多参数持久景观的置信带

Inés García-Redondo[*] Anthea Monod[†] Qiquan Wang[‡]

2025 年 4 月 6 日

摘要

多参数持久同调是拓扑数据分析中一种中心且广泛应用的方法——经典持久同调的推广，它考虑了密度估计，并且是在存在噪声的情况下进行数据分析的有效工具。与经典的单参数对应方法类似，由于其复杂的代数结构，计算和实际使用起来颇具挑战性。在本文中，我们在统计设置下研究了一种流行的多参数持久同调不变量：多参数持续景观。我们推导出了多参数持续景观的功能中心极限定理，并由此计算了置信带，这构成了首个针对多参数持续景观的统计推断方法之一。我们提供了一个置信带的实现，并展示了其在合成数据上的机器学习任务中的应用。

1 介绍

持久同调 (PH) 是来自拓扑数据分析（TDA）的一种方法论，在过去几十年中，作为一种有效的数据分析、统计和机器学习工具，在各个应用领域取得了广泛的成功。它提供了一个强大的框架和计算效率高的算法，用于从数据中提取拓扑描述符，旨在捕捉其潜在的形状和结构。核心思想是构造一系列嵌套的拓扑空间 $\{K_t : t \in T\}$，称为过滤，通常由实线的一个区间 $T \subset \mathbb{R}$ 索引，并且满足当 $t \leq s$ 时 $K_t \subset K_s$。通过跟踪拓扑特征（用同调表征）随着

[*]LSGNT, Imperial College, London, UK; i.garcia-redondo22@imperial.ac.uk

[†]Imperial College, London, UK; a.monod@imperial.ac.uk

[‡]Imperial College, London, UK; qiquan.wang17@imperial.ac.uk

过滤参数变化而出现和消失的情况，计算出适用于数据分析的拓扑不变量。此过程处理数据并输出一个持续图谱，这是一个统计摘要，捕获其拓扑特征的寿命。作为随机对象，持久性图的统计和概率性质一直是 TDA 内部丰富且具有挑战性的研究领域，因为持久性图的代数构造在所有持久性图的空间中诱导出复杂的非欧几里得几何。应对这种复杂几何的一种常见方法是将持久性图向量化或嵌入到具有更适合统计学和机器学习方法的标准几何的欧几里得或其他空间中；最受欢迎的向量化之一被称为持久景观 [Bub15] 。

上述经典的单参数 PH 设置可以扩展到多参数持续同调 (MPH)，其中 T 变成一个多重集合，通常 $\mathbb{R}^d$ 采用乘积序。实际上，这种扩展将额外的结构信息—如密度估计—融入过滤，并且是在有噪声存在的情况下对经典设置的有效适应。MPH 在真实数据环境中计算和使用显著更具挑战性，因为它缺乏与单参数持久性中的持久图相对应的自然不变量。寻找有意义且可处理的不变量仍然是这一领域活跃且不断发展的前沿。多参数持久景观 [Vip20] 是这样一个不变量，它同时也是经典持久景观的一种扩展。因此，已知的持久景观的功能和统计特性可以延续到多参数设置 [Bub15, Vip20] 中。

本文将置信带的统计框架从单参数 [CFL$^+$14a, CFL$^+$14b] 扩展到多参数持久景观。我们的工作是首批为 MPH 开发统计方法的研究之一。我们主要的贡献包括：(i) 多参数持久景观的新泛函中心极限定理 (CLT)；(ii) 基于 CLT 的多参数持久景观的置信区间，以及 (iii) 一种用于在合成数据上实现置信区间的算法，展示它们在分类任务中的应用。

2 背景

我们首先提供我们的工作背景和相关背景。

2.1 持久同调与景观

我们概述了使用多参数持久景观进行持续同调的向量化。

持久模块和同调。在持久同调理论中的中心对象是持久模块，一组由 $\mathbb{R}^d$ 按照乘积序索引的向量空间 $M_\bullet = \{M_{\boldsymbol{x}} : \boldsymbol{x} \in \mathbb{R}^d\}$，以及线性映射 $\varphi_{\boldsymbol{y}}^{\boldsymbol{x}} : M_{\boldsymbol{x}} \rightarrow$

