $L\in RV_0$ 满足 $h(t)=t^{-\beta-d}L(t)$。

证明. 设 $A_r(z)$ 表示以 z 为中心、半径为 r 的球 $\{y\colon |y-z|<r\}$。则，

$$\mathbb{P}(|Y-z|\le r)=\int_{A_r(z)}h(|y|)dy\le\mathrm{Vol}(A_r(z))\cdot\sup_{A_r(z)}h(|y|)\le\frac{\omega_{d-1}r^d}{d}\cdot\sup_{A_r(z)}h(|y|).$$

注意到 下确界$\{|y|\colon\ y\in A_r(z)\}>|z|/2$ 这意味着，对于所有 $y\in A_r(z)$，$2|y|>|z|>t_0$。此外，由于对于所有 $t>t_0\ge t_2$，$L(t/2)\le 2L(t)$，因此可得，

$$h(|y|)=|y|^{-\beta-d}L(|y|)\le 2|y|^{-\beta-d}L(2|y|)=2^{\beta+d+1}h(2|y|).$$

但根据假设，$h(t)=t^{-\beta-d}L(t)$ 在 $[t_0,\infty)$ 上是非递增的。由于 $2|y|>|z|>t_0$，我们得到，对于所有 $y\in A_r(z)$，$h(|y|)\le 2^{\beta+d+1}h(|z|)$。因此，

$$\mathbb{P}(|Y-z|\le r)\le\frac{\omega_{d-1}r^d}{d}\cdot 2^{\beta+d+1}h(|z|).$$

因此，得证。

□

PROOF OF PROPOSITION 2.1。根据命题 A.2， P的径向分量的分布函数属于具有参数 $\beta>0$ 、规范化序列为 $a_n=(An)^{1/\beta}$的弗雷歇分布的吸引域。由 [19], 中的定理 2.1.1 可知，当 $n\to\infty$时，

$$\frac{|Y_{(n-k)}|-(An)^{1/\beta}}{\beta^{-1}(An)^{1/\beta}}\xrightarrow{d}\frac{(E_1+\cdots+E_{k+1})^{-1/\beta}-1}{\beta^{-1}},$$

其中 $E_1,\ldots,E_{k+1}$为独立同分布的标准指数分布。化简各项可得，

$$\frac{|Y_{(n-k)}|}{n^{1/\beta}}\xrightarrow{d}\frac{(E_1+\cdots+E_{k+1})^{-1/\beta}}{A^{-1/\beta}},$$

这证明了该结论的第一部分。几乎必然收敛性可通过考虑一个足够丰富的概率空间得到。

对于第二部分，令 $0<\epsilon<1$ 为任意值。定义 $g:(\mathbb{R}^d)^n\to\mathbb{R}$ 如下：

$$g(y_1,\ldots,y_n)=\frac{1}{|y_n|}\min_{j\neq n}|y_j-y_n|.$$

注意， g关于其前 $n-1$ 个自变量是对称的。而且，

$$\mathbb{P}\left(\min_{j\neq n-k}|Y_{(j)}-Y_{(n-k)}|>|Y_{(n-k)}|/2n^{\epsilon}\right)=\mathbb{P}\left(g(Y_{-(n-k)},Y_{(n-k)})>1/2n^{\epsilon}\right).$$

~ ~令 $Y_{n,1},\ldots,Y_{n,n-1},Z_n$和 $\mathbb{Q}$如第 A.1节所定义。根据命题 A.1，

$$\begin{aligned}&\mathbb{P}\left(\min_{j\neq n-k}|Y_{(j)}-Y_{(n-k)}|>|Y_{(n-k)}|/2n^{\epsilon}\right)\\&\quad=1-\mathbb{P}\left(g(Y_{-(n-k)},Y_{(n-k)})\leq 1/2n^{\epsilon}\right)\\&\quad\geq 1-4\sqrt{k+1}\cdot\mathbb{Q}\left(g(\tilde{Y}_{1,n-1},\ldots,\tilde{Y}_{n,n-1},Z_n)\leq 1/2n^{\epsilon}\right)\\&\quad=1-4\sqrt{k+1}\cdot\mathbb{Q}\left(\min_j|\tilde{Y}_{n,j}-Z_n|\leq|Z_n|/2n^{\epsilon}\right).\end{aligned}$$

我们将证明， $\mathbb{Q}\left(\text{最小}_j|\tilde{Y}_{n,\ j}-Z_n|\leq|Z_n|/2n^{\epsilon}\right)$ 在 $n\to\infty$时趋于0。考虑，

$$\begin{aligned}&\mathbb{Q}\left(\min_j|\tilde{Y}_{n,j}-Z_n|\leq|Z_n|/2n^{\epsilon}\right)\\&\quad=\mathbb{Q}\left(\exists j:|\tilde{Y}_{n,j}-Z_n|\leq|Z_n|/2n^{\epsilon}\right)\\&\quad=\mathbb{E}\left[\mathbb{Q}\left(\min_j|\tilde{Y}_{n,j}-Z_n|\leq|Z_n|/2n^{\epsilon}\middle|Z_n\right)\right]\\&\quad\leq\mathbb{E}\left[n\mathbb{Q}\left(|\tilde{Y}_{n,1}-Z_n|\leq|Z_n|/2n^{\epsilon}\middle|Z_n\right)\right]\\&\quad\leq n\mathbb{Q}(|Z_n|\leq 2t_0)+n\mathbb{E}\left[n\mathbb{Q}\left(|\tilde{Y}_{n,1}-Z_n|\leq|Z_n|/2n^{\epsilon}\middle|Z_n,|Z_n|>2t_0\right)\right].\end{aligned}$$

但根据引理A.3，

$$\begin{aligned}\mathbb{Q}\left(|\tilde{Y}_{n,1}-Z_n|\leq|Z_n|/2n^{\epsilon}\middle|Z_n,|Z_n|>2t_0\right)&\leq\frac{\omega_{d-1}}{d\cdot 2^{-\beta-d-1}}\left(\frac{|Z_n|}{2n^{\epsilon}}\right)^d|Z_n|^{-\beta-d}L(|Z_n|)\\&=\frac{2^{\beta+1}\omega_{d-1}}{d}\frac{|Z_n|^{-\beta}}{n^{d\epsilon}}L(|Z_n|).\end{aligned}$$

结合上述结果，

$$\mathbb{Q}\left(\min_j |\tilde{Y}_{n,j} - Z_n| \le |Z_n|/2n^{\epsilon}\right) \le n\mathbb{Q}(|Z_n| \le 2t_0) + \frac{2^{\beta+1}\omega_{d-1}}{d}\mathbb{E}\left[\frac{|Z_n|^{-\beta}}{n^{d\epsilon-1}}L(|Z_n|)\right].$$

对于第一项，注意到

$$n\mathbb{Q}(|Z_n| \le 2t_0) = n\sum_{j=0}^{k}\binom{n}{j}(F_R(2t_0))^{n-j}(1-F_R(2t_0))^{n-j} \le (k+1)n^{k+1}(F_R(2t_0))^{n-k},$$

当 $n\to\infty$ 时，该项趋于 0，因为 $F_R(2t_0) < 1$。对于第二项，回顾 $|Z_n|$ 的量级为 $n^{1/\beta}$， $L(t)$ 收敛到一个有限常数，且 $d\epsilon > 0$。于是，被积函数在 $n\to\infty$ 时收敛到 0。因此，由有界收敛定理，期望趋于 0。这就证明了该结果。 □

附录 B：定理的证明 2.2

定理 2.2 的证明。我们对 $k=0$，即 $Y_{(n)}$，加以证明。对 $k>0$ 的论证将类似进行。设 $x \in B_{\sqrt{\nu}}(Y_{(n)})$ 为任意的，并设 $v = -x/|x|$。考虑，

$$v^T S_n(x) = \sum_{j=1}^{n}\frac{-x^T(x-Y_j)/|x|}{\nu+|x-Y_j|^2} \le \sum_{j=1}^{n}\frac{-|x|+|Y_{(j)}|}{\nu+|x-Y_{(j)}|^2}.$$

考虑集合 $H_n = \{j : |Y_{(j)}| > |Y_{(n)}|/2 - \sqrt{\nu}\}$。对于 $j\notin H_n$ 和 $x\in B_{\sqrt{\nu}}(Y_{(n)})$，$-|x|+|Y_j| \le -|Y_{(n)}|/2$ $|x| \ge |Y_{(n)}| - \sqrt{\nu}$ $|x-Y_j| \le 3|Y_{(n)}|/2$ $|x| \le |Y_{(n)}| +$ 因为 。类似地，因为 。因此，

$$\begin{aligned} v^T S_n(x) &\le \sum_{j\notin H_n}\frac{-|x|+|Y_{(j)}|}{\nu+|x-Y_{(j)}|^2} + \sum_{j\in H_n}\frac{-|x|+|Y_{(j)}|}{\nu+|x-Y_{(j)}|^2} \\ &\le \sum_{j\notin H_n}\frac{-2|Y_{(n)}|}{4\nu+9|Y_{(n)}|^2} + \sum_{j\in H_n}\frac{1}{2\sqrt{\nu}} \end{aligned}$$

对于$x\in B_{\sqrt{\nu}}(Y_{(n)})$ 和 $v=-x/|x|$。此外，由于 $|Y_{(n)}|\uparrow\infty$，当 n较大时， $4\nu \le 9|Y_{(n)}|^2$并且我们得到，

$$v^T S_n(x) \le -\frac{n-\sharp H_n}{9|Y_{(n)}|} + \frac{\sharp H_n}{2\sqrt{\nu}}.$$

