arXiv:2603.01670v1 [math.PR] 2 Mar 2026

# 行列式点过程离散到连续极限的统计一致性

作者：Hugo Jaquard[1], [a] 和 Nicolas Keriven[1], [b]

[1]法国国家科学研究中心，*IRISA*，a*hugo.jaquard@irisa.fr*；b*nicolas.keriven@cnrs.fr*

我们研究从中进行采样的集合大小趋于无穷时，离散行列式点过程（DPP）趋向连续行列式点过程的极限行为。我们提出了对这一极限的非渐近刻画，通过考察与这些过程相关的统计量的集中性来描述，并将其称为“弱相干性”。这使我们能够将极限过程的统计保证转化为原始离散过程的统计保证。我们的主要结果给出了弱相干性成立的充分条件。特别地，我们的研究涵盖了连续过程的核及其底层空间均不可访问的情形，也涵盖了离散边缘核是其连续对应核的含噪版本的情形。

我们通过若干示例阐释我们的理论。我们证明，离散多元正交多项式系综可以用于构造在误差相同的情况下严格 小于独立采样的核心集。我们提出了一种过程，可以根据从未知密度中抽取的一组点，在未知 流形上实现排斥抽样。最后，我们展示了连续行列式点过程可以作为具有伯努利边的随机 图上的极限得到，即使只观察图结构也不例外。我们还由此得到了一些有趣的附带结果。

**1. 引言。** 行列式点过程（DPP）是一类定义在空间 $\Gamma$ 的随机点集上的概率分布，这些点集可能是无限的，并且点之间可以呈现排斥相互作用；粗略地说，这意味着，与相距较远的两个点相比，彼此靠近的两个点同时被纳入样本的可能性较低。由于在这类模型中兼具独特的理论可处理性和数值可处理性，行列式过程在当前的排斥过程研究领域中占据着特殊地位。O. Macchi 最初在数学物理中提出了这类过程，将其用作描述电子探测时间中观测到的反聚束现象的工具（历史回顾见 [7] ）；此后，它们一直是概率论 [63, 62], 、机器学习 [43], 、空间统计学 [46], 乃至数值分析和信号处理 [10]中持续研究的对象；当然，这份清单远非穷尽。

通常将行列式点过程分为两类。一类是在有限 集合 $\Gamma$ 上定义的行列式点过程，称为离散 行列式点过程。例如，在计算统计学中，离散行列式点过程常用于从包含 $n$ 个 $\Gamma = X_n$ 数据点的数据集中抽取子样本，以应对数据集过大、无法通过特定处理流程处理，或无法对其进行完整的人工标注等情况。与同等规模的独立采样相比，点间排斥有助于获得能够更全面覆盖数据集的子样本。离散行列式点过程易于定义和抽样 [34, 1, 11]。与之相对，当 $\Gamma = \mathcal{X}$ 是一个“连续”空间（例如$\mathbf{R}^d$ 或某个流形）时，我们称之为连续 行列式点过程。连续行列式点过程可借助精细的分析工具进行研究，从而确立其良好的统计性质（例如， [10, 8, 52]），但与离散情形相比，通常更难处理。

尽管这种区别更多是实践层面而非数学层面的，但对实践者而言，抽样离散 DPP 与抽样连续行列式点过程之间通常有着天壤之别。我们能否

在抽样离散行列式点过程时，利用连续行列式点过程的统计保证？更确切地说，我们研究以下这个自然的问题。

> 假设数据 $X_n$ 是从底层空间 $\boldsymbol{X}$ 中抽取的（特别是对于独立同分布（*iid*）数据），当规模 $n \to \infty$ 时，定义在 $X_n$ 上的离散 *DPP* 能否对 $\boldsymbol{X}$ 上的连续行列式点过程进行“良好近似”？

进一步说，能否利用这种近似，将连续行列式点过程的统计保证迁移到离散行列式点过程？如果可以，那么文献中已有的许多离散行列式点过程或许都可以从“极限”的角度重新审视，从而获得新的保证。与此同时，也可以专门构造新的 离散行列式点过程，使其能够良好地近似具有优良性质的连续行列式点过程。

为了更好地说明我们的方法，下面介绍一个具体示例。据我们所知，这是此前少数考虑过这一问题特殊情形的研究之一。我们将在第5节重新讨论该示例。在这个示例中，我们考虑统计学习中的一个标准设定，其目标是找到一个满足 $f_\theta$ 且具有 $\theta \in \Theta$ 的函数 $L$，使得依赖于某些可用数据 $X_n = \{x_1、\ldots\ldots、x_n\}$的损失函数 $\Theta \to \mathbf{R}_{>0}$ 最小。对于经验风险最小化等许多经典学习任务，损失函数具有如下形式

$$L(\theta) = \sum_{i=1}^{n} f_\theta(x_i)$$

当 $n$ 较大时，优化可能会很耗费资源。[1] 为了降低计算负担，可以构建一个核心集，即一个大小为$\mathcal{S} \subseteq X_n$ 的子集，并为其赋予权重 $w_x$，使得

$$L_{\mathcal{S}}(\theta) = \sum_{x \in \mathcal{S}} w_x f_\theta(x) \tag{1.1}$$

与 $L(\theta)$接近，从而只需在这个小得多的核心集 [55]上进行训练。离散 DPP 可用于对 $\mathcal{S}$ 进行抽样；有严格证明表明，与独立采样相比，它的表现不差，且实验效果更佳 [66]。另一方面，关于严格优势的数学保证，目前仅针对一个特定例子得以确立 [8, 9]: 当 $x_i$ 从 $\mathcal{X} = [-1,1]^d$中独立同分布地抽取，目标近似质量为 $L$时，在 $X_n$ 上使用经过精确定义的 DPP，可获得比独立采样方法更小的核心集$\mathcal{S}$，前提是 $n$ 足够大。 [8] 的证明表明，与该过程相关的某个特定量趋近于与连续 *DPP*在$\mathcal{X}$ 上相关的量；连续 DPP 具有显著的解析性质，进而可据此得到上述保证。请注意，在 [8], 中，这一论证并未明确表述为“离散到连续”的极限，而这正是本文的主要目的。基于我们的通用框架，我们在第5 节证明，另一种更简单的离散 DPP 也满足同类保证。

1.1. 贡献总结与文章结构。 现在，我们非正式地概述一下我们的主要贡献。宽泛地说，一个 DPP $\mathcal{P}$ = 行列式点过程$(\mathcal{K}\ \mu)$ 定义在 $\Gamma$ 上，由一个核$\mathcal{K} : \Gamma \times \Gamma \to \mathbf{C}$ 以及一个参考测度 $\mu$ 定义在 $\Gamma$上；精确定义将在第2 节中给出。我们注意到，当集合 $\Gamma$ 为有限集时，该核可以表示为一个矩阵[2] $K \in \mathbf{C}^{|\Gamma| \times |\Gamma|}$。

[1]例如，在 k-均值、支持向量机或低秩近似中。[2]严格来说，这取决于对 $\Gamma = \{x_1, \ldots, x_n\}$ 中各元素的排序，使得 $K_{ij}$是 $(x_i, x_j)$ 处的核值。本文中这种排序是隐含的，且从未产生歧义。

下文中，除了一些更一般的技术性结果外，我们假设 $\mathcal{X}$ 是一个紧空间，并且 $X_n = \{x_1,...,x_n\} \subset \mathcal{X}$ 已根据概率测度 $\mu$独立同分布地抽取。记经验测度为 $\mu_n = \frac{1}{n}\sum_i \delta_{x_i}$，并令 $K_n \in \mathbf{C}^{n\times n}$ 表示某个矩阵（由实践者观测或构造），我们关注 $\mathcal{P}_n =$ 行列式点过程$(K_n, \mu_n)$ 向某个 $\mathcal{P} =$ 行列式点过程$(\mathcal{K}\ \mu)$ 的“收敛”。我们通过一种称为弱相干性的新概念来刻画这种收敛性；该概念旨在保证过程样本的所有线性统计量 的期望具有集中性。对于有界可测函数 $\varphi_r : \mathcal{X}^r \to \mathbf{R}$ 和有限样本 $\mathcal{S} \subseteq \mathcal{X}$，相对于 $\mathcal{S}$的 $r$点线性统计量，以及其在给定点过程 $\mathcal{Q}$下的期望，定义如下：

$$\Lambda^{(\varphi_r)}(\mathcal{S}) := \sum_{\substack{x_1,...,x_r\in\mathcal{S}\\ x_i\neq x_j}} \varphi_r(x_1,...,.x_r), \quad \Phi^{(\varphi_r)}(\mathcal{Q}) := \mathbf{E}_{\mathcal{S}\sim\mathcal{Q}}[\Lambda^{(\varphi_r)}(\mathcal{Q})]$$

也就是说， $\Phi^{(\varphi_r)}(\mathcal{Q})$ 是对$\mathcal{S}$的所有 $r$元素子集中 $\varphi_r$ 的取值求和所得的期望。粗略地说，我们称$\mathcal{P}_n$ 与$\mathcal{P}$ 弱相干，如果对于所有$r \in \mathbf{N}$、 $\varphi_r$： $\mathcal{X}^r \to \mathbf{R}$、 $\delta \in (0,1)$以及 $\varepsilon > 0$，存在 $N(\delta, \varepsilon, \varphi_r)$，使得对于所有$n \geq N(\delta, \varepsilon, \varphi_r)$，

$$\mathbb{P}\left(\left|\Phi^{(\varphi_r)}(\mathcal{P}_n) - \Phi^{(\varphi_r)}(\mathcal{P})\right| \geq \varepsilon\right) \leq \delta,$$

其中，概率 $\mathbb{P}$在本文中及下文均指关于 $K_n$的随机性，$\mu_n$；也就是说，是指随机抽取的 $X_n$ ，有时也指定义 $K_n$的其他随机量所涉及的随机性。我们强调，这并不涉及关于从 中抽取$\mathcal{P}_n$ 或$\mathcal{P}$样本的随机性。详细定义见第3 节。通常，处理行列式点过程时，人们感兴趣的正是计算这些线性统计量 $\Lambda^{(\varphi_r)}(\mathcal{S})$ 及其相关的矩：例如，它们出现在上文方程（1.1）中的核心集估计，或蒙特卡罗方法的求积公式 [56],中，而它们的方差决定了其向均值集中的程度 [9]。因此，弱相干性自然能够将离散行列式点过程的统计性质与其连续极限的统计性质联系起来，下面我们将通过若干示例加以说明。

这里的关键要素是矩阵 $K_n$，它完全决定了离散 DPP$\mathcal{P}_n =$ DPP$(K_n,\ \mu_n)$。一个典型例子是所谓的格拉姆矩阵 $K_n = \mathcal{K}[_{X_n\times X_n} := [\mathcal{K}(x_i,\ x_j)]_{i,\ j=1}^n$，在这种最简单的情形下，我们证明了弱相干性成立。然而，在许多情况下，核$\mathcal{K}$ （有时甚至连点 $x_i$ 本身[3]）都无法获取，人们只能观测到或估计其含噪版本 $K_n$ 。本文的主要结果给出了在这种更困难的情形下弱相干性成立的充分条件。非正式地说，我们证明了以下结论。

**定理 1.1**（弱相干条件，非正式表述）。 设 $K$ 是一个连续核，并且 $K_n$ 是一族随机 $n \times n$ 矩阵。假设对于所有 $\delta \in (0,\ 1)$ 以及$\varepsilon > 0$，都存在 $N_K(\delta,\ \varepsilon)$，使得以下条件之一成立：

i) $K$ 和 $K_n$ 是复值的，并且

$$n \geq N_K(\delta,\varepsilon) \ \Rightarrow\ \mathbb{P}\left(\max_{i,j\in[n]} |[K_n]_{i,j} - \mathcal{K}(x_i,x_j)| \geq \varepsilon\right) \leq \delta; \tag{1.2}$$

[3]在这种离散情形下，对于许多用途而言，来自 $X_n$ 上某个点过程的样本可以用定义在指标 $[n] = \{1,...,n\}$上的点过程来替代；此时无需使用 $x_i$本身，只需矩阵 $K_n$。由于这在很大程度上取决于具体用途，我们在此不再进一步阐述这些技术细节。

ii) $K$ 和 $K_n$ 是实值的，并且

(1.3)
$$n \geq N_K(\delta, \varepsilon) \;\Rightarrow\; \mathbb{P}\left(\max\left(\left\|\frac{K_n - \mathcal{K}_{|X_n \times X_n}}{n}\right\|_F, \left|\operatorname{tr}\left(\frac{K_n}{n}\right) - \operatorname{tr}\left(\frac{\mathcal{K}_{|X_n \times X_n}}{n}\right)\right|\right) \geq \varepsilon\right) \leq \delta,$$

这里回顾一下，$\mathbb{P}$ 指的是对（$K_n, \mu_n$）的随机性。于是，序列 $P_n$ =行列式点过程（$K_n$, $\mu_n$）与 $P$ =行列式点过程（$K$ $\mu$）弱相干。

我们将在第4节陈述该定理的更详细版本。暂且不论复值与实值的差异，我们指出，定理中的条件*i*）严格强于条件*ii*）。不过，我们仍决定在定理的表述中保留这两个条件，因为这两种情形还会产生略有不同的弱相干速率。正如我们将看到的，用于刻画线性统计量 $\Lambda^{(\varphi_r)}(\mathcal{S})$ 的期望集中性的速率 $N(\delta, \varepsilon, \varphi_r)$（以及我们将看到的，其矩的集中速率）可直接与核的集中速率 $N_K(\delta, \varepsilon)$ 联系起来。这使我们能够以非渐近方式刻画离散 DPP 的性质，下面将通过三个例子加以说明。

**1.1.1. 概要。** 本文大致分为两部分：第一部分涉及弱相干性的通论；第二部分则介绍具体示例。全文结构如下。

- 在第2节中，我们回顾行列式点过程的必要背景，重点介绍如何使用相关函数和线性统计量来定义它们。
- 在第 3节中，我们引入弱相干性，并证明它意味着线性统计量的所有矩 都具有集中性。这一框架具有很强的普适性，可扩展到本文关注范围之外的情形，为未来的研究发展提供了有趣的可能。具体而言，它允许考虑任意不一定是行列式点过程的点过程之间的弱相干性，而且也不限于离散到连续极限。

- 在第 4节中，我们详细阐述定理 1.1及其证明。证明中包含若干本身也颇具价值的中间结果，包括关于子矩阵行列式的集中不等式（引理4.10），以及适用于比上文所讨论的 $\mathcal{P}_n$ 和 $\mathcal{P}$ 更一般的点过程的弱相干性结果。