$M_{\boldsymbol{y}}$，对于 $\boldsymbol{x} \leq \boldsymbol{y}$ 称为内部或转换映射。对于本文的其余部分，我们将假设参数取值在一个有界箱体 $B(T_1, \ldots, T_d) := [0, T_1] \times \cdots \times [0, T_d] \subset \mathbb{R}^d$ 中。在每个方向上进行归一化后，我们可以简单地考虑箱体 $[0, T]^d$，其中 $T = \max\{T_i : 1 \leq i \leq d\}$。

持久模块通过从输入数据构造的过滤 $K_\bullet = \{K_{\boldsymbol{x}:\boldsymbol{x}\in\mathbb{R}^d}\}$ 中的拓扑空间取同调群而在 PH 中出现。对于拓扑空间 K，k 同调群 $\mathrm{H}_k(K)$ 包含有关 K 的拓扑特征的信息：对于 $k = 0$，这些是它的连通分量；对于 $k = 1$，它的环路；对于 $k = 2$，它的气泡和空腔，等等，对于更高值的 $k \geq 0$ 也是如此。这些向量空间的维度，$\beta_k = \dim \mathrm{H}_k(K)$，对应于这些拓扑特征的数量。

多参数持久景观。给定一个持久模块 $M_\bullet = \{M_{\boldsymbol{x}} : \boldsymbol{x} \in \mathbb{R}^d\}$，其秩不变量定义为内部映射 $\beta_{\boldsymbol{y}}^{\boldsymbol{x}} = \dim(\mathrm{Im}(\varphi_{\boldsymbol{y}}^{\boldsymbol{x}}))$ 的秩作为映射 $\mathrm{rk} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$，使得

$$\mathrm{rk}(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} \beta_{\boldsymbol{y}}^{\boldsymbol{x}} & \text{if } \boldsymbol{x} \leq \boldsymbol{y}, \\ 0 & \text{otherwise.} \end{cases}$$

可以通过以下方式对这个函数进行重新缩放来获得持久景观。

定义 2.1 (多参数持久景观 [Bub15, Vip20])**.** 给定一个持久模块 $M_\bullet = \{M_{\boldsymbol{x}} : \boldsymbol{x} \in \mathbb{R}^d\}$，其持久景观 $\lambda : \mathbb{N} \times \mathbb{R}^d \to \mathbb{R}$ 被定义为

$$\lambda(k, \boldsymbol{x}) = \sup\{\epsilon > 0 : \beta_{\boldsymbol{x}+\boldsymbol{h}}^{\boldsymbol{x}-\boldsymbol{h}} \geq k \text{ for all } \boldsymbol{h} \geq \boldsymbol{0} \text{ with } \|\boldsymbol{h}\|_\infty \leq \epsilon\}.$$

换句话说，第 k 个持续景观输出的是最大半径（在 ℓ^∞ 度量下，$\mathbb{R}^d$）中的 k 拓扑特征在每个 $\boldsymbol{x} \in \mathbb{R}^d$ 的所有正方向上持续存在的范围）。从 [Vip20] 的引理 20 以及 $\lambda(k, \boldsymbol{x}) \geq 0$ 可知，它是 $\boldsymbol{x}$ 中的一个 1-利普希茨函数对于所有 $k \geq 1$。本文重点研究 $\lambda(\boldsymbol{x}) := \lambda(1, \boldsymbol{x})$，但结果适用于任何固定的 k，关键属性是 $\lambda(\boldsymbol{x})$ 是 1-利普希茨。我们令 $\mathcal{L}$ 为参数在有界盒 $[0, T]^d \subset \mathbb{R}^d$ 中的持久模块的景观空间 $\lambda : [0, T]^d \to \mathbb{R}$，它是 Banach 空间 $L^p([0, T]^d)$,$1 \leq p \leq \infty$ 的一个子集。

