arXiv:2304.01599v1 [stat.ME] 4 Apr 2023

从弯曲环面曲面抽样：新的创世纪

Buddhananda Banerjee
b
banerjee@maths.iitkgp.ac.in
和 Surojit Biswas
surojit23@iitkgp.ac.in

印度理工学院克勒格布尔分校数学系，印度-
721302

Abstract. 环面数据的分布通常被视为圆周分布的扩展，但并未考虑弯曲环面的内蕴几何。Diaconis等人(2013年)[*Diaconis, P., Holmes, S., & Shahshahani, M. (2013).* 流形上的抽样。《现代统计理论与应用进展：纪念 *Morris L. Eaton* 文集》，*10*，*102-125*。] 首次提出了相对于表面积而言，弯曲环面曲面上的均匀分布。然而，所提出的从该分布中进行抽样的接受-拒绝法会拒绝大约一半的数据。我们提出一种概率变换，从同一分布中抽样而不丢失数据。此外，我们采用基于直方图的接受-拒绝抽样，并使用非常薄的包络函数，为若干常见圆周分布中的随机样本开辟了新的生成方式。这一思路可推广至弯曲环面曲面上的分布抽样，并获得较高的接受率。除了降低不同环面分布推断研究中的计算成本外，从弯曲环面的曲面进行均匀抽样，也有助于理解其上的任意未知分布。

关键词：均匀分布；环面；圆形分布；包络；接受-拒绝抽样。

Contents

1. 引言 32. 背景 52.1. 环面的内蕴几何 73. 环面的精确面积均匀抽样（EAU） 83.1. 模拟分析 94. 直方图-接受-拒绝抽样（HAR） 124.1. HAR抽样的所提方法 144.2. 冯·米塞斯分布模拟 165. 推广至环面表面抽样 175.1. 模拟分析 196. 结论 297. 附录 307.1.A1
307.2.A2
317.3.A3
32参考文献 33

1. Introduction

对曲面或一般流形上的数据进行分析，取决于其生成方式和表示形式。这一过程始于从流形上的某个分布中抽样，随后利用这些数据进行统计推断，例如参数估计、假设检验和预测等。在受约束参数空间（如协方差矩阵）上的后验分布中抽取样本、基于充分统计量检验指数族的拟合优度，以及生成用于评估拓扑统计学算法的数据，都是需要从流形上抽样的情形。

由于高维相关数据日益普遍，流形上的统计推断如今受到越来越多的关注。Bhattacharya 和 Patrangenaru（2003）以及 Pennec（2006）探讨了流形上均值和方差估计量的发展。Beran（1979）、Watson（1983）和 Fisher 等（1993）研究了射影空间和球面上的数据。在物理学和化学中，常见一种涉及原子构型的情形，其中角度或原子间距离固定；参见 Fixman（1974）以及 Ciccotti 和 Ryckaert（1986）。上述研究都需要从流形进行蒙特卡罗抽样。文献中可以找到许多成熟的算法，用于从齐性空间和紧群的均匀分布中抽样。例如，如果一个 $n \times n$ 随机矩阵的所有元素都服从标准正态分布，并对其进行 QR 分解，那么 Q 部分在正交群上服从均匀分布（见 Eaton，1983）。Diaconis 和 Shahshahani（1986）研究了紧群上均匀分布的平方根。已有一些优秀算法可用于从 $\mathbb{R}^n$ 凸紧集的边界抽样（参见 B´elisle 等，1993；Boender 等，1991；Lalley 和 Robbins，1987）。对于更一般的流形，Diaconis 等人（2013）开发了一种算法，使得能够从嵌入于 $\mathbb{R}^n$中的子流形上的概率分布进行抽样。他们展示了如何将这些技术用于多种场景，例如评估拓扑统计学算法、对指数族进行拟合优度检验，以及实施 Neyman 光滑度检验。

天体物理学数据示例：活动星系核（AGN）是位于星系核心的致密区域（称为活动星系），是气体落入超大质量黑洞（SBH）的结果。因此，由于角动量守恒，气体形成了以SBH为中心的环面结构。AGN的一个重要特征是它经常表现出显著的红外辐射，据认为这种辐射是由

源自包围中心黑洞的尘埃环面。在统一模型的背景下，环面相对于地球上的观测者具有不同朝向，这解释了活动星系核的观测特征。长期以来，SBH周围的尘埃环面一直被视为理解活动星系核理论的重要组成部分。然而，直到Carilli 等人(2019)利用美国国家科学基金会的Karl G. Jansky甚大天线阵（VLA），首次获得了强射电星系天鹅座A中活动星系核（AGN）厚环面的直接图像，此类数据在适当的物理尺度上才得以直接成像。因此，对这类环面数据进行均匀抽样，对于理解活动星系核的基本行为至关重要。

生物化学数据示例： 蛋白质折叠是蛋白质分子链转变为最终三维形状的物理过程。蛋白质的正确折叠对其功能至关重要，因为折叠不当可能导致多种疾病，例如神经退行性疾病。蛋白质折叠的识别和预测是生物化学或计算生物学领域中一个重大的未解决问题。有关更全面的讨论，（参见 Selkoe， 2003）。由于蛋白质折叠所涉及的物理过程性质复杂，因此从蛋白质的氨基酸序列准确预测其最终三维构象所取得的成果有限。对蛋白质折叠的深入理解无疑将带来临床益处，例如开发成功的药物化合物，以治疗包括前述疾病在内的各种疾病。蛋白质结构的某些部分可能呈现无定形状态，因此需要使用随机模型和概率分布对其进行精确表征。 Hin- gorani 和 O'Donnell (1998) 以及 Hingorani 和 O'Donnell (2000) 指出，实际上，大量具有不同进化起源、参与 DNA 代谢的蛋白质都呈现环面形状。具有这种环面结构的蛋白质数量相对较多。由于蛋白质折叠问题和 DNA结合过程的重要性，统计学家开始关注开发能够利用环面数据准确描述这些复杂现象的统计模型。

本文主要关注如何从弯曲环面曲面上的不同分布中抽取随机样本。在一项开创性研究中，Diaconis 等人 (2013年) 提出了一种从此类曲面获取均匀随机样本的方法。他们所说的均匀，是指获取随机样本的频率与环面曲面上的局部面积成比例。最后，他们通过接受-拒绝抽样实现了这一构想。然而，所提出的抽样方案的缺点是，对于其中一种边缘密度，拒绝率高得惊人，约为 50%。我们提出了一种概率变换，可便于抽取随机样本

从目标边缘分布中抽样，无需对数据进行任何拒绝。为了便于理解，曲面上任何未知分布本质上都需要从该曲面获得一个均匀随机样本。所提方法将使我们能够以很高的效率完成同样的工作。此外，我们还利用基于直方图的接受-拒绝抽样，为一些常见的圆周分布生成随机样本。这一思想的灵感来源于积分的上黎曼和，它能够在圆上为目标密度提供一个非常紧的包络函数。结合弯曲环面的内蕴几何，这一思想可以推广到从弯曲环面的曲面上的不同分布中抽取随机样本，并获得较高的接受率。所提方法从圆上或环面曲面上的分布中生成数据，不仅为概率研究增添了优势，而且还加快了用于推断用途的估计量和检验统计量的模拟研究。

本文组织如下。第2 节首先介绍环面上的几何测度论基础，随后介绍其内蕴几何。在第3节中，介绍了精确面积均匀（EAU）抽样方法。与 Diaconis Diaconis 等 (2013年）提出的方法相比，该方法得到了改进，用于从弯曲环面的曲面上的均匀分布中抽取随机样本。在第4节中，我们首先介绍直方图-接受-拒绝抽样（HAR）算法的思想。该算法为有限支撑连续分布提供了一个非常紧的包络函数，从而带来非常高的接受率。我们在圆上的冯·米塞斯分布中实现了 HAR 算法。随后，我们将这一思想扩展到从弯曲环面的曲面生成样本。每个部分均报告了详细的模拟结果。结论部分之后是附录-7。

2. Background

环面是一种几何对象，表示两个具有各自半径的角变量。只有其角度部分可以用平坦环面表示（$[0, 2\pi) \times [0, 2\pi)$），但涉及半径时，则用弯曲环面表示，见式1。弯曲环面与平坦环面不同胚，因为二者的拓扑性质不同。为了分析表示在弯曲环面曲面上的数据，必须恰当地定义其上的概率分布以及用于推断的统计方法。文献中已有的适用于平坦环面的统计技术并不适用于分析弯曲环面上的数据，因为这些技术没有考虑曲面的拓扑和几何。

这里，我们聚焦于二维弯曲环面，这是一种嵌入在 $\mathbb{R}^3$. 中的黎曼流形。在本报告中，我们将“弯曲环面”这一术语用于二维弯曲环面。

弯曲环面的参数空间为 $\mathcal{S} = \{(\theta_1, \theta_2) : 0 \leq \theta_1, \theta_2 < 2\pi\}$，可用参数方程表示为

$$\begin{aligned} x(\theta_1, \theta_2) &= (R + r\cos\theta_2)\cos\theta_1 \\ y(\theta_1, \theta_2) &= (R + r\cos\theta_2)\sin\theta_1 \\ z(\theta_1, \theta_2) &= r\sin\theta_2, \end{aligned} \tag{1}$$

其中 R, r 分别是水平圆和竖直圆的半径。接下来，我们考虑 Federer Federer (2014) 中关于几何测度论的一些基础知识，以便进一步发展。

定义 1。 函数 $f : \mathbb{R}^m \to \mathbb{R}^n$称为利普希茨函数，当且仅当存在常数 $C > 0$，使得 $|f(x) - f(y)| < C|x - y|$，只要 $x, y \in \mathbb{R}^m$.

定义 2。 集合 $E \subset \mathbb{R}^n$，称为 m-可求 rectifiable 集合，当且仅当它是某个有界子集在 $\mathbb{R}^m$. 中的利普希茨像

定义 3。 设 A 是 $\mathbb{R}^n$ 的任意子集， m维豪斯多夫测度 $\mathbb{H}_m(A)$ 定义为

$$\mathbb{H}_m(A) = \lim_{\delta \to 0} \inf_{\substack{A \subseteq \cup S_i, \\ \operatorname{diam}(S_i) \leq \delta}} \sum \alpha_m \left(\frac{\operatorname{diam}(S_i)}{2} \right)^m,$$

其中 $\alpha_m = \dfrac{\Gamma\left(\frac{1}{2}\right)^m}{\Gamma\left[\frac{m}{2} + 1\right]}$ 是 $\mathbb{R}^n$ 中单位球的体积，下确界取遍 A 的所有可数覆盖 S_i，满足 $\operatorname{diam}(S_i) = \sup\{|x - y| : x, y \in S_i\}$.