- 在第 5、6、7 节中，我们将结果应用于三个不同的例子。– 在第 5 节中，我们重新审视前文 [8]所述例子中建立的核心集保证，并将其应用于一种定义起来更简单的离散 DPP：所谓的离散多元正交多项式系综 [66]。利用我们的结果证明它与 X 上的连续过程 $= [-1, 1]^d$弱相干，我们得到了另一个基于行列式点过程的核心集实例，其效果严格优于独立采样。– 在第 6 节中，我们引入一种定义在紧流形上点云 $X_n \subseteq \mathrm{R}^d$ 上的新型 DPP。该过程基于与 $X_n$相关联的图构建，其思想与 [65]类似。我们证明，它与该流形对应的调和系综弱相干；调和系综是一种连续行列式点过程，已有研究为其建立了多种优于独立采样的保证（例如 [52, 15]）。据我们所知，这是首个具有此类性质的过程。– 最后，在第 7 节中，我们研究一个更为特殊的例子，其中观测到的是潜在位置随机图 [22]。也就是说，这是一个具有独立伯努利边 $a_{ij} \sim$ Bernoulli(K($x_i, x_j$)) 的图，因此其邻接矩阵是格拉姆矩阵的一个（高度）含噪版本，而 $x_i$ 未被观测到。我们证明它

可以构造出 $\mathcal{K}$ 的估计，使其满足定理*1.1*的条件ii)，因此，使用相关 DPP 对图的节点进行抽样，相当于对底层未观测到的潜在空间进行抽样。我们指出，这个例子更多是一个说明性的概念验证，而非实用方法；特别是，图上的行列式抽样通常着眼于使用随机森林的根[2, 65], ，这些根对应于另一种核。研究后者可能的无限图极限，是未来研究的一个重要方向。

第4、5、6 和7 节中的大部分技术论证被归入补充材料。

1.2. 相关工作。 尽管关于行列式点过程已有大量理论与应用研究，并且用离散行列式点过程对独立同分布数据进行子抽样会自然引出诸多问题，行列式点过程的离散到连续分析仍然相对较少受到研究。与我们最为接近的工作无疑是 [8], ，如上所述。尽管如此，仍有若干研究方向与本文存在相似之处。

1.2.1. 同一空间上行列式点过程的极限。 与我们的离散到连续极限不同，对定义在同一空间（连续或离散）上的行列式点过程进行比较，是一个更为成熟的课题。[4] 行列式点过程的极限在经典框架中被表述为依分布收敛，或称弱收敛；为完整起见，我们在附录A节中回顾这一概念。弱收敛的研究通常受到数学物理的推动，其关注点在于定义在同一连续空间上、且其样本量 $m = |\mathcal{S}|$ 趋于无穷的一列过程，以便从统计上描述大量粒子（就行列式点过程而言，即费米子）的位置。这与我们的设定不同：在我们的设定中，底层离散空间的规模 $X_n$ 反而趋于无穷。特别地，就我们的目的而言，我们强调弱收敛是一个纯粹的渐近框架。而弱相干性则提供了一个简单框架，使我们能够获得非渐近结果，这些结果直接关系到实践者通常关心的统计量。[5]

对于相对于同一参考测度定义的行列式点过程，已有几篇开创性论文推导出弱收敛成立的充分条件 $\mu$ [63, 62]。例如，当 $\mathrm{X} = \mathrm{R}^d$ 时，若厄米核 $\mathrm{K}_n : \mathrm{X} \times \mathrm{X} \to \mathrm{R}$ 在弱算子拓扑下收敛到 K，且 $\mathrm{K}_n$ 在 X 的每个紧集上的迹都收敛到 K 的迹，则弱收敛成立。对于更一般的空间（即我们在第 2 节中定义的空间）和核，已有研究表明，只要 $\mathrm{K}_n : \mathrm{X} \times \mathrm{X} \to \mathrm{C}$ 在 $\mathrm{X} \times \mathrm{X}$ 的每个紧集上一致收敛到 K，弱收敛便会成立。需要注意的是，这种核的收敛与我们在定理 1.1 中提出的纯离散判据截然不同。据我们所知，尽管行列式过程的极限仍是一个活跃的研究领域（例如， [38]），这些判据尚未经过重新审视，以适用于离散到连续极限。特别是，我们的示例通常不在现有结果的适用范围内，因为底层空间 X 不同于（随机的！）离散空间 $X_n$，而且核 $K_n$ 和参考测度 $\mu_n$ 在 $X_n$ 之外甚至没有定义。

[4]这一表述可能会令最细心的读者感到困惑，因为样本 $X_n$ 确实属于 $\mathcal{X}$。关键区别在于，现有结果考虑的是由彼此绝对连续的测度定义的行列式点过程，而我们关注的是分别定义在某个有限（离散）点集 $X_n$ 以及 例如，$\mathcal{X} \subseteq \mathbf{R}^d$的博雷尔集上的过程的收敛性。扩展讨论见注 2.3 。[5]我们指出，尽管在 $\mathcal{X}$ 上点过程弱收敛的拓扑在理论上可能是可度量化的，但实际操作起来十分繁琐。据我们所知，文献中尚未出现关于行列式点过程弱收敛的非渐近结果。

1.2.2. 行列式点过程的估计。虽然我们的框架是构造一个离散核$K_n$，用来估计格拉姆矩阵 $\mathcal{K}[X_n \times X_n]$，以确保基于独立同分布数据的离散行列式点过程收敛到连续行列式点过程，但许多研究直接观测从行列式点过程抽取的离散数据，并旨在以参数或非参数方式学习该过程 [43]。特别是，离散行列式点过程的最大似然估计（MLE）受到广泛关注：研究表明，它通常是NP难问题，在某些情形下也难以近似 [33],；不过，在较为受限的设定中，已有研究提出了具有理论保证的简单估计量 [68, 32]。在连续情形中，例如可以提及 [30]，它利用再生核希尔伯特空间的方法进行估计。

虽然这些设置与我们自己的设置不同，但可能会受益于我们开发的工具。对于连续行列式点过程，[60] 旨在根据单个无限样本（通过递增窗口观测得到）估计参数族中的连续行列式点过程，并提供渐近一致性保证：随着 $n$ 观测到的数据点数量增加，似然函数将满足一致性。这与我们数据量递增的设置类似，我们预计，第1.1节4 中描述的定理（命题4.4 和4.8）的一个变体，有助于以非渐近的方式描述最大似然估计量线性统计量的集中性。

**1.3. 记号。**我们用$\mathbf{R}$ 和$\mathbf{C}$ 分别表示实数域和复数域，用 $[n] = \{1,...,n\}$ 表示前 $n$ 个整数；对于可容许的空间 $\Gamma$和测度 $\gamma$，用$\|\cdot\|_{L^p(\Gamma,\gamma)}$ 表示关于测度 $\gamma$的通常 $p$阶勒贝格范数。若空间可由上下文确定且不存在歧义，我们就简记为$\|\cdot\|_{L^p(\gamma)}$。全文中，我们用 $\mathbb{P}$ 表示关于点过程随机性的概率（通常是行列式点过程中的随机性，相对于引言中描述的对象$(X_n, K_n)$而言）；而$\mathbf{P}$、$\mathbf{E}$ 和$\mathbf{Var}$ 表示给定过程样本的概率、期望和方差。

**2. 背景：关于（行列式）点过程。**由于我们研究定义在不同类型（连续或离散）空间上的点过程，因此给出一个一般的测度论定义；我们的设定与行列式点过程理论中已被明确研究过的最一般设定相一致 [62]。对这套精确术语不太熟悉的读者可以记住，这一设定足够灵活，能够处理定义在$\mathbf{R}^d$中的区域、流形或离散空间上的（行列式）点过程，以及文献中出现的所有常见参考测度 $\mu$。关于测度论和拓扑学概念的参考资料，可参见例如 [27]。

**2.1. 相关函数与行列式点过程。**设 $\Gamma$ 为一个第二可数、局部紧豪斯多夫空间。（简单）点过程$\mathcal{P}$ 是定义在 $\Gamma$的局部有限子集上的一个概率分布：子集$\mathcal{S} \subseteq \Gamma$ 满足

$$\#(\mathcal{S} \cap C) < \infty$$

对所有紧集 $C \subseteq \Gamma$ [23]均成立。

**2.1.1. 线性统计量和相关函数。**回顾一下，对于任意有界可测函数 $\varphi_r : \Gamma^r \to \mathbf{R}$ 和 $\mathcal{S} \subseteq \Gamma$，我们将它们的（$r$点）线性统计量及其期望记作

$$\Lambda^{(\varphi_r)}(\mathcal{S}) := \sum_{\substack{x_1,...,x_r \in \mathcal{S} \\ x_i \neq x_j}} \varphi_r(x_1, ..., .x_r), \qquad \Phi^{(\varphi_r)}(\mathcal{P}) := \mathbf{E}_{\mathcal{S}\sim\mathcal{P}} \left[\Lambda^{(\varphi_r)}(\mathcal{S})\right] \tag{2.1}$$

在本文余下部分，我们考虑从点过程 $\mathcal{P}$ 中采样得到的集合 $\mathcal{S}$，并宽泛地将 $\Lambda^{(\varphi_r)}(\mathcal{S})$ 称为点过程 的线性统计量 $\mathcal{P}$ （不过我们强调，它们仅取决于 $\varphi_r$ 和 $\mathcal{S}$）。

点过程的不同性质可以用其相关函数 （也称为乘积密度函数，或联合强度）来描述，只要它们存在。给定 $\Gamma$上的拉东测度 $\gamma$，局部可积函数 $\rho_r : \Gamma^r \to \mathbf{R}$ 称为一个点过程的$r$-点相关函数，如果对任意 有界可测 $\varphi_r$，下式成立：

$$\Phi^{(\varphi_r)}(\mathcal{P}) = \int_{\Gamma^r} \varphi_r \rho_r d\gamma^{\otimes r} := \int_{\Gamma^r} \varphi_r(x_1, ..., x_r)\rho_r(x_1, ..., x_r)d\gamma(x_1)...d\gamma(x_r) \tag{2.2}$$

其中 $\int_{\Gamma^r} \varphi_r \rho_r d\gamma^{\otimes r}$ 是我们为方便起见经常使用的简写记号。[6] 换言之，如果它们存在，相关函数将点过程的任意线性统计量的期望表示为 $L^2$-内积（相对于参考测度） $\gamma$。

在许多情况下，实际上可以定义 一个点过程，依据是其相关函数$(\rho_r)_r$，而它们确实定义一个（唯一）点过程所须满足的一般可容许性条件已被刻画 [49, 50, 51, 24]。

2.1.2. 行列式点过程。 行列式点过程恰好是一类由其相关函数定义的点过程 [54] （对于定义在有限集上的行列式点过程，这一定义可以简化，无需提及相关函数；见下文第 2.2 节）。一个点过程是行列式的 当且仅当存在一个核$K$： $\Gamma \times \Gamma \to \mathbf{C}$，称为该过程的 相关核 （也称边缘核），使得

$$\rho_r(x_1, ..., x_r) = \det\left([K(x_i, x_j)]_{i,j=1}^r\right), \tag{2.3}$$

其中 $[K(x_i, x_j)]_{i,j=1}^r$ 表示元素为 $K(x_i, x_j)$ 的 $r \times r$ 矩阵。本文中，我们用 DPP$(K, \gamma)$ 表示相对于测度 $\gamma$、核为 $K$ 的行列式点过程，并用 $\rho_r[K]$ 表示相应的 $r$- 点相关函数。注意，并非所有对 $(K, \gamma)$ 都能定义一个（唯一的）行列式点过程。以下经典定理是确保这一点的一种方法。

2.1.3. *Macchi-Soshnikov*定理。 存在性条件是在如下假设下陈述的： $K$ 连续且为厄米的，并且其关联的积分算子$T_K$： $L^2(\Gamma; \gamma) \to L^2(\Gamma; \gamma)$是局部迹类的。这里， $T_K$ 由下式定义

$$(T_K f)(x) = \int_\Gamma f(x)K(x, y)d\gamma(y). \tag{2.4}$$

根据非负定算子的Mercer定理，局部迹类条件相当于要求

$$\mathrm{Tr}_\gamma(K) := \int_C K(x, x)d\gamma(x) < \infty \tag{2.5}$$

对所有紧集 $C \subseteq \Gamma$成立。我们指出，若 $\Gamma$ 进一步为紧集，Mercer定理还保证 $T_K$ 是Hilbert-Schmidt算子，且其Hilbert-Schmidt范数等同于

$$\|K\|^2_{L^2(\gamma^{\otimes 2})} := \int_{\Gamma^2} |K(x, y)|^2 d\gamma(x)d\gamma(y) < \infty. \tag{2.6}$$

因此，Macchi-Soshnikov定理断言如下。

[6]为了定义方程（2.2）中的积分，赋予 $\Gamma$ 其Borel $\sigma$-代数。

定理 2.1（Macchi-Soshnikov定理 [54, 63, 62]）。在上述假设下， $K$ 定义一个行列式点过程，当且仅当 $T_K$ 的特征值属于 $[0,1]$。在这种情况下，它是唯一的。

尽管这是文献中最常见的设定，我们指出，核 $K$不必是厄米的即可定义行列式点过程，但该情形下的理论发展较少。对于有限空间 $\Gamma$ （见下文第2.2 节），我们请读者参阅 [59] ，以系统研究具有非厄米核的行列式点过程，包括对那些定义行列式点过程的核的刻画。关于行列式点过程更广泛的背景，参见例如， [23,34, 43, 37, 4, 12]。

在本文余下部分，**我们假设核 $K$ 是连续的**，对我们所考虑的所有行列式点过程均如此，并指出，对于定义在离散空间上的核，情况总是如此。[7] 此外，除非另有说明，我们假设$\mathcal{X}$ 是一个**紧集**、第二可数的豪斯多夫空间。

2.2. 有限空间上的行列式点过程。 在本工作中，相对于测度 $\gamma_n$ 定义、且具有有限支撑 $X_n$ （大小为 $n$）的行列式点过程起着关键作用，通常称为“离散”行列式点过程。于是先前的定义显著简化，并且可以证明如下：核为 $K_n\colon X_n\times X_n\to\mathbf{C}$、相对于测度 $\gamma_n\colon X_n\to\mathbf{R}$、定义在 $X_n$ 上的行列式点过程，是 $2^{X_n}$ 上的一个概率分布，使得

$$\mathbf{P}_{\mathcal{S}}(A\subseteq\mathcal{S})=\det\left([K_n(x,y)]_{x,y\in A}\right)\prod_{x\in A}\gamma_n(x)$$

对所有 $A\subseteq X_n$。换言之，在样本 $\mathcal{S}$ 中观测到 $A$ 的概率，除以一个取决于测度 $\gamma_n$ 的因子后，等于矩阵 $[K_n(x,y)]_{x,y\in\mathcal{S}}$ 中由 $A$ 所索引的行和列构成的子矩阵的行列式。文献中通常对 $\gamma_n$ 采用两种选择，它们只需对核重新归一化即可相互等价：

