次柯西采样：逃离《月之暗面》

Sebastiano Grazzi[1], Sifan Liu[2], Gareth O. Roberts[3], 和 Jun Yang[4]

[1]决策科学系和BIDSA，博科尼大学[2]统计科学系，杜克大学[3]统计系，华威大学[4]数学科学系，哥本哈根大学

arXiv:2601.11066v1 [stat.CO] 16 Jan 2026

摘要

我们介绍了一种基于次柯西投影的马尔可夫链蒙特卡罗算法。次柯西投影是一种几何变换，通过将欧几里得空间映射到超球面的球冠上来推广立体投影；该球冠被称为月之暗面的补集[1]。我们证明，对于次柯西目标分布，即尾部至多与多维柯西分布一样厚重的目标分布，我们提出的方法是一致遍历的，并通过实验证明了其在具有挑战性的高维问题中的性能。我们方法简单且适用范围广，为在现有大多数方法都不可靠的场景中，使用重尾分布开展贝叶斯建模和计算开辟了新的机会。

关键词——重尾分布、Metropolis–Hastings算法、球极投影MCMC、几何投影、稳健贝叶斯推断。

1 引言

1.1 重尾分布的采样

重尾分布的特征是其尾部按多项式而非指数速度衰减，因此与指数分布或高斯分布等轻尾分布相比，它们赋予极端值的概率要大得多。这类分布自然地出现在许多领域，包括金融、保险和气候科学；在这些领域中，极端事件发生得更频繁，影响也更大。对于这类模型，似然通常关于数据和模型参数都呈重尾，这常常会导致重尾后验分布。在完全分离的情况下，二元回归问题也会出现重尾后验分布 [Gelman 等人, 2008, Ghosh 等人，2018] ；稳健估计中也会出现重尾后验分布。正如 De Finetti [1961], 最早指出的，对于重尾模型，随着观测值的幅度增大，极端观测值对后验分布前两个矩（若存在）的贡献会变得可以忽略。后来，这一观察在一个一般框架下得到了形式化

[1]这个术语是为了向平克·弗洛伊德的《月之暗面》（1973年）致敬而选用的。

该框架建立在贝叶斯冲突解决之上，研究存在离群值时先验分布与似然的设定之间相当微妙而复杂的相互作用。关于全面综述，参见 O'Hagan 和 Pericchi [2012]；关于利用重尾分布成功开展稳健贝叶斯估计的应用示例，参见 Gagnon 和 Hayashi [2023],De 以及 Ghosh [2025]，并参阅其中的参考文献。

尽管建模方面取得了这些进展，计算方面仍面临许多挑战。对于重尾后验分布，大多数常用的 Metropolis–Hastings算法并非几何遍历 [Jarner 和 Hansen，2000，Livingstone 等人，2019，Wang 和 Yang， 2025]。对于特定统计模型（尤其是回归问题），最常用的计算方法是基于吉布斯抽样，并采用适当的数据增广方案 [Polson 等人，2013，Ghosh 等人，2018，Liu，2004，Gagnon 和 Hayashi， 2023]。只有在能够从全条件分布中进行抽样时，才能应用这些算法，这限制了可用的模型和先验分布。此外，针对重尾目标分布的吉布斯采样器，其收敛保证有限，实证研究也表明其混合速度较慢 [Ghosh 等人， 2018]。

因此，尽管重尾模型具有公认的统计优势，缺乏广泛适用且可靠的抽样方法仍限制了它们在贝叶斯实践中的应用。本文直接针对这一缺口，提出一种简单、合理且普遍适用的方法，用于从重尾后验分布中抽样。

1.2 我们的贡献概述

从总体上看，所提出的马尔可夫链蒙特卡罗（MCMC）算法首先将原始状态空间 $\mathbb{R}^d$ 映射到单位球面的一个球冠上，从而将 $\mathbb{R}^d$ 上的目标分布变换为一个支撑在球面子集上的分布。随后，使用一种新的随机游走型 Metropolis 算法对球冠上的目标分布进行采样，再通过逆变换将所得样本映射回原始空间 $\mathbb{R}^d$ 。

从球冠到 $\mathbb{R}^d$ 的逆映射称为次柯西投影（SCP）。具体来说，我们将一个 d维单位球面置于 $\mathbb{R}^{d+1}$中，并使其位于 d维超平面 $\{(y,\ 0)\colon\ y \in \mathbb{R}^d\} \subset \mathbb{R}^{d+1}$. 上方。该超平面与球面在原点 $(0,\ \ldots\ldots,\ 0) \in \mathbb{R}^{d+1}$处相切，该点对应球面的南极。我们将一名观测者$o \in \mathbb{R}^{d+1}$ 放在球内，并移除位于观测者上方的球面部分。被移除的球冠在下文中称为月之暗面，其补集称为亮侧。对于球面亮侧上的一点 x ，次柯西投影 $y = \mathrm{SCP}_o(x) \in \mathbb{R}^d$ 定义为经过 o 和 x 的直线与超平面 $\{(y,\ 0)\colon\ y \in \mathbb{R}^d\}$的交点。图1展示了观测者取两种不同位置时的示意图 o。

对于每个观测者 o，映射 SCP_o 是从亮侧到 $\mathbb{R}^d$ 的光滑双射（即一个微分同胚），但它不保持距离和体积。当 $o = (0,\ \ldots\ldots,\ 0,\ 2)$ 时，对应于球面的北极，我们得到经典立体投影[2]，参见例如 Coxeter， 1961]，，而暗侧则退化为单点 o。然而，对于球面内部的其他任意选择 o，次柯西投影与经典立体投影有两个重要区别：

- 暗侧是一个 d 维区域，而非一个点；$\mathbb{R}^d$ 中的无穷远点会映射到其 $(d-1)$ 维边界。这一区别体现在变换的雅可比行列式中（备注 1），使 SCP 能够适用于尾部比经典立体投影所能处理的更重的目标分布（备注 2）。

[2]立体投影也可以定义为以原点为球心（见 [Yang 等人，2024年，图 2]）。这两种定义只相差一个缩放常数，因此等价。

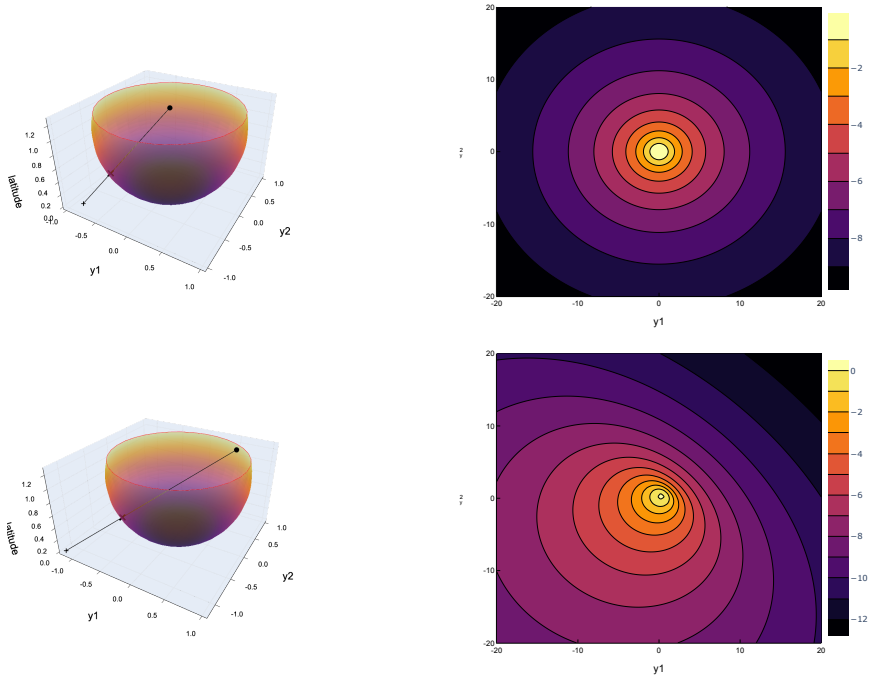

图 1：二维次柯西投影示意图，$o = (0, 0, 1.5)$（上方各图）和 $o = (0.6, 0.6, 1.5)$（下方各图）。右侧各图：投影到 $\mathbb{R}^2$ 的球面亮侧均匀分布的对数密度等高线。

- 观测者 o 充当调节参数，相较于经典立体投影提供了更大的灵活性。具体而言，通过改变 o 的位置，映射 SCP$_o$ 可以偏向某个方向，因此非常适合从不对称或偏斜的重尾分布中进行采样。

我们提出的算法称为次柯西投影采样器 （SCS），是一种作用于亮侧的随机游走型Metropolis算法。我们的主要理论结果证明，对于任意 d 维次柯西分布，SCS 都是一致遍历的；也就是说，这些分布的尾部至多与 d 维柯西分布一样重，参见定义 1。据我们所知，SCS 是首个针对多维次柯西分布、具有可证明的一致遍历性保证的通用型 Metropolis调整算法。

定义 1。 分布 π *in* $\mathbb{R}^d$ 是次柯西分布，当且仅当

$$\sup_{y \in \mathbb{R}^d} \pi(y)(1 + \|y\|^2)^{(d+1)/2} < \infty.$$

除了理论保证之外，我们还提供了选择观察者参数 o 的实用指南，以及移位和缩放常数等其他调节参数的选择指南。我们进一步展示了 SCS 在一系列高维重尾目标分布上的性能，其中包括源自鲁棒贝叶斯二元回归的示例；在这类问题中，众所周知，数据增广吉布斯采样器的混合速度较慢 [Ghosh 等人， 2018]。在这些设置下，与哈密顿蒙特卡罗等常用替代方法相比，SCS 收敛速度快得多，能够更准确地探索后验分布尾部，并实现更高的计算效率。

1.3 柯西分布示例

在详细介绍我们的方法之前，我们先给出一个示例，以展示 SCS 的优势并阐释其背后的直觉。我们将其性能与其他常用方法进行比较

在一个典型的重尾目标分布示例——100维柯西分布上，比较 SCS 与其他流行的马尔可夫链蒙特卡罗方法。对于这一目标分布，立体投影采样器（SPS；Yang 等人，2024年，算法 1）以及标准的 MCMC 算法，如随机游走 Metropolis（RWM）、Metropolis 调整朗之万算法（MALA）和哈密顿蒙特卡洛（HMC），均无法达到几何遍历性 [Yang 等人，2024年；Jarner 和 Hansen，2000年；Roberts 和 Tweedie，1996年；Livingstone等人， 2019], ，导致相应马尔可夫链混合缓慢。图 2 比较了我们提出的 SCS（参数按照第3节中的方法选取）、SPS，以及 RWM 和 HMC 等其他常用算法的轨迹。左图展示了从平稳分布出发的平方范数轨迹，中图展示了链从尾部位置初始化时前两个坐标的变化。在后一种情况下，SPS 被困在初始位置，因此我们省略其结果。根据文献建议，SCS、SPS 和 RWM 的经验接受概率约为 0.234，HMC 的则约为 0.7。

SCS 与 SPS 的行为不同，有一个简单的几何原因可以解释这一点；当观测者 o位于北极时，后者对应于 SCS。对于立体投影（SP），由于多元柯西密度的尾部过重，球面上的变换后密度在北极附近无界。相比之下，对于球面内部 o的任何其他取值，SCP 所诱导的密度在亮侧仍然有界；二维示例见图3。这两个投影之间看似细微的差异会带来重要后果：从尾部初始化时，SPS 往往会被困在北极附近（月之暗面），而 SCS 则能逃离暗侧，并高效地探索目标分布。