设 $K\ge 1$ 为任意的。则在事件 $\{\sharp H_n \le K\}$上， $v^T S_n(x)\downarrow -\infty$ 随 $n\to\infty$ 成立，这由命题2.1 对 $\beta>1$给出。特别地，对任意 $c>0$ 以及对任意 $K\ge 1$，存在一个 $N_0(K)$，使得对于 $n\ge N_0$

$$\begin{aligned} \{\sharp H_n \le K\} &\implies \left\{\inf\left\{(-x/|x|)^T S_n(x) : x\in B_{\sqrt{\nu}}(Y_{(n)})\right\} \le -c\right\} \\ &\implies \left\{\forall x\in B_{\sqrt{\nu}}(Y_{(n)})\exists v\in\mathbb{S}^{d-1} : v^T S_n(x) \le -c\right\}. \end{aligned} \tag{B.1}$$

对于 $K\ge 1$，考虑，

$$\begin{aligned} \mathbb{P}(\sharp H_n > K) &= \mathbb{P}\left(\sharp\{j\le n-1 : |Y_{(j)}| > |Y_{(n))}|/2 - \sqrt{\nu}\} \ge K\right) \\ &\le \frac{1}{K}\mathbb{E}\left(\sharp\{j\le n-1 : |Y_{(j)}| > |Y_{(n))}|/2 - \sqrt{\nu}\}\right) \\ &\le \frac{n}{K}\mathbb{E}\left(\frac{1-F_R(|Y_{(n)}|/2-\sqrt{\nu})}{F_R(|Y_{(n)}|/2)}\right), \end{aligned}$$

其中 F_R 是 Y 在 $Y \sim P$ 下径向分量的分布函数。期望中的项关于 $|Y_{(n)}|$ 单调递减。根据命题 2.1，对于某个常数 $C_{\beta,\ d}$，

$$n\left(\frac{1-F_R(|Y_{(n)}|/2-\sqrt{\nu})}{F_R(|Y_{(n)}|/2)}\right) \longrightarrow C_{\beta,d}\left(\frac{G_k}{2}\right)^{-\beta}$$

以 $n \to \infty, \mathbb{P}-$ 几乎必然成立。这意味着当 n 足够大时，

$$\mathbb{P}(\sharp H_n > K) \leq \frac{1}{K} C'_{\beta,d}\left(\frac{G_k}{2}\right)^{-\beta}$$

对某些 $C'_{\beta,d} > C_{\beta,d}$。右侧关于 K 单调递减。令 $\epsilon > 0$ 为任意值。那么，我们可以选择 $K(\epsilon) \geq 1$，使右侧小于 ϵ。也就是说，给定任意的 $\epsilon > 0$，我们可以选择 $N_1 > N_0$，使得，

$$\mathbb{P}(\sharp H_n > K(\epsilon)) \leq \epsilon,$$

对所有 $n \geq N_1$。将其与 (B.1) 结合可得，对所有 $n \geq N_1$，

$$\mathbb{P}\left(\left\{\forall x \in B_{\sqrt{\nu}}(Y_{(n)}) \exists v \in \mathbb{S}^{d-1} : v^T S_n(x) \leq -c\right\}\right) \geq 1-\epsilon.$$

由于 $\epsilon > 0$ 是任意的，结论随即得出。 □

附录 C：经验得分近似

在本节中，我们将证明第3节中的结果。

C.1. 引理3.1的证明。为证明引理 3.1 将 $\mathbb{S}_n$ 拆分为，

$$\mathbb{S}_n(z) \leq \sup_{x:|x-z|\leq\frac{2|z|}{n}} \frac{2|z|}{n}|S_n(x) - n\mathbb{E}S(x,Y)| + \sup_{x:|x-z|\leq\frac{2|z|}{n}} \frac{2|z|}{n}\left|n\mathbb{E}S(x,Y) - \frac{nx}{|x|^2}\right|$$

我们将在以下两个引理中分别控制每一项。

引理 C.1. 令 $0 < \epsilon < 1$ 为任意值。存在一个 $T_2(\epsilon)$，使得对所有 $n \geq T_2(\epsilon)$，

$$\mathbb{P}\left(\sup_{x:|x-z|\leq 2|z|/n} \frac{2|z|}{n}\left|S_n(x) - n\mathbb{E}S(x,Y)\right| > 48\nu n^{2\epsilon-1} + \frac{|z|^{-\beta}}{n^{\epsilon(d-1)}}L'(|z|) + t\right)$$
$$\leq 5^d \exp\left(-\frac{n^{1-2\epsilon}t^2}{512}\right) + C_d\frac{5^d|z|^{-\beta}}{2^d n^{d\epsilon-1}}L(|z|),$$

对所有 $t > 0$ 和 $|z| > 2n\sqrt{\nu}$，其中 L' 是缓变的，且 $L(t) = t^{\beta+d}h(t)$。

证明。令 $0 < \epsilon < 1$ 为任意值。定义 $\delta_n = 2/n + 1/2n^{\epsilon}$。令 $T_1(\epsilon) > t_0/\sqrt{\nu}$ 满足 对所有 $n \geq T_1(\epsilon)$，。令 (ϵ) 为任意值。在下文中，$|z| > 2n\sqrt{\nu}$。于是我们有：

1. $|z|/2 > n\sqrt{\nu} \geq T_1(\epsilon)\sqrt{\nu} > t_0$。
2. $|z|/2n^{\epsilon} > 1$。 3. $|z| - |z|\delta_n > |z|/2$。

定义集合 $A_\epsilon = \{y: \ |y-z| < |z|\delta_n\}$ 和 $A_\gamma = \{y: \ |y-z| < 2|z|/n\}$。令 $S_1(x, \ y) = S(x, \ y)1(y \notin A_\epsilon)$ 和 $S_2(x, \ y) = S(x, \ y)1(y \in A_\epsilon)$。定义,

$$\Psi_n(x) = n^{-1}|z|\left(S_n(x) - n\mathbb{E}S(x, Y)\right), \quad x \in \mathbb{R}^d, \tag{C.1}$$

那么,

$$\Psi_n(x) = \underbrace{\frac{2|z|}{n}\sum_{j=1}^{n}(S_1(x, Y_j) - \mathbb{E}S_1(x, Y_j))}_{:=\Psi_{1,n}(x)} + \overbrace{\frac{2|z|}{n}\sum_{j=1}^{n}(S_2(x, Y_j) - \mathbb{E}S_2(x, Y_j))}^{:=\Psi_{2,n}(x)} \tag{C.2}$$

我们将分别考虑每一项。

1) 关于 $\Psi_{1、n}$:

1.1) A_γ 的离散化:

首先, 考虑 $\Psi_{1, \ n}$。考虑函数 $s: \ x \mapsto x/(\nu + |x|^2)$。假设 $|x|, \ |y| > r > 1$。那么,

$$\begin{aligned}
|s(x) - s(y)| &= \left|\frac{x}{\nu + |x|^2} - \frac{y}{\nu + |y|^2}\right| \leq \frac{1}{(\nu + |x|^2)(\nu + |y|^2)}\left|\nu(x-y) + x|y|^2 - y|x|^2\right| \\
&\leq \frac{\nu|x-y|}{(\nu + |x|^2)(\nu + |y|^2)} + \frac{|y|^2|x-y|}{(\nu + |x|^2)(\nu + |y|^2)} + \frac{|y||x|^2 - |y|^2|}{(\nu + |x|^2)(\nu + |y|^2)} \\
&\leq \frac{\nu|x-y|}{r^4} + \frac{|x-y|}{r^2} + \frac{|x-y|}{r^3} \leq 3\nu\frac{|x-y|}{r^2}.
\end{aligned} \tag{C.3}$$

假设 $|z|/2n^\epsilon > 1$ $y \notin A_\epsilon$ $x \in A_\gamma$ $|x-y| > ||y-z| - |x-z|| \geq |z|\delta_n - 2|z|/n =$ 并且, 那么, 。方程 (C.) 蕴含对任意 , , $x^1 \ x^2 \in A_\gamma$ 3

$$|S_1(x^1, Y) - S_1(x^2, Y)| \leq \frac{12\nu n^{2\epsilon}}{|z|^2}||x^1 - Y| - |x^2 - Y|| \leq \frac{12\nu n^{2\epsilon}|x^1 - x^2|}{|z|^2}$$

以概率 1。现在, 考虑,

$$\begin{aligned}
&\left|\Psi_{1,n}(x^1) - \Psi_{1,n}(x^2)\right| \\
&= \frac{2|z|}{n}\left|\sum_{j=1}^{n}(S_1(x^1, Y_j) - \mathbb{E}S_1(x^1, Y_j)) - \sum_{j=1}^{n}(S_1(x^2, Y_j) - \mathbb{E}S_1(x^2, Y_j))\right| \\
&\leq \frac{2|z|}{n}\left(\sum_{j=1}^{n}\left|S_1(x^1, Y_j) - S_1(x^2, Y_j)\right| + n\mathbb{E}\left[\left|S_1(x^1, Y) - S_1(x^2, Y)\right|\right]\right) \\
&\leq \frac{2|z|}{n} \cdot 2n \cdot \left(\frac{12\nu n^{2\epsilon}|x^1 - x^2|}{|z|^2}\right) \leq 48\nu n^{2\epsilon}|z|^{-1}|x^1 - x^2|,
\end{aligned}$$

以概率 1。由此可知, 对所有 $x \in A_\gamma$,

$$|\Psi_{1,n}(x)| \leq |\Psi_{1,n}(x) - \Psi_{1,n}(z)| + |\Psi_{1,n}(z)| \leq 48\nu n^{2\epsilon}|z|^{-1}\frac{|z|}{n} + |\Psi_{1,n}(z)|.$$

因此,

$$\sup_{x \in A_\gamma}|\Psi_{1,n}(x)| \leq 48\nu n^{2\epsilon-1} + |\Psi_{1,n}(z)| \tag{C.4}$$