- 计数测度 $\gamma_n=\sum_{x\in X_n}\delta_x$ 的情形，

$$\mathbf{P}_{\mathcal{S}}(A\subseteq\mathcal{S})=\det\left([K_n(x,y)]_{x,y\in A}\right), \tag{2.7}$$

  这是有限空间上行列式点过程最常见的定义 [43];
- 经验测度 $\gamma_n=\mu_n$ 的情形：$=\frac{1}{n}\sum_{x\in X_n}\delta_x$，

$$\mathbf{P}_{\mathcal{S}}(A\subseteq\mathcal{S})=\frac{1}{n^{|A|}}\det\left([K_n(x,y)]_{x,y\in A}\right), \tag{2.8}$$

  为方便起见，本文采用这种约定。

当选择 $\mu_n$ 时，正如本文其余部分所做的那样，厄米核的 Macchi-Soshnikov 条件在 核矩阵 $K_n$ 的特征值属于 $[0,n]$ 时得到满足。这意味着相对于 $\mu_n$ 定义的 算子 $T_{K_n}$ 的特征值属于 $[0,1]$。[8]

注 2.2. 对于我们的目的，我们强调，离散 DPP $\mathrm{P}_n=\mathrm{DPP}(K_n,\ \gamma_n)$ 完全由一个矩阵 $K_n=[K_n(x,\ y)]_{x,\ y\in X_n}$ （表示核）和一个向量 $\gamma_n=[\gamma_n(x)]_{x\in X_n}$ （表示参考测度）确定，其中我们沿用记号 $K_n$ 和 $\gamma_n$。按这种形式书写时， $K_n$ 和 $\gamma_n$ 隐式地定义了$X_n=\{x_1,\ \ldots,\ x_n\}$上的一个排序。特别地，给定核时，甚至不必知道 $X_n$ 的元素

[7]由于它们赋有离散拓扑，因此所有映射都是连续的。[8]关于计数测度，我们要求这些特征值属于 $[0,1]$。

从$\mathcal{P}_n$中抽样时， $X_n$ 中的一个样本对应于指标在 $[n]$中的一个子集。这对于我们在第7节考虑的例子尤为重要，其中 $x_i$没有被观测到，而矩阵 $K_n$ 则由一些辅助的可用数据构造而成。如此一来，离散行列式点过程的线性统计量的期望可以方便地仅用矩阵 $K_n$ 和向量 $\gamma_n$表示：

$$\int_{\Gamma^r} \varphi_r \rho_r[K_n] d\gamma_n^{\otimes r} = \int_{\Gamma^r} \varphi_r(x_1,...,x_r) \det\left([K_n(x_i,x_j)]_{i,j=1}^r\right) d\gamma_n(x_1)...d\gamma_n(x_r)$$

$$= \sum_{i_1,...,i_r=1}^{n} \varphi_r(x_{i_1},...,x_{i_r}) \det\left([K_n]_{\{i_1,...,i_r\}}\right) [\gamma_n]_{i_1}...[\gamma_n]_{i_r}. \tag{2.9}$$

每当采用这种排序时，都会明确说明； $K_n$ 可以指核，也可以指其矩阵，具体取决于哪种记号更方便。

2.2.1. 随机点过程。现在让我们回到促成本研究的问题。考虑在$\mathcal{X}$中随机抽取的一个子集 $X_n$ ；例如，按照概率分布 $\mu$独立同分布地抽取。定义在 $X_n$ 上的行列式点过程$\mathcal{P}_n = \mathrm{DPP}(K_n, \mu_n)$，于是便是一个**随机的离散 DPP**，其中随机性来自 $X_n$ ，最终也可能来自参与计算 $K_n$ 的其他随机量（例如，如第7节所示）。特别地，我们关注$)(\ \Phi^{(\varphi_r \mathcal{P}_n)}$向某个连续行列式点过程$\mathcal{P}$ 的$)(\ \Phi^{(\varphi_r \mathcal{P})}$集中；这种集中性是相对于$\mathcal{P}_n$的随机性而言的。我们强调，这不应与从点过程$\mathcal{P}$ 或$\mathcal{P}_n$中抽取点云$\mathcal{S}$所产生的随机性相混淆。

注 2.3. 另一种做法是研究一个替代性的确定性过程：先抽取 $X_n$ 个独立同分布样本，再用离散 DPP 对其进行子抽样，从而得到定义在 **连续** 空间 $\mathcal{X}$ 上的点过程。然而，这种定义在 $\mathcal{X}$ 上的“先独立同分布、再 DPP”点过程通常并非行列式点过程。这是我们选择考虑 $\mathcal{P}_n$ 是离散的、随机 行列式点过程的原因之一。

**3. 点过程的线性统计量与弱相干性。** 我们首先正式引入点过程序列的弱相干性概念，它保证了其线性统计量的渐近集中性。本节内容具有普遍性，也适用于定义在任意局部紧空间 $\mathcal{X}$ 上的点过程，例如例如，$\mathbf{R}^d$。

3.1. 定义。两个随机点过程序列 $\mathcal{P}_n$ 和 $\mathcal{Q}_n$ （不一定是行列式点过程、离散点过程或连续点过程）若其线性统计量的期望彼此渐近集中，则称它们弱相干，具体定义如下。

定义 3.0.1（弱相干性）。设 $\mathcal{P}_n$ 和 $\mathcal{Q}_n$ 分别是定义在子空间序列 $\mathcal{X}_n \subset \mathcal{X}$ 和 $\mathcal{Y}_n \subset \mathcal{X}$ 上的两个**随机**点过程序列，其中子空间均来自 $\mathcal{X}$。

若对所有 $r \in \mathbf{N}_{>0}$、紧支撑的有界可测函数 $\varphi_r \colon \mathcal{X}^r \to \mathbf{R}$、所有 $\delta \in (0,1)$ 以及所有 $\varepsilon > 0$，都存在$N(\delta, \varepsilon, \varphi_r)$ 使得

$$n \geq N(\delta, \varepsilon, \varphi_r) \Rightarrow \mathbb{P}\left(\left|\Phi^{(\varphi_r)}(\mathcal{P}_n) - \Phi^{(\varphi_r)}(\mathcal{Q}_n)\right| \geq \varepsilon\right) \leq \delta. \tag{3.1}$$

速率 $N(\delta, \varepsilon, \varphi_r)$ 是 $\mathcal{P}_n$ 趋近于 $\mathcal{Q}_n$ 的弱相干集中速率 .

这里，概率 $\mathbb{P}$ 指的是 $\mathrm{P}_n$ 和 $\mathrm{Q}_n$的随机性。换言之，对于线性统计量 $\left|\Phi^{(\varphi_r)}(\mathrm{P}_n) - \Phi^{(\varphi_r)}(\mathrm{Q}_n)\right|$的集中性，若任意小的容许误差水平 $\varepsilon$ 和失败概率 $\delta$ 都能在 $n$ 足够大时达到，则弱相干性成立。请注意，速率 $N(\delta,\varepsilon,\varphi_r)$可以依赖于 $\varphi_r$：它并非对这些测试函数统一适用的速率。在第 4 节，我们的主要结果（定理 4.1）针对这样一种特殊情形，给出了弱相干性成立的充分条件：$\mathrm{P}_n$ 是相对于经验概率测度、基于随机独立同分布数据的离散 DPP，而 $\mathrm{Q}_n = \mathrm{P}$ 是相对于概率测度 $\mu$的固定确定性连续行列式点过程。

注释3.1（弱相干性的灵活性）。如前所述，我们对弱相干性的定义还涵盖了比行列式点过程的离散到连续极限更一般的情形，并允许考虑不一定是行列式点过程、且可以是随机的、离散的或连续的点过程。此外，尽管本文聚焦于概率测度 $\mu$ ，点过程也可以基于一般的拉东测度作为参考测度来定义。[9] 特别地，它足够灵活，可以涵盖行列式点过程二者均为离散的情形或二者均为连续的情形；这一设定在例如基于经验测量进行核估计时十分自然。本文其余部分将针对特定的行列式点过程$\mathcal{P}_n$ 和$\mathcal{P}$来表述我们的结果，但一些技术性结果对一般点过程同样成立。

注释3.2（逼近速率）。根据上下文，将诸如方程（3.1）中的概率误差界重新表述为强调误差速率 $\varepsilon_n = \varepsilon_n(\delta)$ 的形式，可能会更加方便。特别是，机器学习中常见如下保证：对于某些 $n$ 和 $\delta$，以至少 $1-\delta$ 的概率，误差$\left|\Phi^{(\varphi_r)}(\mathcal{P}_n) - \Phi^{(\varphi_r)}(\mathcal{Q}_n)\right|$ 有界于 $\varepsilon_n$。在某些例子中，我们发现这种表述的灵活性略低，因此选择改为相对于 $\varepsilon$ 和 $\delta$ 来表达 $n$ 。当然，通过考察 $N(\cdot)$ ，通常可以在两者之间自由转换；在下文中，我们也确实会不时讨论可达到的误差速率 $\varepsilon_n$ 。

3.2. 矩映射定理。 弱相干性保证了线性统计量期望的集中性。显然，当考虑线性统计量的利普希茨函数时，例如，其期望的这种集中性仍然成立。然而，在许多情况下，这些线性统计量的矩 才是最重要的：例如，其方差 已被用于量化行列式点过程相较于独立同分布抽样的某些优势 [9]。本节将表明，上文定义的弱相干性也 确实蕴含矩的集中性。

给定 $\varphi_r$ 和一个点过程$\mathcal{P}$，我们用 $m_k^{\mathcal{P}}\left(\Lambda^{(\varphi_r)}\right), m_k^{\mathcal{P}}\left(\Lambda^{(\varphi_r)}\right)$ 表示相对于$\mathcal{P}$的 $k$ 阶原始矩和中心矩：

$$m_k^{(\varphi_r)}(\mathcal{P}) = \mathbf{E}_{\mathcal{S}\sim\mathcal{P}}\left[\Lambda^{(\varphi_r)}(\mathcal{S})^k\right], \quad \overline{m}_k^{(\varphi_r)}(\mathcal{P}) = \mathbf{E}_{\mathcal{S}\sim\mathcal{P}}\left[\left(\Lambda^{(\varphi_r)}(\mathcal{S}) - m_1^{(\varphi_r)}(\mathcal{P})\right)^k\right], \tag{3.2}$$

从而在我们先前的记号 $\Phi = m_1$ 下。类比于弱收敛的所谓连续映射定理，我们将下述定理称为矩映射定理。

[9]当集合 $X_n$ 按照相对于拉东测度 $\mu$定义的泊松过程，在$\mathcal{X}$ 中抽样时， $\mu$就会自然出现这种情形。实际上，我们的结果可以推广到这种情况；关于以图为中心的应用，简要讨论请见第8 节。

定理 3.3（矩映射定理）. 设 $\varphi_r : \boldsymbol{X}^r \to \mathbf{R}$ 为具紧支集的有界可测函数。设 $\boldsymbol{P}_n$、$\boldsymbol{Q}_n$ 为随机点过程序列，使得 $\boldsymbol{P}_n$ 以速率 $N(\delta, \varepsilon, \varphi_r)$与 $\boldsymbol{Q}_n$ 弱相干。

于是，存在一个由紧支撑、有界可测函数组成的有限族 $\left(\varphi^i\right)_{i=1}^L$，该函数族依赖于 $\varphi_r$；并存在常数 $s \in \mathbf{R}$，使得对于所有 $\delta \in (0, 1)$和 $\varepsilon > 0$,

$$n \geq \max_{i=1,\ldots,L} N\left(\frac{\delta}{L}, \frac{\varepsilon}{s}, \varphi^i\right) \;\Rightarrow\; \mathbb{P}\left(\left|m_k^{(\varphi_r)}\left(\mathcal{P}_n\right) - m_k^{(\varphi_r)}\left(\mathcal{Q}_n\right)\right| \geq \varepsilon\right) \leq \delta. \tag{3.3}$$

此外，若每个矩 $m^{(\varphi_r)}{}_k\left(\boldsymbol{Q}_n\right)$ 在 $n$ 上均一致有界，且该结论以概率 1 成立，那么对于中心矩，同样的结论也成立，只是常数 $s$ 不同。

这一结果依赖于对 $\left(\Lambda^{(\varphi_r)}\right)^k$ 进行的一个简单（但冗长）的组合展开。我们在补充材料的 3.3节中给出定理 C 的详细版本及其明确常数。我们指出，中心矩所需的一致有界性假设在我们关注的情形中显然成立，因为 $\mathcal{Q}_n = \mathcal{P}$ 是一个固定的确定性行列式点过程。

注 3.4。矩映射定理转化了一个有用的性质：矩的速率，$N(\delta/L,\ \varepsilon/s,\ \varphi_r)$，仅改变其对 $\delta$与 $\varepsilon$ 的依赖，相差乘法常数，相较于弱相干速率 $N(\delta,\ \varepsilon,\ \varphi_r)$。特别地，当反转该界并考察误差速率 $\varepsilon_n$（对于给定的 $n$，$\delta$）时（见注3.2），这通常得到相同速率 $\varepsilon_n$，用于线性统计量的期望及其矩，至多相差乘法常数。

**4. 行列式点过程离散到连续极限的弱相干性。** 我们的主要结果关注独立同分布数据上的离散 DPP 与相应的连续行列式点过程之间的弱相干性。回顾我们考虑紧集上的一个概率测度 $\mu$ $\mathcal{X}$，以及集合 $X_n = \{x_1, \ldots\ldots x_n\} \subseteq \mathcal{X}$，它由 $n$ 个依 $\mu$独立同分布采样的点组成。

我们将给出定理1.1的如下详细版本，从而建立离散到连续弱相干性成立的两个充分条件。

定理 4.1（定理 1.1的详细速率）。设 $\mathcal{K}$ 为核，且 $(K_n)_n$ 是由 $n \times n$ 矩阵构成的序列，使得所有配对 $(\mathcal{K}\ \mu)$ 和 $(K_n, \mu_n)$ 均满足 *Macchi-Soshnikov*定理 的条件。若对于所有 $\delta \in (0, 1)$ 和 $\varepsilon > 0$，存在 $N_K(\delta,\ \varepsilon)$，使得满足以下任一条件：

i）$K$ 和 $K_n$ 均为复值，且

$$n \geq N_K(\delta, \varepsilon) \;\Rightarrow\; \mathbb{P}\left(\max_{i,j \in [n]} \left|[K_n]_{i,j} - \mathcal{K}(x_i, x_j)\right| \geq \varepsilon\right) \leq \delta;$$

ii）$K$ 和 $K_n$ 均为实值，且

$$n \geq N_K(\delta, \varepsilon) \;\Rightarrow\; \mathbb{P}\left(\max\left(\left\|\frac{K_n - \mathcal{K}_{|X_n \times X_n}}{n}\right\|_F, \left|\operatorname{tr}\left(\frac{K_n}{n}\right) - \operatorname{tr}\left(\frac{\mathcal{K}_{|X_n \times X_n}}{n}\right)\right|\right) \geq \varepsilon\right) \leq \delta,$$