豪斯多夫测度是一种外测度，对于子集而言起面积测度的作用，并且在 $\mathbb{R}^n$ 的博雷尔集上满足可数可加性。 Federer (2014) 表明，对于一个 m-可整流集合 A，覆盖可以限制为立方体或球，并且

$$\mathbb{H}_m(A) = \lim_{\epsilon \to 0} \left[\frac{\lambda_n(\{x : \operatorname{dist}(x, A) < \epsilon\})}{\alpha_{n-m}\epsilon^{n-m}} \right],$$

其中 $\lambda_n(dx)$ 是定义在 $\mathbb{R}^n$. 集合上的勒贝格测度。面积公式将在以下定理中介绍（参见 Federer, 2014），它是微分学中变量替换概念的自然延伸。

定理 1。 设 $f : \mathbb{R}^m \to \mathbb{R}^n$ 是一个利普希茨函数，且 $m \leq n$，则

$$\int_A g(f(x)) J_m f(x) \lambda_m dx = \int_{\mathbb{R}^n} g(y) N(f|A, y) \mathbb{H}_m dy \tag{2}$$

只要 $g : \mathbb{R}^n \to \mathbb{R}$ 是博雷尔集， A就是 λ_m 可测的，且 $N(f|A, y)$ 是集合 $\{x \in A : f(x) = y\}$.的基数

本文中，函数 f 是环面的参数化；见公式 1。因此，函数 f 是一一映射，右侧的积分是 g 在集合 $f(A) \in \mathbb{R}^n$上的曲面积分，而左侧则利用雅可比和勒贝格测度在 $\mathbb{R}^m$上计算积分，其中尤其有 $m = 2$以及 $n = 3.$。这表明，从 $J_m f(x)$ 在 $\mathbb{R}^m$ 上的归一化概率密度抽样，再通过 f 将样本映射到弯曲环面上，就能得到一个以面积测度表示的弯曲环面曲面上的样本。

雅可比在计算面积测度时发挥着至关重要的作用。设 $f : \mathbb{R}^m \to \mathbb{R}^n$ 是一个函数，如果存在一个从 $\mathbb{R}^m$ 到 $\mathbb{R}^n$ 的线性映射 L，使得 f 在点 $x \in \mathbb{R}^m$ 处可微，则称其在该点可微，并且

$$\frac{|f(x+h) - f(x)|}{|h|} \to 0 \text{ as } h \to 0.$$

该线性映射在存在时记为 $Df(x)$，并且可以通过偏导数确定

$$D_i(x) = \lim_{h \to 0} \frac{[f(x_1, \cdots, x_i + h, \cdots, x_n) - f(x_1, \cdots, x_n)]}{h}.$$

用矩阵表示法，导数矩阵为 $(Df(x))_{i,j} = D_i f_j(x)$，其中 $1 \leq j \leq n$，且$1 \leq i \leq m.$ 设 $f : \mathbb{R}^m \to \mathbb{R}^n$ 是在点 $x \in \mathbb{R}^m$ 处可微的函数，p 维雅可比矩阵 $J_p f(x)$ 可以通过对导数矩阵取范数得到。对于给定的 $Df(x)$ 的秩，可以使用以下方法求得雅可比矩阵。当 $Df(x)$ 的秩为 $< p$ 时，$J_p f(x) = 0.$当 $Df(x)$ 的秩为 $= p$ 时，$J_p^2 f(x)$ 等于 $p \times p$ 个子矩阵的行列式平方之和，其中这些子矩阵取自 $Df(x)$。通常，当 $p = m$ 或 n 时，$J_p^2 f(x)$ 等于 $p \times p(Df)$ 的转置与 $D^T f(x)$ 及 $Df(x)$ 的乘积的行列式。对于 $p = m = n$，$J_p^2 f(x)$ 是 $Df(x)$ 的行列式的绝对值。

2.1. 环面的内蕴几何。式 1 中二维环面的参数方程是集合 $\{(\theta_1, \theta_2) : 0 < \theta_1, \theta_2 < 2\pi\} \subset \mathbb{R}^2.$ 的利普希茨像。显然，函数$f(\theta_1, \theta_2) = \{(R + r\cos\theta_2)\cos\theta_1, (R + r\cos\theta_2)\sin\theta_1, r\sin\theta_2\}$是从$\mathbb{R}^2$ 到 $\mathbb{R}^3$的可微函数。现在，f 关于 θ_1和 θ_2 的偏导数为

$$\frac{\partial f}{\partial \theta_1} = \{-(R + r\cos\theta_2)\sin\theta_1, (R + r\cos\theta_2)\cos\theta_1, 0\},$$

and

$$\frac{\partial f}{\partial \theta_2} = \{-r\sin\theta_2\cos\theta_1, -r\sin\theta_2\sin\theta_1, r\cos\theta_2\},$$

分别为。因此，导数矩阵为

$$Df(\theta_1,\theta_2)=\begin{bmatrix}-(R+r\cos\theta_2)\sin\theta_1 & -r\sin\theta_2\cos\theta_1\\ (R+r\cos\theta_2)\cos\theta_1 & -r\sin\theta_2\sin\theta_1\\ 0 & r\cos\theta_2\end{bmatrix}.$$

因此，雅可比为

$$J_2^2 f(x)=\det\left[D^T f(x)\cdot Df(x)\right]=\det\begin{bmatrix}(R+r\cos\theta_2)^2 & 0\\ 0 & r^2\end{bmatrix}=r^2(R+r\cos\theta_2)^2 \tag{3}$$

利用上述面积元素平方的表达式（即导数矩阵转置与导数矩阵乘积的行列式），Diaconis等人(2013年) 提出从式 θ_1、 θ_2给出的密度函数中抽取样本，以确保样本相对于弯曲环面曲面上的面积测度是均匀的

$$g(\theta_1,\theta_2)=\frac{(1+(r/R)\cos\theta_2)}{4\pi^2}=g_1(\theta_1)\ g_2(\theta_2), \tag{4}$$

其中

$$g_1(\theta_1)=\frac{1}{2\pi},\ 0\le\theta_1<2\pi, \tag{5}$$

and

$$g_2(\theta_2)=\frac{1}{2\pi}\left[1+\frac{r}{R}\cos\theta_2\right],\ 0\le\theta_2<2\pi. \tag{6}$$

θ_2 的累积分布函数为

$$G_2(\theta_2)=\frac{1}{2\pi}\left[\theta_2+\frac{r}{R}\sin\theta_2\right],\ 0\le\theta_2<2\pi.$$

Diaconis 等人(2013年) 使用接受-拒绝抽样法从密度 $g_2(\theta_2$中生成样本，并在其论文中给出了相应的算法。

3. Exact area uniform (EAU) sampling from torus

在本节中，我们提出了一种从弯曲环面的曲面生成随机均匀样本的新方法，这等价于从方程式 Eq. $g_1(\theta_1$中给出的密度5以及 $g_2(\theta_2$方程式6中给出的密度。具体而言，我们考虑 $a=r/R\in(0,1]$ 并提出使用概率变换从 $g_2(\theta_2$生成样本的方法，该方法在定理-2中进行了描述。下面的算法-1 给出了所提出的新 EAU 抽样方法的伪代码。

定理 2. 令 U 服从 $[0,1]$上的均匀分布， X 服从$[0,2\pi]$上的均匀分布，且 $p(X)=\frac{1}{2}(1+a$ 余弦 $X)$。则随机变量 Y定义为

$$Y = \begin{cases} X & if \quad U < p(X), \quad X < \pi \\ \pi - X & if \quad U > p(X), \quad X < \pi \\ X & if \quad U < p(X), \quad X > \pi \\ 3\pi - X & if \quad U > p(X), \quad X > \pi \end{cases}$$

follows the CDF, $G(y) = \dfrac{(y + a\sin y)}{2\pi}$ *where* $0 < a \leq 1$, *and* $0 < y < 2\pi$.

证明。 该定理的证明可见附录-7。 □

算法 1： EAU 抽样算法

数据： $(W, X) \leftarrow n$ 对来自 $U[0, 2\pi]$. 的随机样本
$a \leftarrow$ 介于 0 和 1 之间的数字;
$pct \leftarrow \dfrac{1 + a\text{余弦 } X}{2}$; `/* define the probability */`
$Y \leftarrow$ 创建一个大小为 n 的零数组;
对于 $i \leftarrow 1$ 到 n **执行**
  $rp[i] \leftarrow$ 从 $Bernoulli(pct[i])$ 中抽取一个伯努利分布随机样本;
  如果 $X[i] < \pi$ **则**
    $a_1 \leftarrow (X[i] * rp[i]) + (\pi - X[i]) * (1 - rp[i])$;
  else
    如果 $X[i] > \pi$ **则**
      $a_2 \leftarrow (X[i] * rp[i]) + (3\pi - X[i]) * (1 - rp[i])$;
    end
  end
  $Y \leftarrow a_1 + a_2$; `/* The random samples follow the desired distribution */`
end
结果： Y 遵循 $F(y) = \dfrac{(y + a\sin y)}{2\pi}$ 其中 $0 < a \leq 1$, $0 \leq y < 2\pi$. (W, Y) 为来自弯曲环面上面积均匀分布的样本。

3.1. **模拟分析**。本小节报告一项全面模拟研究的结果，以比较所提出的 EAU 抽样（算法 1 ）与Diaconis 等人（2013年）提出的现有面积均匀拒绝抽样算法（AUR）。我们考虑的不同取值，从 0.1到 *1*，间隔相等，并计算两种算法在样本量为 *10000* 时的接受率。以下表1显示，EAU 抽样的表现远胜于 AUR 抽样。

图 1 是垂直角边缘分布的抽样数据直方图 θ_2 ，依据公式 6。图 2是使用算法 1生成的数据散点图，该算法在弯曲环面的曲面上采用面积测度，以保持均匀分布，其中 $R = 3$, $r = 1.5$, 因此 $a = 0.5$. 图 3 是.....的投影

将从平坦环面均匀生成的数据映射到弯曲环面的曲面上。从图 3 可以明显看出，与采用 EAU 抽样得到的图 2 相比，负曲率曲面区域的数据概率密度更高。图 2 和图 3 的散点图可能难以区分；一组相对频率条形图可以使差异更加明显。我们将水平圆和垂直圆的象限组合视为 $Q_H \times Q_V$，其中 $Q_H = Q_V = \{1, 2, 3, 4\}$，每个象限的分区长度为 $\pi/2.$ 。图4的条形图表示使用 EAU采样方法在环面上均匀生成的点的相对频率；而图5的条形图表示角参数服从均匀分布时，从平坦环面生成的点的相对频率。在这两幅图中，红线和绿线分别表示正曲率和负曲率区域内各象限组合的面积占弯曲环面总表面积的比例。虽然图4中象限组合的相对频率与各自的面积比例相符，但图5中则不然。这也可以用数据的边缘图表示。图6展示了依据面积测度从弯曲环面曲面上均匀抽取的数据的边缘分布，而图7展示了从其参数空间（即平坦环面）均匀抽取的数据的相应分布。因此，如公式 4 所示，联合密度函数 $g(\theta_1, \theta_2)$ 是弯曲环面曲面上均匀分布的一种更自然的概念，而非平坦环面上的均匀分布；Diaconis 等人（2013）也倡导这一观点。