2.2 中心极限定理和持久景观的自助置信带

我们现在讨论单参数和多参数景观作为随机变量（r.v.）在 Banach 空间中的已知收敛性，这是由中心极限定理给出的。我们还概述了使用自助法构建置信带的方法——这与假设检验是对偶构造。

中心极限定理（CLT） 对于单参数和多参数持久景观均成立 [Bub15, Vip20]：假设 X 是某个概率空间上的一个波莱尔可测随机变量；则 X 的持续景观，记为 $\Lambda = \Lambda(X)$，是一个巴拿赫空间值的随机变量。令 $\{X_i\}$ 为 X 和 Λ_i 的独立同分布副本，然后给定 $\mathbb{E}[\|\Lambda\|] < \infty$ 和 $\mathbb{E}[\|\Lambda^2\|] < \infty$ 我们有 $\sqrt{n}(\bar{\Lambda}_n - I_\Lambda)$，其中 I_Λ 是 Λ 的佩蒂斯积分，弱收敛于 $G(\Lambda)$，一个具有与 Λ 相同协方差结构的高斯随机变量。

对于单参数情况，存在一个更强的函数中心极限定理（也称为 *Donsker* 定理）[CFL$^+$14b]：

$$\{\mathbb{G}_n(\boldsymbol{x})\}_{\boldsymbol{x}\in[0,T]} := \{\sqrt{n}(\overline{\Lambda}_n(\boldsymbol{x}) - I_\Lambda(\boldsymbol{x}))\}_{\boldsymbol{x}\in[0,T]}$$

弱收敛于 $[0, T]$ 上的高斯过程。高斯过程是实值随机变量族，使得该族中的任何有限集合都遵循多元正态分布。弱收敛于极限高斯过程允许构建置信带和假设检验，这可以使用非参数方法（如 bootstrap）来实现。本文的主要贡献是在第 3节将该框架扩展到多参数情况，并在第 4节展示其效用。

自助法是一种推断技术，用于通过检查样本参数与大量重采样计算出的统计量之间的关系来从样本统计量估计总体参数。在本文中，我们重点讨论使用自助法构建持久性景观的信心带。对于给定的功能参数 $\theta : \Omega \to \mathbb{R}$，水平为 α 的参数的置信带由一对函数 $l, u : \Omega \to \mathbb{R}$ 给出，使得 $\mathbb{P}(\theta(\omega) \in [l(\omega), u(\omega)], \forall \omega \in \Omega) \geq 1 - \alpha$。基于公式 [CFL$^+$14b, CFL$^+$14a] 中的单参数方法，我们推导出均值多参数持久景观的置信带，详见算法 1。

3 多参数持续景观的功能性中心极限定理

我们扩展了 [CFL$^+$14a, CFL$^+$14b] 中的理论，以证明多参数持续景观的功能性 CLT。令 $\Lambda_1, \ldots, \Lambda_n$ 是来自多参数持续景观空间 $\mathcal{L}$ 上的概率分布 P

的一个独立同分布样本。回忆一下我们有平均景观 $I_\Lambda = \mathbb{E}_{\Lambda\sim P}[\Lambda(\boldsymbol{x})]$ 和样本均值 $\overline{\Lambda}_n(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n}\Lambda_i(\boldsymbol{x})$。

我们考虑可测函数族 $\mathcal{F} = \{f_{\boldsymbol{x}:\mathcal{L}\to\mathbb{R},\,\lambda\mapsto f_{\boldsymbol{x}(\lambda)=\lambda(\boldsymbol{x})}\}_{\boldsymbol{x}\in[0,T]^d}$ 并定义由该族索引的经验过程：

$$\{\mathbb{G}_n f_{\boldsymbol{x}}\}_{f_{\boldsymbol{x}}\in\mathcal{F}:=\{\mathbb{G}_n(\boldsymbol{x})\}_{\boldsymbol{x}\in[0,T]^d}=\{\sqrt{n}(\overline{\lambda}_n(\boldsymbol{x})-\mu(\boldsymbol{x}))\}_{\boldsymbol{x}\in[0,T]^d}},(1)$$