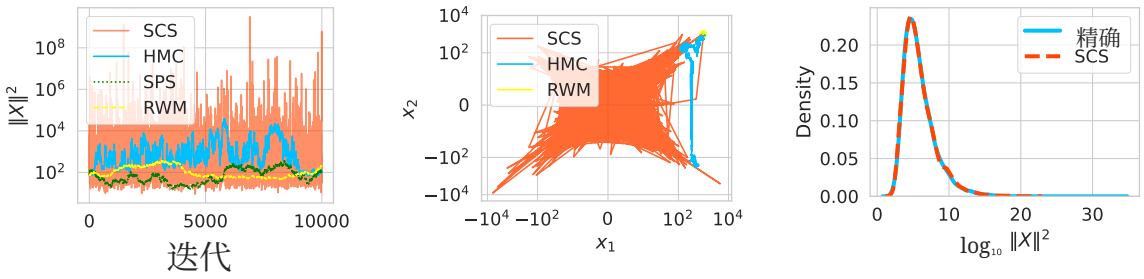

图2：从100维标准柯西分布中采样。左图：SCS（橙色）、HMC（蓝色）、SPS（绿色）和 RWM（黄色）的 $\|X\|^2$ 轨迹图，初始值取自目标分布中的一个样本。中图：SCS（橙色）、HMC（蓝色）和 RWM（黄色）在前两个维度上的轨迹（对数尺度），初始值为 (10_3，10_3，.....)。右图：SCS 样本的 $\|X\|^2$ 密度（红色，由核密度回归估计），并与真实密度（蓝色）进行比较。

1.4 相关工作

我们的工作克服了当前流行马尔可夫链蒙特卡罗方法的局限性：对于重尾目标分布，这些方法并不具备几何遍历性 [Jarner 和 Hansen，2000，第 4] [节；Livingstone等人，2019，定理 2。2],[Wang 和 Yang，2025，命题 2]。已有研究提出了一些变换策略，以恢复几何遍历性/一致遍历性 [Johnson 和 Geyer，2012；Yang 等人， 2024],但前者无法恢复一致遍历性，后者则不适用于不具有有限（$d-1$）阶矩的 d维分布 [Yang 等人，2024，定理 2.1 和定理 2。2][Breˇsar 等，2025，第 3.1 节3]。近年来，人们对重尾分布的采样算法日益关注。例如， [He 等，2022；Mousavi-Hosseini 等，2023；Vasdekis 和 Roberts，2023；He 等，2024；Bertazzi 和 Vasdekis， 2025]为朗之万动力学和分段确定性马尔可夫过程推导了理论保证

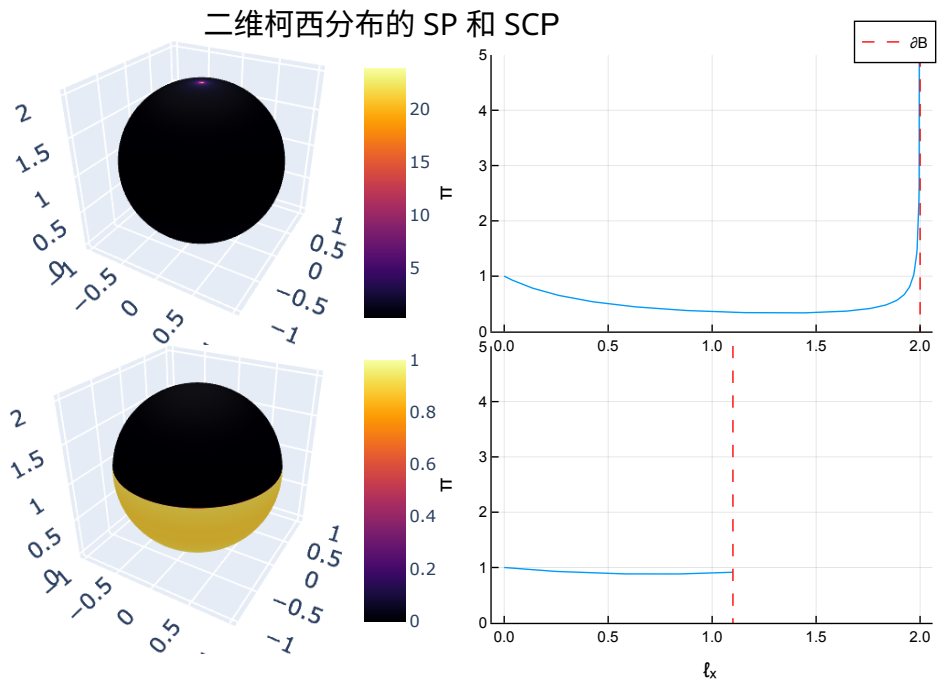

图3：二维柯西分布的 SP（上排图）和观测者位于 $o = (0, 0, 1。1)$ 时的 SCP（下排图）。右侧各图显示沿纬度分量 ℓ_x 计算的未归一化密度。对于 SP，球面上的密度在北极处趋于无穷远；而对于 SCP，密度上下有界。

然而，现有分析要么侧重于连续时间动力学，要么侧重于所谓的未调整设定；后者会使目标分布产生偏差，而且在实际情形中难以量化。我们关注重尾目标分布下的 Metropolis 调整算法，而针对这类算法，大多数研究得出的结论都是否定的 [Roberts and Tweedie，1996；J arner 和 Hansen，2000；Livingstone等人， 2019], 尽管 Jarner and Roberts [2007] 指出，使用重尾MCMC提议分布在一定程度上可以缓解收敛速率缓慢且呈多项式衰减的问题。

1.5 概述与符号说明

本文其余部分安排如下。第2.1节详细介绍次柯西投影。第2.2节介绍我们提出的算法——次柯西投影采样器（SCS）。第2.3节证明所提出的次柯西分布采样器具有一致遍历性。第3节介绍如何自动校准所提方法的调节参数。第4节通过数值实验检验我们的算法在复杂重尾高维分布上的表现。复现数值结果的代码可在 https://github.com/liusf15/sub-cauchy-sampler 获取。更多技术性证明见附录A。

记号。 对于向量 $z \in \mathbb{R}^{d+1}$，令 $h_z := z_{1:d} \in \mathbb{R}^d$ 和 $\ell_z := z_{d+1} \in \mathbb{R}$ 分别表示 z的水平分量（经度）和垂直分量（纬度）。对于 $d \geq 1$和 $s \in \mathbb{R}^{d+1}$，令 $\mathcal{B}^{d+1}(s) := \{x \in \mathbb{R}^{d+1} : \|x - s\| \leq 1\}$, $\mathcal{S}^d(s) := \{x \in \mathbb{R}^{d+1} : \|x - s\| = 1\}$，以及 $\mathcal{B}^{d+1}(s) := \mathcal{B}^{d+1}(s) \setminus \mathcal{S}^d(s)$。对于固定的 $o \in \mathcal{B}^{d+1}(s)$，球面 $\mathcal{S}^d(s)$的亮侧定义为

$$\mathcal{C}_o^d(s) := \{x \in \mathcal{S}^d(s) : x_{d+1} < o_{d+1}\}.$$

当 $s = (0, 0, \ldots\ldots, 0)$，我们省略参数，直接写作 $\mathcal{B}^{d+1}$， $\mathcal{S}^d$， $\mathcal{C}_o^d$。对于所有 $1 \leq i \leq$

$d+1$， e_i 表示 $\mathbb{R}^{d+1}$中的第 i个标准基向量。

2 次柯西分布的采样

2.1 次柯西投影

设观测者为 $o \in \mathcal{B}^{d+1}(e_{d+1})$。 d 维的次柯西投影 SCP_o： $\mathcal{C}_o^d(e_{d+1}) \to \mathbb{R}^d$ 将任意点 $x \in \mathcal{C}_o^d(e_{d+1})$ 映射到 $y \in \mathbb{R}^d$，使得 x， o，并且 $\bar{y} = (y,\ 0) \in \mathbb{R}^{d+1}$ 共线，即

$$x = t\,o + (1-t)\,\bar{y}, \quad \text{for some } t \in [0,1].$$

通过进一步平移和重新缩放 y 为 $yR+\mu$，其中 $\mu \in \mathbb{R}^d$ 和 $R>0$，我们将投影推广到以（μ，R）为中心、半径为 R 的球面上。我们将该投影记为SCP_θ，其中θ =（o， μ， R）。次柯西投影、其逆变换及其雅可比行列式的显式表达式如下。推导见附录A.1。

命题1. 固定 $\theta = (o、\ \mu、\ R)$，其中 $o \in \mathcal{B}^{d+1}(e_{d+1})$ $\mu \in \mathbb{R}^d$，且 $R>0$。次柯西投影的形式为

$$y = \mathrm{SCP}_\theta(x) = R\left(\frac{\ell_o}{\ell_o-\ell_x}h_x - \frac{\ell_x}{\ell_o-\ell_x}h_o\right) + \mu, \qquad x \in \mathcal{C}_o^d(e_{d+1}),$$

其逆变换为

$$\begin{pmatrix} h_x \\ \ell_x \end{pmatrix} = \mathrm{SCP}_\theta^{-1}(y) = \begin{pmatrix} M\hat{y} + (1-M)h_o \\ (1-M)\ell_o \end{pmatrix}, \qquad \hat{y} = (y-\mu)/R, \qquad y \in \mathbb{R}^d,$$

其中

$$M = \frac{-\big(\langle \hat{y}-h_o, h_o\rangle - \ell_o(\ell_o-1)\big) + \sqrt{\big(\langle \hat{y}-h_o, h_o\rangle - \ell_o(\ell_o-1)\big)^2 - (\|\hat{y}-h_o\|^2+\ell_o^2)(\|h_o\|^2+\ell_o^2-2\ell_o)}}{\|\hat{y}-h_o\|^2+\ell_o^2}. \tag{1}$$

正向映射SCP_θ 的雅可比行列式为

$$J_\theta(y) = R^d\left(\frac{M\|\hat{y}-h_o\|^2 + \langle \hat{y}-h_o, h_o\rangle + \ell_o - \ell_o^2(1-M)}{M^d\ell_o}\right). \tag{2}$$

当观测者的纵向分量位于中心时，雅可比行列式会大大简化，即当 $h_o = (0,\ 0,\ \ldots\ldots,\ 0) \in \mathbb{R}^d$时，尤其是在以下两个特殊情形中。例 2 是将 $o \in \mathcal{B}^{d+1}(e_{d+1})$ 扩展到还包含边界点 $(0,\ 0,\ \ldots\ldots,\ 0,\ 2) \in \mathbb{R}^{d+1}$ 而得到的。

例 1. *(*柯西投影*)* 若 $h_o = (0,\ 0,\ \ldots\ldots,\ 0) \in \mathbb{R}^d$ 且 $\ell_o = 1$，则

$$J_\theta(y) = R^d(\|\hat{y}\|^2+1)^{(d+1)/2}, \quad where\ \hat{y} = (y-\mu)/R. \tag{3}$$

在这种情况下， $J_\theta(y)^{-1}$ 与标准 d 维柯西分布的密度成正比（当 $\mu = (0,\ 0,\ \ldots\ldots,\ 0)$ $R=1$ 时）。因此，SCP_θ的逆映射将 d 维柯西分布映射为月亮亮侧上的均匀分布。

例 2. （立体投影）如果 $h_o = (0,\ 0,\ \ldots,\ 0) \in \mathbb{R}^d$和 $\ell_o = 2$，那么我们就得到经典立体投影，其形式为

$$J_\theta(y) = R^d\left(\frac{\|\hat{y}\|^2+4}{4}\right)^d, \quad \hat{y} = (y-\mu)/R. \tag{4}$$

此处 $J_\theta(y)^{-1}$ 与自由度为 d 的标准多元 *Student* t 密度成正比（当 $\mu = (0、\ 0、\ \ldots、\ 0)$ $R = \sqrt{d/2}$ 时）。

注 1。 式 (3) 中的雅可比行列式的倒数比式 (4) 中的倒数衰减得更慢，慢了 $\|\hat{}y\|^{d-1}$ 倍，随着 $\|y\| \to \infty$。这使我们能够改进 *Yang* 等人，*2024*，定理 *2.1]* 对具有次柯西尾部的目标分布的一致遍历性结论。

2.2 次柯西投影采样器

这里详细介绍我们提出的次柯西投影采样器（SCS）。SCS 是一种作用于月球亮侧的随机游走型 Metropolis 算法。Metropolis 接受-拒绝机制确保链满足细致平衡条件，并且对 π 保持不变。在边界附近必须格外小心，因为落在"暗侧"（即观测者 o 上方）的提议点会被重新定位到"亮侧"，同时不破坏对称性和细致平衡。为简便起见，我们将算法定义在以 0 为中心的单位球面上。