$\mathbf{a}=\frac{r}{R}$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
EAU	100	100	100	100	100	100	100	100	100	100
AUR	50.35	51.07	49.91	49.68	50.14	49.82	49.88	49.41	49.96	49.41

TABLE 1. 弯曲环面上均匀分布的接受率比较表。

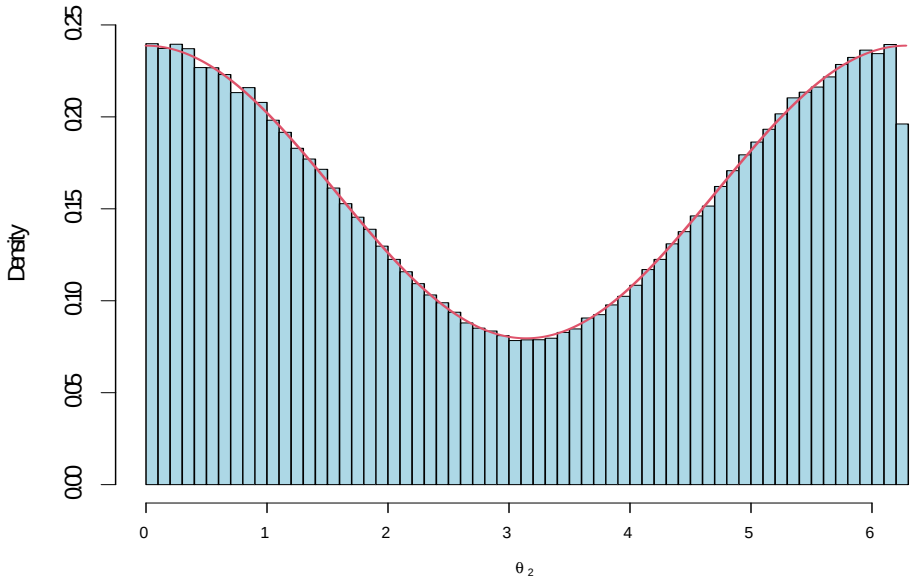

FIGURE 1. 定理对应样本的直方图 2 概率密度 $\frac{1}{2\pi}(1+a\text{余弦 }\theta_2)$。 —

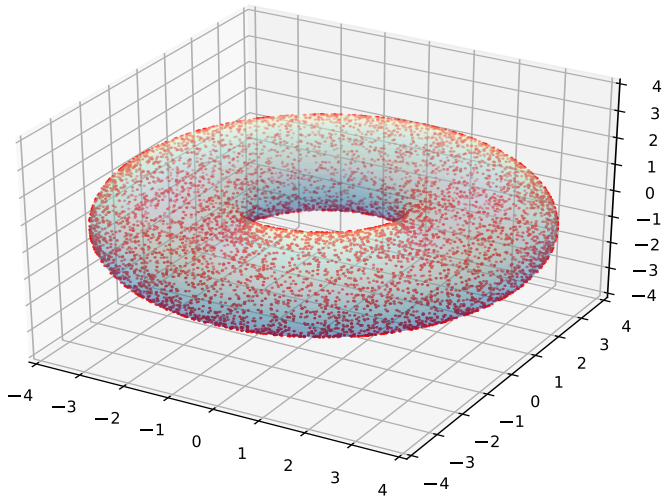

FIGURE 2. 根据面积测度，在弯曲环面上服从均匀分布的 EAU采样方法所得样本的散点图。

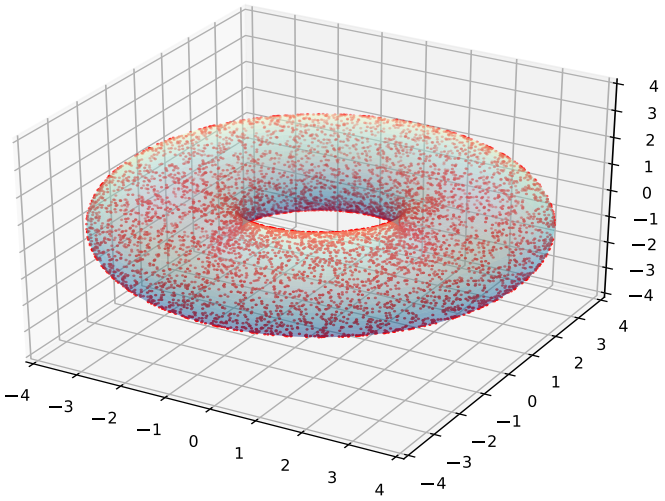

FIGURE 3. 从平坦环面均匀抽取角参数时环面上各点的散点图

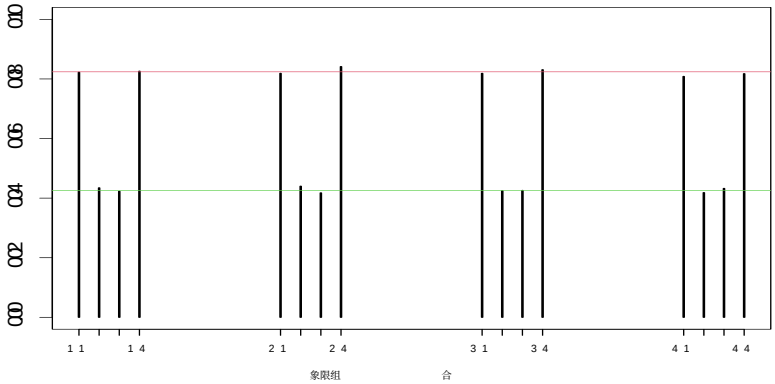

FIGURE 4. 使用 EAU采样方法在环面上均匀生成点的列联表示意图。红线和绿线分别表示与正曲率和负曲率成比例的面积。

4. HISTOGRAM-ACCEPTANCE-REJECTION (HAR) SAMPLING

在许多科学和技术应用中，对角度或圆周数据进行建模至关重要。对于理论和实践研究而言，模拟角分布也非常重要。接受-拒绝采样是从角分布中抽样最常用的方法之一，例如冯·米塞斯分布、心形分布和 Kato-Jones分布

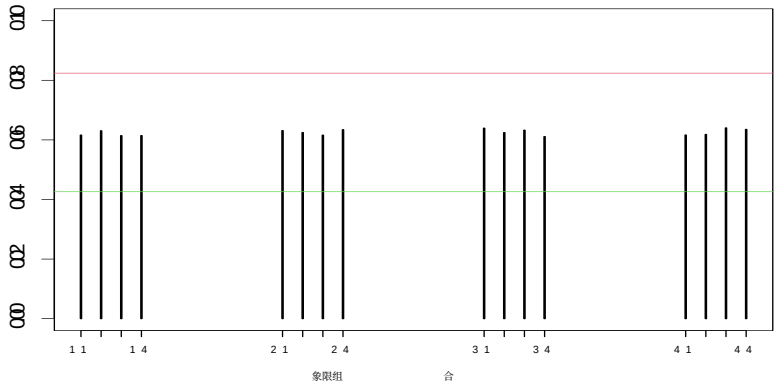

FIGURE 5. 使用角参数的均匀分布绘制环面上各点的列联表示意图。红线和绿线分别表示与环面的正曲率部分和负曲率部分成比例的面积。

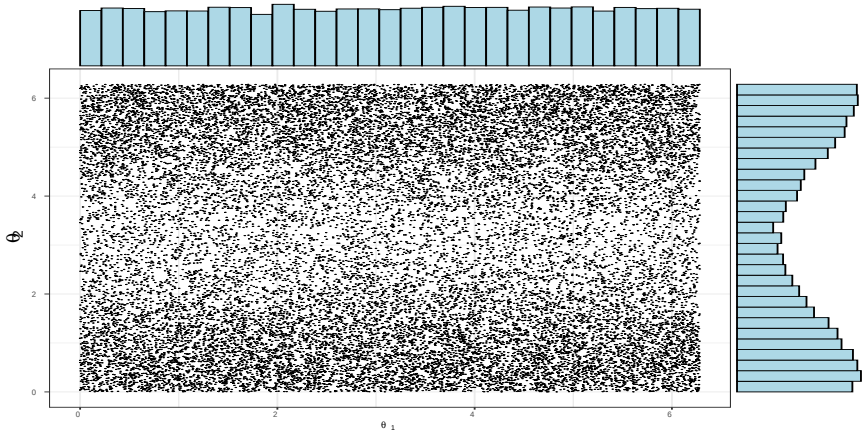

FIGURE 6. 根据面积测度，从弯曲环面曲面上均匀抽取的数据的边缘分布图。

等。接受-拒绝抽样法的主要挑战在于找到合适的包络函数，因为接受概率在很大程度上取决于它。在进行大规模模拟时，如果仅仅由于包络函数的选择导致大量样本被拒绝，计算成本就会很高。在此，我们提出一种新的方法，称为直方图-接受-拒绝（HAR）抽样，用于构造目标分布的包络，其灵感来自黎曼上和的构造，因此可以大幅减少被拒绝的样本数量。HAR采样方法可用于在有界区间上黎曼可积的概率密度函数。特别是，本文聚焦于一些常见的圆周分布，它们天然满足上述条件。在黎曼积分中，黎曼上和，也就是

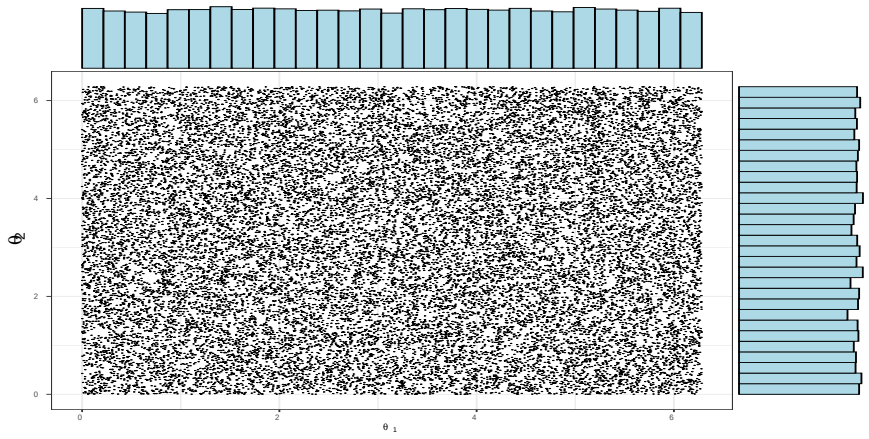

FIGURE 7. 根据勒贝格测度，从参数空间中均匀抽取的数据的边缘分布图。

非负可积函数的支配阶梯函数下方的面积大于实际被积函数下方的面积。但可以将支配阶梯函数归一化，使其总面积为 1，这样就可以将其视为合法的概率密度函数，最终用于作为预先指定的目标分布的提议概率密度。在下一小节中，我们将详细介绍这种方法。