我们可以将其重写为由 $[0,T]^d$ ($\mathcal{L}$ 中的函数域) 索引，以简化表示。一个由 $\mathcal{F}$ 弱收敛 到 Y 索引的随机过程 $\{Y_n\}_{f\in\mathcal{F}}$，在 $\ell^\infty(\mathcal{F})$ 中表示为 $Y_n \rightsquigarrow Y$，如果对于每一个有界连续函数 $g:\ell^\infty(\mathcal{F})\to\mathbb{R}$ 都有 $\lim_{n\to\infty}\mathbb{E}^*[g(Y_n)] = \mathbb{E}[g(Y)]$；其中 $\mathbb{E}^*$ 表示外期望。在这里使用外部期望是一个技术细节，这是在 Y_n 不是波莱尔可测的情况下所必需的，但这对本文其余部分来说并不关键。

定理 3.1 (多参数持久景观的一致收敛)**.** 令 $\mathbb{G}$ 是一个以 $\boldsymbol{x}\in[0,T]^d$ 为索引的高斯过程，均值为零，协方差函数为 $\kappa(\boldsymbol{x},\boldsymbol{y}) = \int\lambda(\boldsymbol{x})\lambda(\boldsymbol{y})dP(\lambda) - \int\lambda(\boldsymbol{x})dP(\lambda)\int\lambda(\boldsymbol{y})dP(\lambda).$。则 $\mathbb{G}_n \rightsquigarrow \mathbb{G}$，其中 $\mathbb{G}_n$ 是由来自 (1) 的独立同分布样本定义的经验过程。

我们需要来自 [Kos08] 的定理 2.5 中的以下引理，以便证明定理 3.1。此引理提供了函数族 $\mathcal{F}$ 上的一个充分条件，以确保由该函数族索引的经验过程收敛到一个高斯过程。给定 Q 上的概率 $\mathcal{L}$，我们用它来定义函数类 $\mathcal{F}$ 上的距离 $L^2(Q)$ 为 $\|f-g\|_{Q,2}^2 = \int|f-g|^2\,dQ$。令 $N(\mathcal{F},L_2(Q),\epsilon)$ 为 $\mathcal{F}$ 的覆盖数，即在这种度量下的最小 ϵ-网的大小。最后，回想一下，给定一组可测函数 $\mathcal{F}=\{f:\Omega\to\mathbb{R}\}$，该集合的可测包络是“最小”的可测函数 $F:\Omega\to\mathbb{R}$，使得对于几乎每一个 $\omega\in\Omega$ 都有 $f(\omega)\le F(\omega)$。

引理 3.2. 设 $\mathcal{F}$ 是一类可测函数，满足

$$\int_0^1\sqrt{\log\sup_Q N(\mathcal{F},L_2(Q),\epsilon\|F\|_{Q,2})}\,d\epsilon<\infty$$

其中 F 是 $\mathcal{F}$ 的一个可测包络，上确界取遍所有具有 $\|F\|_{Q,2}>0$ 的有限离散概率测度 Q。如果 $\mathbb{E}[F^2]<\infty$，则 $\mathbb{G}_n \rightsquigarrow \mathbb{G}$，其中 $\{\mathbb{G}_n f\}_{f\in\mathcal{F}}$ 是由 $\mathcal{F}$ 指数的经验过程，而 $\mathbb{G}$ 是具有零均值和协方差函数的高斯过程。$\kappa(\boldsymbol{x},\boldsymbol{y}) = \int\lambda(\boldsymbol{x})\lambda(\boldsymbol{y})dP(\lambda) - \int\lambda(\boldsymbol{x})dP(\lambda)\int\lambda(\boldsymbol{y})dP(\lambda).$

定理 3.1. 定义 2.1直接暗示任何景观 $\lambda \in \mathcal{L}$ 在任意 $\boldsymbol{x} \in [0,T]^d$ 处的值被上限限制为最大的球体（在 ℓ^∞ 度量下的 $\mathbb{R}^d$）在 $[0,T]^d$ 中的半径，其等于 $T/2$。从这一观察中，我们得出结论，$F : \mathcal{L} \to \mathbb{R}, F(\lambda) = T/2$ 是 $\mathcal{F} = \{f_{\boldsymbol{x}:\mathcal{L}\to\mathbb{R},\, \lambda\mapsto\lambda(\boldsymbol{x})}\}_{\boldsymbol{x}\in[0,T]^d}$ 的可测包络。