对于给定点 $x \in \mathcal{S}^d$，首先通过在单位球面于 x 处的切空间上采样一个高斯随机变量，并将其投影回 $\mathcal{S}^d$，来提议一个新点 x'；参见算法 1。这个初始提议与 [Yang 等人，2024，算法 1] 中的 SPS 提议分布相同，使我们能够利用已有的理论结果。如果初始提议点 x' 落在亮侧，我们令 $x^\star = x'$。否则，若 x' 落在暗侧，即 $\ell_{x'} > \ell_o - 1$，我们通过以下 向外步进 函数将其带回亮侧

$$x^\star := \cos(K\alpha)x + \sin(K\alpha)u, \tag{5}$$

其中 $K := \arg\min_k \{k \in \mathbb{N} : k\alpha > \phi + \gamma\}$，$u := (x' - (\langle x, x'\rangle)x)/\sqrt{1 - (\langle x, x'\rangle)^2}$，$\alpha = \arccos(\langle x, x'\rangle)$，且 $\phi := \arccos(x_{d+1}/\sqrt{x_{d+1}^2 + u_{d+1}^2})$，$\gamma := \arccos(\ell_o - 1/\sqrt{x_{d+1}^2 + u_{d+1}^2})$；示意图见图 4。

请注意，步进扩展步骤不会带来额外的计算成本，因为式 (5) 有闭式表达式。全文中，我们令 $\ell_o \geq 1$，并在附录 A.4 中证明，对于合适的 ℓ_o 取值，随着维数增加，初始提议点 x' 以高概率位于亮侧，因此式 (5) 仅会偶尔被调用。随后以如下概率接受提议点

$$\alpha(y, y^\star) = \frac{J_\theta(y^\star)\pi(y^\star)}{J_\theta(y)\pi(y)} \wedge 1, \tag{6}$$

其中 y，$y^\star$ 分别是通过 SCP$_\theta$投影到欧几里得空间中的 x，$x^\star$。如果提议点被拒绝，则将新状态设为 x；转移的完整描述见算法 2。

由于提议在球面上是对称的，且公式 (5) 中的向外扩展函数相对于提议密度是对称的，因此 SCS 满足细致平衡条件，并且以下命题成立。证明见附录 A.2。

命题 2。 如果 $\pi(y)$ 在 $\mathbb{R}^d$中为正且连续，则 *SCS* 在 $\mathbb{R}^d$ 中产生一个具有不变分布 π的遍历马尔可夫链。

算法 1 单位球面上的随机游走提议

要求： 当前状态 $x \in \mathcal{S}^d$，步长 $h > 0$。

1: 采样 $\tilde{\delta} \sim \mathcal{N}(0, h^2 I_{d+1})$

2: 设置 $\delta = \tilde{\delta} - \langle x, \tilde{\delta}\rangle x$

3: **返回** $x' = \frac{x+\delta}{\|x+\delta\|}$

算法 2 次柯西投影采样器（SCS）转移（在单位球面上 e)

要求： 当前状态 $x \in \mathcal{S}^d$，调节参数 $\theta = (o,\ R,\ \mu)$，步长 $h >$ 0.
1: 使用算法 1 提议 $x' \in \mathcal{S}^d$ 。2: 如果 $\ell_{x'} > \ell_o - 1$ (x' 位于暗侧)，则按公式 (5) 设置 $x^\star$；否则设置 $x^\star = x'$。3: 按公式 (6) 计算 $\alpha(y,\ y^\star)$，其中 $y = \mathrm{SCP}_\theta($ $x + e_{d+1})$，$y^\star = \mathrm{SCP}_\theta(x^\star + e_{d+1})$4: 以概率 $\alpha(y,\ y^\star)$ 设置 $x^{\mathrm{new}} = x^\star$；否则设置 $x^{\mathrm{new}} = x$。5: **返回**下一状态 x^{new}。

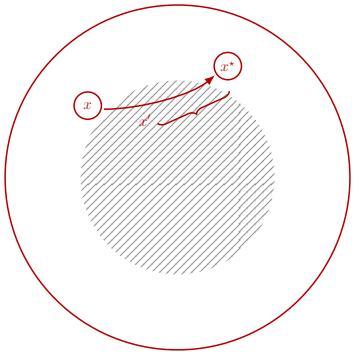

图 4：步进扩展机制的俯视图。阴影区域表示暗侧。从 x 开始，点 x' 被提议至暗侧，而向外扩展函数将其带回亮侧 $x^\star$。

2.3 一致遍历性

我们的主要结果表明，SCS 链收敛到次柯西目标分布 π 的速度是指数级的，且收敛速度不依赖于初始值。形式上，对于所有 $y \in \mathbb{R}^d$，若马尔可夫链满足以下条件，则称其一致遍历：

$$\|\mathcal{P}^n(y,\cdot) - \pi\| \le C\rho^n,$$

存在常数 $C > 0$、$0 < \rho < 1$，使得其中 $\|\cdot\|$ 表示全变差距离，而 $\mathcal{P}^n(y,\ \cdot)$表示以 y 为初始状态、经过 n 步后的马尔可夫链分布。一致遍历性是一种强（但仅具定性意义）的性质，在 Metropolis–Hastings 链中鲜少成立。

定理 1。若 $\pi(y)$ 在 $\mathbb{R}^d$上为正且连续，则对于任意固定的 $\ell_o \in [1, 2)$，*SCS* 一致遍历当且仅当

$$\sup_{y\in\mathbb{R}^d} \pi(y) J_\theta(y) < \infty,$$

其中 $J_\theta(y)$ 由式 (2)给出。

定理 1 表明，对于一大类重尾／轻尾分布，所提出的 SCS 都是一致遍历的；这类分布包括尾部不比（多元）柯西分布更重的分布；参见推论 1。据我们所知，SCS 是首个具有此类可证明收敛保证的通用型 Metropolis 调整的马尔可夫链蒙特卡罗方法。

推论 1 (SCS 对次柯西分布的一致遍历性)**.** 如果 $\pi(y)$ 在 $\mathbb{R}^d$ 中是次柯西、正且连续的，那么任意固定的 $\ell_o \in [1, 2)$ 对应的 *SCS* 都是一致遍历的。

证明。 根据定理 1，只需证明 $\sup_{y\in\mathbb{R}^d} \pi(y) J_\theta(y) < \infty$。由于

$$\sup_{y\in\mathbb{R}^d} \pi(y) J_\theta(y) \le \left[\sup_{y\in\mathbb{R}^d} \pi(y)(1 + \|y\|^2)^{(d+1)/2}\right] \cdot \left[\sup_{y\in\mathbb{R}^d} (1 + \|y\|^2)^{-(d+1)/2} J_\theta(y)\right],$$

且第一项由定义 1 可知是有限的，因此只需证明

$$\sup_{y\in\mathbb{R}^d}(1+\|y\|^2)^{-(d+1)/2}J_\theta(y)<\infty.$$

由于 R、μ 和 $\ell_o<2$ 均为常数，只需在公式（2）中代入 $J_\theta(y)$，利用 π 的连续性，并取 $\|y\|\to\infty$ 时的极限即可验证这一点。 □

备注 2. 推论 *1* 对 *Yang* 等人 *[2024,* 的 *SPS* 一致遍历性结果作出了重大改进。该结果仅适用于尾部比自由度为 $\nu\geq d$ 的 *Student* t 更轻的 d 维目标分布。特别是，当 *Student* t 满足 $\nu<d$ *[* 时，*SPS* 并非几何遍历（*Yang* 等人，*2024*年，定理 *2* 的证明*1][*；*Breˇsar* 等，*2025*年，第 *3.1* 节*3]*）；另见第 *1.3* 节中的数值比较。

3 调节参数

本节将介绍一种有理论依据的方法，用于在运行 MCMC 算法之前选择 SCP 的调节参数。给定观测者的纬度 $\ell_o\in[1,2)$，目标是¯选择参数 $\theta=(h_o$、μ、$R)$，使变换到亮侧的目标分布 π $(\mathcal{C}_o^d e_{d+1})$ 接近均匀分布。这是因为亮侧的密度均匀时，对于任意步长 $h>0$，算法 2 的接受概率都等于 1，从而实现快速混合并降低自相关。

本文中，纬度 ℓ_o 固定（关于简单几何论证所启发的指导，请参见附录 A.4），而 SCS 的步长 $h>0$ 经过调节，以达到某个平均经验接受概率 [（Gelman 等人，1996；Roberts 等人， 1997], ），从而使总迭代次数成为唯一由用户指定的参数。

备注 3. 用于调节马尔可夫链蒙特卡罗超参数的常见自适应方案 *[*（例如 *Gilks* 等人，*1998*；*Bell* 等人， *2024]*）基于对目标分布前两个矩的估计与匹配。对于重尾分布，这种策略可能会失效，因为这类分布不一定具有有限矩。

对于给定的 $\ell_o\in[1,2)$，令 $\bar{\theta}:=(h_o,\mu,R)\in\Theta$ 其中 $\Theta=\{h_o\in\mathbb{R}^d\colon\|h_o\|_2^2+(\ell_o-1)^2\leq 1\}\times\mathbb{R}^d\times\mathbb{R}^+$。令 $q_{\bar{\theta}}$ 表示 $y=\mathrm{SCP}_\theta(x)$的分布，其中 $x\sim\mathrm{Unif}(\mathcal{C}_o^d(e_{d+1}))$ 且¯ ¯ $\theta=(\ell_o,\theta)$。我们通过最小化 $q_{\bar{\theta}}$ 与π之间的 Kullback-Leibler（KL）散度来优化 θ，即 $\min_{\bar{\theta}\in\Theta}\mathrm{KL}(q_{\bar{\theta}}\|\pi)$。 这种方法属于变分推断的框架[（Blei 等人， 2017], ），其目标是用来自变分族¯的分布 $\theta\in\Theta\}$ 来近似目标分布 π 。

目标函数可表示为

$$\begin{aligned}\mathrm{KL}(q_{\bar{\theta}}\|\pi)&=\mathbb{E}_{q_{\bar{\theta}}}[\log q_{\bar{\theta}}(y)-\log\pi(y)]\\&=\mathbb{E}_{\mathrm{Unif}(\mathcal{C}_o^d(e_{d+1}))}[\log q_{\bar{\theta}}(\mathrm{SCP}_\theta(x))-\log\pi(\mathrm{SCP}_\theta(x))]\\&=\mathbb{E}_{\mathrm{Unif}(\mathcal{C}_o^d(e_{d+1}))}[-\log J_\theta(x)-\log\pi(\mathrm{SCP}_\theta(x))].\end{aligned}\tag{7}$$

目标函数中的期望可使用从 $\mathrm{Unif}(\mathcal{C}_o^d(e_{d+1}))$ 中抽取的蒙特卡洛样本进行近似。随后，通过自动微分计算梯度，并按照黑盒变分推断中的标准做法 [Ranganath et al., 2014]，使用随机梯度下降优化参数 $\bar{\theta}$ 。

如果目标分布 π 属于变分族，那么采用最优 θ 将 π 精确变换为亮侧上的均匀分布 $\mathcal{C}_o^d(e_{d+1})$ 。当 π 不属于变分族时，¯亮侧上的诱导密度不会完全均匀。不过，位置参数 μ 和尺度参数 R 使变分近似能够

捕捉目标分布的大部分，而观测者 o 则提供了额外的自由度，以适应偏度。尺度参数 R 也可以推广，使不同维数具有不同的尺度，不过本文不探讨这一扩展。我们在第4.1节中展示了变分近似性能的一个示例。