4.1. HAR抽样的所提方法。 设 U 服从 $[0,1]$上的均匀分布。假设 $f(x)$ 和 $p(x)$ 分别为目标概率密度函数和所提概率密度函数，且具有共同的有限支撑集 $[a,b]$。将区间 $[a,b]$划分为若干分区，如下： $a = a_0 < a_1 < a_2 < \cdots < a_{i-1} < a_i < \cdots < a_{k-1} < a_k = b$。现在，对于 $i = 1, \cdots, k$，令 $A_i = [a_{i-1}, a_i]$， $B = a_i - a_{i-1} = \frac{(b-a)}{k}$，且 $P(A_i) = \int_{A_i} f(y)dy$。对于第 i^{th} 个单元，令 ~~H_i~~ $= \max_{x \in A_i} f(x)$， $Y_i = a_{i-1} + BU$。因此，整个支撑集上的所提概率密度为

$$p(y) = \sum_{i=1}^{k} \frac{H_i \mathbf{I}_{A_i}(y)}{B \sum_{i=1}^{k} H_i} \tag{7}$$

满足条件 $p(y \mid y \in A_i) = \frac{1}{B}$。我们选择$i^{th}$ 单元的数量 $M_i = \frac{BH_i}{P\ A_i()}$，使得 $M_i \geq$ ~~$\max$~~ $\frac{f}{p(y|y \in A_i)\ (y|y \in A_i)}$。考虑 i^{th} 单元的 X 累积分布函数（CDF）：

$$P(X \leq x) = P(Y_i \leq x \,|\, Y_i \text{ accepted})$$

$$= \frac{P\left(Y_i \le x,\ U < \dfrac{f(Y_i)/P(A_i)}{M_i(\frac{1}{B})}\right)}{P\left(U < \dfrac{f(Y_i)/P(A_i)}{M_i(\frac{1}{B})}\right)} \tag{8}$$

现在考虑分子

$$\begin{aligned} P\left(Y_i \le x,\ U < \frac{f(Y_i)/P(A_i)}{M_i(\frac{1}{B})}\right) &= \int P\left(Y_i \le x,\ U < \frac{f(Y_i)/P(A_i)}{M_i(\frac{1}{B})}\middle| Y_i = y\right)\left(\frac{1}{B}\right) dy \\ &= \int \mathbf{I}_{(y\le x)} P\left(U < \frac{f(y)/P(A_i)}{M_i(\frac{1}{B})}\right)\left(\frac{1}{B}\right) dy. \end{aligned} \tag{9}$$

同样,

$$P\left(U < \frac{f(Y_i)/P(A_i)}{M_i(\frac{1}{B})}\right) = \int_{A_i} \frac{f(y)/P(A_i)}{M_i(\frac{1}{B})}\left(\frac{1}{B}\right) dy \tag{10}$$

利用式9和式10 代入式8 可得

$$P(Y_i \le x \,|Y_i \text{ accepted}) = \frac{\displaystyle\int \mathbf{I}_{(y\le x)} P\left(U < \frac{f(y)/P(A_i)}{M_i(\frac{1}{B})}\right)\left(\frac{1}{B}\right) dy}{\displaystyle\int_{A_i} \frac{f(y)/P(A_i)}{M_i(\frac{1}{B})}\left(\frac{1}{B}\right) dy.} \tag{11}$$

X 在整个范围内的累积分布函数为

$$P(X \le x) = \int_a^x f(t)\ dt = \left(\frac{1}{M}\int_a^x f(t)\ dt\right) \Big/ \left(\frac{1}{M}\right), \tag{12}$$

其中，我们可以选择 M 使得

$$M \ge \max_x \frac{f(x)}{p(x)} = \max_i \left\{\max_{x\in A_i} \frac{f(x)}{p(x)}\right\} = \max_{1\le i\le k} \left[H_i \Big/ \left(\frac{H_i}{B\ \sum_{i=1}^k H_i}\right)\right] = B\sum_{i=1}^k H_i.$$

注意,

$$\begin{aligned} \frac{1}{M}\int_a^x f(t)\ dt &= \sum_{i=1}^k \left[\frac{1}{M}\int_{A_i} \mathbf{I}_{(t\le x)}\ \frac{f(t)}{p(t)}\ p(t)\ dt\right] \\ &= \sum_{i=1}^k \left(\frac{P(A_i)}{M}\right) \left[\frac{\displaystyle\int_{A_i} \mathbf{I}_{(t\le x)}\ P\left(U < \frac{f(t)/P(A_i)}{M_i(\frac{1}{B})}\right)\ \left(\frac{1}{B}\right)\ dt}{\dfrac{1}{1/M_i}}\right], \end{aligned} \tag{13}$$

详情见附录-7。因此，利用方程11 和方程13 代入方程12 可得

$$\frac{\frac{1}{M}\int_a^x f(t)\ dt}{\frac{1}{M}} = \sum_{i=1}^{k} P(A_i)\left[\frac{\int_{A_i} \mathbf{I}_{(t\le x)}\ P\left(U < \frac{f(t)/P(A_i)}{M_i(\frac{1}{B})}\right)\left(\frac{1}{B}\right)\ dt}{\int_{A_i} \frac{f(t)/P(A_i)}{M_i(\frac{1}{B})}\left(\frac{1}{B}\right) dt.}\right]. \tag{14}$$

因此 $\int_a^x f(t)\ dt = \sum_{i=1}^{k} P(A_i)\left[P(Y_i \le x \,|\, Y_i \text{已接受})\right]$, 可通过以下算法实现：2 即所提出的 HAR 抽样方法的伪代码。

算法 2： HAR采样算法

数据：支撑集为 $[a,b]$ 的目标概率密度函数 $f(x)$。$n \leftarrow$ 要生成的随机样本数；$np \leftarrow$ 等长分区的数量；$pt \leftarrow$ 长度为 $(np+1)$ 的 $[a,b]$ 中的一个序列；$H \leftarrow f(pt)$；`/* Heights of the probability density function at pt */`$H_m \leftarrow$ 每个分区最大高度的序列；$bl \leftarrow \frac{b-a}{np}$ ；`/* Bin length */`$p_m \leftarrow \frac{H_m}{\sum H_m}$ ；

`/* Probability vector */`$count = 0$；$y \leftarrow$ 初始化一个大小为零的向量；当 $count \le n$ 时执行$u \leftarrow$从 $U[0,1]$ 中随机抽取一个数；$ml \leftarrow$ 根据概率向量 p_m 从多项分布中抽取一个随机样本；$x \leftarrow pt[ml] + u * bl$；$px \leftarrow f(x)\ H_m[ml]$；$rp \leftarrow$ 从 $Bernoulli(px)$ 中抽取一个随机数；如果 $rp = 1$ ，则$count = count + 1$；$y[count] \leftarrow x$；结束 结束 结果：从支撑集为 $[a,b]$ 的概率密度函数 $f(x)$ 中获得的样本。

4.2. 冯·米塞斯分布的模拟。概率密度函数

$$f(\theta) = \frac{e^{\kappa\cos(\theta-\mu)}}{2\pi I_0(\kappa)}, \tag{15}$$

其中 $0 \leq \theta < 2\pi$, $0 \leq \mu < 2\pi$, $\kappa > 0$, 且 $I_0(\kappa)$ 是在 κ 处取值的零阶修正贝塞尔函数，定义了冯·米塞斯分布。这是一种应用广泛的概率分布，可用于对角度或圆形数据建模。由式 15可知，集中参数为 κ 的概率分布关于均值方向 μ呈连续且对称的分布。由于冯·米塞斯分布的累积分布函数（CDF）没有闭式解，因此从该分布中抽样颇具挑战性，正如 Mardia 等人 (2000) 所讨论的那样。因此，逆变换抽样等传统抽样方法无法用于该分布。

由 Best 和 Fisher (1979) 提出的冯·米塞斯分布抽样程序是一种广为人知的方法，它采用传统的拒绝抽样技术，并以包裹柯西分布作为包络函数，我们称之为 vMBFR 抽样。然而，vMBFR 抽样方法的拒绝率会因参数而异，且较高。相比之下，本文提出的 HAR 采样方法的拒绝率要低得多。

在本研究中，我们进行了样本量为 $n = 50000$ 的模拟，以比较 HAR 和 vMBFR 算法的样本接受率。表-2 和表-3 展示了集中参数取不同值时的接受率κ，平均方向参数为 $\mu = 0$。在样本接受率方面，本文提出的 HAR 采样方法优于现有的 vMBFR 采样方法。图-8 展示了使用 HAR 抽样技术从冯·米塞斯分布中抽取样本点的直方图，

κ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
HAR	99.96	99.92	99.87	99.85	99.81	99.77	99.72	99.71	99.67	99.65
vMBFR	99.76	99.06	97.90	96.67	95.04	93.23	91.88	89.88	88.12	86.94

TABLE 2. 冯·米塞斯分布接受率比较表 n.

κ	2	3	4	5	10	20	40	60	80	100
HAR	99.48	99.21	99.02	98.91	98.462	97.76	96.96	96.31	96.76	95.15
vMBFR	76.95	72.37	69.96	69.46	67.46	66.64	66.43	65.96	65.94	65.69

TABLE 3. 冯·米塞斯分布接受率比较表 n.

同样，除了冯·米塞斯分布外，还可以将 HAR 采样方法用于心形分布、Kato-Jones分布等，以获得更高的接受率。

5. GENERALIZATION TO SAMPLING FROM THE SURFACE OF TORUS

尽管许多研究人员已经提出并研究了平环面上的几种分布，但只有弯曲环面曲面上的分布得到了研究

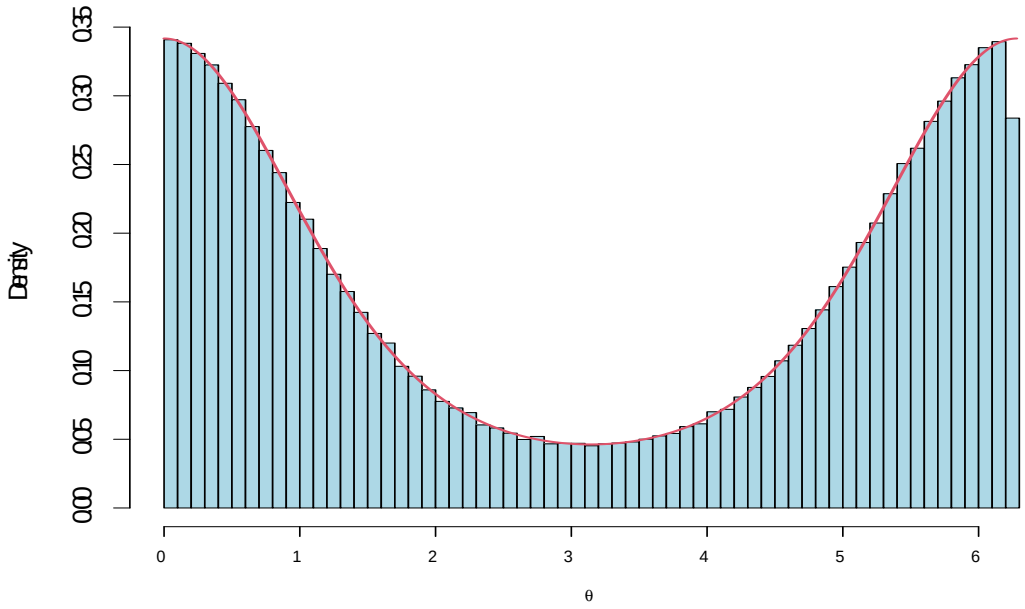

FIGURE 8. 由 HAR 抽样方法生成的冯·米塞斯分布样本数据直方图。

作者：Diaconis等人 (2013年）。他们特别研究了式 1 中给出的环面上的均匀分布，并将其作为流形的一个示例。本节中，我们介绍了弯曲环面曲面上一些常见分布的推广，将圆形分布扩展到更高维度。