那么，序列 $\boldsymbol{P}_n =$ 行列式点过程$(K_n, \mu_n)$ 与 $\boldsymbol{P}\ =$ 行列式点过程$(K\ \mu)$弱相干。特别地，对于任意可测且有界的 $\varphi_r : \boldsymbol{X}^r \to \mathbf{R}$，我们有

$$N(\delta, \varepsilon, \varphi_r) = \max\left(N_K\left(\frac{\delta}{4}, \mathcal{K}_\infty\right),\ N_K\left(\frac{\delta}{4}, \frac{\varepsilon}{c}\right),\ \frac{2^{r+5}\beta \log\left(\frac{4}{\delta}\right)}{\varepsilon^2},\ \frac{C_r}{\varepsilon}\right), \tag{4.1}$$

在假设 *i)* 下为 $c = 2^r(r \times r!)\|\varphi_r\|_{L^1(\mu_n^{\otimes r})}K_\infty^{r-1}$，在假设 *ii)* 下为 $c = 2^r(r \times r!)\|\varphi_r\|_{L^\infty(\mu_n^{\otimes r})}K_\infty^{r-1}$。两种情况下，我们都有：

$$\beta = \max_{\{x_1,\dots,x_r\}\in\mathcal{X}^r} \det\left([\mathcal{K}(x_j,x_k)]_{j,k=1}^r\right)\|\varphi_r\|_\infty, \quad C_r = 2^{r+1}\left(\mu^{\otimes r}(\mathrm{supp}(\varphi_r))\beta + \beta^2\right),$$

其中 $\|\varphi_r\|_\infty =$ 最大值$_{z\in\mathcal{X}^r}|\varphi_r(z)|$。

请注意，在条件*ii)*下， $K_n$ 对$\mathcal{K}$ 的近似条件更宽松，但我们要求核为实值。我们认为复值核也应有类似的结果，但我们的证明技巧无法推广到这种情况。我们还指出，我们的证明确实可以推广到由非厄米 核定义的行列式点过程序列（若此类序列存在），且速率完全相同。

请注意，对于格拉姆矩阵 $K_n = \mathcal{K}[X_n\times X_n]$，弱相干性 始终 以速率 $N_K(\cdot) = 0$成立。因此，在任何情况下，只要能够访问核 $\mathcal{K}(x_i,x_j)$，就始终可以构造一个与行列式点过程 $(\mathcal{K}\ \mu)$弱相干的离散 DPP。本定理的核心在于对 $K_n$ 施加的稳定性条件，这些条件仍能保证弱相干性成立。正如我们将在第5、6和7节的示例中看到的，确实存在许多应用场景，其中真实核$\mathcal{K}$ 无法访问；在这种情况下，控制其近似误差便成为实践者的主要目标。

注 4.2。根据注3.2 并考察给定 $n,\delta$ 时的速率 $\varepsilon_n$，项 $2^{r+5}\beta\log(\frac{4}{\delta})\,\varepsilon_n^2$ 施加了不快于 $\varepsilon_n = O\left(\sqrt{\frac{\log(\frac{1}{\delta})}{n}}\right)$ 的速率限制。这是大多数关于 $n$ 独立同分布变量的统计量的典型收敛速率（例如，由 McDiarmid 集中不等式得到）。这一速率可能达到（例如，对于格拉姆矩阵），或者根据具体例子，核的收敛速率 $N_K$ 可能导致严格更慢的速率 $\varepsilon_n$。

定理 4.1 的证明概要。 设 $\varphi_r$： $\mathcal{X}^r \to \mathbf{R}$ 是紧集 $\mathcal{X}^r$上的任意有界可测函数。由于我们处理的是行列式点过程，线性统计量可以用行列式表示，因此首先需要证明：对于任意 $\delta\in(0,\ 1)$ 以及$\varepsilon > 0$，存在 $N(\delta,\ \varepsilon,\ \varphi_r)$ 使得在此之后，以至少 $1-\delta$ 的概率，

$$\left|\int_{X_n^r}\varphi_r\rho_r[K_n]d\mu_n^{\otimes r} - \int_{\mathcal{X}^r}\varphi_r\rho_r[\mathcal{K}]d\mu^{\otimes r}\right| \le \varepsilon\,.$$

这里我们回顾一下方程（2.2）和（2.9）中的简写记号。

应用三角不等式，我们得到

$$\underbrace{\left|\int_{X_n^r}\varphi_r\rho_r[K_n]d\mu_n^{\otimes r} - \int_{\mathcal{X}^r}\varphi_r\rho_r[\mathcal{K}]d\mu^{\otimes r}\right|}_{E_n} \le \underbrace{\left|\int_{X_n^r}\varphi_r\left(\rho_r[K_n]-\rho_r[\mathcal{K}]\right)d\mu_n^{\otimes r}\right|}_{E_n^K}$$
$$+\underbrace{\left|\int_{X_n^r}\varphi_r\rho_r[\mathcal{K}]d\mu_n^{\otimes r} - \int_{\mathcal{X}^r}\varphi_r\rho_r[\mathcal{K}]d\mu^{\otimes r}\right|}_{E_n^\mu},$$

因此，我们只需证明每个加项的集中性。按照经典方法，我们将推导出 $N(\delta,\ \varepsilon,\ \varphi_r)$，使得对于任意 $n \ge N(\delta,\ \varepsilon,\ \varphi_r)$，

$$\mathbb{P}\left(E_n^K \le \frac{\varepsilon}{2}\right) \ge 1-\frac{\delta}{2} \quad \text{and} \quad \mathbb{P}\left(E_n^\mu \le \frac{\varepsilon}{2}\right) \ge 1-\frac{\delta}{2}.$$

然后由并集界即可得出结果。

□

我们称 $E_n^K$ 为核误差，并称 $E_n^\mu$ 为测度误差，其集中性将在第 4.1 节和4.2 节分别证明。前者源于 $K_n$ 与格拉姆矩阵 $\mathcal{K}|_{X_n\times X_n}$ 之间的偏差；特别地，当 $K_n = \mathcal{K}|_{X_n\times X_n}$ 时，前者将消失。我们将利用针对行列式的新型稳定性界来控制前者。后者源于 $\mu$ 与经验测度 $\mu_n$ 之间的偏差，并将利用针对独立同分布变量的经典集中不等式来控制。

本节其余部分将专门介绍这些技术计算。随后，我们将在第5、6 和7节介绍我们结果的应用。

**4.1. 核误差的集中界。** 本节将在非常一般的设定下对核误差 $E_K^n$ 进行界定。我们考虑两个行列式点过程随机序列的一般情形：$\mathcal{P}_n = \mathrm{DPP}(K_n^{\mathcal{P}}, \nu_n)$ 和 $\mathcal{Q}_n = \mathrm{DPP}(K_n^{\mathcal{Q}}, \nu_n)$，它们定义在同一个拉东测度 $\nu_n$ 上，取值空间为第二可数局部紧豪斯多夫空间 $\mathcal{X}$。需要强调的是，此处的 $\mathcal{X}$ 不一定是紧集，且 $\nu_n$ 不一定是归一化的。

备注 4.3. 有趣的是，本节的结果直接给出了关于弱相干性的一般条件，适用于相对于同一测度定义的 DPP 序列（该测度可以是随机的，和/或依赖于 $n$），因为在这种情况下，测度误差 $E_\mu^n$ 会消失。由于这些结果也适用于非紧空间 $\mathcal{X}$ 和非概率测度 $\nu_n$，因此也可用于相对于 $\mathbf{R}^d$ 上的勒贝格测度或计数测度 $\sum_{x\in X_n}\delta_x$ 定义的 DPP 序列；我们将在第8节简要讨论这些更一般设定下的应用。

我们固定一个有界、可测且紧支撑的函数 $\varphi_r : \mathcal{X}^r \to \mathbf{R}$，并旨在为以下形式的一般核误差建立集中界

$$\mathcal{E}_n^K = \left| \int_C \varphi_r \left( \rho_r[K_n^{\mathcal{P}}] - \rho_r[K_n^{\mathcal{Q}}] \right) d\nu_n^{\otimes r} \right|.$$

定理 $E_n^K = \int_{X_n^r} \varphi_r \left(\rho_r[K_n] - \rho_r[\mathcal{K}]\right) d\mu_n^{\otimes r}$ 证明中的核误差4.1 是一个特殊情形，其中 $\nu_n = \mu_n$。

**4.1.1. 技术性假设。** 我们在一般设定下陈述结果，并给出定理4.1这一特殊情形的推论。处理一般情形需要若干须谨慎给出的定义，这些定义在定理*4.1*的设定下可以放心忽略。特别地，出现在我们的界中的某些量依赖于 $\varphi_r$ 的支集以及测度 $\nu_n$ 的支集：为避免在这些项的定义中引入大量技术细节，并且不失一般性，我们假设对某个 $C_{\varphi_r} \subseteq \mathcal{X}$ 有 $\mathrm{supp}(\varphi_r) = C_{\varphi_r}^r$，并用 $C_{\varphi_r,n} = C_{\varphi_r} \cap \mathrm{supp}(\nu_n)$ 表示该区域与 $\mathrm{supp}(\nu_n)$ 的交集。为方便起见，我们进一步假设 $C_{\varphi_r,n}$ 是紧集。在定理4.1的假设下，我们直接有 $\mathrm{supp}(\nu_n) = C_{\varphi_r,n} = X_n$，且这一情形总是成立。

4.1.2. **一致逼近下的集中。** 我们首先给出 $\mathcal{E}_n^K$ 的一个上界，该上界取决于 $K_n^{\mathcal{P}}$ 与 $K_n^{\mathcal{Q}}$ 之间的最大差值，这对应于定理 4.1的第一个假设。该结果的证明将在本节末尾详细给出；在此之前，我们先讨论其推论并介绍一些必要的引理。

命题 4.4。 在当前假设下，有

$$\mathcal{E}_n^K \le (r \times r!)\|\varphi_r\|_{L^1(\nu_n^{\otimes r})} a_n^{r-1} \max_{x,y \in C_{\varphi r,n}} \left| K_n^{\mathcal{P}}(x,y) - K_n^{\mathcal{Q}}(x,y) \right|, \tag{4.2}$$

其中 $a_n = \max\left(\max_{x,\ y \in C_{\varphi r,\ n}} |K_n^{\mathcal{P}}(x,\ y)|,\ \max_{x,\ y \in C_{\varphi r,\ n}} |K_n^{\mathcal{Q}}(x,\ y)|\right)$。

根据命题 4.4，核在$\mathcal{X}$ 的所有紧集上以最大范数集中趋于 $K_n^{\mathcal{Q}}$，是 DPP($K_n^{\mathcal{P}}$，$\nu_n$) 与 DPP($K_n^{\mathcal{Q}}$，$\nu_n$) 具有弱相干性的充分条件；方程（4.2）则给出了各阶 $r$ 的 $r$-点线性统计量的集中速率，这些速率显式取决于核的集中速率。在我们的离散到连续情形中，

$$\nu_n = \mu_n \quad C_{\varphi r\ n} = X_n \quad K_n^{\mathcal{P}} =, \ , \quad \text{以及}$$

$K_n \qquad K_n^{\mathcal{Q}} =$
$\mathcal{K}$，由此得到以下结果。

推论 4.5（定理 4.1 的核误差，$i$））。 令 $\mathcal{K}_\infty = \max_{x,\ y \in X_n} |\mathcal{K}(x,\ y)|$。在定理 *4.1 i)* 的假设下，对于任意 $\delta \in (0,\ 1)$ 以及 $\varepsilon > 0$，

$$n \ge \max\left( N_K\left(\frac{\delta}{4}, \mathcal{K}_\infty\right),\ N_K\left(\frac{\delta}{4}, \frac{\varepsilon}{c}\right)\right) \ \Rightarrow\ \mathbb{P}\left(E_n^K \ge \varepsilon/2\right) \le \frac{\delta}{2}, \tag{4.3}$$

其中

$$c = 2^r (r \times r!)\|\varphi_r\|_{L^1(\mu_n^{\otimes r})} \mathcal{K}_\infty^{r-1}.$$

推论 4.5 的证明。 我们首先注意到，对于 $n \ge N_K\left(\frac{\delta}{4}, \mathcal{K}_\infty\right)$，且以至少为 $1 - \frac{\delta}{4}$ 的概率，

$$\max_{i,j \in [n]} |[K_n]_{i,j}| \le 2\mathcal{K}_\infty,$$

从而 $E_n^K \le 2^{r-1}(r \times r!)\|\varphi_r\|_{L^1(\mu_n^{\otimes r})} \mathcal{K}_\infty^{r-1} \left(\max_{i,j \in [n]} |[K_n]_{i,j} - \mathcal{K}(x_i, x_j)|\right)$。取 $n \ge N_K\left(\frac{\delta}{4}, \frac{\varepsilon}{c}\right)$ 为

$$c = 2^r (r \times r!)\|\varphi_r\|_{L^1(\mu_n^{\otimes r})} \mathcal{K}_\infty^{r-1}$$

则可保证，由命题 4.4 并以至少为 $1 - \frac{\delta}{2}$ 的概率，有 $E_n^K \le \frac{\varepsilon}{2}$。对两个条件中任一不成立的概率应用并界，即得该结果。 □

命题 4.4 在 $\nu_n = \mu_n$ 情形下的论证依赖于如下估计，其证明推迟至第 B.1 节（附录）。

命题 4.6. 设 $A$ 与 $B$ 为两个具有复系数的 $n \times n$ 矩阵，且 $I_r = \{i_1, \ldots, i_r\} \subseteq [n]$ 一组 $r$ 个互异指标。则

$$|\det(A_{I_r}) - \det(B_{I_r})| \le r! \sum_{j=1}^{r} \left(\max_{k,l \in [n]} |A_{k,l}|\right)^{j-1} \left(\max_{k,l \in [n]} |A_{k,l} - B_{k,l}|\right) \left(\max_{k,l \in [n]} |B_{k,l}|\right)^{r-j}, \tag{4.4}$$