令 Q 是在 $\mathcal{L}$ 上的测度，并且 $\boldsymbol{x}, \boldsymbol{y} \in [0,T]^d$。$L^2(Q)$ 距离满足 $f_{\boldsymbol{x},\, f_{\boldsymbol{y}\in\mathcal{F}}}$

$$\begin{aligned}\|f_{\boldsymbol{x}-f_{\boldsymbol{y}}}\|_{Q,2}^2 &= \int_{\mathcal{L}} |f_{\boldsymbol{x}(\lambda)-f_{\boldsymbol{y}(\lambda)|^2\, dQ(\lambda)}} \\ &= \int_{\mathcal{L}} |\lambda(\boldsymbol{x}) - \lambda(\boldsymbol{y})|^2 \, dQ(\lambda) \\ &\le \int_{\mathcal{L}} \|\boldsymbol{x}-\boldsymbol{y}\|_\infty^2 \, dQ(\lambda) = \|\boldsymbol{x}-\boldsymbol{y}\|_\infty^2 .\end{aligned}$$

我们考虑一个由 $0 = t_0 < t_1 < \cdots < t_N = T$ 表示的区间 $[0,T]$ 的划分 G，使得 $|t_i - t_{i+1}| = \epsilon\|F\|_{Q,2} = \frac{\epsilon T}{2}$，其中 F 如上定义为 $\mathcal{F}$ 的可测包络。令 G^d 为在这种划分下获得的相应网格 $[0,T]^d$。我们声称 $\{f_{\boldsymbol{t}:\boldsymbol{t}\in G^d\}}$ 是 $\mathcal{F}$ 的一个 $(\epsilon T/2)$-网：对于任意的 $f_{\boldsymbol{x}\in\mathcal{F}}$，存在某个 $\boldsymbol{t} = (t_{i_1}, \ldots, t_{i_d}) \in G^d$ 使得对于 $1 \le k \le d$ 成立 $t_{i_k} \le x_k \le t_{i_k+1}$。然后，我们有 $\|f_{\boldsymbol{x}-f_{\boldsymbol{t}}}\|_{Q,2} \le \|\boldsymbol{x}-\boldsymbol{t}\|_\infty = \frac{\epsilon T}{2}$，证明这个函数族是一个 $(\epsilon T/2)$-网。

我们可以将长度为 $\epsilon T/2$ 的 $2/\epsilon$ 个区间放入 $[0,T]$ 中，这意味着对于任何概率测度 Q，$(2/\epsilon)^d$ 是 $N(\mathcal{F}, L_2(Q), \epsilon\|F\|_{Q,2})$ 的上界，因此也是 Q 上确界的上界。这意味着

$$\int_0^1 \sqrt{\log \sup_Q N(\mathcal{F}, L_2(Q), \epsilon\|F\|_{Q,2})}\, d\epsilon = \int_0^1 \sqrt{d(\log 2 - \log \epsilon)}\, d\epsilon < +\infty.$$

由于可测包络的平方显然是可积的，使用引理 3.2，我们得出结论，由 $\mathcal{F}$ 索引的经验过程在 (1) 中收敛到定理 3.1中所述的高斯过程。 □

4 持久景观的置信带：方法与模拟

我们现在给出一个用于推导多参数持久景观置信带的算法；其实现及随后的实验可在以下位置获得:`https://github.com/inesgare/bands-mph-landscapes`。

我们通过一个实际用例展示了置信区间，该用例涉及从三个不同的拓扑空间中抽取 $N = 500$ 个样本点：一个球面（半径为 $R = 3$），一个环面（半

Algorithm 1 自助置信区间计算（标准和乘数自助法）

Require: 数据集或多重参数景观 $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$，自助样本数量 B，置信水平 α