应用第2小节讨论的微分几何原理，根据式3，我们可以确定环面的面积元为 $dA = r(R + r\cos\theta_2)$。因此，环面的面积显然为

$$A = \int_0^{2\pi}\int_0^{2\pi} r(R + r\cos\theta_2)\, d\theta_1\, d\theta_2 = 4\pi^2 rR \tag{16}$$

现在，考虑 $h(\theta_1,\ \theta_2$关于 θ_1 和 θ_2的联合概率密度函数，并由式16 得到恒等式

$$\begin{aligned} 4\pi^2 rR &= \int_0^{2\pi}\int_0^{2\pi} h(\theta_1,\, \theta_2)(2\pi r)(2\pi R)\, d\theta_1\, d\theta_2 \\ &= \frac{1}{C}\int_0^{2\pi}\int_0^{2\pi} h(\theta_1,\, \theta_2)\left(1 + \frac{r}{R}\cos\theta_2\right)(2\pi r)(2\pi R)\, d\theta_1\, d\theta_2, \end{aligned}$$

其中 C 是归一化常数，这意味着

$$\begin{aligned} 1 &= \frac{1}{C}\int_0^{2\pi}\int_0^{2\pi} h(\theta_1,\,\theta_2)\left(1+\frac{r}{R}\cos\theta_2\right)\,d\theta_1\,d\theta_2 \\ &= \frac{1}{C}\int_0^{2\pi}\int_0^{2\pi} h_1(\theta_1\,|\,\theta_2)\,\left[h_2(\theta_2)\left(1+\frac{r}{R}\cos\theta_2\right)\right]\,d\theta_1\,d\theta_2. \end{aligned} \tag{17}$$

现在在几何 通常，我们可以考虑从联合概率密度函数中抽样

$$h^*(\theta_1,\theta_2) = \frac{1}{C}\,h_1(\theta_1\,|\,\theta_2)\,\left[h_2(\theta_2)\left(1+\frac{r}{R}\cos\theta_2\right)\right],\ \text{where } 0\le\theta_1,\theta_2<2\pi \tag{18}$$

作为一种从弯曲环面曲面进行抽样的方案，参数空间或平坦环面上的联合密度预先指定为 $h(\theta_1,\ \theta_2)$。特别地，当 θ_1 和θ_2 相互独立分布时，则有 $h_1(\theta_1\,|\,\theta_2)\ \propto h_1(\theta_1)$。因此，公式 18 将化简为

$$h^*(\theta_1,\theta_2)\propto h_1(\theta_1)\,\left[h_2(\theta_2)\left(1+\frac{r}{R}\cos\theta_2\right)\right],\ \text{where } 0\le\theta_1,\theta_2<2\pi. \tag{19}$$

在以下小节中，我们将展示该方法，以便从弯曲环面曲面上的不同分布中抽取样本，同时保持与面积成比例的支配测度。

5.1. 模拟分析。 在此，我们开展广泛的模拟研究，以从弯曲环面曲面上的不同分布中抽取样本。

（I）均匀分布：第 3.1 小节中的模拟分析是第 5 节所述方法的一个特殊情形，其中 θ_1 和 θ_2 关于勒贝格测度在平坦环面上服从相互独立的均匀分布，因此，其联合概率密度函数为

$$h^*(\theta_1,\theta_2) = g(\theta_1,\theta_2) = \frac{1}{2\pi}\frac{(1+\frac{r}{R}\cos\theta_2)}{2\pi},$$

由公式 4，公式 19，这意味着

$$h_1(\theta_1) = \frac{1}{2\pi},\ \text{and}\ h_2(\theta_2) = \frac{(1+\frac{r}{R}\cos\theta_2)}{2\pi},\ \text{where } 0\le\theta_1,\theta_2<2\pi.$$

(II) 冯·米塞斯分布：要从弯曲环面的曲面上的边缘冯·米塞斯分布中抽取样本，令 θ_1 和 θ_2 分别独立地服从平坦环面上的冯·米塞斯分布，其集中参数分别为 κ_1、 κ_2，位置参数分别为 μ_1、 μ_2。因此，现在有必要从目标分布中抽取样本

$$h^*(\theta_1,\theta_2) = \frac{e^{\kappa_1\cos(\theta_1-\mu_1)}}{2\pi I_0(\kappa_1)}\left[\frac{e^{\kappa_2\cos(\theta_2-\mu_2)}}{C}\left(1+\frac{r}{R}\cos\theta_2\right)\right], \tag{20}$$

其中 $0\le\theta_1,\ \theta_2<2\pi, 0\le\mu_1,\ \mu_2<2\pi, \kappa_1,\ \kappa_2>0$，这意味着

$$h_1(\theta_1) = \frac{e^{\kappa \cos(\theta_1 - \mu_1)}}{2\pi I_0(\kappa_1)} \text{ and } h_2(\theta_2) = \frac{e^{\kappa_2 \cos(\theta_2 - \mu_2)}}{C} \left(1 + \frac{r}{R} \cos\theta_2\right),$$

with

$$C = 2\pi \left[I_0(\kappa_2) + \frac{r}{R} \cos\mu_2 \; I_1(\kappa_2)\right]$$

是归一化常数（参见附录-7）。因此，式20 中的联合概率密度函数代表弯曲环面曲面上的冯·米塞斯分布。现在令 $\mu_1 = \mu_2 = 0$， $\kappa_1 = \kappa_2 = 1$，且 $r = 1.5$， $R = 3$. 则下图9 是使用 HAR 采样方法从垂直角 θ_2 的分布中抽取的样本数据直方图。

图 10显示了相对于面积测度，从弯曲环面曲面上的冯·米塞斯分布中抽取的点的散点图；图 11 则显示了从平坦环面上的冯·米塞斯分布中抽取样本，并将其投影到弯曲环面曲面上的相同结果。

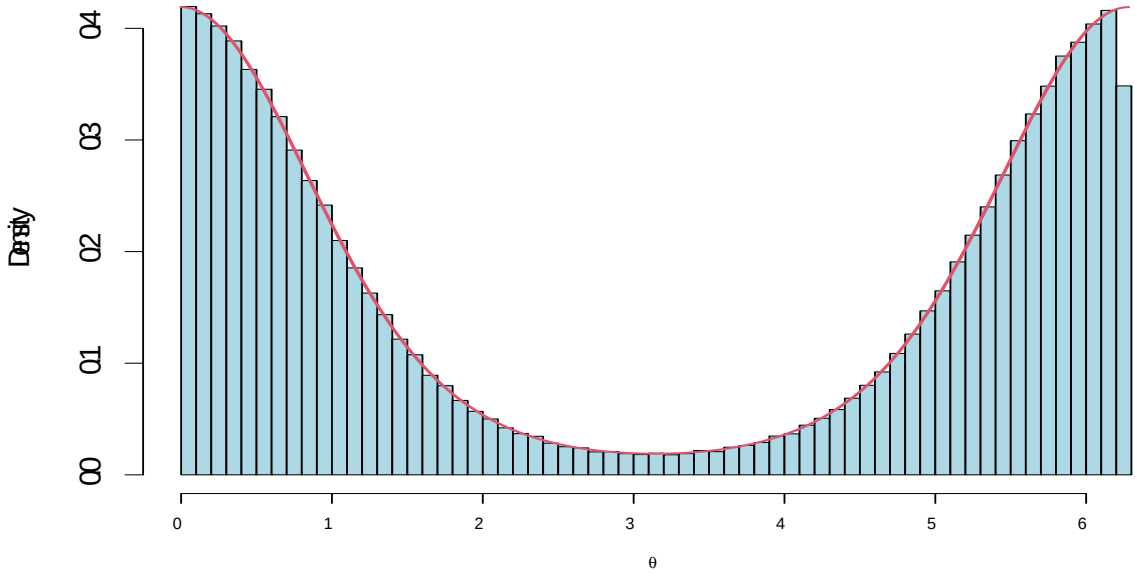

FIGURE 9. 来自概率密度的直方图 $\frac{e^{\cos\theta}\left(1+\frac{r}{R}\cos\theta\right)}{2\pi\left[I_0(1)+\frac{r}{R}I_1(1)\right]}$.

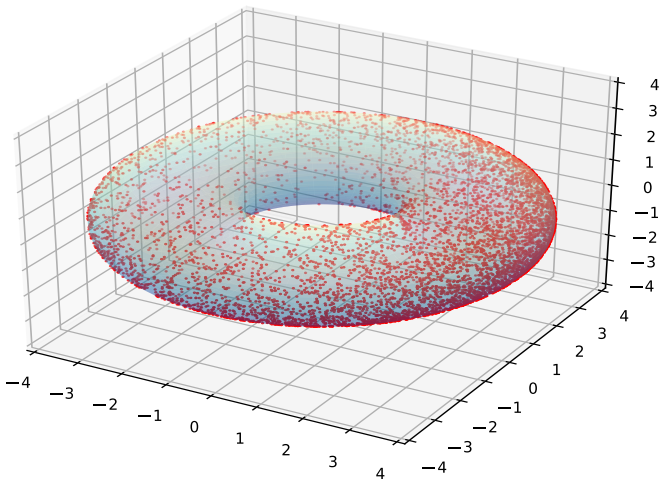

FIGURE 10. 使用面积测度绘制的环面上冯·米塞斯分布点的散点图。

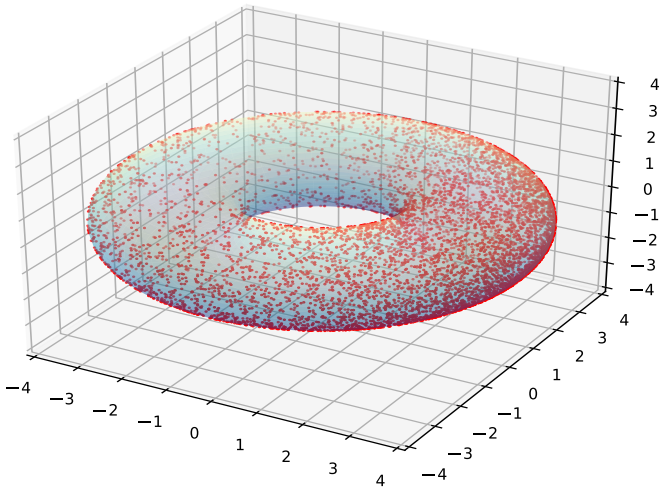

FIGURE 11. 从扁平环面的冯·米塞斯分布中抽取角参数时环面上各点的散点图。

（III）包裹柯西分布：包裹柯西分布是众所周知的圆周分布之一，其概率密度函数见式 21。

$$f_{wc}(\theta) = \frac{1}{2\pi} \frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu)}, \tag{21}$$