其中 $M_{I_r}$ 表示将 $M$ 限制于由 $I_r$ 所索引的行与列而得到的 $r \times r$ 矩阵。

该证明可推广，得到命题 4.7，该命题适用于任意测度 $\nu_n$；我们在附录的B.1 节中详述相关差异。

命题 4.7。 对于任意给定的 $x_1, ..., x_r \in C_{\varphi,\ n}$，记 $b(x_1, ..., x_r)$ 为差

$$b(x_1,...,x_r) = \left|\det\left(\left[K_n^{\mathcal{P}}(x_j,x_k)\right]_{j,k=1}^{r}\right) - \det\left(\left[K_n^{\mathcal{Q}}(x_j,x_k)\right]_{j,k=1}^{r}\right)\right|. \tag{4.5}$$

那么，在我们一直采用的假设下，有

$$b(x_1,...,x_r) \le r! \sum_{j=1}^{r} M_{\mathcal{P}}^{j-1} \left(\max_{x,y\in C_{\varphi r,n}} |K_n^{\mathcal{P}}(x,y) - K_n^{\mathcal{Q}}(x,y)|\right) M_{\mathcal{Q}}^{r-j},$$

其中 $M_{\mathcal{P}} =$ 最大值$_{x,\ y\in C_{\varphi r,\ n}} |K_n^{\mathcal{P}}(x,y)|$ 以及 $M_{\mathcal{Q}} =$ 最大值$_{x,\ y\in C_{\varphi r,\ n}} |K_n^{\mathcal{Q}}(x,y)|$。

利用命题4.7，我们最终可以推出命题 4.4。

命题 4.4的证明。 利用三角不等式，我们得到

$$\begin{aligned}\mathcal{E}_n^K &\le \int_{\mathrm{supp}(\varphi_r)} |\varphi_r(x_1,...,x_r)|\, b(x_1,...,x_r) d\nu_n^{\otimes r}(x_1,...,x_r)\\ &\le \|\varphi_r\|_{L^1(\nu_n^{\otimes r})} \max_{x_1,...,x_r\in C_{\varphi r,n}} \left|\det\left(\left[K_n^{\mathcal{P}}(x_j,x_k)\right]_{j,k=1}^{r}\right) - \det\left(\left[K_n^{\mathcal{Q}}(x_j,x_k)\right]_{j,k=1}^{r}\right)\right|,\end{aligned}$$

其中 $b$ 的定义与命题 4.7中相同，我们随后可以应用该命题。由此可得

$$\begin{aligned}\mathcal{E}_n^K \le r!\|\varphi_r\|_{L^1(\nu_n^{\otimes r})} &\left(\sum_{j=1}^{r} \left(\max_{x,y\in C_{\varphi r,n}} |K_n^{\mathcal{P}}(x,y)|\right)^{j-1} \left(\max_{x,y\in C_{\varphi r,n}} |K_n^{\mathcal{Q}}(x,y)|\right)^{r-j}\right)\\ &\times \max_{x,y\in C_{\varphi r,n}} \left|K_n^{\mathcal{P}}(x,y) - K_n^{\mathcal{Q}}(x,y)\right|,\end{aligned}$$

由此我们得到不等式（4.2）。 □

4.1.3. 基于平均逼近的集中性。 对于实值核，我们可以放宽命题 4.4中的一致逼近要求。我们的结果基于与上一节相同的技术假设。

命题 4.8。 在当前假设下，有

$$\mathcal{E}_n^K \le (r\times r!)\|\varphi_r\|_{L^\infty(\nu_n^{\otimes r})} (a_n')^{r-1} \max\left(\left\|K_n^{\mathcal{P}} - K_n^{\mathcal{Q}}\right\|_{L^2(\nu_n^2)}, \left|\mathrm{Tr}_{\nu_n}\left(K_n^{\mathcal{P}}\right) - \mathrm{Tr}_{\nu_n}\left(K_n^{\mathcal{Q}}\right)\right|\right), \tag{4.6}$$

其中 $a_n' =$ 最大值 $\left(\|K_n^{\mathcal{P}}\|_{L^2(\nu_n^2)}\ \|K_n^{\mathcal{Q}}\|_{L^2(\nu_n^2)}, \text{迹}_{\nu_n}\left(K_n^{\mathcal{P}}\right), \text{迹}_{\nu_n}\left(K_n^{\mathcal{Q}}\right)\right)$，而给定测度和核时，迹 迹 的定义与方程 (2.5)中相同。

有趣的是，该界涉及局部迹和 $L^2$ 范数，它们也出现在 Macchi-Soshnikov 定理的假设中。此外，正如我们先前所述，复值核也很可能满足类似结果，但我们目前的证明无法推广到这种情形。我们首先讨论其对离散到连续极限的影响，并引入一个关键引理来给出证明。

对于定理 4.1 的离散到连续情形， $C_{\varphi r,\ n} = X_n$ 和 $\nu_n = \mu_n$。在这种情况下，我们识别出常见的弗罗贝尼乌斯范数和矩阵迹差异

$$\left\|K_n - \mathcal{K}_{|X_n\times X_n}\right\|_{L^2(\mu_n^2)} = \left\|\frac{K_n - \mathcal{K}_{|X_n\times X_n}}{n}\right\|_F,$$

$$\left|\mathrm{Tr}_{\mu_n}(K_n) - \mathrm{Tr}_{\mu_n}\left(\mathcal{K}_{|X_n\times X_n}\right)\right| = \left|\mathrm{tr}\left(\frac{K_n}{n}\right) - \mathrm{tr}\left(\frac{\mathcal{K}_{|X_n\times X_n}}{n}\right)\right|$$

并且，与推论4.5类似，可得到核误差的以下界。

推论 4.9（定理 4.1的核误差， $ii$）） 。 设 $\mathcal{K}_\infty = \max_{x,\ y} |\mathcal{K}(x,\ y)|$。在定理 *4.1ii*) 的假设下，对任意 $\delta \in (0,\ 1)$ 以及 $\varepsilon > 0$,

$$n \geq \max\left(N_K(\delta, \mathcal{K}_\infty),\ N_K\left(\delta, \frac{\varepsilon}{c}\right)\right) \ \Rightarrow\ \mathbb{P}\left(E_n^K \geq \varepsilon/2\right) \leq \frac{\delta}{2}, \tag{4.7}$$

其中

$$c = 2^r (r \times r!) \|\varphi_r\|_{L^\infty(\mu_n^{\otimes r})} \mathcal{K}_\infty^{r-1}.$$

命题 4.8 的证明中的主要内容是以下引理 4.10，该引理涉及子矩阵行列式之和，其证明推迟至附录 B.2 节。我们认为这一结果可能具有独立的研究价值。

引理 4.10（行列式集中引理）。 设 $A$ 和 $B$ 是两个具有实系数的 $n \times n$ 矩阵。那么，

$$\sum_{\substack{I_r \subseteq [n] \\ |I_r| = r}} |\det(A)_{I_r} - \det(B)_{I_r}| \leq (r \times r!) M_{A,B}^{r-1} \max\left(\|A - B\|_F, |\mathrm{tr}\,(A) - \mathrm{tr}\,(B)|\right),$$

其中 $M_{A,\ B} = \max\left(h(A,\ B)\ t(A,\ B)\right)$，对于 $h(A,\ B) = \max\left(\|A\|_F,\ \|B\|_F\right)$ 且 $t(A,\ B) = \max\left(\mathrm{tr}(A),\ \mathrm{tr}(B)\right)$。

更一般地，以下结论成立，详见附录 B.2 节。

引理 4.11（行列式集中引理（一般情形））。 假设核 $K_n^{\mathcal{P}}$ 和 $K_n^{\mathcal{Q}}$ 为实值，并令

$$d = \int_{\mathrm{supp}(\varphi_r)} \left|\det\left(\left[K_n^{\mathcal{P}}(x_j, x_k)\right]_{j,k=1}^r\right) - \det\left(\left[K_n^{\mathcal{Q}}(x_j, x_k)\right]_{j,k=1}^r\right)\right| d\nu_n^{\otimes r}(x_1, ..., x_r). \tag{4.8}$$

于是，在我们的当前假设下，成立

$$d \leq (r \times r!) M_{\mathcal{P},\mathcal{Q}}^{r-1} \max\left(\left\|K_n^{\mathcal{P}} - K_n^{\mathcal{Q}}\right\|_{L^2(\nu_n^2)}, \left|\mathrm{Tr}_{\nu_n}\left(K_n^{\mathcal{P}}\right) - \mathrm{Tr}_{\nu_n}\left(K_n^{\mathcal{Q}}\right)\right|\right),$$

其中$M_{\mathcal{P},\ \mathcal{Q}} = \max\ \left(h(K_n^{\mathcal{P}}, K_n^{\mathcal{Q}})\ t(K_n^{\mathcal{P}}, K_n^{\mathcal{Q}})\right)$；对此，给定两个核 $K_1$ 与 $K_2$，$h(K_1, K_2) = \max\ \left(\|K_1\|_{L^2(\nu_n^2)}\ \|K_2\|_{L^2(\nu_n^2)}\right)$ 以及 $t(K_1, K_2) = \max\left(\mathrm{Tr}_{\nu_n}(K_1), \mathrm{Tr}_{\nu_n}(K_2)\right)$。

我们现在可以证明命题4.8。

证明命题4.8。由三角不等式，我们有

$$\mathcal{E}_n^K \leq d \|\varphi_r\|_{L^\infty(\nu_n^{\otimes r})},$$

其中 $d$ 的定义如方程（4.8）中给出。随后，应用引理4.10得到该结果，类似于命题4.4的证明。 □

4.2. 测度误差的集中界。 我们现在关注向0 集中的测度误差 $E_n^\mu$。我们首先给出一个一般结果，从概率上界定相对于测度 $\mu_n^{\otimes r}$与 $\mu^{\otimes r}$的积分之差。其证明位于第B.3 节（附录），基于对经典集中不等式的细致应用。

命题 4.12. 设 $f:\boldsymbol{X} \to \mathbf{R}$ 为有界可测函数，使得$|f(x)| \le \beta$对所有 $x \in \boldsymbol{X}^{r}$ 成立。则对任意 $\widetilde{\varepsilon} > 0$，以至少 $1 - 2\exp\left(\frac{-2\widetilde{\varepsilon}^2}{nb_n}\right)$的概率（关于 $X_n$），

$$\left| \int_{X_n^r} f d\mu_n^{\otimes r} - \int_{\mathcal{X}^r} f d\mu^{\otimes r} \right| \le \varepsilon + \frac{M}{n}, \tag{4.9}$$

其中 $M = 2^{r-1}\left(\beta + \|f\|_{L^1(\mu^{\otimes r})}\right)$ 是一个常数，并且

$$b_n = \left(\beta \frac{\sum_{l=1}^{r} \binom{r}{l} n^{l-1}}{n^r}\right)^2 \sim \frac{\beta^2}{n^2}. \tag{4.10}$$

注意，$\frac{M}{n}$ 项是确定性的，概率界仅涉及 $\widetilde{\varepsilon}$项。

为了将这一结果应用于测度误差 $E_n^{\mu}$，我们考虑函数

$$f(x_{i_1}, ..., x_{i_r}) = \varphi_r(x_{i_1}, ..., x_{i_r}) \det\left([\mathcal{K}(x_{i_j}, x_{i_k})]_{i,k=1}^{r}\right).$$

在 $\varphi_r$ 有界且$\mathcal{K}$ 连续（因此在紧集$\mathcal{X} \times \mathcal{X}$上有界）的当前假设下，$f$也有界。我们进一步将 $\varphi_r$的下界和上界分别记为 $\varphi_{min}, \varphi_{max} \in \mathbf{R}$，于是

$$\varphi_{min} \le \varphi_r(x) \le \varphi_{max} \tag{4.11}$$

对于所有 $x \in \mathrm{supp}(\varphi_r)$。接下来，我们很容易得到如下结果，具体内容详见附录的 B.4 节。

推论 4.13。 设 $\boldsymbol{K}:\boldsymbol{X} \times \boldsymbol{X} \to \mathbf{C}$ 是一个有界可测核，并且

$$\beta = \max_{\{x_1,...,x_r\} \in \mathcal{X}^r} \det\left([\mathcal{K}(x_j, x_k)]_{j,k=1}^{r}\right) \max(-\varphi_{min}, \varphi_{max}). \tag{4.12}$$

那么，存在常数 $C_r$，使得对于任意 $\delta \in (0, 1)$ 和 $\varepsilon > 0$，

$$n \ge \max\left(\frac{2^{r+5}\beta \log\left(\frac{4}{\delta}\right)}{\varepsilon^2}, \frac{C_r}{\varepsilon}\right) \Rightarrow \mathbb{P}\left(E_n^{\mu} \ge \varepsilon/2\right) \le \frac{\delta}{2}, \tag{4.13}$$

其中 $C_r = 2^{r+1}\left(\mu^{\otimes r}(C)\beta + \beta^2\right)$。

综上所述，定理 4.1 的证明至此完成。

**5. 优于泊松的方差 $[-1, 1]^d$。**现在，我们将把我们的结果应用于若干示例。在第一个示例中，我们重新探讨所谓的多元正交多项式系综在核心集构造中的应用；这是一类既可定义在连续空间、也可定义在离散空间上的行列式点过程。对于定义在连续空间$\mathcal{X} = [-1, 1]^d$上的行列式点过程，其某些具体形式已被证明具有优于独立抽样的保证；其中一些还被用于提出一种离散行列式点过程替代过程，并具有严格优于独立抽样的核心集保证 [10, 8, 9]。本节将表明，对于独立同分布样本 $X_n \subseteq \mathcal{X}$ [67]，，使用离散多元正交多项式系综也能获得同类保证，而无需借助替代过程。在本节余下部分，$m$ 表示这些行列式点过程的样本量，并保持固定。我们取$\mathcal{X} = [-1, 1]^d$，并指出，我们的推导对于$\mathbf{R}^d$中的任意紧区域同样成立。

5.1. 多元正交多项式系综。为了定义多元多项式系综的核，我们首先定义与连续测度和有限支撑测度相关的正交多项式。我们考虑一个概率测度$\mu$，它相对于勒贝格测度绝对连续，并取子集 $X_n = \{x_1,...,x_n\} \subseteq \mathcal{X}$ 中的 $n$ 个点，这些点按照 $\mu$ 独立同分布抽取，其对应的经验测度为 $\mu_n$。