概率为 1。

1.2）Ψ_1中的集中性，n

注意，由于 $|y-z| \geq |z|\delta_n > |z|/2n^\epsilon$ 对所有 $y \notin A_\epsilon$均成立，因此 $|S_1(z, Y)| \leq 2n^\epsilon|z|^{-1}$ 是一个有界随机向量。通过欧几里得网得到的标准集中不等式（例如参见 [12]中的命题 28）可得：

$$\mathbb{P}(|\Psi_{1,n}(z)| > t) = \mathbb{P}\left(\frac{1}{n}\left|\sum_{j=1}^{n}(S_1(z,Y_j) - \mathbb{E}S_1(z,Y_j))\right| > \frac{t}{2|z|}\right)$$

$$\leq 5^d \exp\left(-\frac{n}{8}\cdot\frac{t^2}{4|z|^2}\frac{1}{(2n^\epsilon|z|^{-1})^2}\right) = 5^d\exp\left(-\frac{n^{1-2\epsilon}t^2}{128}\right). \tag{C.5}$$

1.3) 合并观测值

利用 (C.5) 和 (C.4)，对所有满足 z 的 $|z| \geq 2n\sqrt{\nu}$，可得：

$$\mathbb{P}\left(\sup_{x\in A_\gamma}|\Psi_{1,n}(x)| > 48\nu n^{2\epsilon-1} + t/2\right) \leq 5^d\exp\left(-\frac{n^{1-2\epsilon}t^2}{128}\right). \tag{C.6}$$

2) Ψ_2的界n：

现在，我们考虑 Ψ_2，n。由 (C.2) 可得：

$$|\Psi_{2,n}(x)| \leq \frac{2|z|}{n}\left|\sum_{j=1}^{n}S_2(x,Y_j)\right| + 2|z|\cdot\mathbb{E}|S_2(x,Y)|. \tag{C.7}$$

2.1) 控制 (C.7) 中的期望

对于任意 $x \in A_\gamma$，首先考虑期望 $\mathbb{E}|S_2(x, Y)|$。于是，

$$\begin{aligned}\mathbb{E}|S_2(x,Y)| &= \int_{A_\epsilon}|S(x,y)|P(dy)\\ &= \int_{A_\epsilon^1}|S(x,y)|P(dy) + \int_{A_\epsilon^2}|S(x,y)|P(dy),\end{aligned}$$

其中 $A_\epsilon^1 = \{y : |x-y| \leq \sqrt{\nu}\} \cap A_\epsilon$ 且 $A_\epsilon^2 = A_\epsilon \backslash A_\epsilon^1$。

在 A_ϵ^1 上，我们使用 $|S(x, y)|$ 的全局界，即 $|S(x, y)| \leq 1/2\sqrt{\nu} =: M$。因此，

$$\int_{A_\epsilon^1}|S(x,y)|P(dy) \leq MP(\{Y : |x-Y| \leq \sqrt{\nu}\} \cap A_\epsilon) \leq MP(\{Y : |x-Y| \leq \sqrt{\nu}\})$$

$$\leq MC_d(\sqrt{\nu})^d|x|^{-\beta-d}L(|x|),$$

其中最后一个不等式由引理 A. 3 $x \in A_\gamma$ $|x| > |z| -$ 得出，$2|z|/n > |z|/2 > t_0$ $\sqrt{\nu} < n\sqrt{\nu} < |x|/2$ $|x|^{-\beta-d}L\ |x|\ = h\ |x|$ 且因为对所有，以及。回顾一下，()()。利用假设 h 在 t_0之后单调不增，我们得到，

$$\sup_{x\in A_\gamma}\int_{A_\epsilon^1}|S(x,y)|P(dy) \leq \frac{MC_d\nu^{d/2}}{2^{-\beta-d}}|z|^{-\beta-d}L(|z|/2) \leq \frac{MC_d\nu^{d/2}}{2^{-\beta-d-1}}|z|^{-\beta-d}L(|z|). \tag{C.8}$$

另一方面，考虑：

$$\int_{A_\epsilon^2}\frac{|x-y|}{\nu+|x-y|^2}P(dy) \leq \int_{A_\epsilon^2}\frac{1}{|x-y|}h(|y|)dy < \infty,$$

因为在 A_ϵ^2 上有 $|x-y| > \sqrt{\nu}$。对于所有 $y \in A_\epsilon^2$, $|y| > |z| - |z|\delta_n > |z|/2$。因此，对于所有 $y \in A_\epsilon^2$, $h(|y|) \leq h(|z|/2)$。此外，在 A_ϵ^2 上，注意到对于任意 $x \in A_\gamma$, $\sqrt{\nu} < |x-y| < 2|z|/n + |z|\delta_n$。

Reca所有满足 $\delta_n = 2/n + 1/2n^\epsilon$ 的情况。由于 $0 < \epsilon < 1$，我们有 $2|z|/n + |z|\delta_n \le 9|z|/2n^\epsilon$。然后，

$$\int_{A_\epsilon^2} \frac{1}{|x-y|} h(|y|)dy \le h(|z|/2) \int_{A_\epsilon^2} \frac{1}{|x-y|} dy \le 2^{\beta+d+1} h(|z|) \int_{\sqrt{\nu} \le |x-y| \le \frac{9|z|}{2n^\epsilon}} \frac{1}{|x-y|} dy$$

$$= 2^{\beta+d+1} h(|z|) \begin{cases} \omega_d \int_{\sqrt{\nu}}^{\frac{9|z|}{2n^\epsilon}} r^{d-2}\, dr, & d \ge 2, \\ 2 \int_{\sqrt{\nu}}^{\frac{9|z|}{2n^\epsilon}} \frac{1}{r}\, dr, & d = 1, \end{cases}$$

$$\le 2^{\beta+d+1} h(|z|) \begin{cases} \omega_d \left(\frac{9|z|}{2n^\epsilon}\right)^{(d-1)}, & d \ge 2, \\ 2 \ln \frac{9|z|}{2n^\epsilon}, & d = 1, \end{cases} \tag{C.9}$$

其中 ω_d 是 $(d-1)$–维单位球面在 $\mathbb{R}^d$中的表面积。由 (C.8) 和 (C.9)，我们得到，

$$2|z| \cdot \sup_{x \in A_\gamma} \mathbb{E}|S_2(x,Y)| \le 8 \left(\frac{|z|}{2}\right)^{-\beta-d+1} L(|z|) \begin{cases} MC_d\nu^{d/2} + \omega_d \left(\frac{9|z|}{2n^\epsilon}\right)^{(d-1)}, & d \ge 2, \\ MC_d\nu^{d/2} + 2\ln \frac{9|z|}{2n^\epsilon}, & d = 1, \end{cases}.$$

注意，由于 $|z| > 2n\sqrt{\nu}$，通过取足够大的 n，项 $|z|/n^\epsilon$ 可以任意大。因此，令 $T_2(\epsilon) > T_1(\epsilon)$ 满足：对所有 $n > T_2(\epsilon)$，存在一个缓变函数 L' 使得，

$$|z| \cdot \sup_{x \in A_\gamma} \mathbb{E}|S_2(x,Y)| \le \frac{|z|^{-\beta}}{n^{\epsilon(d-1)}} L'(|z|). \tag{C.10}$$

2.2) 控制(C.7)中的平均值

注意到 $|\sum_{j=1}^n S_2(x,Y_j)| > 0$ 当且仅当存在 Y_j 使得 $Y_j \in A_\epsilon$。回顾 $A_\epsilon = \{y : |y-z| \le |z|\delta_n\}$。由于 $\delta_n < 1/2$，由引理 A.3，我们得到，

$$\mathbb{P}\left(\sup_{x \in A_\gamma} \left|\sum_{j=1}^n S_2(x,Y_j)\right| > \frac{nt}{4|z|}\right) \le nP(|Y-z| \le |z|\delta_n) \le nC_d(|z|\delta_n)^d |z|^{-\beta-d} L(|z|).$$

但同时还有 $\delta_n = 2/n + 1/2n^\epsilon \le 5/2n^\epsilon$。因此，

$$\mathbb{P}\left(\sup_{x \in A_\gamma} \left|\sum_{j=1}^n S_2(x,Y_j)\right| > \frac{nt}{4|z|}\right) \le C_d \frac{5^d |z|^{-\beta}}{2^d n^{d\epsilon-1}} L(|z|). \tag{C.11}$$

2.3) 合并观测值

合并（C.10）与（C.11），再结合（C.7），得到:

$$\mathbb{P}\left(\sup_{x \in A_\gamma} |\Psi_{2,n}(x)| > \frac{|z|^{-\beta}}{n^{\epsilon(d-1)}} L'(|z|) + t/2\right) \le C_d \frac{5^d |z|^{-\beta}}{2^d n^{d\epsilon-1}} L(|z|). \tag{C.12}$$

3) 合并 $\Psi_{1,\ n}$ 和 $\Psi_{2,\ n}$

假设 $n \ge T_2$（ϵ）。合并（C.6）与（C.12），对于所有 $|z| > 2n\sqrt{\nu}$，

$$\mathbb{P}\left(\sup_{x:|x-z| \le 2|z|/n} \frac{2|z|}{n} \left|S_n(x) - n\mathbb{E}S(x,Y)\right| > 48\nu n^{2\epsilon-1} + \frac{|z|^{-\beta}}{n^{\epsilon(d-1)}} L'(|z|) + t\right)$$

$$\le 5^d \exp\left(-\frac{n^{1-2\epsilon}t^2}{512}\right) + C_d \frac{5^d |z|^{-\beta}}{2^d n^{d\epsilon-1}} L(|z|).$$