其中 $0 \leq \theta < 2\pi$、 $0 \leq \mu < 2\pi$ 和 $0 \leq \rho < 1$。

现在我们介绍定义在弯曲环面曲面上的包裹柯西分布。设 θ_1 和 θ_2 在平坦环面上分别独立服从集中参数为 ρ_1， ρ_2、位置参数为 μ_1， μ_2 的包裹柯西分布。因此，相应的联合密度为

$$h^*(\theta_1,\theta_2) = \frac{(2\pi)^{-1}(1-\rho_1^2)}{1+\rho_1^2-2\rho_1\cos(\theta_1-\mu_1)} \left[\frac{(C\ 2\pi)^{-1}(1-\rho_2^2)}{1+\rho_2^2-2\rho_2\cos(\theta_2-\mu_2)} \left(1+\frac{r}{R}\cos\theta_2\right)\right], \tag{22}$$

其中 $0 \leq \theta_1$， $\theta_2 < 2\pi$、 $0 \leq \mu_1$， $\mu_2 < 2\pi$、 $0 \leq \rho_1$， $\rho_2 < 1$，这意味着

$$h_1(\theta_1) = f_{wc}(\theta_1) \text{ and } h_2(\theta_2) = \frac{1}{C} f_{wc}(\theta_2) \left(1+\frac{r}{R}\cos\theta_2\right),$$

其归一化常数为

$$C = \int_0^{2\pi} \left[f_{wc}(\theta_2)\ \left(1+\frac{r}{R}\cos\theta_2\right)\right]\ d\theta_2.$$

因此，方程22中的联合概率密度函数代表弯曲环面曲面上的包裹柯西分布。现在，假设 $\mu_1 = \mu_2 = 0$， $\rho_1 = \rho_2 = 0.3$，且 $r = 1.5$， $R = 3.$，那么下图12 是使用 HAR 采样方法从垂直角 θ_2 的分布中获得的抽样数据直方图。

图13展示了相对于面积测度，从弯曲环面曲面上的包裹柯西分布中抽取的点的散点图；图14 则展示了从平坦环面上的包裹柯西分布中抽取样本，并将其投影到弯曲环面曲面上的相同散点图。

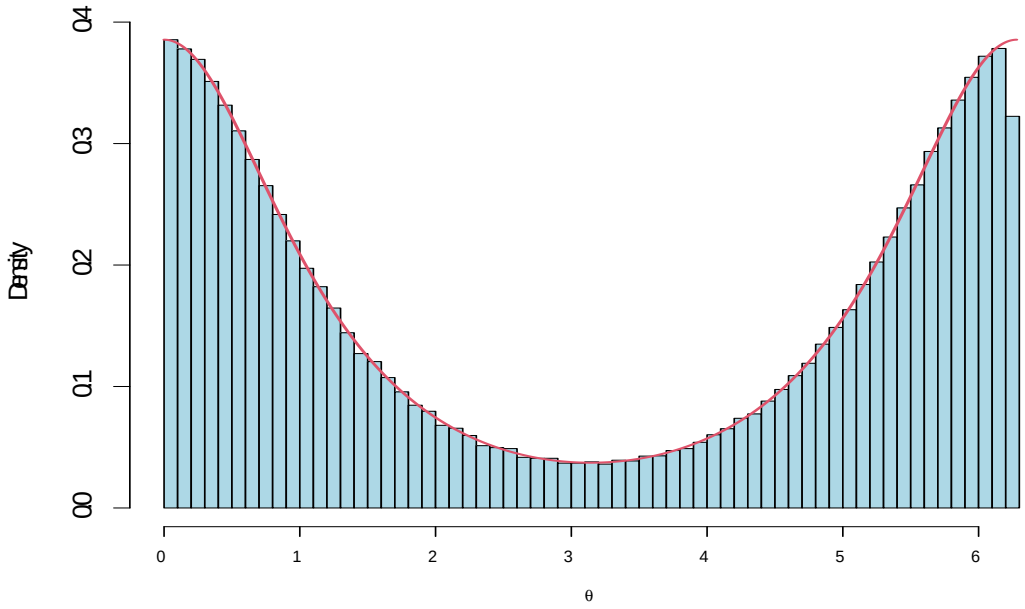

FIGURE 12. 概率密度 $\frac{1}{C} f_{wc}(\theta)\ \left(1 + \frac{r}{R}\cos\ \theta\right)$ 的数据直方图。 –

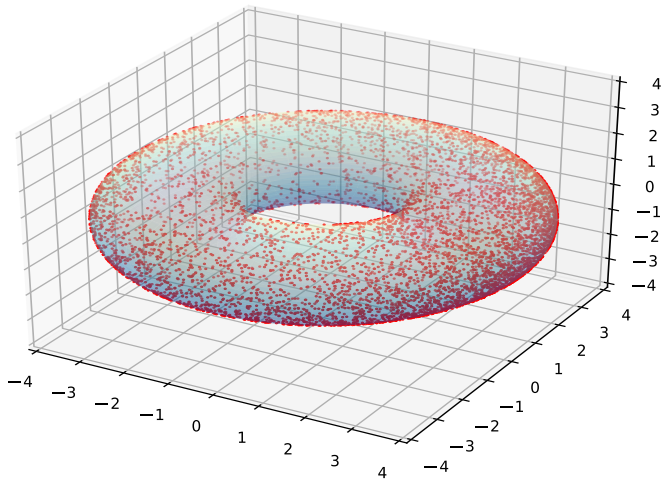

FIGURE 13. 使用面积测度对包裹柯西分布进行抽样得到的环面上点的散点图。

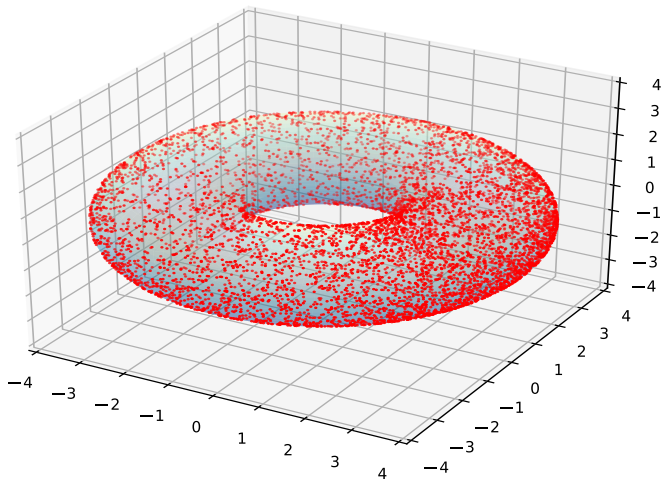

FIGURE 14. 从平坦环面上的包裹柯西分布抽取角参数时，环面上各点的散点图。

(IV) Kato 和 Jones 分布： Kato 和 Jones 分布是一个四参数圆周分布族，由Kato和Jones(2010年) 首次提出，他们使用 M¨obius 变换，将 Θ 变换为 Θ，其中

$$
\begin{aligned}
e^{i\Theta} &= e^{i\mu}\,\frac{e^{i\bar{\Theta}}+\rho e^{i\nu}}{\rho e^{i(\bar{\Theta}-\nu)}+1} \qquad \text{or} \\
\Theta &= \mu+\nu+2\arctan\left[\left(\frac{1-\rho}{1+\rho}\right)\tan\left(\frac{\bar{\Theta}-\nu}{2}\right)\right], \qquad (23)
\end{aligned}
$$

其中 $0 \le \mu,\ \nu < 2\pi$ 和 $0 \le \rho < 1$。

现在，他们取随机变量 Θ，令其服从参数为 $\mu = 0$ 和κ的冯·米塞斯分布，然后使用式 23 中的变换，得到所需的概率密度：

$$
f_{kj}(\theta) = \frac{(1-\rho^2)(2\pi I_0(k))^{-1}}{1+\rho^2-2\rho\cos(\theta-\gamma)}\exp\left[\frac{\kappa\{\xi\cos(\theta-\eta)-2\rho\cos\nu\}}{1+\rho^2-2\rho\cos(\theta-\gamma)}\right], \qquad (24)
$$

其中 $0 \le \mu,\ \nu < 2\pi$ 和 $0 \le \rho < 1,\ \kappa > 0$，以及 $\gamma = \mu+\nu$，$\xi = \sqrt{\rho^4+2\rho^2\cos(2\nu)+1}$，$\eta = \mu + \arg(\rho^2\cos(2\nu)+1+i\rho^2\sin(2\nu))$

现在我们介绍弯曲环面的曲面上的 Kato 和 Jones 分布。d Jones 分布位于弯曲环面的曲面上。令 θ_1 和 θ_2 独立地服从 Kato 和 Jones 分布，其集中参数为 $\rho_1, \rho_2, \kappa_1, \kappa_2$，位置参数为

$\mu_1, \mu_2, \nu_1, \nu_2$，分别位于平坦环面上。因此，相应的联合密度为

$$h^*(\theta_1,\theta_2) = \frac{(1-\rho_1^2)(2\pi I_0(\kappa_1))^{-1}}{1+\rho_1^2-2\rho_1\cos(\theta_1-\gamma_1)} \exp\left[\frac{\kappa_1\{\xi_1\cos(\theta_1-\eta_1)-2\rho_1\cos\nu_1\}}{1+\rho_1^2-2\rho_1\cos(\theta_1-\gamma_1)}\right] \times \left[\frac{(1-\rho_2^2)(C\ 2\pi I_0(\kappa_2))^{-1}\left(1+\frac{r}{R}\cos\theta_2\right)}{1+\rho_2^2-2\rho_2\cos(\theta_2-\gamma_2)} \exp\left[\frac{\kappa_2\{\xi_2\cos(\theta_2-\eta_2)-2\rho_2\cos\nu_2\}}{1+\rho_2^2-2\rho_2\cos(\theta_2-\gamma_2)}\right]\right], \quad (25)$$

其中 $0 \le \mu_j,\ \nu_j < 2\pi$，以及 $0 \le \rho_j < 1, \kappa_j > 0$，和 $\gamma_j = \mu_j + \nu_j, \xi_j = \sqrt{\rho_j^4 + 2\rho_j^2\cos(2\nu_j) + 1}, \eta_j = \mu_j + \arg(\rho_j^2\cos(2\nu_j) + 1 + i\rho_j^2\sin(2\nu_j))$，对于 $j = 1,\ 2$，这意味着

$$h_1(\theta_1) = f_{kj}(\theta_1) \text{ and } h_2(\theta_2) = \frac{1}{C} f_{kj}(\theta_2)\left(1+\frac{r}{R}\cos\theta_2\right),$$