4 数值模拟

4.1 多元偏斜 t分布

这里我们考虑由 [Branco 和 Dey（2001）、Azzalini 和 Capitanio（ 2003]）引入的多元偏斜 t 分布。根据 Azzalini 和 Capitanio， [2003], a d 维偏斜t 分布的位置参数为 $\xi \in \mathbb{R}^d$、尺度矩阵为 $\Omega \in \mathbb{R}^{d\times d}$、偏度为 $\alpha \in \mathbb{R}^d$、自由度为 ν，可表示为 $\xi + V^{-1/2}Z$ 的分布，其中 $V \sim \chi^2_\nu/\nu$，且 Z 服从密度为

$$2\,\varphi_d(z;0,\Omega)\,\Phi(\alpha^\top \omega^{-1} z),$$

其中 $\omega =$ 为 diag($\Omega_{11}^{1/2},\ldots,\Omega_{dd}^{1/2}$)， Φ 表示标准正态累积分布函数， $\varphi_d(y;\ \mu,\Sigma)$ 表示均值为 μ 、方差为 Σ 的 d 维高斯密度，在 y 处取值。我们固定 $d = 100$、 $\xi = \mathbf{0}$、$\Omega = I_d$ 和 $\alpha =$(100, -100, 0, .. . ,0)。我们将 SCS（球面纬度为 $\ell_o = 1.1$、步长为 $h = 0.1$ ）与标准哈密顿蒙特卡洛采样器进行比较；后者每次迭代执行 10 个步长为 0.1 的蛙跳步。两个采样器均以 (1,. . ., 1) 为初始值，运行 500,000 次迭代，丢弃前 100 个样本，并每隔 50 个样本保留一个用于评估。SCS 和哈密顿蒙特卡洛的平均接受率分别约为 0.4 和 0.8。我们计算概率为 0.01、0.02、0.05、0.1、0.2、0.5、0.8、0.9、0.95、0.98 和 0.99 时的经验边缘分位数，并将其与相应的精确分位数进行比较；后者由直接从目标分布中抽取的 10^7 个独立样本估计。

图5展示了前四个坐标的 Q–Q 图，阴影区域表示10次独立重复实验之间的变异性。当 $\nu = 2$ （上方各图）时，两个采样器都能捕捉分布的主体部分，但 HMC 在极端尾部的准确性较低。当 $\nu = 1$（下方各图）时，对应于偏斜柯西分布，HMC 未能充分探索该分布，而 SCS 仍表现良好，分位数估计准确至第1和第99百分位数。值得注意的是，SCS 是无梯度的，而 HMC 每次迭代需要进行10次梯度计算，因此 SCS 的效率高得多。

为了评估第3节所述变分方法的表现，我们报告 h_o 与 α之间的余弦，以及 μ 与 ξ之间的相对差异。在目标分布中， α 和 ξ 分别控制偏度和位置参数。它们在变分族中的对应项是 $-h_o$ 和 μ，作用类似。因此，我们预期使用第3节的变分方法时， h_o 和 α 会趋于对齐，而 μ 会接近 ξ 。

在本实验中，我们固定 $d = 100$， $\nu = 2$。我们使用与之前相同的 α ，并将 ξ 的各元素设置为在 -10 和 10 之间等距分布。使用样本量为 2000年 的蒙特卡洛样本和学习率为 0.01 的 Adam优化器 [Kingma， 2014] 对 KL散度 方程（7）进行优化。如图6所示，随着梯度下降步数增加，cos(h_o， α) 趋近于 -1 ，而 $\|\mu - \xi\|_2/\|\xi\|_2$ 趋近于 0。

为进一步考察瞬态行为，图7展示了每种采样器各一条链的轨迹图；两条链均从 (500，... ...， 500) 冷启动，且 $d = 10$ 和 $\nu = 2$。即使经过 100,000 次迭代，HMC 仍未能到达典型集，并继续在重尾区域中游走，而 SCS 则迅速收敛至典型集。这表明，当链从尾部很远处初始化时，SCS 相较于 HMC 具有显著优势。

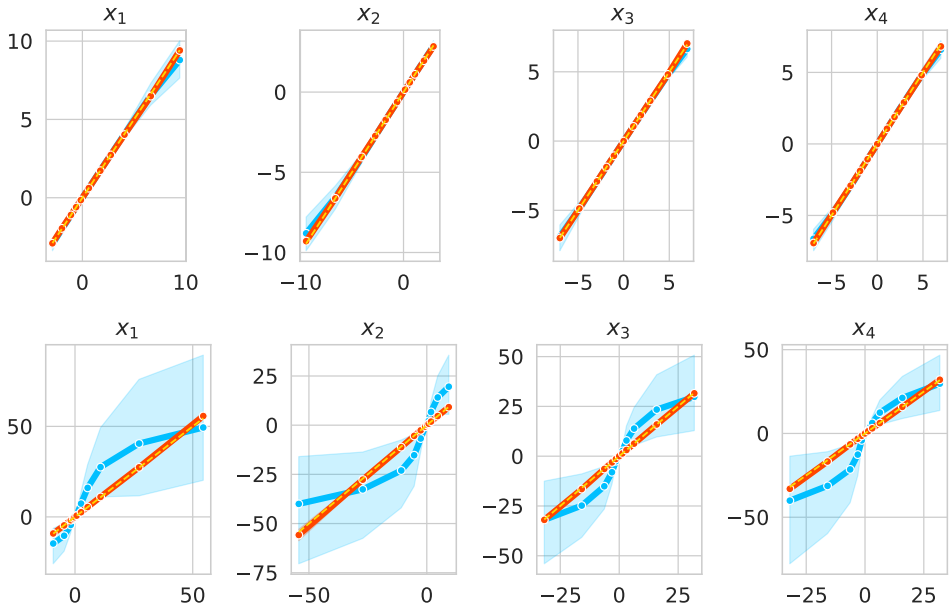

图5：在对一个 100 维多元偏斜 t分布进行抽样时，SCS（红色）和哈密顿蒙特卡洛（蓝色）的 Q–Q图，其中自由度为 2（上）或 1（下）。阴影区域显示了 10 次独立重复实验的变化。

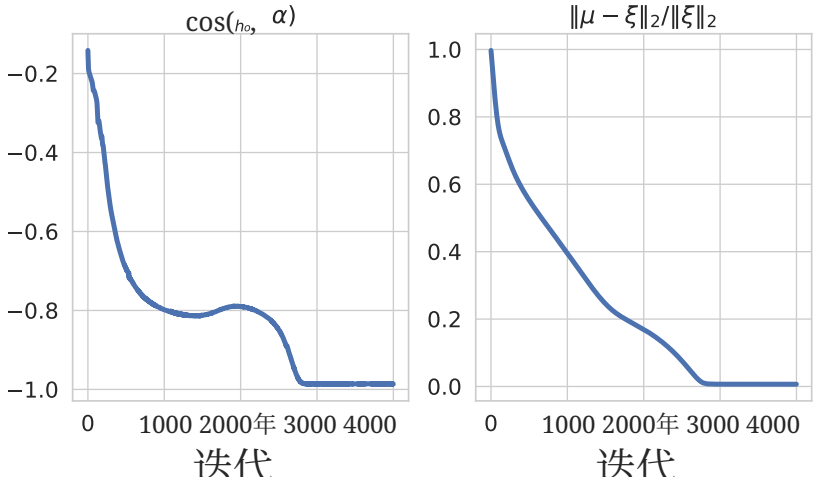

图6：优化 KL 散度 (7) 时，h_o 与 α 之间的余弦（左）以及 μ 与 ξ之间的相对平方距离（右）随梯度下降迭代次数的变化。

4.2 鲁棒贝叶斯二元回归

这里我们考虑贝叶斯二元回归问题，其中二元响应 $y_i \in \{0, 1\}$ 建模为

$$\mathbb{P}(y_i = 1 \mid x_i, \beta) = F(x_i^\top \beta), \quad 1 \le i \le n,$$

含有协变量 $x_i \in \mathbb{R}^d$ 和回归系数 $\beta \in \mathbb{R}^d$。当 $F(x) = \frac{1}{1+e^{-x}}$时，该表述可化为逻辑回归模型；当 $F = \Phi$是标准正态分布的累积分布函数时，则可化为 Probit模型。这两种模型都已知对离群值敏感，因此促使人们使用重尾链接函数，例如学生 t 分布[的累积分布函数[Albert and Chib, 1993, Liu, 2004], ，这被称为 robit回归。当 β 具有正态

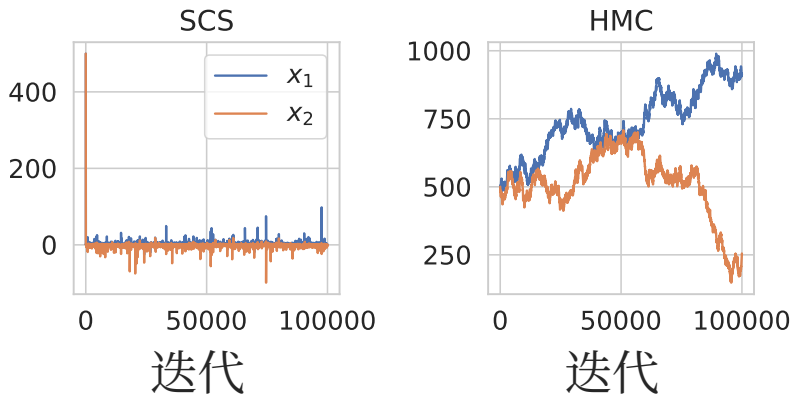

图 7：对一个具有 2 个自由度的 10 维多元偏态 t 分布进行采样时，SCS（左）和哈密顿蒙特卡洛（右）前两个坐标的轨迹图。

先验分布时，在适当条件下，已证明采用数据增广的吉布斯采样器（DA-Gibbs）具有几何遍历性 [Roy,2012,Mukherjee etal., 2023]。

Gelman 等人 [2008] 建议将独立 t分布作为逻辑回归及其他广义线性模型（例如多项逻辑回归和概率单位模型）中回归系数的默认先验分布。这类先验分布编码了较弱的先验信息，而且重要的是，当响应变量线性可分时，能够稳定估计值。最近，Ghosh 等人 [2018] 建立了柯西先验下后验均值存在的充要条件，并针对采用柯西先验的逻辑回归开发了一种 DA-Gibbs。

然而，Ghosh 等人 [2018] 报告称，在存在完全分离的情况下，采用重尾先验分布的 DA-Gibbs 混合极其缓慢。尽管他们承认，无转向采样器（NUTS； [Hoffman 等人， 2014]）等替代采样器可以改善混合，但他们指出：“柯西先验下后验分布具有极重尾这一棘手问题，若不改变先验分布便无法解决。”在附录 A.5 中，我们描述了数据分离情况下的 DA-Gibbs，并通过启发式论证说明链的混合缓慢。具体而言，我们证明，当 $n \to \infty$ 时， β 的后验标准差为 $O(\sqrt{n}\log\ n)$ 量级，而在众数附近，DA-Gibbs 的步长为 $O(\sqrt{\log n})$ 量级。因此，步长远小于后验分布的主体部分，导致自相关较高且混合缓慢。

我们展示了在数据存在完全分离时，所提出的 SCS 在贝叶斯逻辑回归中的表现。我们设定 $d = 20$ 和 $n = 50$。协变量 x_i 独立同分布地取自标准正态分布，响应变量设为 $y_i = \mathbf{1}\{x_{i,1} > 0\}$，使数据沿第一个协变量完全分离。我们在二元回归中同时考察 logit 链接函数和 robit 链接函数（自由度为 2 的学生 t 分布 t）。我们采用 Gelman 等人提出的默认先验分布 [2008], ：首先将协变量标准化，使其均值为 0、标准差为 0.5，然后赋予彼此独立的学生 t t 先验分布，尺度参数为 2.5。我们比较三种采样器：SCS（使用 $\ell_o = 1.1$）、采用数据增广的吉布斯采样器，以及标准 HMC。对于 HMC，每次迭代执行 10 个蛙跳步，并调整步长，使平均接受率的目标值为 0.8。所有方法均独立重复 20 次。每条链运行 100 次预热迭代。SCS 和 Gibbs 运行 5,000,000 次迭代，而 HMC 运行 1,000,000 次迭代，使实际运行时间大致处于同一量级。抽稀后，每条链保留 10,000 个样本用于计算分位数。参考分位数由 NUTS 得出，使用 20 条并行链，每条链运行 5,000,000 次迭代。

我们报告了三种算法对逻辑回归（图8）和 robit回归（图9）所采样的 β_1 的 Q–Q 图。阴影带表示每种方法的 20 条独立链之间的变异性。与 Ghosh 等人 [2018], 中的观察一致，Gibbs