对于一个缓变的 L'。证明至此完成。 □

引理 C.2. 设 $Y \sim P$ 其中 P 满足假设 2.1 且 $0 < \beta \leq 1$。对于所有 ν，存在一个 $T_3(\nu)$，使得对于所有 $n \geq T_3(\nu)$ 及 $|z| > 2n\sqrt{\nu}$，

$$\sup_{x:|x-z|\leq \frac{2|z|}{n}} \frac{2|z|}{n} \left| n\mathbb{E}S(x,Y) - \frac{nx}{|x|^2} \right| \leq RV_{-\beta}(|z|).$$

证明。设 $T_3(\nu) \geq 4$ 使得 $T_3 \geq 2t_0/\sqrt{\nu}$。设 $n > T_3$ 为任意的。则 $|z| > 2n\sqrt{\nu}$ 对所有 $|x - z| < 2|z|/n$ 均蕴含 $|x| > |z|/2 > \max\{2t_0, 2\sqrt{\nu}\}$。对任意这样的 x，定义

$$g(y) = \frac{|x|(x-y)}{\nu + |x-y|^2}, \quad y \in \mathbb{R}^d.$$

则，$|x|\mathbb{E}S(x,Y) = \mathbb{E}g(Y)$，且，

$$\left| |x|\mathbb{E}S(x,Y) - \frac{x}{|x|} \right| \leq \left| \mathbb{E}g(Y)1_{|Y|<|x|/2} - \frac{x}{|x|} \right| + \mathbb{E}|g(Y)|1_{|Y|\geq|x|/2}. \tag{C.13}$$

我们首先考虑 $\mathbb{E}|g(Y)|1_{|Y|\geq|x|/2}$。我们将区域 $|y| \geq |x|/2$ 进一步分成三部分：

$$\{y : |y| \geq |x|/2\} = A_1 \cup A_2 \cup A_3,$$

其中，

$$A_1 := \{|y| \geq |x|/2, |x-y| > |x|/2\}, \quad A_2 := \{\sqrt{\nu} < |x-y| \leq |x|/2\},$$
$$A_3 := \{|x-y| \leq \sqrt{\nu}\},$$

并分别考虑每一部分。

对于 $y \in A_1$，注意到 $|g(y)| \leq 2$。于是，

$$\mathbb{E}|g(Y)|1_{Y\in A_1} \leq 2P(A_1) \leq 1 - F_R(|x|/2),$$

其中，F_R 是径向分量 $|Y|$ 的分布函数，且 $1 - F_R \in RV_{-\beta}$。因此，我们有，

$$\mathbb{E}|g(Y)|1_{Y\in A_1} \leq RV_{-\beta}(|x|) \tag{C.14}$$

对于 $y \in A_2$，注意到 $t_0 < |x|/2 \leq |y| \leq 3|x|/2$。由于 h 在 $[t_0, \infty)$ 上单调不增，

$$\mathbb{E}|g(Y)|1_{Y\in A_2} = \int_{A_2} \frac{|x||x-y|}{\nu + |x-y|^2} h(|y|)\, dy \leq |x|h(|x|/2) \int_{c<|u|<|x|/2} \frac{|u|}{\nu + |u|^2}\, du.$$

Now,

$$\int_{\sqrt{\nu}<|u|<|x|/2} \frac{|u|}{\nu + |u|^2}\, du \leq \int_{\sqrt{\nu}<|u|<|x|/2} \frac{1}{|u|}\, du \leq \begin{cases} C_d|x|^{d-1} & d \geq 2 \\ C_1 \ln|x|, & d = 1, \end{cases}$$

对于某些依赖于维数的常数 C_d。这意味着，

$$\mathbb{E}|g(Y)|1_{Y\in A_2} \leq \begin{cases} C_d|x|^d h(|x|/2) & d \geq 2 \\ C_1|x|h(|x|/2)\ln|x|, & d = 1. \end{cases}$$

由于对于 $\beta > 0$，$h \in RV_{-\beta-d}$ 成立，且 $\ln \in RV_0$，即 $\ln(|x|)$ 是缓变的，右侧 $\in RV_{-\beta}$ 对所有 d 均成立。我们得到，

$$\mathbb{E}|g(Y)|1_{Y\in A_2} \leq RV_{-\beta}(|x|) \tag{C.15}$$

最后，考虑 A_3。那么，$|g(y)|$ 被 $|x|/2\sqrt{\nu}$ 所界定。我们得到，

$$\mathbb{E}|g(Y)|1_{Y\in A_3} \leq \frac{|x|}{2\sqrt{\nu}} \mathbb{P}(A_3).$$

注意，$|x| \geq \max\{2t_0, 2\sqrt{\nu}\}$ 意味着 $x - \sqrt{\nu} > t_0$。因此，h 在 $|x| - \sqrt{\nu} \leq |y| \leq |x| + \sqrt{\nu}$ 上非递增。我们得到，

$$\mathbb{P}(A_3) = \int_{|x-y| \leq \sqrt{\nu}} h(|y|)dy \leq C_d' h(|x| - \sqrt{\nu}).$$

因此，

$$\mathbb{E}|g(Y)|1_{Y \in A_3} \leq RV_{-\beta-d}(|x|). \tag{C.16}$$

于是，方程（C.14）、（C.15）和（C.16）表明，

$$\mathbb{E}|g(Y)|1_{|Y| \geq |x|/2} \leq \mathbb{E}|g(Y)|1_{Y \in A_1} + \mathbb{E}|g(Y)|1_{Y \in A_2} + \mathbb{E}|g(Y)|1_{Y \in A_3} \leq RV_{-\beta}(|x|). \tag{C.17}$$

现在，考虑 $|y| < |x|/2$。对于 $y \neq x$，我们可以将 g 分解为：

$$g(y) = \frac{x}{|x|} + \underbrace{\left(\frac{|x|(x-y)}{|x-y|^2} - \frac{x}{|x|}\right)}_{:=g_1(y)} - \underbrace{\left(\frac{\nu|x|(x-y)}{|x-y|^2(\nu + |x-y|^2)}\right)}_{:=g_2(y)}$$

对于 $|y| < |x|/2$，可得 $|x-y| > |x|/2$。接着不难看出 $|g_2(y)| \leq 8\nu/|x|^2$。此外，$|g_1(y)| \leq 14\frac{|y|}{|x|}$。因此，

$$\left|\mathbb{E}g(Y)1_{|Y|<|x|/2} - \frac{x}{|x|}\right| \leq \frac{14}{|x|}\mathbb{E}\left(|Y|1_{|Y|<|x|/2}\right) + \frac{8\nu}{|x|^2}.$$

对于 $0 < \beta \leq 1$，Karamata定理 [10] 意味着 $|x| \mapsto |x|^{-1}\mathbb{E}\left(|Y|1_{|Y|<|x|/2}\right) \in RV_{-\beta}$。因此，

$$\left|\mathbb{E}g(Y)1_{|Y|<|x|/2} - \frac{x}{|x|}\right| \leq RV_{-\beta}(|x|) + \frac{8\nu}{|x|^2} \leq RV_{-\beta}(|x|). \tag{C.18}$$

结合（C.17）与（C.18），可得：

$$\left||x|\mathbb{E}S(x,Y) - \frac{x}{|x|}\right| \leq RV_{-\beta}(|x|).$$

这就证明了结论。 □

证明 引理 3.1。结合引理 C.1 和 C.2 以及正则变函数的单调等价性 [10,定理 1.5。3]。 □

C.2. 定理3.2的证明。

定理 3.2 的证明。定义 g：$(\mathbb{R}^d)^n \to \mathbb{R}$ 如下：

$$g(y_1, \ldots, y_n) = \sup_{\{x: |x-y_n| \leq 2|y_n|/n\}} \left|\sum_{j=1}^{n-1} S(x, y_j) - \frac{(n-1)x}{|x|^2}\right|.$$

~ ~那么，g在其前 $n-1$ 个参数上是对称的。令 $\{Y_{n,1}, \ldots, Y_{n,n-1}, Z_n\}$ 如第A.1节中的定义所示。注意，

$$\begin{aligned}\sup_{x \in \mathbb{R}^d} \left|S_n(x) - \hat{S}_n(x)\right| &= \sup_{x \in \mathcal{F}_n} \left|\sum_{j \neq n-k} S(x, Y_{(j)}) - \frac{(n-1)x}{|x|^2}\right| \\ &= g(Y_{-(n-k)}, Y_{(n-k)}).\end{aligned}$$

那么，根据命题A.1，

$$\begin{aligned}
&\mathbb{P}\left(\sup_{x\in\mathbb{R}^d}\left|S_n(x)-\hat{S}_n(x)\right|>n^{1-1/\beta-\delta}\mid A_n\right)\\
&\quad=\mathbb{P}\left(g(Y_{-(n-k)},Y_{(n-k)})>n^{1-1/\beta-\delta}\mid |Y_{(n-k)}|>2n\sqrt{\nu}\right)\\
&\quad\leq 4\sqrt{k+1}\mathbb{Q}\left(g(\tilde{Y}_{n,1},\ldots,\tilde{Y}_{n,n-1},Z_n)>n^{1-1/\beta-\delta}\mid |Z_n|>2n\sqrt{\nu}\right).
\end{aligned}$$

再次，我们将证明右侧在 $n\to\infty$ 时趋于0。为此，注意到：

$$\begin{aligned}
&n^{1/\beta-1+\delta}g(\tilde{Y}_{n,1},\ldots,\tilde{Y}_{n,n},Z_n)\\
&\quad=n^{1/\beta-1+\delta}\sup_{|x-Z_n|\leq 2|Z_n|/n}\left|\sum_{j=1}^{n-1}S(x,Y_{n,j})-\frac{(n-1)x}{|x|^2}\right|\\
&\quad\leq n^{1/\beta-1+\delta}\frac{n-1}{2|Z_n|}\sup_{|x-Z_n|\leq 2|Z_n|/(n-1)}\frac{2|Z_n|}{n-1}\left|\sum_{j=1}^{n-1}S(x,Y_{n,j})-\frac{(n-1)x}{|x|^2}\right|\\
&\quad\leq n^{\delta}\left(\frac{n^{1/\beta}}{|Z_n|}\right)\left(\frac{n-1}{2n}\right)\mathbb{S}_{n-1}(Z_n),
\end{aligned}$$