5.1.1. 正交多项式。对于度数族 $(\beta_1, ..., \beta_d)$，考虑定义在 $\mathbf{R}^d$ 上的单项式求值函数集合，其形式为 $x \mapsto x(1)^{\beta_1}...x(d)^{\beta_d}$，其中$x = (x(1) ..., x(d)) \in \mathbf{R}^d$，并按照度数 $\beta_i$ 的分级字典序对这些函数排序。我们将这一列函数记作 $(\mathbf{M}_i)_i$，并将 $M_i$ 它们限制在 $X_n$的点上所得的函数记作[10]。对 $L^2(\mathbf{X}; \mu)$（分别为 $L^2(X_n; \mu_n)$）中的这些函数取前 $m$ 个，应用格拉姆-施密特正交化过程，可得到一列多元正交多项式 $(\mathbf{P}_i)_{i=1}^m$（分别为 $(P_i)_{i=1}^m$）。更明确地说，我们分别有如下结果：

$$\begin{cases} \mathcal{P}_1 = \frac{\mathcal{M}_1}{\|\mathcal{M}_1\|_\mu} \\ \mathcal{P}'_{k+1} = \mathcal{M}_{k+1} - \frac{\langle \mathcal{M}_{k+1}, \mathcal{P}_k \rangle_\mu}{\langle \mathcal{P}_k, \mathcal{P}_k \rangle_\mu} \mathcal{P}_k \\ \mathcal{P}_{k+1} = \frac{\mathcal{P}'_{k+1}}{\|\widetilde{\mathcal{P}}_{k+1}\|_\mu}, \end{cases} \tag{5.1}$$

$$\begin{cases} P_1 = \frac{M_1}{\|M_1\|_{\mu_n}} \\ P'_{k+1} = M_{k+1} - \frac{\langle M_{k+1}, \mathcal{P}_k \rangle_{\mu_n}}{\langle P_k, P_k \rangle_{\mu_n}} P_k \\ P_{k+1} = \frac{P'_{k+1}}{\|P_{k+1}\|_{\mu_n}}, \end{cases} \tag{5.2}$$

其中我们使用了以下记号

$$\langle \mathcal{P}, \mathcal{Q} \rangle_\mu = \int_{\mathcal{X}} \mathcal{P}(x)\mathcal{Q}(x) d\mu(x), \quad \|\mathcal{P}\|_\mu = \langle \mathcal{P}, \mathcal{P} \rangle_\mu, \tag{5.3}$$

$$\langle P, Q \rangle_{\mu_n} = \sum_{x \in X_n} \frac{1}{n} P(x)Q(x), \quad \|P\|_\mu = \langle P, P \rangle_{\mu_n}. \tag{5.4}$$

在非常宽松的可容许性条件下——例如，$\mu(A) > 0$ 对某个开子集$A \subseteq \mathcal{X}$成立——这些过程确实定义了（唯一的）正交多项式序列。下文中我们始终作此假设。

5.1.2. 多元正交多项式系综。连续和离散多元正交多项式系综是分别针对测度 $\mu$ 和$\mu_n$定义的行列式点过程，其核为$\mathcal{K}$ 和 $K_n$，由相对于这些测度的前 $m$ 个正交多项式参数化：

$$\mathcal{K}(x, y) = \sum_{i=1}^m \mathcal{P}_i(x)\mathcal{P}_i(y) \quad \forall x, y \in \mathcal{X}, \tag{5.5}$$

$$[K_n]_{j,k} = \sum_{i=1}^m P_i(x_j)P_i(x_k) \quad \forall j, k \in [n], \tag{5.6}$$

[10]也就是说，$M_i$ 是从 $X_n$ 到$\mathbf{R}$的函数。

其中 $m$ 是某个正整数。注意，这些行列式点过程确实存在，因为我们满足Macchi-Soshnikov定理的假设。[11] 此外，它们是所谓的投影行列式点过程 [34], ，因此来自 $\mathcal{S}$ 的样本 DPP($\mathcal{K}$ $\mu$) 或 DPP($K_n, \mu_n$) 的基数 $m$几乎必然成立。

现在我们可以开始证明这两个多元正交多项式系综的弱一致性。在整个分析过程中，我们强调 $m$ 是固定的，而我们令 $n \to \infty$。

5.2. 正交多项式系综的弱一致性。我们工作的第一步是注意到，相对于 $\mu_n$ 的正交多项式会集中到其连续对应物附近。

命题5.1。在第5.1节的设定下，存在三个常数$A$， $B$， $C > 0$，使得对于任意 $\delta \in (0, 1)$ 以及 $\widetilde{\varepsilon} > 0$,

$$n \geq \frac{A^2 B^2 \log\left(\frac{C}{\widetilde{\delta}}\right)}{\widetilde{\varepsilon}^2} \;\Rightarrow\; \mathbb{P}\left(\max_{i\in[m]} \max_{x\in X_n} |\mathcal{P}_i(x) - P_i(x)| \geq \widetilde{\varepsilon}\right) < \widetilde{\delta}. \tag{5.7}$$

直观上，这种集中现象理应发生，但据我们所知，先前文献中尚未出现过此类结果。[12] 我们的证明推迟至补充材料的D.1节，其中速率是通过应用 McDiarmid 不等式推导出来的。

因此显然， $K_n$ 趋近于$\mathcal{K}$，我们可以应用定理4.1， $i$)。

定理5.2（多元正交多项式系综的弱相干性）。在第5.1节的设定下，行列式点过程($K_n$， $\mu_n$) 与行列式点过程($\mathcal{K}$ $\mu$)弱相干，其速率为

$$N_K(\delta, \varepsilon) = \frac{A^2 B^2 m^2 M^2 \log\left(\frac{C}{\delta}\right)}{\varepsilon^2},$$

其中 $A$， $B$ 和 $C$ 是命题5.1中的相同常数，并且

$$M = 2 \max_{i\in[m]} \max_{x\in\mathcal{X}} |\mathcal{P}_i(x)|.$$

具体的弱相干速率 $N_K$ 在补充材料的D.2 节中推导。根据注4.2，我们注意到这给出的误差速率为

$$\varepsilon_n = \mathcal{O}\left(m\sqrt{\frac{\log\left(\frac{1}{\delta}\right)}{n}}\right).$$

5.3. 优于独立的保证。 在这个例子中，弱相干性实际上可用于通过矩映射定理 行列式点过程($\mathcal{K}$ $\mu$) 的性质转化为 行列式点过程($K_n$， $\mu_n$)的性质。定理 3.3。事实上，关于 1点线性统计量的方差行列式点过程($\mathcal{K}$ $\mu$) 已由 [8], 研究，而矩映射定理使我们能够将其与.....联系起来

[11]根据构造，这些正交多项式相对于内积$\langle .,. \rangle_\mu$（分别为$\langle .,. \rangle_{\mu_n}$）构成标准正交基，因此 $T\mathcal{K}$（分别为 $K_n$）的特征值均属于{0,1}（分别属于{0,$n$}）。[12]尽管人们对关于<style>正交多项式的渐近性质越来越感兴趣，尤其是所谓的 Christoffel 函数（ $\mu_n$的倒数），其为 $K_n$对角线的倒数，并已应用于例如在$\mathbf{R}^d$ [45]上概率密度函数的支撑集估计。

与和 行列式点过程($K_n$, $\mu_n$)相关的方差联系起来。事实证明，该方差严格小于独立采样时的方差；由于它控制着线性统计量向其期望集中的程度，较低的方差意味着所需核心集更小（见下文注释 5.5）。我们先回顾 [8], 关于连续多元正交多项式系综的一个结果。

这一结果依赖于$\mathcal{K}$卓越的解析性质，并且是在较宽松的技术假设下陈述的： $\varphi$ 足够正则，且 $\mu$ 属于 Nevai 类。[13]

定理 5.3 ([8]). 假设我们处于第*5.1* 节的设定下，且对于某个 $p \in \mathbf{N}_{>0}$, $m = p^d$。[14] 进一步假设 $\mu$ 属于 *Nevai* 类，并考虑一个~有界可测函数 $\widetilde{\varphi} : X \to \mathbf{R}$，使得 $\varphi : x \mapsto \widetilde{\varphi}(x)\, \mathcal{K}(x,x)$ 是 *Lipschitz* 连续的。那么，

$$\mathbf{Var}_{\mathcal{S}\sim \mathrm{DPP}(\mathcal{K},\mu)}\left[\Lambda^{(\varphi)}(\mathcal{S})\right] \in \mathcal{O}\left(\frac{1}{m^{1+1/d}}\right), \tag{5.8}$$

其中， $\Lambda^{(\varphi)}(\mathcal{S})$ 表示相对于 $\varphi$的DPP($K$ $\mu$) 的1点线性统计量。

根据矩映射定理（定理 3.3），可得以下结论。

推论 5.4. 在定理 *5.3*的假设下，对于任意 $\delta \in (0,1)$ 且$\varepsilon > 0$ 足够小，有至少 $1-\delta$ 的概率满足

$$\mathbf{Var}_{\mathcal{S}\sim \mathrm{DPP}(K_n,\mu_n)}\left[\Lambda^{(\varphi)}(\mathcal{S})\right] \leq \mathbf{Var}_{\mathcal{S}\sim \mathrm{DPP}(\mathcal{K},\mu)}\left[\Lambda^{(\varphi)}(\mathcal{S})\right] + \varepsilon \tag{5.9}$$

*as* $s$一旦 $n \geq N(\delta,\ \varepsilon,\ \varphi)$，其中 $N$如方程(4.1) 所示，且对于 $N_K$如定理*5.2*所述 .

特别地，我们可以将速率实例化为 $\varepsilon_n = \mathcal{O}\left(m\sqrt{\frac{\log(\frac{1}{\delta})}{n}}\right)$。为便于比较， $X_n$ 上泊松点过程的相应方差属于$\mathcal{O}\left(\frac{1}{m}\right) + \mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right)$（参见例如， [8]中的附录 S1）。这意味着，当固定的 $m$足够大且随着 $n \to \infty$时，行列式点过程($K_n$, $\mu_n$)的线性统计量的方差严格优于独立采样的相应方差。

注释 5.5。方差低于独立采样所带来的一个具体结果是，在用于构建核心集时，DPP($K_n, \mu_n$) 经证明优于独立采样。具体而言，随着 $m$ 增大， $L_{\mathcal{S}}$ 与 $L(\theta)$ 之间差值的下降速率会更快（公式 (1.1)）；关于这一现象的全面描述，请参阅 [9] 。这一理论结果也得到了大量实证证据 [67, 9] 的支持。事实上，正如我们在引言中所讨论的， [8, 9], 提出了一种满足此类保证的、定义在 $X_n$ 上的另一种 DPP；该文还通过证明其线性统计量的方差趋近于 DPP(K $\mu$) 的方差，得到了这一结论。[15] 第二个例子的主要缺点是，该过程的核构造复杂得多；关于两种方法所提供保证的差异，我们将在补充材料的 E 节中讨论。就实际表现而言， [9] 中的实证结果表明，离散正交多项式系综的表现实际上更好。

[13]这是一个常见的正则性条件：当密度 $p_1\ x_1\ \cdots p_d\ x_d$ $p_i$ $p$ $\mu$ $[-1,1]$ $p\ x_1$ $x_d$ = [10] 分解为（，.....）（）（）时，例如只需每一项在其上为正即可；更详尽的讨论参见例如。[14]这最后一个假设可以放宽，但代价是需要一个更为技术性的陈述。[15]对于该过程，很可能同样可以得到更强的弱一致性结果。

5.3.1. 示例：用于核心集的离散正交多项式系综与独立同分布抽样的比较。根据注释5.5，我们通过一个受控示例，展示行列式点过程$(K_n,\ \mu_n)$在核心集构造中的表现，并与独立同分布抽样进行比较。关于更全面的评估，请参阅 [67, 9]。我们均匀抽取 $n$ 个点 $X_n = \{x_1,\dots,x_n\}$，取自 $[-1,1]^2$，并考虑1-均值问题，即，

$$\operatorname*{argmin}_{\theta\in\mathbf{R}^2} \underbrace{\sum_{i=1}^{n} \|x_i-\theta\|^2}_{L(\theta)}.$$

很容易计算出这个问题的最优解。对于核心集，我们强调

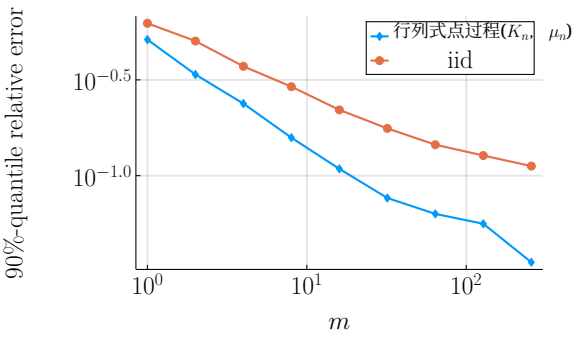


图 1：$m$ 的90%分位数相对误差与其在双对数尺度下的函数关系。

我们的目标其实是构造一个子集$\mathcal{S} \subseteq X_n$ 和一个 $L$ 的估计量 $L_{\mathcal{S}}$，使得$L(\theta)$和 $L_{\mathcal{S}}(\theta)$对 $\theta$的所有取值都相近。我们将比较以下两个 $L(\theta)$估计量的表现：

$$L_{\mathcal{S}}^{\mathrm{iid}}(\theta) = \sum_{i=1}^{n} \frac{\|x_i-\theta\|^2\epsilon_i}{mp_i}, \quad L_{\mathcal{S}}^{\mathrm{DPP}}(\theta) = \sum_{i=1}^{n} \frac{\|x_i-\theta\|^2\epsilon_i}{[K_n]_{i,i}/n}.$$

对于独立同分布估计，从 $X_n$ 中按照以下敏感度分数 [66]有放回地独立同分布抽取一个包含 $m$ 个点的子样本$\mathcal{S}$：

$$p_i \propto \frac{1}{n}\left(1+\frac{\|x_i\|^2}{v}\right), \quad v = \frac{1}{n}\sum_{i=1}^{n}\|x_i\|^2.$$

这是独立同分布数据 [44, 5]的最优估计。对于基于行列式点过程的估计，S 从行列式点过程$(K_n,\mu_n)$中抽取，而$[K_n]_{i,i}/n$ 分母是 $x_i$属于 S 的概率。在这两种情况下，$\epsilon_i$ 表示 $x_i$ 在样本中的出现次数（对于行列式点过程样本，该值始终为 0 或 1；但对于独立同分布样本，由于采用有放回抽样，该值可能更大）。$L_{\mathcal{S}}^{\mathrm{iid}}(\theta)$是 $L(\theta)$的无偏估计量，这是经典的 [5]，；而$L_{\mathcal{S}}^{\mathrm{DPP}}(\theta)$则可通过检查方程 (2.2) 和 (2.3) 得出。对于我们的 Julia 实现，[16] 我们采用 [11]提供的精确且快速的行列式点过程采样器。[17]