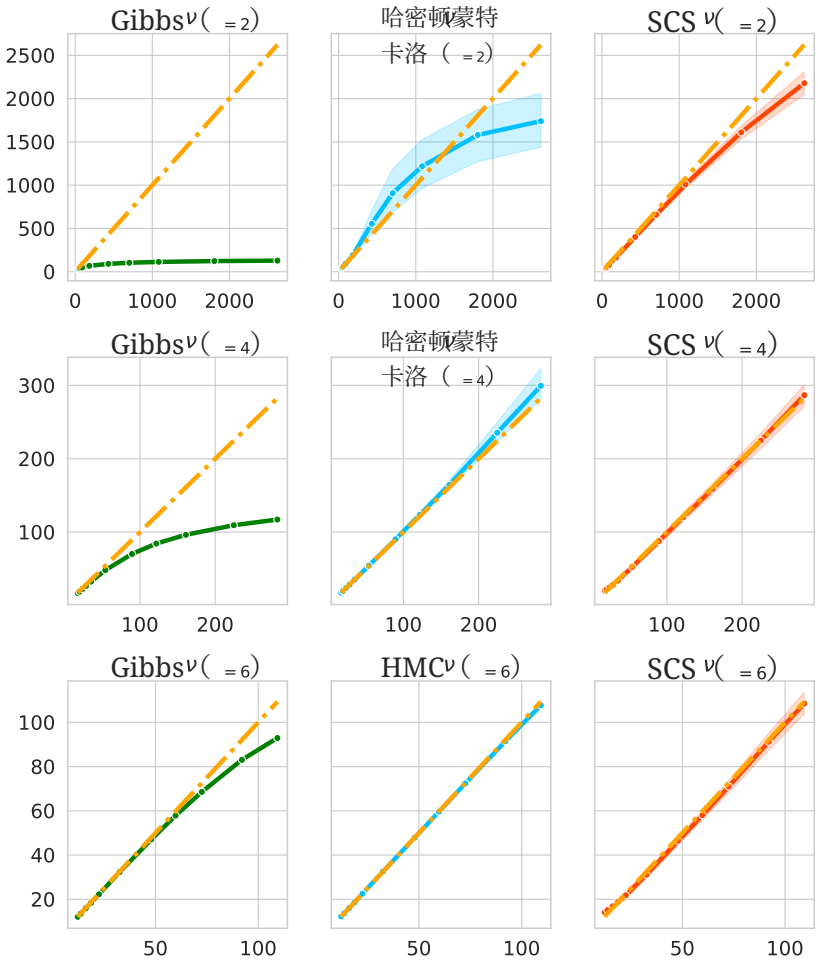

图8：逻辑回归中由吉布斯采样器（左侧列）、HMC（中间列）和 SCS（右侧列）得到的 β_1 的 Q–Q图。先验分布为自由度等于 2（上）、4（中）或 6（下）的 t 分布。

在先验分布具有重尾时表现不佳，但随着自由度增加，其性能会有所提升。当 $\nu = 4$ 或 6 时，哈密顿蒙特卡洛更加准确；但当 $\nu = 2$ 时，它难以在尾部进行采样。即使在极重尾情形下，所提出的 SCS 方法也表现得更加可靠。

5 讨论

我们提出了一种新的 Metropolis–Hastings算法——次柯西投影采样器，该算法对于次柯西目标分布是一致遍历的。投影的调节参数通过变分方法自动设定，因此只要怀疑后验分布具有重尾，所提出的 SCS 就具有广泛的适用性。我们的数值实验表明，在这些情形下，SCS 优于若干流行方法。最后，我们概述了若干尚待解决的问题以及未来研究方向。

1. 最优尺度与自适应马尔可夫链蒙特卡罗：我们提出了一种变分方法，用于选择在我们的各项模拟中表现良好的调节参数。然而，这一调节规则是否最优仍是一个悬而未决的问题。最优尺度理论[Roberts 等人 1997]提供了一种广受欢迎的框架，用于在 Metropolis–Hastings 算法中自适应地调整超参数。然而，在我们针对重尾目标分布和几何

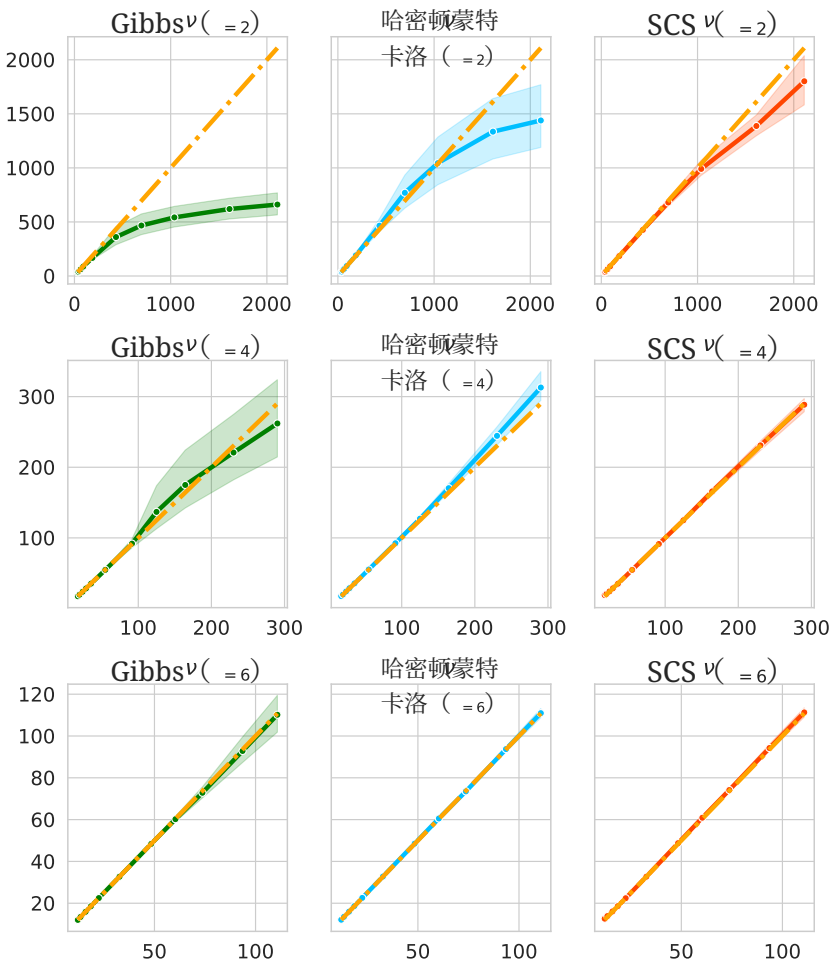

图9：robit回归中 β_1 的Q–Q图。图注同图8。

经过变换的 MCMC 算法，其尺度极限的推导颇具挑战性 [Yang 等人， 2020, 2024]。因此，我们将对此问题进行严谨处理留待未来工作。

2. 定量收敛结果：近年来，关于 MCMC 算法混合时间的定量界及其对维数的依赖，一直是理论研究的活跃领域 [Dwivedi 等人，2019年，Chen 等人，2020年，Yang 和 Rosenthal，2023年，Andrieu 等人，2024年，Milinanni 等人， 2026]。未来一项重要的研究方向，是为 SCS 在广泛的次柯西目标分布类别上建立快速混合保证。

3. 针对次柯西分布的基于梯度的 *MCMC*：SCS 是一种随机游走型算法，不利用目标分布的梯度信息。未来一项重要的研究方向，是将 SCS 与稳健的基于梯度的 MCMC 方法中的思想相结合 [Livingstone 和 Zanella，2022年，Bell 等人， 2026], ，或将其与纳入边界条件的 MCMC 框架相结合 [Bierkens 等人，2023a,b]。

致谢

作者感谢 Nicolas Chopin 就贝叶斯稳健回归提供了有益的讨论，也感谢 Wenkai Xu 在本研究早期提供了有益的讨论。SG 感谢欧盟（欧洲研究委员会）通过启动基金“PrSc-HDBayLe”（项目编号：101076564）提供的支持。JY 的研究得到丹麦独立研究基金（DFF）通过 Sapere Aude 启动基金（5251-00032B）提供的支持。GOR 得到 PINCODE（EP/X028119/1）、EP/V009478/1 以及英国研究与创新署（UKRI）资助的 OCEAN（EP/Y014650/1）的支持。

参考文献

J. H. Albert 和 S. Chib。二元及多类别响应数据的贝叶斯分析。《美国统计学会杂志》，88(422):669–679，1993年。C. Andrieu、A. Lee、S. Power 和 A. Q. Wang。Metropolis 马尔可夫链的显式收敛界：等周性、谱隙与剖面函数。《应用概率年刊》，34(4):4022–4071，2024年。A. Azzalini 和 A. Capitanio。由对称性扰动生成的分布，着重讨论多元偏斜 t 分布。英国皇家统计学会会刊 *B*辑：统计方法学，65(2):367–389，2003年。C. Bell、K. Łatuszyński 和 G. O. Roberts。自适应立体投影 MCMC。*arXiv* 预印本 *arXiv:2408.11780*，2024年。C. Bell、J. Yang、K. Łatuszyński、G. O. Roberts 和 J. S. Rosenthal。立体投影 Barker MCMC 提议：效率与稳健性，尽在掌握。准备中的手稿，2026年。A. Bertazzi 和 G. Vasdekis。使用时间变换马尔可夫过程进行采样。*arXiv* 预印本 *arXiv:2501.15155*，2025年。J. Bierkens、S. Grazzi、F. v. d. Meulen 和 M. Schauer。用于稀疏和局部推断问题的黏滞 PDMP 采样器。统计与计算，33(1):8，2023a。J. Bierkens、S. Grazzi、G. Roberts 和 M. Schauer。带边界条件的 PDMP 采样器：方法与应用，2023b。网址：https://arxiv.org/abs/2303.08023。

D. M. Blei、A. Kucukelbir 和 J. D. McAuliffe。变分推断：面向统计学家的综述。《美国统计学会杂志》，112(518):859–877，2017年。M. D. Branco 和 D. K. Dey。多元偏斜椭圆分布的一般类别。多元分析杂志，79(1):99–113，2001年。M. Bre˘sar、A. Mijatovi´c 和 G. Roberts。马尔可夫链遍历平均的中心极限定理，以及重尾分布采样算法的比较。*arXiv* 预印本 *arXiv:2512.18255*，2025年。Y. Chen、R. Dwivedi、M. J. Wainwright 和 B. Yu。Metropolis 化哈密顿蒙特卡罗的快速混合：多步梯度的优势。《机器学习研究杂志》，21(92):1–72，2020年。H. S. M. Coxeter。几何导论。1961年。S. De 和 J. Ghosh。针对线性回归模型重尾误差的稳健贝叶斯模型平均。应用统计学杂志，第1–27页，2025年。出版商：泰勒与弗朗西斯。B. De Finetti。用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，第1卷，第99–210页。加利福尼亚大学出版社，伯克利，加利福尼亚州，美国，1961年。R. Dwivedi、Y. Chen、M. J. Wainwright 和 B. Yu。对数凹采样：Metropolis-Hastings算法速度很快。《机器学习研究杂志》，20(183):1–42，2019年。P. Gagnon 和 Y. Hayashi。贝叶斯 Student-t 线性回归的理论性质。统计与概率快报，193:109693，2023年。A. Gelman、G. O. Roberts 和 W. R. Gilks。高效的 Metropolis 跳转规则。贝叶斯统计 *5*，5:599–608，1996年。A. Gelman、A. Jakulin、M. G. Pittau 和 Y.-S. Su。适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(4):1360，2008年。J. Ghosh、Y. Li 和 R. Mitra。柯西先验分布在贝叶斯逻辑回归中的应用。贝叶斯分析，13(2):359–383，2018。W. R. Gilks、G. O. Roberts 和 S. K. Sahu。通过再生实现自适应马尔可夫链蒙特卡罗。《美国统计学会杂志》，93(443):1045–1054，1998年。Y. He、K. Balasubramanian 和 M. A. Erdogdu。通过变换的未校正 Langevin 算法进行重尾采样。*arXiv* 预印本 *arXiv:2201.08349*，2022年。Y. He、T. Farghly、K. Balasubramanian 和 M. A. Erdogdu。重尾采样中离散化 Itô 扩散的均方分析。《机器学习研究杂志》，25(43):1–44，2024年。M. D. Hoffman、A. Gelman 等。无转向采样器：在哈密顿蒙特卡罗中自适应设定路径长度。《机器学习研究杂志》，15(1):1593–1623，2014年。S. F. Jarner 和 E. Hansen。Metropolis 算法的几何遍历性。《随机过程及其应用》，85(2):341–361，2000年。