其中 $\mathbb{S}_{n-1}(Z_n)$ 的定义见（3.1）。根据命题2.1， $n^{1/\beta}/|Z_n|$ 收敛到一个非退化随机变量，记为 G。此外，中间项收敛到 $1/2$。

考虑 $\mathbb{S}_{n-1}(Z_n)$。取 $0<\epsilon<1/2-\delta$。这样的 ϵ 总是存在，因为 $\delta<1/2$。令$T_0(\epsilon)$ 为引理3.1中的 $T_0(\epsilon)$。对于所有 $n\geq T_0(\epsilon)$，定义事件 $A_n=\{|Z_n|>2n\sqrt{\nu}\}$。那么，在事件 A_n上，根据引理3.1，

$$n^{\delta}\mathbb{S}_{n-1}(Z_n)\leq 48\nu n^{2\epsilon+\delta-1}+\frac{|Z_n|^{-\beta}}{n^{\epsilon(d-1)-\delta}}L'(|Z_n|)+n^{\delta}g(|Z_n|)+n^{\delta}R_n(|Z_n|).$$

其中， L'、 g和 R_n 与引理3.1的陈述中相同。注意， Z_n 按 $n^{1/\beta}$缩放，且 $\delta<1/2$。因此，根据连续映射定理，对于任意选择的 ϵ ，右侧的第二项和第三项在 时趋于0。第一项由于 ϵ的选取而趋于 0 。此外，由引理3.1的陈述可得，

$$\mathbb{Q}\left(R_n(|Z_n|)>tn^{-\delta}\mid |Z_n|\right)\leq 5^d\exp\left(-\frac{n^{1-2\epsilon-2\delta}t^2}{512}\right)+C_d\frac{5^d|Z_n|^{-\beta}}{2^d n^{d\epsilon-1}}L(|Z_n|)$$

其中 $L(t)=t^{\beta+d}h(t)$ 当 $n\to\infty$ 时收敛到一个有限常数。再次，由于 $|Z_n|$ 按 $n^{1/\beta}$ 的尺度变化，最右侧的项对任意 $\epsilon>0$ 都趋于0。此外，右侧的指数项对所有 $t>0$ 都趋于0，因为 $\epsilon+\delta<1/2$。

上述论证意味着，在事件 A_n 条件下，$n^{\delta}\mathbb{S}_{n-1}(Z_n)$ 依概率趋于0 。这意味着

$$\mathbb{Q}\left(g(\tilde{Y}_{n,1},\ldots,\tilde{Y}_{n,n-1},Z_n)>n^{1-1/\beta-\delta}\mid |Z_n|>2n\sqrt{\nu}\right)\longrightarrow 0.$$

因此，

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}^d}\left|S_n(x)-\hat{S}_n(x)\right|>n^{1-1/\beta-\delta}\mid A_n\right)\longrightarrow 0$$

如 $n\to\infty$。 □

附录 D：证明引理5.2

本节将证明引理5.2。首先，我们将控制右侧偏离。

引理 D.1. 令 $y \in [x_n^+, Y_{(n)})$ 为任意但固定的值。定义 $T_y^{\uparrow}$ 为满足对所有 $t > 0$ 的随机变量

$$\mathbb{P}(T_y^{\uparrow} > t) = \exp\left(- \int_0^t \lambda_n(y+u, +1)\, du \right).$$

于是，当 E_n 这一事件发生且 n 足够大时

$$\mathbb{E}(T_y^{\uparrow}) \le \sqrt{\nu} + \frac{\sqrt{\nu\pi}}{2} \frac{\Gamma(\nu/2)}{\Gamma((\nu+1)/2)} =: C_\nu. \tag{D.1}$$

假设 $\gamma_n(x) \equiv 0$，或由以下式子定义(5.3)。令 $0 < c < 2\sqrt{\nu}$ 为常数。于是，当事件 E_n 发生且 n 足够大时

$$\mathbb{E}(T_y^{\uparrow}) \ge c e^{-(\nu+1)} \left(1 - 2c(\nu+1)\overline{\gamma_n}\right). \tag{D.2}$$

证明。由式 (5.4) 可知，对 $x > x_n^+$，

$$\lambda_n(x, +1) = (\nu+1) S_n(x) + \gamma_n(x) \ge (\nu+1)(S(x, Y_{(n)}))_+.$$

令 $y \ge x_n^+$ 为任意但固定的值。于是，对所有 $t > 0$

$$\mathbb{P}(T_y^{\uparrow} > t) \le \exp\left(-(\nu+1) \int_0^t (S(y+u, Y_{(n)}))_+ \, du \right).$$

因此，

$$\mathbb{E}(T_y^{\uparrow}) = \int_0^\infty \mathbb{P}(T_y^{\uparrow} > t)\, dt \le \int_0^\infty \exp\left(- \int_0^t \frac{(\nu+1)(y+u-Y_{(n)})_+}{\nu + (y+u-Y_{(n)})^2} \, du \right) dt.$$

计算右侧的积分可得，对于 $y \in [x_n^+, Y_{(n)})$

$$\mathbb{E}(T_y^{\uparrow}) \le (Y_{(n)} - y) + \frac{\sqrt{\nu\pi}}{2} \frac{\Gamma(\nu/2)}{\Gamma((\nu+1)/2)} \le \sqrt{\nu} + \frac{\sqrt{\nu\pi}}{2} \frac{\Gamma(\nu/2)}{\Gamma((\nu+1)/2)},$$

因为 $Y_{(n)} - y \le Y_{(n)} - x_n^+ \le \sqrt{\nu}$。

对于下界，令 $0 < c < 2\sqrt{\nu}$ 为一个常数。于是，在事件 E_n 上，$Y_{(n)} + c \le Y_{(n)} + Y_{(n)}/n$ 且对于 $u \in [x_n^+, Y_{(n)} + c]$, $S_n(x) < \hat{} S_n(x) + n^{1-1/\beta-\delta}$。由于 $\gamma_n \le (\nu+1)\gamma_n$，我们得到，对于 $u > x_n^+$

$$\begin{aligned} \lambda_n(u, +1) = (\nu+1) S_n(u) + \gamma_n &\le (\nu+1) \left(\hat{S}_n(u) + n^{1-1/\beta-\delta} + \overline{\gamma_n} \right) \\ &\le (\nu+1) \left(\frac{n-1}{x_n^+} + \frac{1}{2\sqrt{\nu}} + n^{1-1/\beta-\delta} + \overline{\gamma_n} \right). \end{aligned}$$

这与 $x_n^+ > x_n^-$ 以及（5.5）给出

$$\lambda_n(u, +1) \le (\nu+1) \left(\frac{1}{2\sqrt{\nu}} + 2\overline{\gamma_n} \right).$$

对于任意 $y \in (x_n^+, Y_{(n)}]$，马尔可夫不等式意味着，

$$\begin{aligned}\mathbb{E}(T_y^{\uparrow}) \geq c\mathbb{P}(T_y^{\uparrow} > c) &= c\exp\left(-\int_y^{y+c} \lambda_n(u,+1)\,du\right)\\ &\geq c\exp\left(-c(\nu+1)\left(\frac{1}{2\sqrt{\nu}} + 2\overline{\gamma_n}\right)\right)\\ &\geq ce^{-(\nu+1)}\left(1 - 2c(\nu+1)\overline{\gamma_n}\right),\end{aligned}$$

由于当 n 较大时，$c < 2\sqrt{\nu}$ 和 $cn^{1-1/\beta-\delta} < 1$，且 $e^{-x} \geq 1 - x$。因此，得证。 □

引理 *D.2.* 假设 $\gamma_n(x) \equiv 0$，或由 *(5.3)* 定义。令 $y \in [x_n^+, Y_{(n)})$ 任意但固定，且 $0 < c < 2\sqrt{\nu}$ 为常数。考虑 $T_n(y,+1)$，其中 $T_n(\overline{y},v)$ 按照 *(5.1)* 定义。那么，当 n 较大时，在事件 E_n 上，

$$2ce^{-(\nu+1)}\left(1 - 2c(\nu+1)\overline{\gamma_n}\right) \leq \mathbb{E}(T_n(y,+1)) \leq \frac{2\sqrt{\nu} + C_\nu}{\tilde{p}_n} + \sqrt{\nu},$$

其中 $1 \geq \tilde{p}_n \geq \exp(-\sqrt{\nu}(\nu+1)\overline{\gamma_n})$。

证明。令 $y \in [x_n^+, Y_{(n)})$ 任意但固定。我们关注 $\mathbb{E}(T_n(y,+1))$。从 $(y,+1)$ 开始，令 $T_y^{\uparrow}$ 表示过程在首次跳跃之前向 y 右侧移动的距离。因此可得，

$$\mathbb{E}(T_n(y,+1)) = \mathbb{E}\left(T_y^{\uparrow} + \mathbb{E}(T_n(y + T_y^{\uparrow}, -1) \mid T_y^{\uparrow})\right).$$

这里，$T_y^{\uparrow}$的分布满足，

$$\mathbb{P}(T_y^{\uparrow} > t) = \exp\left(-\int_0^t \lambda_n(y+u,1)\,du\right), \quad t \geq 0.$$