Ensure: 置信带 $\ell_n, u_n : [0,T]^d \to \mathbb{R}$

计算 $\overline{\lambda}_n = n^{-1} \sum_{i=1}^{n} \lambda_i$

for $b = 1$ to B **do**

  if using standard bootstrap **then**

    抽取一个自助样本 $\{\lambda_1^*, \ldots, \lambda_{\lfloor n/2 \rfloor}^*\}$ 通过有放回抽样

    设定 $\hat{\theta}_b^* = \sup_{\boldsymbol{x} \in [0,T]^d} \left| \sqrt{n}(\overline{\lambda}_{\lfloor n/2 \rfloor}^*(\boldsymbol{x}) - \overline{\lambda}_n(\boldsymbol{x})) \right|$

  else if using multiplier bootstrap **then**

    生成一组独立同分布的乘子 $\xi_1, \ldots \xi_n \sim N(0,1)$

    集合 $\hat{\theta}_b^* = \sup_{\boldsymbol{x} \in [0,T]^d} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} \xi_i (\lambda_i^*(\boldsymbol{x}) - \overline{\lambda}_n(\boldsymbol{x})) \right|$

  end if

end for

定义 $\tilde{Z}(\alpha) = \inf \left\{ z : B^{-1} \sum_{b=1}^{B} I(\hat{\theta}_b^* > z) \leq \alpha \right\}$

设定 $\ell_n(\boldsymbol{x}) = \overline{\lambda}_n(\boldsymbol{x}) - \frac{\tilde{Z}(\alpha)}{\sqrt{n}}$ 和 $u_n(\boldsymbol{x}) = \overline{\lambda}_n(\boldsymbol{x}) + \frac{\tilde{Z}(\alpha)}{\sqrt{n}}$

表 1: **五折交叉验证后的平均准确率** 对于使用标准或乘数自助法的 MBD 分类器，在单参数（SPH）或多重参数（MPH）持久性中。模型是通过对每个形状类别的 n 个子样本进行训练的：球体、环面和克莱因瓶，如图 1所示。

	SPH Standard	SPH Multiplier	MPH Standard	MPH Multiplier
$n = 100$	0.97 ± 0.02	0.93 ± 0.03	1.00 ± 0.00	1.00 ± 0.00
$n = 500$	0.92 ± 0.01	0.88 ± 0.02	0.997 ± 0.003	0.985 ± 0.004

径为 $R = 3$,$r = 0.7$）以及一个克莱因瓶，所有这些都嵌入在 $\mathbb{R}^3$ 中。为了引入噪声，我们应用了尺度为 0.1 的高斯噪声和椒盐噪声，其中 0.5% 的点被随机移位最多 0.5 单位。使用多视角包的 [LS24] 来计算多参数景观，构建了二维过滤器，其中一个参数是尺度（用于 Vietoris–Rips 过滤），另一个参数是一个核密度估计值（用于超水平集过滤）。我们考虑一维同调的第一个景观，即捕捉形状内最大的环的景观。输入点云的三个实例及其相应的平均景观和使用标准自助法（参见算法 1）的置信带对于 $n = 100$ 样本，在图 1中进行了说明。

为了展示置信带的实用性，我们还训练了一个最大深度带（MDB）功能分类器。景观的深度由其在每个形状的置信带内出现的频率定义。我们根据最大化此深度的类别对景观进行分类。我们在每类中有 $n = 100$ 和 $n = 500$ 个景观进行了分类，并在经过五折验证后，在表 1中报告了准确率。

5 讨论

本文建立了多参数持久景观置信带的理论基础和方法论，推进了拓扑数据分析这一新兴子领域的统计工具的发展。置信带具有内在的统计价值；我们进一步展示了它们在分类任务中的实际效用。我们的结果显示，该方法实现了近乎完美的准确率，且 MPH 的表现优于 SPH。然而，随着子样本数量的增加，准确率略有下降，这导致了更窄的置信带。

我们工作的主要限制在于计算多参数持续景观的计算瓶颈，这使得我们的方法本质上受到 MPH 挑战的约束。未来研究的一个关键方向是降低计算成本，然后将此方法应用于实际数据。