给定 S 和估计量 $L_{\mathcal{S}}$，我们计算相对误差$|L_{\mathcal{S}}(\theta) - L_\theta| / L_\theta$ 的最坏值，这些值是通过从$[-1,1]^d$中均匀选取的 $\theta$ 的 $n_\theta = 100$ 个随机值获得的。对于 S 的 $n_S = 100$ 次抽取，重复此操作。然后，我们计算这 $n_S$ 次抽取所得相对误差的 90%分位数；也就是说，90% 的抽取核心集 S 所对应的相对误差不超过 $\epsilon$的值 $\epsilon$。我们报告了不同样本量 $m = |\mathsf{S}| \in \{1, 2, 4, 8, 16, 32, 64, 128, 256\}$下的这些 90%分位数，结果见图1，并对 $X_n$的 100 次实现取平均。通过从行列式点过程$(K_n,\mu_n)$中抽样，确实可以获得更快的速率。

**6. 通过图在未知流形上进行排斥抽样。**我们的第二个示例涉及使用行列式点过程在流形上进行随机采样，尤其是所谓的调和系综。对于某些流形族，已有研究证明该过程具备严格优于独立性保证的性质（例如 [52, 15] 及参考文献

[16]可访问于https://gitlab.irisa.fr/hjaquard/discrete-to-continuous-dpps。[17]已在https://github.com/dahtah/Determinantal.jl中实现。

见其中），而且正如预期的那样，其定义相当复杂，依赖于与流形相关的几何对象$\mathcal{X}$。本节介绍一个作用于从$X_n$ $\mathcal{X}$ 中抽取的点上的行列式点过程，并证明它与调和系综弱相干，因此即使在流形和采样密度均未知的情况下，也能在$\mathcal{X}$ 上实现可证明的排斥抽样，只要 $n$ 足够大即可。我们的结果仅基于定理4.1的应用；该定理结合近期在图流形逼近 [28] 和流形上的密度估计 [72], 方面的进展，消除了大部分难点。

**6.1. 图与流形上的调和系综。** 本节余下部分中，我们考虑$\mathcal{X}$，它是$\mathbf{R}^d$ 中的一个光滑、连通、紧致且可定向的 $d_{\mathcal{X}}$ 维黎曼子流形，且无边界。[18] 关于黎曼流形，我们所用的所有定义和性质均可参见例如经典教科书 [31]。

**6.1.1. 流形上的调和系综。** 紧致、连通且可定向的黎曼流形具有由其体积形式定义的典范测度 $\omega$，调和系综即相对于该测度定义。此外，这些流形还具有所谓的*Laplace-Beltrami*算子$\mathcal{L}$，它对任意光滑函数$f\colon\ \mathcal{X}\to\mathbf{R}$ 定义为

$$\mathcal{L}f=\operatorname{div}(\nabla f), \tag{6.1}$$

其中，div 和 $\nabla$ 分别表示（黎曼）散度和梯度。算子$\mathcal{L}$ 具有纯点谱；又由于它是非正定的，因此可得：1/ $\mathcal{L}$ 的特征值 $\lambda_i$ 构成一个离散序列，我们可以将其排列为

$$0\le\lambda_1\le\lambda_2\le\dots \tag{6.2}$$

并且，2/ 我们可以定义一个由与这些特征值对应的 $L^2(\mathcal{X};\omega)$上的光滑归一化特征函数 $\phi_i\in L^2(\mathcal{X};\omega)$组成的标准正交基。为简化表述，我们假设 $\mathcal{L}$ 的所有特征值都是单重的：尽管这一假设（至少在理论上）有所限制，但在下文中以特征投影代替特征函数即可消除这一限制。[19] 我们的所有结果也都可以推广到这种情形。对于任意 $m\in\mathbf{N}_{>0}$，我们便可定义调和系综，其阶数为 $m$，定义在$\mathcal{X}$上；它是相对于测度 $\omega$ 、由以下核定义的行列式点过程

$$\widetilde{\mathcal{K}}(x,y)=\sum_{i=1}^{m}\phi_i(x)\phi_i(y)\quad\forall x,y\in\mathcal{X}. \tag{6.3}$$

根据 Macchi-Soshnikov 定理，这确实定义了一个行列式点过程，也是引入调和系综的通常方式。就我们的目的而言，用稍有不同的方式定义它更为方便。设 $\mu$ 是一个相对于 $\omega$绝对连续、且密度 $p$为正的概率测度。那么，调和系综等价地定义为相对于测度 $\mu$ 、具有以下核的行列式点过程

$$\mathcal{K}(x,y)=\frac{1}{\sqrt{p(x)}}\widetilde{\mathcal{K}}(x,y)\frac{1}{\sqrt{p(y)}}. \tag{6.4}$$

通过检查方程（2.2）和（2.3），即可轻易得知这定义了同一个行列式点过程；相关讨论见补充材料的第 F.1 节。

接下来，我们在如下技术假设下进行。

[18]请注意，当 $\mathcal{X}$ 赋予通常拓扑时，它是一个第二可数局部紧豪斯多夫空间。[19]进一步的动机是，我们依赖于在同一假设下陈述的结果 [28]。正如 [28], 中所解释的，这些结果也可以推广。

假设6.1。测度 $\mu$ 关于测度 $\omega$绝对连续，其密度为 $p: \mathcal{X} \to [p_{min}, p_{max}]$，对某个正的 $p_{min} > 0$，其为 $\kappa \in$ 阶赫尔德 $(0, 1]$。

现在我们来描述一个定义在集合 $X_n$ 上的行列式点过程，该集合采样自$\mathcal{X}$，下文将证明它与$\mathcal{X}$上的调和系综弱相干。其构造基于图，而图通常被用作流形的离散近似 [13],

6.1.2. 与从流形采样的图相关联的调和系综。我们考虑$\mathcal{X}$ 上的一个概率测度 $\mu$，它关于 $\omega$绝对连续，以及从 $\mu$独立同分布采样的 $X_n = \{x_1, ..., x_n\}$。然后，定义完全图 $G_n$，其节点为 $X_n$，边为 $E_n = \{\{x, y\}; x, y \in X_n\}$，权重为

$$w_n(x, y) = \exp\left(\frac{-\|x - y\|^2_{\mathbf{R}^d}}{4h_1(n)^2}\right) \tag{6.5}$$

对任意 $x, y \in X_n \subseteq \mathcal{X}$以及某个正的 $h_1(n)$，其中$\|.\|_{\mathbf{R}^d}$表示$\mathbf{R}^d$中的欧几里得范数。这里，$h_1(n)$是一个带宽参数，它应随 $n$适当地缩放。一个自然的想法是，为了近似$\mathcal{K}$，利用 $G_n = (X_n, E_n, w_n)$的图拉普拉斯算子 来构造 $-\mathcal{L}$ 的近似，并计算其前 $m$ 个特征向量以构建核 $K_n$，然后考虑与 $\mu_n$相关联的行列式点过程。事实证明，这一构造存在两个问题。

1. 正如人们所预期的那样，图拉普拉斯算子及其特征向量确实会收敛到依赖于 $\mu$ as $n \to \infty$ 的对象。另一方面，[21] 发起的一系列重要研究表明，对图拉普拉斯算子进行适当的重整化后，可以渐近地恢复算子 $\mathcal{L}$（而非依赖于密度的变体）。在我们的情形中，这一构造归功于 [28], ，具体如下。首先，我们考虑一个 $n \times n$ 归一化加权邻接矩阵

$$W_{i,j} = \frac{w_n(x_i, x_j)}{d_n(x_i) d_n(x_j)} \tag{6.6}$$

其中 $d_n(x) = \sum_{i=1}^n w_n(x, x_i)$，并定义其对应的对角度矩阵，其元素为 $D_{i,i} = \sum_{j=1}^n W_{i,j}$。然后我们定义归一化拉普拉斯矩阵

$$L_n = \frac{I - D^{-1}W}{h_1(n)^2},$$

它是半正定的，并将其最小特征值对应的前 $m$ 个特征向量记为 $(\widetilde{u}_i)_{i=1}^m$。[20] 我们还考虑由以下公式给出的重新加权特征向量族

$$u_i(x) = \frac{\widetilde{u}_i(x)^2}{\sqrt{\frac{|\mathbf{S}^{d-1}| h_1(n)^d}{d} \sum_{j=1}^n \frac{\widetilde{u}_i(x_j)}{\#(B(x_i, h_1(n)) \cap X_n)}}}, \tag{6.7}$$

其中 $|\mathbf{S}^{d-1}|$ 表示（$d-1$）维欧几里得球面的体积，并用 $\#(B(x_i, h_1(n)) \cap X_n)$ 表示 $h_1(n)$ 维欧几里得球面内的数据点数量

[20]注意，$L_n$ 并非对称矩阵，但它与一个对称矩阵相似，因此具有相同的谱：$L_n = D^{-1/2}\left(\frac{I - D^{-1/2}WD^{-1/2}}{h_1 \ n}()^2\right)D^{1/2}$。这就是我们称其具有特征值且为半正定的原因。

邻域 $x_i$。 [28] 的结果指出，存在某个常数$h_1(n) > 0$使得，如果 $h_1(n) \leq H_1$ 足够小，且随着 $n$缓慢减小，使得 $h_1(n) \geq \left(\frac{\log(n)}{n}\right)^{4d}$ $x^{+8}$，这些 $u_i$会集中到$\phi_i$ 的特征函数$\mathcal{L}$ 上，并在 $L^\infty$ 意义下覆盖 $X_n$。我们在补充材料的第 F.1 节回顾其精确结果。

2. 调和系综在 $\mathcal{X}$ 上的核 $\mathcal{K}$ 是利用未知的密度 $p$ 定义的。第二个问题可以通过引入密度估计器 [57] 来解决：对于 $h_2(n) > 0$，我们随后考虑由下式定义的 $p$ 的密度估计器 $e[p]$ : $\mathcal{X} \to \mathbf{R}$

$$e[p](x) = \frac{1}{nh_2(n)^{d_\mathcal{X}}} \sum_{i=1}^{n} l\left(\frac{\|x_i - x\|_{\mathbf{R}^d}}{h_2(n)}\right), \tag{6.8}$$

其中某个函数 $l$ : $\mathbf{R}_{\geq 0} \to \mathbf{R}$ 满足温和的可积性条件以及在无穷远处趋于零的条件（见下文假设 6.2）。这里， $h_2(n)$ 是第二个带宽参数，它应随 $n$ 适当地缩放。特别地， [72] 中证明了存在常数 $h_2(n) > 0$ ，使得若 $h_2(n) \leq H_2$ 且 $h_2(n)$ 随 $n$ 下降得足够缓慢，以致 $\left(\log(n)\, nh_2(n)^{d_\mathcal{X}}\right)$ 趋于 0，则当 $n \to \infty$ 时，密度估计器 $e[p]$ 在 $L^\infty$ 意义下趋近于$p$ 。我们在补充材料的 F.1 节中回顾他们的精确结果。

这两个常数 $H_1$ 和 $H_2$ 取决于概率密度 $p$ 以及$\mathcal{X}$的几何性质。关于 $l$，我们更具体地作如下假设。

假设 6.2。 函数 $l$ 在方程（6.8）中应定义为使得 t:

1. 它在$\mathbf{R}_{\geq 0}$上有界；2. 它在任意形如 $[0, \alpha]$ 的紧集上黎曼可积，其中 $\alpha > 0$；3. 存在一个 $t_0$ ，使得对任意 $t \geq t_0$ ，都有 $l(t) \leq \frac{1}{t^\beta}$ ，其中 $\beta > d_\mathcal{X}$；4. 它满足 $\int_{\mathbf{R}^{d_\mathcal{X}}} l(\|v\|_{\mathbf{R}^{d_\mathcal{X}}}) dv = 1$。

构造工作已接近完成，但还需采取最后一个预防步骤，以确保我们定义的行列式点过程存在。考虑测度 $\omega_n = \frac{1}{n} \sum_{x \in X_n} \frac{1}{e[p]}(x)\, \delta_x$，它是 $\omega$在 $X_n$上的一个近似；以及向量 $v_1$、……、 $v_m \in \mathbf{R}^n$，它们是通过对向量 $u_i$相对于内积$\langle f, g \rangle_{\omega_n} = \sum_{x \in X_n} \omega_n(x) f(x) g(x)$应用格拉姆-施密特正交化过程得到的。最后，我们定义核

$$[K_n]_{a,b} = \frac{1}{\sqrt{e[p](x_a)}} \left(\sum_{i=1}^{m} v_i(x_a) v_i(x_b)\right) \frac{1}{\sqrt{e[p](x_b)}} \quad \forall a, b \in [n], \tag{6.9}$$

并将由 $K_n$ 关于测度 $\mu_n$ 所定义的离散调和系综称为行列式点过程。[21] 为清楚起见， $K_n$ 其构造总结于算法 1中。需要强调的是，与多元正交多项式系综的核不同，此处的核由额外的带宽参数 $h_1(n)$ 和$h_2(n)$参数化。特别地，要声称一列行列式点过程具有弱相干性，我们总会指定相应的一列带宽参数。

[21]请注意，这一构造中的某些部分可能可以修改，而不影响我们后续的结果。例如，方程（6.7）中的逐点重新归一化在 $X_n$上起到密度估计器的作用，可以用核密度估计量 $e[p]$替代；针对后者， [28] 的结果可能也可以作相应调整。也可以考虑更稀疏的 $\varepsilon$-图或 $k$-近邻图；但据我们所知，现有文献中尚未出现适用于这些情形的 [28] 结果变体。

**算法 1** 离散调和系综的核

**要求：** $m \geq 0$, $X_n = \{x_1, \ldots, x_n\}$, $h_1(n)$, $h_2(n) > 0$, $l$ 一个核密度函数

1：构建权重为 $w_n(x_i, x_j) = \exp\left(\frac{\|x_i - x_j\|^2}{4h_1(n)}\right) l\left(\frac{\|x_i - x_j\|_{\mathbb{R}^d}}{h_2(n)}\right)$ $\forall i, j \in [n]$ 的完全图 2：计算度 $d_n(x) = \sum_{i=1}^n w_n(x, x_i)$ $\forall x \in X_n$ 3：计算代理归一化邻接矩阵 $W_{i,j} = \frac{w_n(x_i, x_j)}{\sqrt{d_n(x_i) d_n(x_j)}}$ 4：计算其归一化拉普拉斯矩阵 $L_n = \frac{I - D^{-1/2} W D^{-1/2}}{h_1(n)}$ 5：计算前 $k$ 个特征向量 $u_i$ of $L_n$ 6：计算重新归一化的向量 $\widetilde{u}_i(x) = \frac{u_i(x)}{\sqrt{h_1(n)^{d-1} \sum_{x_j \in X_n} u_i(x_j)^2 l\left(\frac{\|x_i - x_j\|_{\mathbb{R}^d}}{h_2(n)}\right)}}$ $\forall x \in X_n$ 7：计算核密度估计量 $e[p](x) = \frac{1}{n h_2(n)^d} \sum_{i=1}^n l\left(\frac{\|x_i - x\|_{\mathbb{R}^d}}{h_2(n)}\right)$ $\forall x \in X_n$ 8：通过相对于 $\omega_n = \frac{1}{n} \sum_{x \in X_n} \frac{1}{e[p](x)} \delta_x$ 对 $u_i$ 进行正交归一化，计算 $v_i$ 9：对于 $1 \leq a, b \leq n$，执行 10：$[K_n]_{a,b} \leftarrow \frac{1}{\sqrt{e[p](x_a)} e[p](x_b)} \sum_{i=1}^k v_i(x_a) v_i(x_b)$ 11：结束循环