现在，注意到 $S_n(x) > 0$ 对所有 $x \geq Y_{(n)}$ 都成立。此外，由于 $\gamma_n(x) \equiv 0$ 对 $x \geq Y_{(n)}$ 成立，我们得到 $\lambda_n(x,-1) = 0$ 对 $x \geq Y_{(n)}$ 成立。这意味着，如果 $y + T_y^{\uparrow} > Y_{(n)}$，那么在过程回落到 $Y_{(n)}$ 之前不会发生跳跃。因此，对任意 $t > 0$，

$$2t \leq t + T_n(y+t,-1) \leq t + (y + t - Y_{(n)})_+ + T_n(Y_{(n)},-1) \leq 2t + T_n(Y_{(n)},-1),$$

因为 $y < Y_{(n)}$。因此，

$$2\mathbb{E}(T_y^{\uparrow}) \leq \mathbb{E}\left(T_n(y,+1)\right) \leq 2\mathbb{E}(T_y^{\uparrow}) + \mathbb{E}(T_n(Y_{(n)},-1)).$$

将（D.1）代入上式可得，

$$2\mathbb{E}(T_y^{\uparrow}) \leq \mathbb{E}(T_n(y,+1)) \leq 2C_\nu + \mathbb{E}(T_n(Y_{(n)},-1)). \tag{D.3}$$

下界由（D.2）即可得出。

为了得到上界，我们需要控制 $\mathbb{E}(T_n(Y_{(n)},-1))$。我们将使用 (D.3) 中与 y 无关的上界。考虑 $T_n(Y_{(n)},-1)$，即从 $(Y_{(n)},-1)$ 出发，首次到达 $(x_n^+,-1)$ 的时间。令 $\tilde{p}_n$ 表示一步之内从 $(Y_{(n)},-1)$ 到达 $(x_n^+,-1)$ 的概率。那么，以概率 $\tilde{p}_n$，$T_n(x_n^+,-1) = (Y_{(n)} - x_n^+)$。另一方面，以概率 $(1-\tilde{p}_n)$，该过程移动 $m_n < (Y_{(n)} - x_n^+)$ 距离并改变方向。

随后还需要 $T_n(Y_{(n)}-m_n,\ +1)$ 时间才能到达 $(x_n^+,\ -1)$。但请注意，$Y_{(n)}-m_n\in(x_n^+,\ Y_{(n)})$。因此，由 (D.3) 可得

$$\begin{aligned}&\mathbb{E}(T_n(Y_{(n)},-1))\\&\quad=\tilde{p}_n\cdot(Y_{(n)}-x_n^+)+(1-\tilde{p}_n)\mathbb{E}\left(m_n+\mathbb{E}(T_n(Y_{(n)}-m_n,+1))\mid m_n<Y_{(n)}-x_n^+\right)\\&\quad\le(Y_{(n)}-x_n^+)+(1-\tilde{p}_n)\left((Y_{(n)}-x_n^+)+2C_\nu+\mathbb{E}(T_n(Y_{(n)},-1))\right).\end{aligned}$$

结合D.3)，可得：

$$\mathbb{E}(T_n(y,+1))\le\left(\frac{2}{\tilde{p}_n}+1\right)(Y_{(n)}-x_n^+)+\frac{1}{\tilde{p}_n}C_\nu\le\frac{2\sqrt{\nu}+C_\nu}{\tilde{p}_n}+\sqrt{\nu},\tag{D.4}$$

因为 $(Y_{(n)}-x_n^+)<\sqrt{\nu}$。

最后，考虑 $\tilde{p}_n$，即一步之内从 $(Y_{(n)},-1)$ 到达 $(x_n^+,-1)$ 的概率。由于$S_n(x)>0$对 $x\ge x_n^+$，对于 $u\in(x_n^+,Y_{(n)}]$, $\lambda_n(u,-1)=(\nu+1)(-S_n(u))_+ +\gamma_n(u)\le(\nu+1)\gamma_n$。于是，

$$\tilde{p}_n=\exp\left(-\int_{x_n^+}^{Y_{(n)}}\lambda_n(u,-1)\,du\right)\ge\exp\left(-(\nu+1)(Y_{(n)}-x_n^+)\overline{\gamma_n}\right)\ge1-\sqrt{\nu}(\nu+1)\overline{\gamma_n},$$

因为 $Y_{(n)}-x_n^+\le\sqrt{\nu}$。证毕。 □

引理 D.3. 假设 $\gamma_n(x)\equiv0$ 或由 (5.3)定义。对于足够大的 n，在事件E_n上，存在常数 C_1 和 C_2 使得

$$C_1+O_p(n^{1-1/\beta})\le\mathbb{E}(T_n(x_n^+,+1))\le C_2+O_p(n^{1-1/\beta})$$

如 $n\to\infty$。

证明。回顾一下，$\overline{\gamma_n}=n^{1-1/\beta}/(G_k+c_\beta)$。由引理D.2 可得：

$$\mathbb{E}(T_n(x_n^+,+1))\ge2ce^{-(\nu+1)}\left(1-2c(\nu+1)\overline{\gamma_n}\right).$$

因此，下界得证。对于上界，引理D.2 给出：

$$\mathbb{E}(T_n(x_n^+,+1))\le\sqrt{\nu}+(2\sqrt{\nu}+C_\nu)\exp(\sqrt{\nu}(\nu+1)\overline{\gamma_n}).$$

当 $\beta=1$时，结论显然成立。对于 $\beta<1$，$\gamma_n\to0$，结论由 $e^x<1+2x$ 的上界在 x 较小时推出。 □

引理 D.3 证明，对于典范 ZZP 和子采样 ZZP，右侧偏离在 $n\to\infty$ 时均保持为 $O_p(1)$。现在我们分别证明典范 ZZP 和子采样 ZZP 中关于 η_n 的 $O_p(1)$ 上界。首先，考虑典范情形。

引理 D.4. 假设 $\gamma_n\equiv0$。对于足够大的 n，在事件 E_n 上，存在一个常数 C_3，使得

$$\mathbb{E}(\eta_n)\le C_3+O_p(n^{1-1/\beta})$$

当 $n\to\infty$ 时。

证明。设 $T_n^{\downarrow}$ 为从 $(x_n^-,-1)$ 出发的过程的首次跳跃时间。若 $T_n^{\downarrow}<W_n$，则该过程会在到达 x_n^- 之前发生跳跃。对于较大的 n，在事件 E_n 上，$S_n(x)<0$ 对于 $x\in(x_n^-,x_n^+)$。当 $\gamma\equiv0$时，在过程到达 $(x_n^+,1)$ 之前不会发生其他跳跃。因此，

$$\eta_n=\begin{cases}2T_n^{\downarrow}, & T_n^{\downarrow}<W_n,\\ W_n, & T_n^{\downarrow}\ge W_n.\end{cases}$$

这意味着，

$$\mathbb{E}(\eta_n) = 2\mathbb{E}(\min\{T_n^{\downarrow}, W_n\}) - W_n\mathbb{P}(T_n^{\downarrow} \geq W_n) = 2\int_0^{W_n} \mathbb{P}(T_n^{\downarrow} > t)\, dt - W_n p_{\tau,n}.$$

然而，根据引理 5.1 和定理 2.4, $W_n p_{\tau,n} = O_p(n^{(1-1/\beta)\nu})$。因此，我们只需关注第一项。

考虑，

$$\begin{aligned}\int_0^{W_n} \mathbb{P}(T_n^{\downarrow} > t)\, dt &= \int_0^{W_n} \exp\left(-\int_0^t \lambda_n(x_n^+ - u, -1)\, du\right) dt \\ &= \int_0^{W_n} \exp\left((\nu+1)\int_{x_n^+ - t}^{x_n^+} S_n(u)\, du\right) dt\end{aligned}$$

因为 $\lambda_n(x,-1) = (\nu+1)(-S_n(x))_+$ 且 $S_n(x) \leq 0$ 对于 $x \in (x_n^-, x_n^+)$成立。不过，还有，

$$\begin{aligned}&\exp\left((\nu+1)\int_{x_n^+ - t}^{x_n^+} S_n(u)du\right) \\ &= \left(\frac{\nu + (x_n^+ - Y_{(n)})^2}{\nu + (x_n^+ - t - Y_{(n)})^2}\right)^{(\nu+1)/2} \cdot \prod_{j=1}^{n-1}\left(\frac{\nu + (x_n^+ - Y_{(j)})^2}{\nu + (x_n^+ - t - Y_{(j)})^2}\right)^{(\nu+1)/2} \\ &\leq \left(\frac{\nu + (x_n^+ - Y_{(n)})^2}{\nu + (x_n^+ - t - Y_{(n)})^2}\right)^{(\nu+1)/2} \cdot \prod_{j=1}^{n-1}\left(\frac{\nu + (x_n^+ - Y_{(j)})^2}{\nu + (x_n^- - Y_{(j)})^2}\right)^{(\nu+1)/2} \\ &= p_n\left(\frac{\nu + (x_n^- - Y_{(n)})^2}{\nu + (x_n^+ - t - Y_{(n)})^2}\right)^{(\nu+1)/2},\end{aligned}$$

其中，回忆一下 $p_n = \pi^{(n)}(x_n^-)/\pi^{(n)}(x_n^+)$。我们得到，

$$\begin{aligned}\int_0^{W_n} \mathbb{P}(T_n^{\downarrow} > t)\, dt &\leq p_n\left(\nu + (x_n^- - Y_{(n)})^2\right)^{(\nu+1)/2} K(\nu) \\ &= K(\nu)\frac{p_n}{(Y_n - x_n^-)^{-(\nu+1)}}\left(1 + \frac{\nu}{(Y_n - x_n^-)^2}\right)^{(\nu+1)/2}\end{aligned}$$