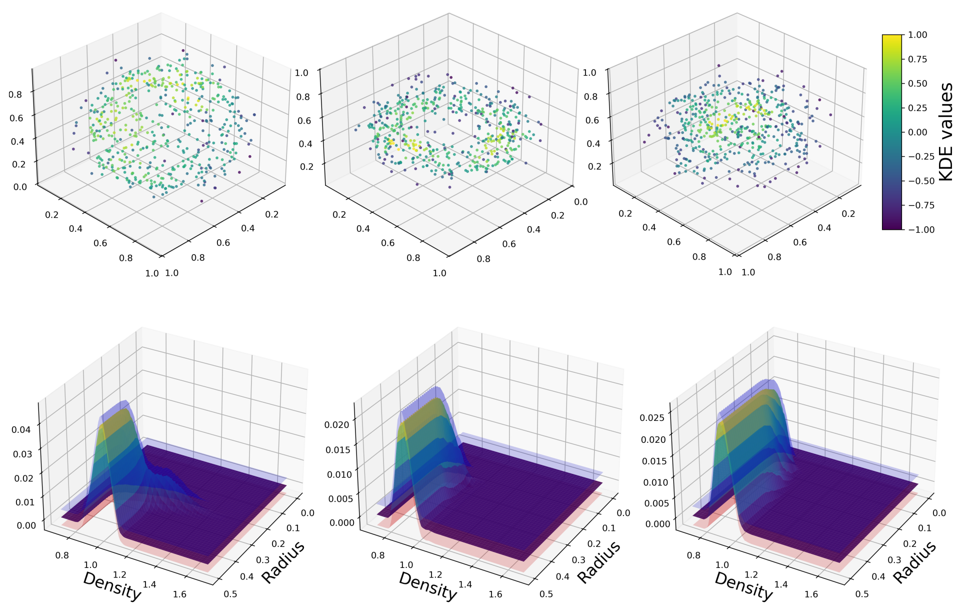

图 1: **输入点云并用标准自助法置信区间平均地貌**。*以上*：半径为 $R = 3$ 的球面上有 $N = 500$ 个点（左），一个具有半径 $R = 3$ 和 $r = 0.7$ 的环面（中），以及一个克莱因瓶（右），并加入了高斯噪声和椒盐噪声。点的颜色由核密度估计值决定。*以下*：标准 Bootstrap 的 $n = 100$ 个样本的平均样本景观和置信带。

致谢。 我们感谢 Jonathan Lau 的有趣讨论，这些讨论激发了这项工作。I.G.R. 受伦敦几何与数论学院——帝国理工学院博士生奖学金资助，该奖学金由 UKRI EPSRC [EP/S021590/1] 提供支持。A.M. 由 UKRI EPSRC 资助 [EP/Y028872/1], 智能的数学基础：AI 的“埃尔朗根纲领”。Q.W. 受癌症研究英国和帝国理工学院融合科学博士生奖学金 (2021 届, PIs Monod/Williams) 资助，该奖学金由癌症研究英国在 [CANTAC721\10021] 奖项下提供支持。

参考文献

[Bub15] Peter Bubenik. Statistical Topological Data Analysis using Persistence Landscapes. *Journal of Machine Learning Research*, 16(3):77–102, 2015.

[CFL+14a] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Aarti Singh, and Larry Wasserman. On the Bootstrap for Persistence Diagrams and Landscapes, January 2014.

[CFL+14b] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman. Stochastic Convergence of Persistence Landscapes and Silhouettes. In *Proceedings of the Thirtieth Annual Symposium on Computational Geometry*, SOCG'14, pages 474–483, New York, NY, USA, June 2014. Association for Computing Machinery.

[Kos08] Michael R. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. Springer, New York, NY, 2008.

[LS24] David Loiseaux and Hannah Schreiber. Multipers: Multiparameter Persistence for Machine Learning. *Journal of Open Source Software*, 9(103):6773, November 2024.

[Vip20] Oliver Vipond. Multiparameter Persistence Landscapes. *Journal of Machine Learning Research*, 21(61):1–38, 2020.