其归一化常数为

$$C = \int_0^{2\pi}\left[f_{kj}(\theta_2)\ \left(1+\frac{r}{R}\cos\theta_2\right)\right]\ d\theta_2.$$

因此，式25中的联合概率密度函数代表弯曲环面曲面上的 Kato 和 Jones 分布。现在考虑 $\mu_1 = \mu_2 = \nu_1 = \nu_2 = 0,\ \rho_1 = \rho_2 = 0.3,\ \kappa_1 = \kappa_2 = 1$，以及 $r = 1.5,\ R = 3.$，那么下面的图15 是使用 HAR采样方法从垂直角 θ_2 的分布中抽取的抽样数据直方图。图16 展示了相对于面积测度，从弯曲环面曲面上的冯·米塞斯分布中抽取的样本点散点图；图17则展示了从平坦环面上的冯·米塞斯分布中抽取样本并投影到弯曲环面曲面上的相同结果。

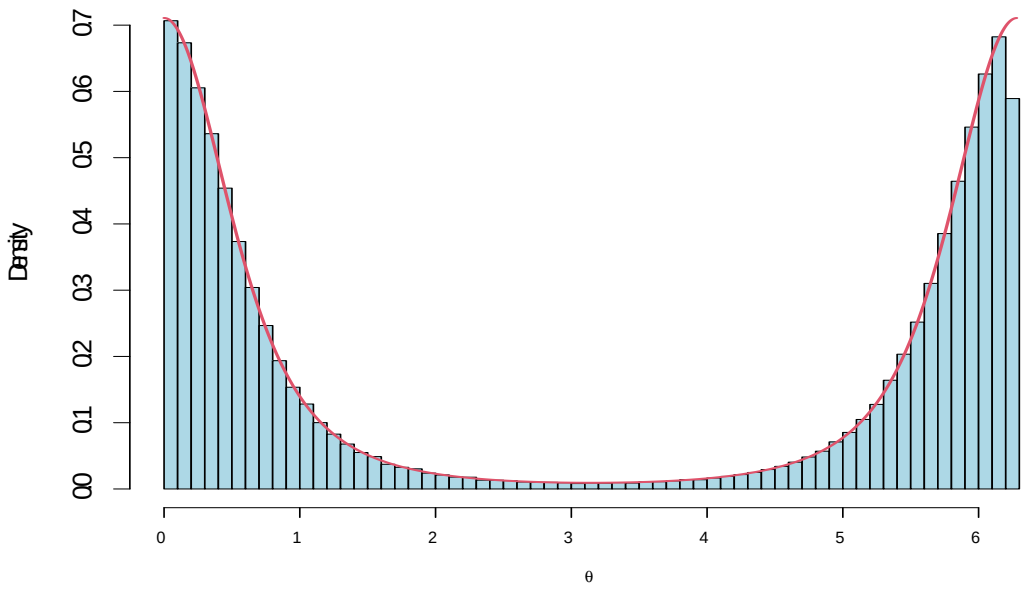

FIGURE 15. 概率密度 $\frac{1}{C} f_{kj}(\theta) \left(1 + \frac{r}{R} \cos\ \theta\right)$ 的数据直方图

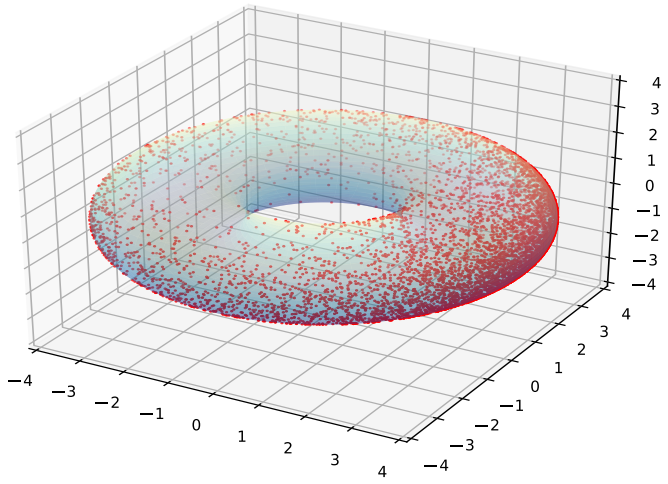

FIGURE 16. 使用面积测度对 Kato 和 Jones 分布进行抽样得到的环面上点的散点图。

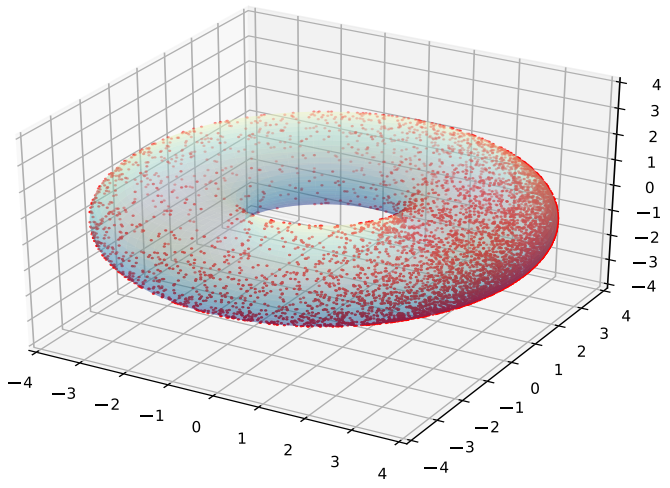

FIGURE 17. 角参数取自平坦环面上的 Kato 和 Jones 分布时，环面上各点的散点图。

备注： 算法-2 可更高效地用于对大量目标密度函数进行批量抽样 $f(x)$，其支撑集为 $[a, b]$ 并满足必要条件。为实现这一点，我们选择最左侧的分区，从该特定单元的包络密度中抽取所需数量的随机样本，并应用接受-拒绝法。接受的样本保留在该单元中，而拒绝的样本则传递到下一个单元。接下来，在下一个单元中，我们计算所需的随机样本数量，并减去该单元中已有的样本数量。然后，我们从包络密度中生成相应数量的随机样本，再次应用接受-拒绝法。接受的样本保留在该单元中，拒绝的样本则移至下一个单元。我们从左至右对所有分区重复这一过程，随后对生成的样本进行随机排列，从而以批次形式获得目标分布的随机样本。表-4，以及表-5 比较了改进的HAR方法批量抽样在不同样本量下的接受率，以及现有vMBFR方法在冯·米塞斯分布下的接受率。

该算法——3 给出了所提出的批量抽样改进的HAR方法的伪代码。

算法 3： HAR 批量采样算法

数据： 目标概率密度函数 $f(x)$，其支撑集为 $[a,b]$。

$na \leftarrow$ 所需的随机样本数;

$np \leftarrow$ 长度相等的分区数量;

$pt \leftarrow$ 一个长度为（$np+1$）的序列，位于 $[a,b]$;

$H \leftarrow f(pt)$; /* Heights of the probability density function at pt */

$H_m \leftarrow$ 每个部分的最大高度序列;

$bl \leftarrow \frac{b-a}{np}$; /* Bin length */

$A \leftarrow \sum_{H_m} H_m * bl$; /* Total Area which is larger than one. */

$p_m \leftarrow \frac{H_m}{\sum H_m}$;

$n \leftarrow \lfloor na * A \rfloor + np$; /* Proposed number of random samples. */

$ng \leftarrow 0$; /* Initializing the count for number of sample generated. */

$y_0 \leftarrow$ 初始化为大小为零的数组;

$y \leftarrow$ 初始化为大小为零的数组;

对于 $i \leftarrow 1$ 至 np **执行**

- $n_i \leftarrow$ 最大值$\{\lfloor n * p_m[i] \rfloor -$ 长度$(y_0)\ 0\}$;
- **如果** $n_i > 0$ **则**
  - $d_i \leftarrow$ 从 $U[0,1]$ 中抽取 n_i 个随机样本;
- **else**
  - 大小为零的$d_i \leftarrow$ 数组;
- **end**
- $ng \leftarrow ng + d_i$;
- $x_i \leftarrow$ marge$(pt[i] + (d_i * bl), (y_0 + bl))$;
- $px \leftarrow \frac{f(x_i)}{H_m[i]}$;
- $rp \leftarrow$ 对于每个 px ，从相应的 $Bernoulli(px)$ 中随机抽取一个样本;
- **若** $rp = 1$ **则**
  - $y_1 \leftarrow x_i$;
- **else**
  - **if** $rp = 0$ **then**
    - $y_0 \leftarrow x_i$;
  - **end**
- **end**
- $y \leftarrow$ 边缘（y，y_1）; /* Desires random sample upto ith cell. */

end

$y \leftarrow$ y 的一个随机排列;

结果： 从所需的概率密度函数 $f(x)$ 中抽取的样本，其支撑集为 $[a,b]$。

κ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
HAR	99.76	99.76	99.73	99.75	99.72	99.74	99.81	99.77	99.73	99.79
vMBFR	99.76	99.06	97.90	96.67	95.04	93.23	91.88	89.88	88.12	86.94

TABLE 4. 批次冯·米塞斯分布的接受率比较表。

κ	2	3	4	5	10	20	40	60	80	100
HAR	99.87	99.81	99.87	99.72	99.68	99.67	99.85	99.93	99.89	99.92
vMBFR	76.95	72.37	69.96	69.46	67.46	66.64	66.43	65.96	65.94	65.69

TABLE 5. 批次冯·米塞斯分布的接受率比较表。

6. CONCLUSION

要理解弯曲环面曲面上的任何分布，就必须从该曲面获得均匀随机样本。本文介绍一种概率变换，无需拒绝任何数据，便可从弯曲环面曲面上的均匀分布抽取随机样本。在此情境下，所提 EAU 采样方法较接受-拒绝采样有显著改进。我们还介绍了利用 HAR 抽样从一些常见圆周分布生成随机样本的新方法。我们将提供积分上黎曼和的阶梯函数转换为有效的概率密度函数，从而在圆周上为目标密度构造了一个非常窄的包络函数。对于平坦环面上边缘密度预先指定的情形，我们构造了一个等效的抽样方案，用于从弯曲环面的表面积抽样。最后，我们将这一思路推广到弯曲环面曲面上的不同分布，在纳入其内蕴几何的同时，以较高接受率抽取随机样本。所提的这些从圆周和弯曲环面曲面生成数据的新方法，不仅为概率研究带来计算优势，也加快了统计推断中的模拟实验。

7. Appendix

7.1. **A1.** 这里给出定理2的证明。

证明。 **第一部分：** 在这种情况下，我们考虑 $U < p(x)$，对于 $X > \pi$ 或 $X < \pi$，我们有 $Y = X$.因此，我们有

$$\begin{aligned} P(Y \leq y) &= \int_0^y \frac{1}{2\pi} P(U < p(x))\, dx \\ &= \frac{1}{2\pi} \int_0^y (1 + a\cos x)\, dx. \end{aligned}$$