对于某个常数 $K(\nu)$。由定理 2.4 和引理 5.1可得该结果。 □

引理 D.5. 假设 γ_n 如 (5.3) 中所定义。对于充分大的 n，在事件 E_n 上，存在常数 C_3 使得

$$\mathbb{E}(\eta_n) \leq C_3 + O_p(n^{1-1/\beta})$$

当 $n \to \infty$ 时。

证明。我们将沿用 [34] 中的论证。设 $Z_0 = (X_0,\ V_0) = (x_n^+,\ -1)$ 和 $T_0 = 0$。设$(F_k)_{k\geq 1}$为一列独立同分布的指数(1) 随机变量。假设对于某个 $k \geq 0$，$Z_k = (X_k,\ V_k)$和 T_k 已经定义。定义 τ_{k+1} 为

$$F_{k+1} = \int_0^{\tau_{k+1}} \lambda_n\left(X_k + V_k s, V_k\right) ds,$$

其中 $\lambda_n(x,v) = (\nu+1)(vS_n(x))_+ + \gamma_n(x)$。设 $T_{k+1} = T_k + \tau_{k+1}$。定义 $X_{k+1} = X_k + T_{k+1}\cdot V_k$ 和 $V_{k+1} = -V_k$。重复此过程，得到一个骨架 $(Z_k,T_k)_{k\geq 0} = ((X_k,V_k)\ T_k)_{k\geq 0}$。定义一个初始条件为 $(X_0^n,V_0^n) = (x_n^+,-1)$ 的连续时间过程 $Z_t^n = (X_t^n,V_t^n)$:

$$X_t^n = X_k + (t-T_k)\cdot V_k, \quad V_t^n = V_k; \quad t\in(T_k,T_{k+1}].$$

那么，Z_t^n是一个速率为 $\lambda_n(x,v)$ 的 Zig-Zag 过程。根据构造，$V_{2k-1} = -V_{2k} = 1$，即从 $(x_n^+,-1)$ 开始，该过程在向左移动和向右移动之间交替。此外，时间 τ_{2k-1} 和 $W\tau_{2k}$ 分别表示这些移动持续的时间（因此也是长度）。

考虑 η_n。注意，η_n 至多是向左行进距离的两倍。事实上，若 $X_\eta^n = x_n^+$，该过程向左移动的时间与向右移动的时间一样长；若 $X_\eta^n = x_n^-$，则向左移动的时间比向右移动的时间更长。令 N_η 为时刻 η_n 之前向左移动的次数。那么，

$$\eta_n \leq 2\sum_{k=1}^{N_\eta}\tau_{2k-1} \quad \text{where} \quad N_\eta = \max\{k : T_{2k-1} < \eta_n\}.$$

但对于所有 $k\leq N_\eta$，$\tau_{2k-1}\leq W_n$。因此，$\eta_n \leq 2W_nN_\eta$。

现在，对于较大的 n，在事件 E_n 上，$S_n(x)\leq 0$ 对所有 $x\in(x_n^-,x_n^+)$ 均成立。那么，$\lambda_n(x,+1) = \gamma_n(x)$ 对所有 $x\in(x_n^-,x_n^+)$ 均成立。定义:

$$J_k = \left\{F_{2k} < \int_{x_n^-}^{x_n^+}\gamma_n(u)\,du\right\}, \quad N = \min\{k\geq 1, J_k^c\}.$$

那么，N 是一个几何随机变量，其满足

$$\mathbb{E}(N) = 1/\mathbb{P}(J_1^c) \leq e^{(\nu+1)W_n\overline{\gamma_n}}.$$

由于 $\gamma_n\geq 0$，在事件 J_k^c 上，我们有 $\tau_{2k} > W_n$。因此，如果 $T_{2k-1} < \eta_n$和 J_k^c发生，那么在过程到达（x_n^+，$+1$）之前不会再发生其他跳跃。因此，必有$N_\eta \leq N$。结合前面的观测，我们得到，

$$\eta_n \leq 2W_nN.$$

当 $\beta = 1$时，对上述等式两边取期望，可得，

$$\mathbb{E}(\eta_n) \leq 2W_ne^{(\nu+1)W_n\overline{\gamma_n}}.$$

结论由定理2.4和等式（5.6）得出。

现在，假设 $0<\beta<1$。那么，根据定理2.4，$W_n\to\infty$ 当 $n\to\infty$时。令 n 足够大，使得 $W_n > 1$。在事件 E_n上令 $0 < \delta_n = W_n^{1/(1+\nu)} < W_n$。令事件 B 为 $\{W_1 < \delta_n, W_1 < F_2/(\nu+1)\ \gamma_n\}$。那么，在事件 B上，由于 $W_1 < \delta_n < W_n$，该过程在第一步不会向左离开。然而，在事件 B上，同样可以推出，$W_2 > W_1$。为此，假设 $W_2\leq W_1$。那么，由于 $\lambda_n(x, +1) = \gamma_n(x) \leq (\nu+1)\ \gamma_n$ 对于$x\in(x_n^-)$，我们得到，

$$F_2 = \int_0^{W_2}\lambda_n(x_n^+ - W_1 + u, +1)\,du \leq \int_0^{W_1}\gamma_n(x_n^+ - W_1 + u)\,du \leq (\nu+1)W_1\overline{\gamma_n},$$

这是一个矛盾。因此，$W_2 > W_1$。于是，在事件 B 上，过程在第一步向右退出，因此，$\eta_n = 2W_1$ 在事件 B 上。我们有，

$$\mathbb{E}(\eta_n) = \mathbb{E}(1_B\eta_n) + \mathbb{E}(1_{B^c}\eta_n) \leq 2\mathbb{E}(1_BW_1) + 2W_n\mathbb{E}(1_{B^c}N).$$

但同样，在事件 B^c 上，如果 $F_2 < (\nu+1)\delta_n\gamma_n$ 那么， $N > 1$。因此我们得到，

$$\begin{aligned}\mathbb{E}(\eta_n) &\leq 2\mathbb{E}(1_B W_1) + 2W_n\mathbb{P}(B^c)(1+\mathbb{E}(N)) \\ &\leq 2\mathbb{E}(1_B W_1) + 2W_n\mathbb{P}(B^c)\left(1+e^{(\nu+1)W_n\overline{\gamma_n}}\right).\end{aligned}$$

首先考虑，

$$\begin{aligned}\mathbb{P}(B) = \mathbb{P}(W_1 < \delta_n, W_1 < F_2/(\nu+1)\overline{\gamma_n}) &= \mathbb{E}\left(e^{-(\nu+1)W_1\overline{\gamma_n}}1(W_1<\delta_n)\right) \\ &\geq \mathbb{E}\left((1-(\nu+1)W_1\overline{\gamma_n})1(W_1<\delta_n)\right) \\ &= \mathbb{P}(W_1<\delta_n) - (\nu+1)\overline{\gamma_n}\mathbb{E}(W_1 1(W_1<\delta_n)).\end{aligned}$$

此外，注意到 $\mathbb{E}(1_B W_1) \leq \mathbb{E}(W_1 1(W_1<\delta_n))$。因此，重新整理后可得

$$\begin{aligned}\mathbb{E}(\eta_n) &= 2W_n\mathbb{P}(W_1>\delta_n)(1+e^{(\nu+1)W_n\overline{\gamma_n}}) \\ &\quad + 2\left(1+(\nu+1)W_n\overline{\gamma_n}(1+e^{(\nu+1)W_n\overline{\gamma_n}})\right)\mathbb{E}(W_1 1(W_1<\delta_n))\end{aligned}$$

由（5.6），在集合 E_n上，当 n较大时，

$$W_n\overline{\gamma_n} \leq 4 + n^{1-1/\beta}\sqrt{\nu}/G_k,$$

因为 $c_\beta = 0$ 当 $\beta < 1$时。右侧在 $n\to\infty$时趋于4 。因此，当 n较大时， $W_n\gamma_n \leq 5$ 在事件 E_n 上。我们得到，

$$\mathbb{E}(\eta_n) \leq 2W_n\mathbb{P}(W_1>\delta_n)(1+e^{5(\nu+1)}) + 2(1+5(\nu+1)(1+e^{5(\nu+1)}))\mathbb{E}(W_1 1(W_1<\delta_n))$$

现在注意，当 n 足够大时，在事件 E_n 上，对于所有 $x\in(x_n^-, x_n^+)$，都有 $\lambda_n(x,-1) = -(\nu+1)S(x, Y_{(n)})$。令 $t \leq W_n$。于是，

$$\mathbb{P}(W_1>t) = \mathbb{P}\left(F_1 > \int_0^t \lambda_n(x_n^+ - u, -1)\,du\right) = \left(\frac{\nu+(Y_{(n)}-x_n^+)^2}{\nu+(Y_{(n)}-x_n^+ + t)^2}\right)^{(\nu+1)/2}$$

那么，回想起 $\delta_n = W_n^{1/(1+\nu)}$，我们得到，

$$W_n\mathbb{P}\left(W_1 > W_n^{1/(1+\nu)}\right) \leq W_n\left(\frac{\nu+(Y_{(n)}-x_n^+)^2}{W_n^{2/(1+\nu)}}\right)^{(\nu+1)/2} = \left(\nu+(Y_{(n)}-x_n^+)^2\right)^{(\nu+1)/2}.$$

此外，

$$\begin{aligned}\mathbb{E}(W_1 1(W_1<\delta_n)) &\leq \int_0^{\delta_n}\mathbb{P}(W_1>t)dt \\ &\leq (\nu+(Y_{(n)}-x_n^+)^2)^{(\nu+1)/2}\int_0^\infty(\nu+t^2)^{-(\nu+1)/2}\,dt \\ &\leq K(\nu)(\nu+(Y_{(n)}-x_n^+)^2)^{(\nu+1)/2},\end{aligned}$$

对于某个常数 $K(\nu)$。结合这些观测值并回想起 $|Y_{(n)}-x_n^+| \leq \sqrt{\nu}$，可得，

$$\mathbb{E}(\eta_n) \leq (2K(\nu) + (2+10(\nu+1)K(\nu))(1+e^{(\nu+1)5}))\,(2\nu)^{(\nu+1)/2}.$$