S. F. Jarner 和 G. O. Roberts。重尾马尔可夫链蒙特卡洛算法的收敛性。《斯堪的纳维亚统计学杂志》，34(4):781–815，2007年。J. E. Johndrow、A. Smith、N. Pillai 和 D. B. Dunson。不平衡分类数据的马尔可夫链蒙特卡洛。《美国统计学会杂志》，2019年。L. T. Johnson 和 C. J. Geyer。通过变量变换使随机游走 Metropolis算法具有几何遍历性。《统计学年鉴》，第3050–3076页，2012年。D. P. Kingma。Adam：一种随机优化方法。*arXiv*预印本 *arXiv:1412.6980*，2014年。C. Liu。Robit回归：逻辑回归和概率单位回归的一种简单稳健替代方法。《从不完整数据视角进行应用贝叶斯建模与因果推断》，第227–238页，2004年。S. Livingstone 和 G. Zanella。Barker提议：兼顾稳健性与效率的基于梯度的马尔可夫链蒙特卡罗。《皇家统计学会杂志*B*辑：统计方法论》，84(2):496–523，2022年。S. Livingstone、M. Betancourt、S. Byrne 和 M. Girolami。关于哈密顿蒙特卡罗的几何遍历性。《伯努利》，25(4A):3109–3138，2019年。S. P. Meyn 和 R. L. Tweedie。《马尔可夫链与随机稳定性》。Springer科学与商业媒体，2012年。F. Milinanni、T. Farghly 和 J. Yang。用于重尾采样的球极投影MCMC快速混合。准备中的手稿，2026年。A. Mousavi-Hosseini、T. K. Farghly、Y. He、K. Balasubramanian 和 M. A. Erdogdu。迈向对朗之万蒙特卡罗的完整分析：超越 Poincar´e不等式。载于第三十六届学习理论年会，第1–35页。PMLR，2023年。S. Mukherjee、K. Khare 和 S. Chakraborty。高维robit回归的数据增强算法的收敛性质。《电子统计学杂志》，17(1):19–69，2023年。A. B. Owen。无限不平衡逻辑回归。《机器学习研究杂志》，8(4)，2007年。A. O’Hagan 和 L. Pericchi。贝叶斯重尾模型与冲突解决：综述。《巴西概率与统计学杂志》，26(4):372–401，2012年。doi：10.1214/ 11-BJPS164。URL https://doi.org/10.1214/11-BJPS164。N. G. Polson、J. G. Scott 和 J. Windle。利用 p´olya–Gamma 潜变量进行逻辑模型的贝叶斯推断。《美国统计学会杂志》，108(504):1339–1349，2013年。R. Ranganath、S. Gerrish 和 D. Blei。黑箱变分推断。载于《人工智能与统计学》，第814–822页。PMLR，2014年。G. O. Roberts 和 A. F. Smith。吉布斯采样器和 Metropolis-Hastings算法收敛的简单条件。《随机过程及其应用》，49(2):207–216，1994年。

G. O. Roberts 和 R. L. Tweedie。多维 Hastings 算法和 Metropolis算法的几何收敛与中心极限定理。*Biometrika*，83(1):95–110，1996年。G. O. Roberts、A. Gelman 和 W. R. Gilks。随机游走 Metropolis算法的弱收敛与最优尺度。《应用概率年刊》，7(1):110–120，1997年。V. Roy。稳健贝叶斯二元回归模型的马尔可夫链蒙特卡罗算法收敛速率。《电子统计学杂志》，6:2463–2485，2012年。L. Tierney。用于探索后验分布的马尔可夫链。《统计学年鉴》，22 (4):1701–1728，1994年。G. Vasdekis 和 G. O. Roberts。加速zig-zag。《应用概率年刊》，33(6A): 4693–4746，2023年。Z. Wang 和 J. Yang。用于重尾采样的立体投影多重尝试 Metropolis算法。*arXiv*预印本 *arXiv:2505.12487*，2025年。J. Yang 和 J. S. Rosenthal。基于定量界推导的马尔可夫链蒙特卡罗复杂度结果。《应用概率年刊》，33(2):1459–1500，2023年。J. Yang、G. O. Roberts 和 J. S. Rosenthal。一般目标分布下随机游走 Metropolis算法的最优尺度。《随机过程及其应用》，130 (10):6094–6132，2020年。J. Yang、K. Latuszy´nski 和 G. O. Roberts。立体投影马尔可夫链蒙特卡罗。《统计学年鉴》，52(6):2692–2713，2024年。

A 附录

A.1 次柯西投影与雅可比行列式

对于固定的观测者 $o=(h_o,\ \ell_o)\in\mathcal{B}^{d+1}(e_{d+1})\subset\mathbb{R}^{d+1}$，单位球面$x=(h_x,\ \ell_x)\in\mathcal{C}^{d+1}(e_{d+1})\subset\mathbb{R}^{d+1}$上的一点以及重缩放因子 $R>0$、 $\mu\in\mathbb{R}^d$，令 $\hat{}y=(y-\mu)/R$为平面上的重缩放点。正向映射

$$y=\mathrm{SCP}_\theta(x)=R\left(\frac{\ell_o}{\ell_o-\ell_x}h_x-\frac{\ell_x}{\ell_o-\ell_x}h_o\right)+\mu,\qquad h_x=x_{1_d},\,\ell_x=x_{d+1},$$

即为连接 o、 x 与 $(\hat{}y,\ 0)$ 的直线。为求逆映射，我们利用

$$\begin{pmatrix}h_x\\ \ell_x-1\end{pmatrix}=M\begin{pmatrix}\hat{y}\\ -1\end{pmatrix}+(1-M)\begin{pmatrix}h_o\\ \ell_o-1\end{pmatrix},$$

对于某个 $M\in[0,1]$。由于 $\|h_x\|^2+(\ell_x-1)^2=1$，对上式两边取平方范数，可得

$$1=M^2\left(\|\hat{y}-h_o\|^2+\ell_o^2\right)+\|h_o\|^2+(\ell_o-1)^2+2M\left(\langle\hat{y}-h_o,h_o\rangle-\ell_o(\ell_o-1)\right)$$

该方程有一个正解 M，可将其显式写成 y、 o、 ℓ_o和 R 的形式（见（1））。因此，逆映射 的形式为

$$\begin{pmatrix}h_x\\ \ell_x\end{pmatrix}=[\mathrm{SCP}_\theta^{-1}(y)]=\begin{pmatrix}M\hat{y}+(1-M)h_o\\ (1-M)\ell_o\end{pmatrix}.$$

我们来推导映射 $y=\pi_o(x)$ 的雅可比行列式。设 $\bar{y}=(\hat{}y,\ 0)$， θ_1 为 $\bar{y}-o$ 与 $x-e_{d+1}$ 之间的夹角， θ_2 为 $o-\bar{y}$ 与 e_{d+1}、 $a=\|x-o\|$ 与 $b=\|\bar{y}-x\|$ 之间的夹角。那么，雅可比行列式定义为

$$\begin{aligned}J(y)=R^d\left(\frac{a+b}{a}\right)^d\frac{cos(\theta_1)}{\cos(\theta_2)}&=\frac{R^d}{M^d}\frac{\langle\bar{y}-h_o,x-e_{d+1}\rangle}{\langle o-\bar{y},e_{d+1}\rangle}\\&=R^d\left(\frac{\langle\hat{y}-h_o,M\hat{y}+(1-M)h_o\rangle+\ell_o-\ell_o^2(1-M)}{M^d\ell_o}\right)\\&=R^d\left(\frac{M\|\hat{y}-h_o\|^2+\langle\hat{y}-h_o,h_o\rangle+\ell_o-\ell_o^2(1-M)}{M^d\ell_o}\right).\end{aligned}$$

对于 $h_o=0_d,\ \ell_o=2$ （例 2），我们有

$$M=\frac{4}{\|\hat{y}\|^2+4}$$

由此得到雅可比行列式为.....的立体投影

$$J_\theta(y)=R^d\frac{1}{M^d}=R^d\left(\frac{\|\hat{y}\|^2+4}{4}\right)^d.$$

对于 $h_o=0_d,\ \ell_o=1$ （例 1），我们得到柯西投影，其.....为

$$M=\frac{1}{\sqrt{\|y\|^2+1}}$$

和雅可比行列式

$$J_\theta(y)=R^d\,\frac{1}{M^d}=R^d\frac{\|\hat{y}\|^2+1}{M^{d-1}}=R^d(\|\hat{y}\|^2+1)^{(d+1)/2}.$$

A.2 命题 2 的证明

我们先介绍以下引理。

引理 1. 给定单位球面 $\mathcal{S}^{d+1} \subset \mathbb{R}^{d+1}$上的两个点 x 和 y，存在唯一一个同时经过 x 和 y的大圆，可写为

$$c_o(\theta) = \cos(\theta)x + \sin(\theta)u, \quad \theta \in [0, 2\pi),$$

其中 $u = \frac{y-(x^Ty)x}{\sqrt{1-(x^Ty)^2}}$。假设 x和 y之间的夹角不大于 $\pi/2$，其中 $x_{d+1} < \ell_o - 1$ 且 $y_{d+1} > \ell_o - 1$。此大圆与纬度 $\ell_o - 1$的两个交点为:

$$p_- := x\cos(\phi - \gamma) + u\sin(\phi - \gamma)$$

它位于该大圆上连接 x 和 y 的较短弧上（即球面线性插值 或球面线性插值），并且

$$p_+ := x\cos(\phi + \gamma) + u\sin(\phi + \gamma),$$

它位于较长的互补弧上，并且

$$\phi := \arccos\left(\frac{x_{d+1}}{\sqrt{x_{d+1}^2 + u_{d+1}^2}}\right), \quad \gamma := \arccos\left(\frac{\ell_o - 1}{\sqrt{x_{d+1}^2 + u_{d+1}^2}}\right).$$

证明。对于 $c_o(\theta)$，可验证 $c_o(0) = x$且 $c_o(\alpha) = x(x^Ty) + y - (x^Ty)x = y$，其中$\alpha :=$ $\arccos(x^Ty) < \pi/2$，根据我们的假设。此外，u 是单位向量，因为

$$u = \frac{y - (x^Ty)x}{\sqrt{1-(x^Ty)^2}} = \frac{y - (x^Ty)x}{\|y - (x^Ty)x\|}.$$

这两个交点应满足

$$c_o(\theta)_{d+1} = x_{d+1}\cos\theta + u_{d+1}\sin\theta = \ell_o - 1,$$

由此可得解 $\theta = \phi \pm \gamma$。最后，$\phi - \gamma$ 对应的点位于较短弧上。 □

根据引理 1，很容易验证 SCS 提议分布是对称的（示意见图 4），因此 SCS 满足详细平衡条件，这意味着平稳分布 π 的存在。根据 [Meyn 和 Tweedie，2012年，定理 13.0.1]，，若马尔可夫链是非周期的、 π-不可约且正 Harris 常返，则命题 2 成立。根据 [Roberts 和 Smith，1994年，定理 3(i)] ，利用接受概率不等于零这一事实，以及作为马尔可夫核的提议分布是非周期的这一性质，可知 SCS 是非周期的。此外，根据 πRoberts 和 Smith，1994年，定理 3(ii)] ，利用作为马尔可夫核的提议分布是 π-不可约这一事实，可知 π-不可约性成立。由于 π 是有限的，该链是正常返的 [Tierney，1994年，第1712]页。Harris 常返性直接来自 [Tierney，1994年，推论 2]。