因此，积分为

$$P(Y \leq y) = \frac{1}{4\pi} (y + a\sin y)$$

第二部分： 在这种情况下，我们考虑 $U > p(x)$。因此，我们有

$$Y = \begin{cases} \pi - X & \text{when} \quad X < \pi \\ 3\pi - X & \text{when} \quad X > \pi \end{cases}$$

现在，当 $0 < X < \pi$ 时，则 $0 < Y < \pi$。所以，我们有

$$\begin{aligned} P(Y \leq y) &= \int_{\pi - y}^{\pi} \frac{1}{2\pi} P(U > p(x))\, dx \\ &= \frac{1}{4\pi} \int_{\pi - y}^{\pi} (1 - a\cos x)\, dx. \end{aligned}$$

因此，积分为

$$P(Y \leq y) = \frac{1}{4\pi} (y + a\sin y)\,.$$

现在，当 $\pi < X < 2\pi$ 时， $\pi < Y < 2\pi$。因此，我们有

$$\begin{aligned} P(Y \leq y) &= \int_0^{\pi} \frac{1}{2\pi} P(U > p(x))\, dx + \int_{3\pi - y}^{2\pi} \frac{1}{2\pi} P(U > p(x))\, dx \\ &= \frac{1}{4\pi} \int_0^{\pi} (1 - a\cos x)\, dx + \frac{1}{4\pi} \int_{3\pi - y}^{2\pi} (1 - a\cos x)\, dx \end{aligned}$$

因此，积分为

$$P(Y \leq y) = \frac{1}{4\pi} (y + a\sin y)\,.$$

将**第一部分**和**第二部分**中的两个概率相加，得到

$$P(Y \leq y) = \frac{1}{2\pi} (y + a\sin y)\,.$$

因此， $G(y) = \frac{1}{2\pi}(y + a\sin\ y)$，定理得证。 □

7.2. A2.

$$
\begin{aligned}
\frac{1}{M}\int_a^x f(t)\ dt &= \sum_{i=1}^{k}\left[\frac{1}{M}\int_{A_i}\mathbf{I}_{(t\le x)}\ \frac{f(t)}{p(t)}\ p(t)\ dt\right] \\
&= \sum_{i=1}^{k}\left[\frac{H_i}{M}\int_{A_i}\mathbf{I}_{(t\le x)}\ \frac{f(t)}{H_i\,p(t)}\ p(t)\ dt\right] \\
&= \sum_{i=1}^{k}\left[\frac{H_i}{M}\int_{A_i}\mathbf{I}_{(t\le x)}\ \frac{f(t)/P(A_i)}{H_i/P(A_i)}\ \left(\frac{\int_{A_i}p(t)\,dt}{p(t)}\right)\ \left(\frac{p(t)}{\int_{A_i}p(t)\,dt}\right)\ dt\right] \\
&= \sum_{i=1}^{k}\left[\frac{B\,H_i}{M}\int_{A_i}\mathbf{I}_{(t\le x)}\ \frac{f(t)/P(A_i)}{B\,H_i/P(A_i)}\ \left(\frac{1}{\frac{1}{B}}\right)\ \left(\frac{1}{B}\right)\ dt\right] \\
&= \sum_{i=1}^{k}\left[\frac{B\,H_i}{M}\int_{A_i}\mathbf{I}_{(t\le x)}\ \frac{f(t)/P(A_i)}{M_i\ \left(\frac{1}{B}\right)}\ \left(\frac{1}{B}\right)\ dt\right] \\
&= \sum_{i=1}^{k}\left[\frac{B\,H_i}{B\sum_{i=1}^{k}H_i}\int_{A_i}\mathbf{I}_{(t\le x)}\ \frac{f(t)/P(A_i)}{M_i\ \left(\frac{1}{B}\right)}\ \left(\frac{1}{B}\right)\ dt\right] \\
&= \sum_{i=1}^{k}\left[\frac{H_i}{\sum_{i=1}^{k}H_i}\frac{1}{M_i}\left(\frac{1}{1/M_i}\right)\int_{A_i}\mathbf{I}_{(t\le x)}\ \frac{f(t)/P(A_i)}{M_i\ \left(\frac{1}{B}\right)}\ \left(\frac{1}{B}\right)\ dt\right] \\
&= \sum_{i=1}^{k}\left[\frac{H_i}{\sum_{i=1}^{k}H_i}\frac{P(A_i)}{B\,H_i}\left(\frac{1}{1/M_i}\right)\int_{A_i}\mathbf{I}_{(t\le x)}\ \frac{f(t)/P(A_i)}{M_i\ \left(\frac{1}{B}\right)}\ \left(\frac{1}{B}\right)\ dt\right] \\
&= \sum_{i=1}^{k}\left(\frac{P(A_i)}{M}\right)\left[\frac{\int_{A_i}\mathbf{I}_{(t\le x)}\ P\left(U < \frac{f(t)/P(A_i)}{M_i(\frac{1}{B})}\right)\ \left(\frac{1}{B}\right)\ dt}{\frac{1}{1/M_i}}\right]
\end{aligned}
$$

(26)

7.3. A3. 我们将求出环面曲面上冯·米塞斯分布的归一化常数 C。为此，我们将利用以下两个贝塞尔函数恒等式。

$$\frac{1}{2\pi}\int_0^{2\pi} e^{\kappa\cos\theta}\cos p\theta\ d\theta = I_p(\kappa), \tag{27}$$

and

$$\frac{1}{2\pi}\int_0^{2\pi} e^{\kappa\cos\theta}\sin n\theta\ d\theta = 0. \tag{28}$$

现在，常数 C 可以写为

$$\begin{aligned} C &= \int_0^{2\pi}\left[e^{\kappa\cos(\theta-\mu)}\left(1+\frac{r}{R}\cos\theta\right)\right]\ d\theta \\ &= \int_0^{2\pi} e^{\kappa\cos(\theta-\mu)}\ d\theta + \int_0^{2\pi}\left(1+\frac{r}{R}\cos\theta\right)\ d\theta \\ &= C_1 + C_2, \end{aligned} \tag{29}$$

其中，利用式 27可得

$$C_1 = \int_0^{2\pi} e^{\kappa\cos(\theta-\mu)}\ d\theta = 2\pi I_0(\kappa), \tag{30}$$

and

$$C_2 = \frac{r}{R}\int_0^{2\pi} e^{\kappa\cos(\theta-\mu)}\cos\theta\ d\theta.$$

现在，对于 C_2，令 $\theta-\mu=t \implies d\theta=dt$，且 $\theta=\mu+t$. 因此， C_2 变为

$$\begin{aligned} C_2 &= \frac{r}{R}\int_{-\mu}^{2\pi-\mu} e^{\kappa\cos t}\cos(\mu+t)\ dt \\ &= \frac{r}{R}\int_{-\mu}^{2\pi-\mu} e^{\kappa\cos t}(\cos t\cos\mu-\sin t\sin\mu)\ dt \\ &= \frac{r\cos\mu}{R}\int_{-\mu}^{2\pi-\mu} e^{\kappa\cos t}\cos t\ dt + \frac{r\sin\mu}{R}\int_{-\mu}^{2\pi-\mu} e^{\kappa\cos t}\sin t\ dt \end{aligned}$$

现在，利用式 27和式 28 可写为

$$C_2 = \frac{r}{R}\cos\mu\ 2\pi I_1(\kappa). \tag{31}$$

将数值代入式30和式31 ，再代入式29 ，得到

$$C = 2\pi\left[I_0(\kappa)+\frac{r}{R}\cos\mu\ I_1(\kappa)\right].$$

References

B´elisle, C. J., Romeijn, H. E., and Smith, R. L. (1993). 用于生成多元分布的随机游走算法。 *Mathematics of Operations Research*, 18(2), 255–266. Beran, R. (1979). 方向数据的指数模型。 *The Annals of Statistics*, pages 1162–1178. Best, D. and Fisher, N. I. (1979). 冯·米塞斯分布的高效模拟。 *Journal of the Royal StatisticalSociety: Series C (Applied Statistics)*, 28(2), 152–157. Bhattacharya, R. and Patrangenaru, V. (2003). 流形上内在和外在样本均值的大样本理论。 *The Annals of Statistics*, 31(1), 1–29. Boender, C., Caron, R. J., McDonald, J. F., Kan, A. R., Romeijn, H. E., Smith, R. L., Telgen, J., and Vorst, A. (1991). 用于在有界多面体边界上生成均匀点的Shake-and-bake算法。*Operations research*, 39(6), 945–954. Carilli, C., Perley, R., Dhawan, V., and Perley, D. (2019). 对天鹅座A活动星系核环面进行成像。 *The Astrophysical Journal Letters*, 874(2), L32. Ciccotti, G. and Ryckaert, J.-P. (1986). 刚性分子的分子动力学模拟。 *Computer Physics Reports*, 4(6), 346–392. Diaconis, P. and Shahshahani, M. (1986). 关于紧群上均匀分布的平方根。 *Proceedingsof the American Mathematical Society*, 98(2), 341–348. Diaconis, P., Holmes, S., Shahshahani, M., *et al.* (2013). 从流形上抽样。 *Advances in modern statistical theory and applications: a Festschrift in honor of Morris L. Eaton*, 10, 102–125. Eaton, M. L. (1983). 多元统计学: 向量空间方法。 *JOHN WILEY & SONS, INC., 605 THIRDAVE., NEW YORK, NY 10158, USA*, 1983, 512 . Federer, H. (2014). *Geometric measure theory*. Springer. Fisher, N. I., Lewis, T., and Embleton, B. J. (1993). *Statistical analysis of spherical data*. 剑桥大学出版社. Fixman, M. (1974). 约束的经典统计力学：一个定理及其在聚合物中的应用。 *Proceedingsof the National Academy of Sciences*, 71(8), 3050–3053. Hingorani, M. M. and O' Donnell, M. (1998). 环状蛋白质：围绕DNA转圈。 *Current Biology*, 8(3), R83–R86. Hingorani, M. M. and O' Donnell, M. (2000). DNA代谢中的环状体故事。 *Nature Reviews Molecular CellBiology*, 1(1), 22–30. Kato, S. and Jones, M. (2010). 一类与莫比乌斯变换相关并由其应用引出的圆上的分布。 *Journal of the American Statistical Association*, 105(489), 249–262. Lalley, S. and Robbins, H. (1987). 平面上的渐近极小极大随机搜索策略。 *Proceedings ofthe National Academy of Sciences*, 84(8), 2111–2112. Mardia, K. V., Jupp, P. E., and Mardia, K. (2000). *Directional statistics*, volume 2. Wiley Online Library. Pennec, X. (2006). 黎曼流形上的内在统计学：几何测量的基本工具。 *Journal ofMathematical Imaging and Vision*, 25, 127–154. Selkoe, D. J. (2003). 致命方式下的蛋白质折叠。 *nature*, 426(6968), 900–904. Watson, G., S. (1983). *Statistics on spheres*. Wiley, New York.