结论得证。 □

引理 5.2的证明。该结果可通过综合引理D.3、引理D.4与引理D.5得到。

□

致谢。作者感谢Alberto Bordino就一致大数定律所作的有益讨论。

经费支持。SA 获得英国工程与自然科学研究委员会博士生奖学金 EP/W523793/1 的资助，以及丹麦独立研究基金通过 Sapere Aude 启动基金（5251-00032B）提供的资助。SG 感谢欧盟通过欧洲研究委员会（ERC）Starting Grant“PrSc-HDBayLe”（项目编号：101076564）提供的支持。GOR 获得 PIN-CODE（EP/X028119/1）、EP/V009478/1 以及英国研究与创新署资助项目 OCEAN（EP/Y014650/1）的支持。

参考文献

[1] AGRAWAL, S., BIERKENS, J. 和 ROBERTS, G. O.（2024）。用于贝叶斯推断的 Zig-Zag 算法的大样本尺度分析。arXiv:2411.14983 [stat]。https://doi.org/10.48550/arXiv.2411. 14983[2] ANDRIEU, C. 和 LIVINGSTONE, S.（2021）。马尔可夫蒙特卡罗的 Peskun–Tierney 排序：超越可逆情形。《统计年鉴》49 1958–1981。[3] ASCOLANI, F. 和 ZANELLA, G.（2025）。用于贝叶斯有限混合模型的快速非可逆采样器。arXiv 预印本 arXiv:2510.03226。[4] BANDEIRA, A. S.、MAILLARD, A.、NICKL, R. 和 WANG, S.（2023）。关于高斯先验中的自由能垒以及冷启动马尔可夫链蒙特卡罗在高维单峰分布中的失效。《皇家学会哲学汇刊》A 辑 381。[5] BARNETT, V. D.（1966）。似然方程具有多个根时最大似然估计量的评估。《生物统计学》 **53** 151 165 10 2307 2334061 –。 [6] 2016 https://doi.org/。/ BIERKENS, J.（）。非可逆 Metropolis-Hastings。《统计与计算》26 1213–1228。[7] BIERKENS, J.、FEARNHEAD, P. 和 ROBERTS, G. O.（2019）。用于大数据贝叶斯分析的 Zig-Zag 过程与超高效采样。《统计年鉴》47 1288–1320。https://doi.org/10. 1214/18-AOS1715[8] BIERKENS, J.、GRAZZI, S.、MEULEN, F. V. D. 和 SCHAUER, M.（2023）。用于稀疏和局部推断问题的黏滞 PDMP 采样器。《统计与计算》33 8。[9] BIERKENS, J.、ROBERTS, G. O. 和 ZITT, P.-A.（2019）。zigzag 过程的遍历性。《应用概率年鉴》29 2266–2301。编号：4。出版者：数学统计学会。https: //doi.org/ 10 1214 18 1453 。/ -AAP BINGHAM, N. H.、GOLDIE, C. M. 和 TEUGELS, J. L.（ [10] 1987）。《正则变差》。剑桥大学出版社，剑桥。https://doi.org/ 10 1017 9780511721434 。/CBO BOONSTRA, H. J.、VAN [11] 2021 DEN BRAKEL, J. 和 DAS, S.（）。荷兰小区域流动趋势的多层时间序列建模。《皇家统计学会会刊 A 辑：社会统计学》 **184** 985 1007 10 1111 12700 –。https://doi.org/。/ [12] 2025 rssa。BORDINO, A. 和 BERRETT, T. B.（）。基于样本协方差矩阵的完全随机缺失检验。《统计年鉴》 **53** 2204 2229 10 1214 25 2540 –。 [13] 2016 **351** https://doi.org/。/ -AOS BOVIER, A. 和 DEN HOLLANDER, F.（）。《亚稳态：势理论方法》。施普林格。[14] BOVIER, A.、ECKHOFF, M.、GAYRARD, V. 和 KLEIN, M.（2004）。可逆扩散过程中的亚稳态。I. 容量与退出时间的精确渐近式。《欧洲数学学会杂志》6 399–424。[15] BROOKS, S. P.、GIUDICI, P. 和 ROBERTS, G. O.（2003）。可逆跳跃马尔可夫链蒙特卡罗提议分布的高效构造。《皇家统计学会会刊 B 辑（统计方法学）》65 3–55。[16] CHEWI, S.（2023）。对数凹采样。书稿见 https://chewisinho. github. i o 9 17–18。[17] DAVIS, M. H.（2018）。马尔可夫模型与优化。劳特利奇。[18] DE FINETTI, B.（1961）。关于拒绝离群值的贝叶斯方法。载于《第四届伯克利数学统计与概率研讨会论文集》1 99–210。加州大学出版社，加利福尼亚州伯克利，美国。[19] DE HAAN, L. 和 FERREIRA, A.（2006）。《极值理论》。施普林格，纽约，纽约州。https://doi.org/ 10.1007/0-387-34471-3[20] DE, S. 和 GHOSH, J.（2025）。具有重尾误差的线性回归模型的稳健贝叶斯模型平均。《应用统计学杂志》1–27。出版者：泰勒-弗朗西斯出版集团。[21] DIACONIS, P.、HOLMES, S. 和 NEAL, R. M.（2000）。非可逆马尔可夫链采样器的分析。《应用概率年鉴》726–752。[22] FERGUSON, T. S.（1978）。样本量为 3 和 4 时柯西分布参数的最大似然估计。《美国统计学会会刊》73 211–213。https: //doi.org/10.1080/01621459.1978.10480031

[23] FERNANDEZ, C. 和 STEEL, M. F. J.（1999）。多元 Student-t 回归模型：陷阱与推断。《生物统计学》86 153–167。[24] FONSECA, T. C. O.、FERREIRA, M. A. R. 和 MIGON, H. S.（2008）。Student-T 回归模型的客观贝叶斯分析。《生物统计学》95 325–333。[25] GAGNON, P. 和 DESGAGNÉ, A.（2024）。贝叶斯与频率学派方法在稳健性方面的一个根本差异。arXiv 预印本 arXiv:2408.10478。[26] GAGNON, P. 和 HAYASHI, Y.（2023）。贝叶斯 Student-t 线性回归的理论性质。《统计与概率快报》

193 109693 10 1016 2022 109693 。https://doi.org/。/

[27] 2024
j.spl。 。GAGNON, P. 和 MAIRE, F.（）。渐近 Peskun 序及其在提升采样器中的应用。《伯努利》30 2301–2325。[28] GELMAN, A.、JAKULIN, A.、PITTAU, M. G. 和 SU, Y.-S.（2008）。用于逻辑回归及其他回归模型的弱信息默认先验分布。《应用统计年鉴》2 1360–1383。https://doi.org/

10 1214 08 191 。/ -AOAS GEYER, C. J. 和 THOMPSON, E. A.（1995）。带有

[29]
祖先推断应用的退火马尔可夫链蒙特卡罗。《美国统计学会会刊》90 909–920。https://doi.org/10.2307/2291325[30] GRAMACY, R. B. 和 PANTALEO, E.（2010）。用于多元推断且包含缺失数据的收缩回归，以及其在投资组合平衡中的应用。《贝叶斯分析》5 237–262。https://doi.org/10.1214/ 10-BA602[31] HE, D.、SUN, D. 和 HE, L.（2021）。Student-t 线性回归的客观贝叶斯分析。《贝叶斯分析》

16 129 145 10 1214 20 1198 –。https://doi.org/。/ -BA

[32] 1989
LANGE, K. L.、LITTLE, R. J. A. 和 TAYLOR, J. M. G.（）。使用 t 分布的稳健统计建模。《美国统计学会会刊》84 881–896。https://doi.org/10.1080/ 01621459.1989.10478852[33] LEE, S.、RAMIL, M. 和 SEO, I.（2025）。欠阻尼朗之万过程的 Eyring-Kramers 定律。arXiv 预印本 arXiv:2503.12610。[34] MONMARCHÉ, P.（2016）。分段确定性模拟退火。《ALEA：拉丁美洲概率与数理统计杂志》

13 357 398 10 30757 13 15 –。

[35] 2004
https://doi.org/。/alea.v - NEAL, R. M.（）。改进 MCMC 估计量的渐近方差：非可逆链更优。arXiv 预印本 math/0407281。[36] O'HAGAN, A. 和 PERICCHI, L.（2012）。贝叶斯重尾模型与冲突消解：综述。《巴西概率与统计杂志》

26 372 401 10 1214 11 164 –。

[37] 2025
https://doi.org/。/ -BJPS PEUTREC, D. L.、MICHEL, L. 和 NECTOUX, B.（）。某些分段确定性马尔可夫过程的 Eyring–Kramers 型公式。《亨利·庞加莱研究所年鉴：概率与统计》

61 803 821 10 1214 23 1445 –。https://doi.org/。/ -AIHP REEDS, J. A.（）。

[38] 1985
柯西位置似然方程根的渐近数量。《统计年鉴》

13 775 784 10 1214 1176349554 –。https://doi.org/。/aos/ RESNICK,

[39] 2008
S. I.（）。极值、正则变差与点过程 4。施普林格科学与商业媒体。[40] ROBERTS, G. O. 和 ROSENTHAL, J. S.（2004）。一般状态空间马尔可夫链与 MCMC 算法。《概率调查》

1 10 1214 154957804100000024 。https://doi.org/。/

[41] 2010
SUN, Y.、SCHMIDHUBER, J. 和 GOMEZ, F.（）。通过插入涡旋改进马尔可夫链蒙特卡罗的渐近性能。《神经信息处理系统进展》23。[42] VASDEKIS, G. 和 ROBERTS, G. O.（2022）。关于一维 Zig-Zag 过程的多项式遍历性的注记。《应用概率杂志》59 895–903。https://doi.org/10.1017/jpr.2021.97