A.3 定理 1 的证明

首先，如果 $\sup_{y\in\mathbb{R}^d} \pi(y)J(y) = \infty$，则 SCS 甚至不是几何遍历的（参见 [Roberts 和 Tweedie，1996，命题 5.1]）。这意味着定理 1 中的条件是必要的。因此，只需证明 $\sup_{y\in\mathbb{R}^d} \pi(y)J(y) < \infty$ 也是一个充分条件。

回顾一下， $\mathcal{S}^{d+1}$ 表示以原点为中心的单位球面，即 $\mathcal{S}^{d1} := \{z \in \mathbb{R}^{d+1} : \sum_{i=1}^{d+1} z_i^2 = 1\}$，并令 $\pi_S(z)$ 表示单位球面上的变换后的目标分布。注意，当

$\ell_o < 2$时，球面上有一部分未被使用，我们将其称为球面的“暗侧”，下文用 DS(0) 定义。我们还将定义大小为 δ 的“近暗侧”，记为 DS(δ)。当 $\ell_o = 2$ 时，我们对 DS(δ) 的定义就退化为 [Yang 等人，2024，定理 2 的证明1] 中“北极圈”的定义，记为 AC(δ)。

1. 我们将球面的“暗侧”定义为

$$\mathrm{DS}(0) := \{(x, \ell_x - 1) \in \mathcal{S}^{d+1} : \ell_x \geq \ell_o\}.$$

2. 类似地，我们可以定义“大小”为 $\delta \geq 0$ 的“近暗侧”，即

$$\mathrm{DS}(\delta) := \{(x, \ell_x - 1) \in \mathcal{S}^{d+1} : \ell_x \geq \ell_o - \delta\},$$

其中包含“大小”为 δ 的“亮侧”边界的一小块区域。

3. 与 [Yang 等人，2024年，定理 2 的证明.1], 相同，我们用 HS(z,0) := $\{z' \in \mathcal{S}^{d+1} : z^T z' \geq 0\}$ 表示来自 z 的“半球”。并用 HS(z, ϵ') := $\{z' \in \mathcal{S}^{d+1} : z^T z' \geq \epsilon'\}$ 表示将“大小”为 $\epsilon' \geq 0$ 的半球边界“切除”后剩下的区域。

证明的其余部分是对 [Yang 等人，2024年，定理 2 的证明.1], 的推广；该证明是针对 $\ell_o = 2$ 的特殊情形，而此处推广到任意给定的 $\ell_o \in [1, 2]$。我们通过证明球面“亮侧”的紧化是一个小集，来证明一致遍历性，并将其记为 DS^c(0)。更准确地说，我们将证明三步转移核满足以下小化条件：

$$P^3(z, A) \geq \epsilon \pi_S(A), \quad z \in \overline{\mathrm{DS}^c(0)}$$

对于所有可测集合 $A \subset \overline{\mathrm{DS}^c(0)}$，其中 $\epsilon > 0$。建立三步小化条件的理由在于，这有助于球面上的马尔可夫链到达球面“亮侧”的任意点，并且“停两步”路径周围存在足够的概率质量，确保链能够避免停在过于靠近“暗侧”边界的地方；这一区域可能会带来问题。

现在我们几乎完全遵循 [Yang 等人，2024年，定理 2 的证明。1]。考虑三步路径 $z \to z_1 \to z_2 \to z' \in A$，其中 $z \in \mathrm{DS}^c(0)$ 是起点，z_1 和 z_2 是球面上的两个中间点，而 $z' \in A$ 是终点。记 $q(z, \cdot)$ 为提议密度，则有

$$p(z, z') \geq q(z, z') \left(1 \wedge \frac{\pi_S(z')}{\pi_S(z)}\right).$$

那么，三步转移核可以下界为

$$\begin{aligned} P^3(z, A) \geq \int_{z' \in A} \int_{z_2 \in \mathcal{S}^d} \int_{z_1 \in \mathcal{S}^d} & q(z, z_1) q(z_1, z_2) q(z_2, z') \\ & \cdot \left(1 \wedge \frac{\pi_S(z_1)}{\pi_S(z)}\right) \left(1 \wedge \frac{\pi_S(z_2)}{\pi_S(z_1)}\right) \left(1 \wedge \frac{\pi_S(z')}{\pi_S(z_2)}\right) \nu(\mathrm{d}z_1) \nu(\mathrm{d}z_2) \nu(\mathrm{d}z') \end{aligned} \tag{8}$$

其中 $\nu(\cdot)$ 表示 $\mathcal{S}^{d+1}$ 上的勒贝格测度。

根据我们假设的条件，$\pi(y) > 0$ 在 $\mathbb{R}^d$中连续，这意味着 $\pi_S(z)$在球面 DS^c(0)的“亮侧”的任意紧子集上都为正且连续。然后，利用紧集上的连续函数有界这一事实，如果排除 DS(ϵ)，则 $\pi_S(z)$有一个严格大于 0 的下界。也就是说，

$$\inf_{z \notin \mathrm{DS}(\epsilon)} \pi_S(z) \geq \delta_\epsilon > 0, \forall \epsilon \in (0, 1),$$

其中 $\delta_\epsilon \to 0$ 如 $\epsilon \to 0$。

此外，由于 $\mathcal{S}^{d+1}$ 的表面积有限，且 $\pi_S(z) \propto \pi(y)J(y)$，我们知道，$\sup_{y\in\mathbb{R}^d} \pi(y)J(y) < \infty$ 意味着 $M := \sup \pi_S(z) < \infty$。因此，我们有

$$\begin{aligned} P^3(z,A) \geq \int_{z'\in A} \int_{z_2\in\mathcal{S}^d\setminus \mathrm{DS}(\epsilon)} \int_{z_1\in\mathcal{S}^d\setminus \mathrm{DS}(\epsilon)} q(z,z_1)q(z_1,z_2)q(z_2,z') \\ \cdot \left(1\wedge \frac{\delta_\epsilon}{M}\right)^2 \left(\frac{\pi_S(z')}{M}\right) \nu(\mathrm{d}z_1)\nu(\mathrm{d}z_2)\nu(\mathrm{d}z'). \end{aligned} \tag{9}$$

接下来，我们考虑这一项 $q(z,\ z_1)q(z_1,\ z_2)q(z_2,\ z')$。对于我们的算法，提议分布可以覆盖整个半球（边界除外）；稍微“切去”一部分边界后，提议密度就有下界。也就是说，我们有

$$q(z,z') \geq \delta'_{\epsilon'} > 0, \quad \forall z' \in \mathrm{HS}(z,\epsilon'), \epsilon' \in (0,1),$$

其中 $\delta'_{\epsilon'} \to 0$，且 $\epsilon' \to 0$。

因此，如果 $z_1 \in \mathrm{HS}(z,\ \epsilon')$ 且 $z_2 \in \mathrm{HS}(z',\ \epsilon')$，那么 $q(z,\ z_1) \geq \delta'_{\epsilon'}$，并且 $q(z_2,\ z') \geq \delta'_{\epsilon'}$。于是我们有

$$\begin{aligned} &\int_{z_2\in\mathrm{HS}(z',\epsilon')\setminus \mathrm{DS}(\epsilon)} \int_{z_1\in\mathrm{HS}(z,\epsilon')\setminus \mathrm{DS}(\epsilon)} q(z,z_1)q(z_1,z_2)q(z_2,z')\nu(\mathrm{d}z_1)\nu(\mathrm{d}z_2) \\ &\geq (\delta'_{\epsilon'})^3 \int_{z_2\in\mathrm{HS}(z',\epsilon')\setminus \mathrm{DS}(\epsilon)} \int_{z_1\in\mathrm{HS}(z,\epsilon')\setminus \mathrm{DS}(\epsilon)} \mathbf{1}_{z_1\in\mathrm{HS}(z_2,\epsilon')}\nu(\mathrm{d}z_1)\nu(\mathrm{d}z_2). \end{aligned} \tag{10}$$

注意，由于“亮侧”不小于一个半球，因此只需 3 步，就足以从“亮侧”上的任意 z 到达任意点 z'。此外，由于球面的“暗侧”不大于一个半球，显然在 $\mathcal{S}^{d+1}$ 上的勒贝格测度 ν 下，我们可以定义以下正常数 C:

$$C := \inf_{z,z'\in\mathrm{DS}^c(0)} \int_{z_2\in\mathrm{HS}(z',0)\setminus \mathrm{DS}(0)} \int_{z_1\in\mathrm{HS}(z,0)\setminus \mathrm{DS}(0)} \mathbf{1}_{z_1\in\mathrm{HS}(z_2,0)}\nu(\mathrm{d}z_1)\nu(\mathrm{d}z_2) > 0.$$

现在考虑以下关于 $\epsilon \geq 0$ 和 $\epsilon' \geq 0$的函数

$$C_{\epsilon,\epsilon'} := \inf_{z,z'\in\mathrm{DS}^c(0)} \int_{z_2\in\mathrm{HS}(z',\epsilon')\setminus \mathrm{DS}(\epsilon)} \int_{z_1\in\mathrm{HS}(z,\epsilon')\setminus \mathrm{DS}(\epsilon)} \mathbf{1}_{z_1\in\mathrm{HS}(z_2,\epsilon')}\nu(\mathrm{d}z_1)\nu(\mathrm{d}z_2).$$

显然，$C \geq C_{\epsilon,\ \epsilon'}$。它们之间的差异可由以下界定

$$\begin{aligned} C - C_{\epsilon,\epsilon'} \leq &\sup_{z\in\mathrm{DS}^c(0)} \nu(\{z_1 : z_1 \in (\mathrm{HS}(z,0)\setminus \mathrm{HS}(z,\epsilon')) \cap (\mathrm{DS}(\epsilon)\setminus \mathrm{DS}(0))\}) \\ &+ \sup_{z'\in\mathrm{DS}^c(0)} \nu(\{z_2 : z_2 \in (\mathrm{HS}(z',0)\setminus \mathrm{HS}(z',\epsilon')) \cap (\mathrm{DS}(\epsilon)\setminus \mathrm{DS}(0))\}) \\ &+ \sup_{z_2\in\mathrm{DS}^c(0)} \nu(\{z_1 : z_1 \in \mathrm{HS}(z_2,0)\setminus \mathrm{HS}(z_2,\epsilon')\}). \end{aligned} \tag{11}$$

那么显然，上界关于 ϵ 和 ϵ' 连续，并且当 $\epsilon \to 0$和 $\epsilon' \to 0$时趋于零。因此，存在 $\epsilon > 0$ 和 $\epsilon' > 0$ 使得 $C - C_{\epsilon,\ \epsilon'} \leq \frac{1}{2}C$，即

$$C_{\epsilon,\epsilon'} \geq \frac{1}{2}C > 0.$$

使用这样的 ϵ和 ϵ'，我们有

$$\int_{z_2\in\mathrm{HS}(z',\epsilon')\setminus \mathrm{DS}(\epsilon)} \int_{z_1\in\mathrm{HS}(z,\epsilon')\setminus \mathrm{DS}(\epsilon)} q(z,z_1)q(z_1,z_2)q(z_2,z')\nu(\mathrm{d}z_1)\nu(\mathrm{d}z_2) \geq \frac{1}{2}C(\delta'_{\epsilon'})^3 > 0.$$

于是我们有

$$P^3(z,A) \geq \left(1 \wedge \frac{\delta_\epsilon}{M}\right)^2 \frac{1}{2} C(\delta'_{\epsilon'})^3 \int_{z' \in A} \left(\frac{\pi_S(z')}{M}\right) \nu(\mathrm{d}z')$$

注意，$\pi_S(z') > 0$ 关于 ν 几乎处处成立，当且仅当 $z' \in \mathrm{DS}^c(0)$。我们定义

$$g(z') := \begin{cases} \left(1 \wedge \frac{\delta_\epsilon}{M}\right)^2 \frac{1}{2} C(\delta'_{\epsilon'})^3 \frac{\pi_S(z')}{M}, & \forall z' \in \mathrm{DS}^c(0) \\ 0, & \forall z' \in \mathrm{DS}(0). \end{cases}$$

注意，$g(z) \propto \pi_S(z)$，因此，我们得出

$$P^3(z,A) \geq \int_{z' \in A} g(z') \nu(\mathrm{d}z') \geq \epsilon \pi_S(A),$$

其中 $\epsilon = \left(1 \wedge \frac{\delta_\epsilon}{M}\right)^2 \frac{1}{2} C(\delta'_{\epsilon'})^3 \frac{1}{M} > 0$。这证明了所需的小化条件，并完成了证明。

A.4 纬度的选择 ℓ_o

第3节中介绍的变分方法适用于固定纬度 $\ell_o \in [1,2)$，因为亮侧上密度的支集取决于该参数。我们对 ℓ_o 的选择受到高维超球面简单几何性质的指导。对于 $o \in \mathcal{B}^{d+1}(e_{d+1})$，令

$$\mathcal{D}_o^{d+1} = \mathcal{S}^{d+1}(e_{d+1}) \setminus \mathcal{C}_o^{d+1} = \{x \in \mathbb{R}^{d+1} \colon \|h_x\|^2 + (\ell_x - 1)^2 = 1; \ell_x \geq \ell_o\}$$

为球面的暗侧，即 SCP 中被移除的部分。以下结果表明，当 $\ell_o \geq 1 + \mathcal{O}(\log d/\sqrt{d})$时，被移除的表面区域相对于球面而言，其表面积在渐近意义下可以忽略不计。

命题 3。对于任意 $d \geq 1$ 和 $\ell_o \in (1,2)$,

$$\frac{|\mathcal{D}_o^{d+1}|}{|\mathcal{S}^{d+1}|} \leq \frac{e^{1/2}}{2} \sqrt{d+1} \left[1 - (\ell_o - 1)^2\right]^{\frac{d-1}{2}}, \tag{12}$$

其中 $|\cdot|$ 是 d 维表面积。

证明。设 X 和 Y 分别是相互独立的卡方随机变量，其 d 和 1 个自由度。利用表面积之比可写成角度之比这一事实，我们关注的概率为

$$\frac{\text{surface area of “dark side”}}{\text{surface of the sphere}} = \mathbb{P}\left(X \leq \frac{1 - (\ell_o - 1)^2}{(\ell_o - 1)^2} Y\right).$$

我们先从 $\mathbb{P}(X \leq cY)$ 的精确公式入手，其中 c为某个常数。利用 F-分布，可得 $\frac{X/d}{Y/1} \sim F_{d,1}$ 这意味着

$$\mathbb{P}(X < cY) = \mathbb{P}\left(\frac{X}{Y} < c\right) = \mathbb{P}\left(d\, F_{d,1} < c\right) = \mathbb{P}\left(F_{d,1} < \frac{c}{d}\right) = F_F\left(\frac{c}{d}; d, 1\right),$$

其中 $F_F(x; d, 1)$ 表示具有 $(d,1)$ 个自由度的 F-分布的累积分布函数。 F-分布的累积分布函数可用正则化不完全贝塔函数表示:

$$F_F(x; d, 1) = I_{\frac{d\,x}{d\,x+1}}\left(\frac{d}{2}, \frac{1}{2}\right),$$

关于 $I_z(a, b) = B(z; a, b) \int_0^z t^{a-1}(1-t)^{b-1}\, dt$。代入 $x = \frac{c}{d}$即可得到精确公式:

$$\mathbb{P}(X < cY) = I_{\frac{d\,(c/d)}{d\,(c/d)+1}}\left(\frac{d}{2}, \frac{1}{2}\right) = I_{\frac{c}{c+1}}\left(\frac{d}{2}, \frac{1}{2}\right).$$

接下来，我们为足够大的维数 d推导一个上界。根据马尔可夫不等式，对于任意满足矩母函数存在条件的$t>0$，

$$\mathbb{P}(X<cY)=\mathbb{P}\left(e^{-tX}\geq e^{-t\,cY}\right)\leq\mathbb{E}\left[e^{-tX+t\,cY}\right]=M_X(-t)\,M_Y(tc).$$

对于 $Z\sim\chi^2_k$，矩母函数为

$$M_Z(u)=(1-2u)^{-k/2},\quad u<\tfrac{1}{2}.$$

因此，对于 $0<t<1/(2c)$，

$$\mathbb{P}(X<cY)\leq(1+2t)^{-d/2}\,(1-2c\,t)^{-1/2}.$$

选择 $t^*=\frac{d-c}{2c(d+1)}$ 可得到此形式下最紧的界

$$\mathbb{P}(X<cY)\leq\left(1+\frac{1}{c}\frac{d-c}{d+1}\right)^{-d/2}\cdot\left(1-\frac{d-c}{d+1}\right)^{-1/2}=\sqrt{\frac{d+1}{1+c}}\left[\frac{d+1}{d}\cdot\frac{c}{c+1}\right]^{d/2}.$$

利用 $(1+1/d)^d<e$，可得

$$\mathbb{P}(X<cY)\leq\sqrt{\frac{d+1}{c+1}}e^{1/2}\left(\frac{c}{c+1}\right)^{d/2}=\sqrt{d+1}e^{1/2}\sqrt{\frac{1}{c+1}\left(1-\frac{1}{c+1}\right)}\left(\frac{c}{c+1}\right)^{\frac{d-1}{2}}.$$

由于 $\sqrt{\frac{1}{c+1}\left(1-\frac{1}{c+1}\right)}\leq\frac{1}{2}$，可得

$$\mathbb{P}(X<cY)\leq\frac{e^{1/2}}{2}\sqrt{d+1}\left(\frac{c}{c+1}\right)^{\frac{d-1}{2}}.$$

最后，代入 $c=\frac{1-(\ell_o-1)^2}{(\ell_o-1)^2}$，即可得到所需结果。 □

命题 3 表明，通过限制 $\ell_o\in(1+\frac{c}{\sqrt{d}},\ 2)$，随着维数增加，$c\sim\log d$比值 (12) 会迅速缩小，使 SPS 提议分布落在暗侧的可能性越来越低。然而，对于柯西分布，例 1 表明，最优选择对应于 $\ell_o=1$。在实践中，我们发现 $\ell_o=1.1$ 在所有模拟中都是一个合理的选择，因此将该参数的精细调优留待未来研究。

A.5 DA-Gibbs 的慢混合

对于贝叶斯逻辑回归，人们已经注意到，当两个类别不平衡时，DA-Gibbs 的表现也很差。Johndrow 等人 [2019] 研究了 Owen [2007] 引入的无限不平衡渐近状态下 DA-Gibbs 的混合行为。Johndrow 等人 [2019] 认为，混合缓慢是由于后验分布的宽度与 Gibbs 的步长之间存在差异（随着样本量 $n\to\infty$ 增加，步长以快于后验分布集中速度的速率趋近于 0）。当两个类别完全可分时，我们也观察到了类似现象，这解释了 DA-Gibbs 在 Ghosh 等人 [2018]的实证结果中表现不佳的原因。

考虑一个自由度为 ν 的采用 Student t先验分布的一维逻辑回归

$$\mathbb{P}(Y_i=1\mid\beta)=\sigma(x_i\beta),\qquad\sigma(t)=\frac{1}{1+e^{-t}}$$

并假设以下数据是可分的

$$(x_i,\,y_i)=\begin{cases}(n^{-1/2},1) & i\leq n/2\\(-n^{-1/2},0) & i>n/2\end{cases}$$

对于 $i = 1$，2、.....、n。在吉布斯采样器中，学生 t 先验分布 $t_\nu(0, \tau^2)$ 表示为高斯混合尺度分布：

$$\beta \mid \lambda \sim N\big(0, \frac{\tau^2}{\lambda}\big), \quad \lambda \sim \mathrm{Gamma}\big(\frac{\nu}{2}, \frac{\nu}{2}\big),$$

并使用 P´olya-Gamma 潜变量实现逻辑斯谛链接函数 [Polson 等人， 2013]。

定义 重缩放后的系数 $\eta = \beta/\sqrt{n}$。在 η 参数化下，对数似然为

$$\log p(y \mid \eta) = \frac{n}{2} \log \sigma(\eta) + \frac{n}{2} \log(1 - \sigma(-\eta)) = n \log \sigma(\eta),$$

其关于 η 的导数为

$$\frac{\partial}{\partial \eta} \log p(y \mid \eta) = n\sigma(-\eta).$$

β 中的 Student t 先验会在 η 中诱导出以下先验：

$$\pi_\eta(\eta) \propto \left(1 + \frac{n\eta^2}{\nu\tau^2}\right)^{-(\nu+1)/2}.$$

此外，我们有

$$\frac{\partial}{\partial \eta} \log \pi_\eta(\eta) = -\frac{(\nu+1)n\eta}{\nu\tau^2 + n\eta^2}.$$

结合似然和先验分布，对数后验的导数为

$$\frac{\partial}{\partial \eta} \log p(\eta \mid y) = n\sigma(-\eta) - \frac{(\nu+1)n\eta}{\nu\tau^2 + n\eta^2}.$$

现在，我们研究当 n 较大时，β 中后验标准差的尺度以及步长的尺度。

- **后验宽度的尺度**。我们首先指出，当 n 较大时， η 中的后验标准差为 $O(\sqrt{\log\ n})$ 量级。当 η 较大时， $\sigma(-\eta) \sim e^{-\eta}$，且 $\frac{\partial}{\partial \eta} \log \pi_\eta(\eta) \sim -\frac{\nu+1}{\eta}$，因此众数 $\hat{\eta}_n$ 满足

$$n e^{-\hat{\eta}_n} \sim \frac{\nu+1}{\hat{\eta}_n} \quad \Longrightarrow \quad \hat{\eta}_n \sim \log n.$$

在众数处再次求导，得到

$$-\frac{\partial^2}{\partial \eta^2} \log p(\eta \mid y) \sim n\sigma(-\hat{\eta}_n) + \frac{\nu+1}{\hat{\eta}_n^2} \sim \frac{\nu+1}{\hat{\eta}_n} = O(\frac{1}{\log n}).$$

因此， η 中的后验标准差约为 $O(\sqrt{\log n})$。用 β 表示时，后验标准差为 $O(\sqrt{n \log\ n})$ 量级。

- **步长的尺度**。在 P´olya-Gamma 吉布斯采样器中， [Polson 等人， 2013]，的 P´olya-Gamma 变量 ω_i 具有条件分布

$$\omega_i \mid \eta \sim \mathrm{PG}(1, x_i\beta) = \mathrm{PG}(1, \pm\eta),$$

均值为

$$\mathbb{E}[\omega_i \mid \eta] = \frac{1}{2|\eta|} \tanh\left(\frac{|\eta|}{2}\right) \sim \frac{1}{2|\eta|} \quad (|\eta| \to \infty).$$

此外，给定 P´olya-Gamma 变量后， β 的条件方差为

$$\mathrm{Var}\left[\beta \mid \omega\right] = \left(X^\top \Omega X + \tau^{-2}\lambda\right)^{-1}.$$

注意，$X^\top \Omega X = \sum_i \omega_i x_i^2 = \frac{1}{n}\sum_i \omega_i$. 在 η 的众数附近，我们有 $\mathbb{E}[\omega_i \mid \eta] \sim \frac{1}{2}$ 对数 n。因此 $X^\top \Omega X \sim \frac{1}{2}$ 对数 n。由于 $\lambda \mid \beta \sim$ 伽马分布$\left(\frac{\nu+1}{2}, \frac{\nu+\beta^2/\tau^2}{2}\right)$，我们有 $\mathbb{E}[\lambda \mid \beta] = \frac{\nu+1}{\nu+\beta^2/\tau^2}$. 在后验众数附近，$\beta^2 \sim n$(对数 n)^2, 因此 $\lambda/\tau^2 \sim \frac{1}{n}$(对数 n)^2 相较于 $X^\intercal \Omega X$可以忽略。所以我们有

$$\mathrm{Var}\left[\beta \mid \omega\right] \sim (X^\intercal \Omega X)^{-1} \sim \log n,$$

这意味着 β 中的步长为 $O(\sqrt{\log n})$ 量级。

总之，β中的后验宽度为 $O(\sqrt{n \log n})$，但β中的 Pólya-Gamma Gibbs 步长为 $O(\sqrt{\log n})$。步长比后验宽度小 $\sqrt{n}$ 倍，导致 Gibb 采样器混合缓慢。