6.2. 调和系综的弱相干性。 我们将证明，当 $h_1(n)$和 $h_2(n)$ 适当缩放时，离散调和系综与 $\mathcal{X}$ 上的调和系综弱相干。以下预备结果在补充材料的第 F.1 节中证明。

命题 6.3。 考虑由 $K_n$ 定义的核

$$\left[\widetilde{K_n}\right]_{a,b} = \sum_{i=1}^m v_i(x_a) v_i(x_b) \quad \forall a, b \in [n].$$

那么，$\mathrm{DPP}(K_n, \mu_n)$ 和 $\mathrm{DPP}(K_n, \omega_n)$ 定义相同的行列式点过程。令

$$e = \widetilde{\varepsilon} + \alpha + \beta, \quad \text{for} \quad \alpha = h_1(n)^{1/2} \quad \text{and} \quad \beta = \left(\frac{\log(n)}{n h_2(n)^{d_X}}\right)^{\kappa/2}, \tag{6.10}$$

并假设 $h_1(n) \leq H_1$和 $h_1(n) \geq \left(\frac{\log(n)}{n}\right)^{4d}$，$\alpha \leq 1$，$h_2(n) \leq H_2$ 以及$\beta < \min\left(1, \frac{p_{min}}{2}\right)$。进一步假设满足假设 *6.1* 和 *6.2*。那么，存在 $A, B, C > 0$，使得对于任意 $\widetilde{\varepsilon} > 0$，且概率至少为 $1 - \widetilde{C}\left(\frac{1}{n^2} + \exp\left(\frac{-2\widetilde{\varepsilon}^2 n}{\widetilde{B}^2}\right)\right)$，

$$\max_{a,b \in [n]} \left| \left[\widetilde{K_n}\right]_{a,b} - \widetilde{\mathcal{K}}(x_a, x_b) \right| \leq m \widetilde{A} e \left(\widetilde{A} e + \widetilde{M}\right),$$

其中

$$\widetilde{M} = \max_{i \in [m]} \left(2 \max_{x \in \mathcal{X}} |\phi_i(x)|\right),$$

*The* 且 $C$ 取决于 $m$ 常数 $A$ 和 $B$ 取决于密度 $p$ 以及 $X$ 的几何结构，并 .

核差的概率界依赖于 [28] 和 [72]的结果；当 $h_2(n)$ 随 $n$衰减得足够缓慢时，则能证明它们的集中性。下面我们对命题6.3的两个方面作一些说明。

~1. 由于DPP($K_n$, $\omega_n$) 根据 Macchi-Soshnikov定理存在，因此DPP($K_n$, $\mu_n$)也存在。2. 关于 $\alpha$ 和 $\beta$ 的附加假设进一步限制了带宽参数 $h_1(n)$ 和 $h_2(n)$的选取，除此之外还需满足 [28] 和 [72]中的条件。
由于命题 6.3 中的集中界依赖于 $e$，因而同时依赖于 $h_1(n)$ 和$h_2(n)$，因此有必要指定带宽参数的可容许序列，以确保 $e$ 随 $n$适当地减小。

定义 6.3.1（可容许带宽）。两个带宽序列 $h_1(n) \leq H_1$ 与 $h_2(n) \leq H_2$ 是 $C$-可容许 的，如果它们满足：对任意 $\varepsilon > 0$，存在$N_h(\varepsilon, C)$，使得

$$n \geq N_h(\varepsilon, C) \Rightarrow \begin{cases} \min\left(h_1(n)^{1/2}, \left(\frac{\log(n)}{nh_2(n)^{d_{\mathcal{X}}}}\right)^{\kappa/2}\right) \leq \max\left(\frac{\varepsilon}{C}\right), \\ h_1(n)^{1/2} < 1/3, \\ \left(\frac{\log(n)}{nh_2(n)^{d_{\mathcal{X}}}}\right)^{\kappa/2} < \min\left(\frac{1}{3}, \frac{p_{min}}{2}\right). \end{cases} \tag{6.11}$$

特别地，方程（6.11）右端最上方的条件确保 $\alpha + \beta$趋于 0 （当 $n \to \infty$ 时）。底部的两个条件是命题 6.3 之假设的更严格变体。

命题6.3 随后可用于证明 $K_n$ 向 $\mathcal{K}$ 在可容许带宽参数下的集中性。下面的常数由（非常）粗略的界得到；关于更精确的集中性结果，请参见补充材料第F.2 节中的证明。

定理 6.4（调和系综的弱相干性）。假设 *6.1* 和 *6.2* 均成立。那么，存在某个 $A > 0$ 使得，对于 $3mA$-可容许的带宽参数 $h_1$（$n$）和 $h_2$（$n$），行列式点过程($K_n$, $\mu_n$) 与 行列式点过程($\mathcal{K}$ $\mu$)弱相干，其弱相干速率为

$$N_K(\delta, \varepsilon) = \max\left(\sqrt{\frac{2\widetilde{C}^2}{\delta}}, \frac{9A^2\widetilde{B}^2m^2\log\left(\frac{2\widetilde{C}}{\delta}\right)}{\varepsilon^2}, N_h(\varepsilon, C)\right). \tag{6.12}$$

~常数 $A$ 取决于 $A$ （因此也取决于 $X$ 的几何）、密度 $p$ 和核 $K$。

注释 6.5. 已经针对某些流形族 X，得到了调和系综行列式点过程 DPP(K $\mu$) 优于独立抽样的定量结果，但这些结果相当技术性。我们将在第 8 节简要讨论其中一些结果如何推广到离散调和系综行列式点过程 DPP($K_n, \mu_n$)。我们不强调这些定量结果，而是着重指出一个简单的方法论启示。一方面，DPP($K_n, \mu_n$) 是定义在有限集上的行列式点过程，其点 $X_n$ 是从一个未知流形 X 中根据未知概率分布 $\mu$ 抽取的；该分布与 DPP(K $\mu$) 弱相干，并且可以利用标准离散DPP采样算法轻松抽样。另一方面，即便只是构建 DPP(K $\mu$) 的核，也需要了解 Laplace-Beltrami算子 L 的特征结构，而这在大多数实际场景中并不现实。对于依赖于流形上的蒙特卡罗积分等方法的应用，DPP($K_n, \mu_n$) 提供了一种可行且很可能具有可证明优势的独立抽样替代方案。深入研究这些保证是未来工作的重要方向。

6.2.1. 示例：离散调和系综与独立同分布在 $\mathcal{S}^2$ 上进行蒙特卡罗的比较。为了佐证注释 6.5，我们展示 行列式点过程$(K_n,\ \mu_n)$ 在球面 $\mathcal{S}^2 \subseteq \mathbf{R}^3$, 上进行蒙特卡洛积分时的表现，并与独立同分布抽样进行比较。我们的目标是估计

$$I := \int_{\mathcal{S}^2} f d\omega \ \left(= \frac{4\pi}{3}\right), \quad \text{where} \quad f(x,y,z) = z^2,$$

仅凭在$\mathcal{S}^2$ 中依据某个概率测度 $\mu$ （其密度为 $p$）独立抽取的点样本 $X_n = \{x_1,...,x_n\}$ 即可。在实践中，我们均匀抽取这些 $x_i$，但在估计 $I$ 时并不利用这一信息。类似地，我们也不假定知道流形$\mathcal{X}$ $=\mathcal{S}^2$。我们强调，这种设定与流形上的常规数值积分设定有本质区别，后者通常假定已知流形；关于这种设定下的蒙特卡罗积分，可参见 [29], ；关于基于行列式点过程的技术，可参见 [52, 48]。接下来，我们将计算蒙特卡罗估计

$$I_n := \int_{\mathcal{S}^2} f d\omega_n,$$

这里我们回顾一下， $\omega_n$ 是由 $e[p]$计算得到的 $\omega$ 的近似值。我们强调，对于给定的 $X_n$, $I_n$ 是一个固定的确定量。在命题 6.3的设定下， $|I - I_n| \leq A_f(\widetilde{\varepsilon} + \beta)$ 以至少 $C_f\left(\frac{1}{n^2} + \exp\left(\frac{2-\widetilde{\varepsilon}^2 n}{B_f^2}\right)\right)$ 的概率成立，其中 $A_f$、 $B_f$和$C_f > 0$为常数（关于这一现象的更多细节，请参见命题 6.3的证明），因此，对 $I_n$ 的估计确实会导出对 $I$的估计。

在我们的实验中，我们取 $n = 3000$ 个点。密度估计 $e[p]$ 使用归一化示性函数 $l = \frac{\mathbf{1}_{[0,1]}}{\pi}$ 和 $h_2(n) = \left(\frac{\log(n)}{n}\right)^{1/4}$ 构造。为计算近似特征函数 $v_i$，我们取 $h_1(n) = \left(\frac{\log(n)}{n}\right)^{1/16}$。

作为基线，我们考虑一种简单策略：子样本 $\mathcal{S}$ 中的 $m$ 个点从 $X_n$ 中独立有放回地抽取，抽取 $x_i$ 的概率 $p_i$ 与 $e[p](x_i)$ 成正比，估计量为

$$I_{\mathcal{S}}^{\text{iid}} = \sum_{i=1}^{n} \frac{f(x_i)\epsilon_i}{n m p_i e[p](x_i)}. \tag{6.13}$$

注意到 $\mathbf{E}_{\mathcal{S}}(\epsilon_i) = mp_i$，通过直接计算可得，这是 $I_n$ 的无偏估计量。我们将这一基线策略与基于行列式点过程的策略进行比较；在后者中，点 $X_n$ 从离散调和系综中抽样。相应的估计量定义为

$$I_{\mathcal{S}}^{\text{DPP}} = \sum_{i=1}^{n} \frac{f(x_i)\epsilon_i}{n e[p](x_i)[K_n]_{i,i}/n}, \tag{6.14}$$

这里我们在分母中保留 $n/n$ 因子，以凸显其与 $I_{\mathcal{S}}^{\text{iid}}$ 的定义相似。利用方程（2.2）和（2.3），可以直接得到 $I_{\mathcal{S}}^{\text{DPP}}$ 是 $I_n$ 的无偏估计量。

对于给定的样本 S 和估计量 $I_{\mathcal{S}}$，我们计算两种估计量的相对误差 $\frac{|I_{\mathcal{S}} - I|}{I}$。我们对样本 S 的 $n_{\mathcal{S}} = 1000$ 次抽样计算这些误差，并在图2中记录两种估计量在不同 $m \in$ {1,2, 4, 8, 16, 32, 64, 128} 取值下的平均误差，该平均值是对 $X_n$ 的 10 次实现取平均得到的。与独立同分布对应方法相比，基于DPP的估计量在定性上对 $m$ 的取值更为敏感，我们观察到两种不同的情形。1/ 对于较小的 $m$， $I_{\mathcal{S}}^{\text{DPP}}$ 的相对误差以更快的速率下降。2/ 对于较大的 $m$，情况则不再如此，相对

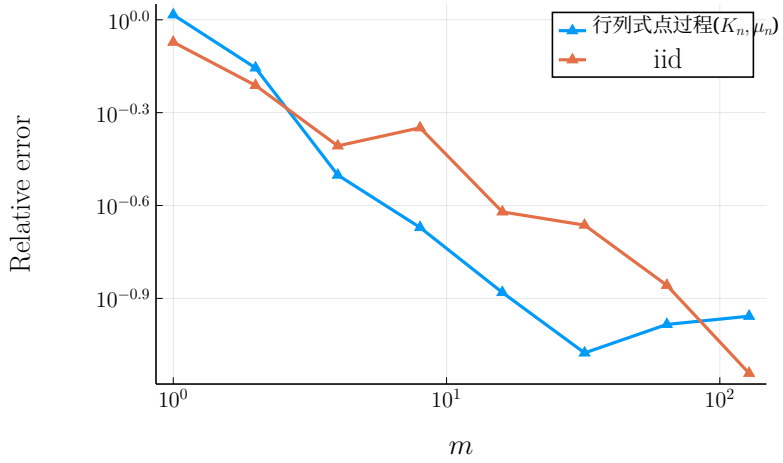


图 2：相对误差作为 $m$ 的函数（对数-对数尺度）。

误差对于 $I_S^{\mathrm{DPP}}$ 实际上会增大。在这种情形下， $m$ 相对于 $n$ 已不再足够小，因此前 $m$ 个特征函数 $\phi_i$ 无法通过 $v_i$ 忠实地近似$\mathcal{L}$ ， DPP($K_n, \mu_n$) 的表现也不再与 DPP($\mathcal{K}$ $\mu$) 相似

注记 6.6. 对于我们在Julia中的实现，[22] 我们使用格拉姆-施密特过程的朴素实现来得到 $v_i$'s，它们来自 $u_i$'s。该计算存在数值不稳定性，且计算出的 $u_i$'s可能并非标准正交，尤其是当 $m$的值较大时。特别地， $K_n$ 的特征值可能超过1。为缓解该问题，我们在实现中进一步将 $K_n$ 除以其最大特征值，以便能够1/确保我们的行列式点过程始终存在，以及2/依赖现有的行列式点过程采样器。我们指出，该问题可通过更精细的工程处理加以解决，但对我们的定性观察影响很小。

在 $e[p]$的计算中，我们需要知道流形$\mathcal{X}$ 的维数（此处为$d_{\mathcal{X}} = 2$）。为简便起见，我们在流程中假定其已知，但我们强调，该值可以根据点云 $X_n$ 本身加以估计；关于这一经典问题的近期综述，参见例如 [14]。特别地，对于性态良好的流形，维数估计的极大极小风险随 $n$ [40]呈指数下降。

**7. 随机图上行列式点过程的离散到连续极限**。最后一个例子涉及潜在位置随机图 [22]，在这类图中，点 $X_n$ 的位置不可用，但可以通过这些点之间的随机边隐式表示。我们将展示，即使在这种具有挑战性的情形下，也可以利用弱相干性构造离散到连续极限。与前面的例子相比，这里有两个重要的技术差异：首先，核 $K_n$ 除了 $X_n$ （随机边）之外，还多了一层随机性。其次，也是更关键的一点，在这里 $[K_n]_{ij}$ 将不收敛到$\mathcal{K}(x_i,\ x_j)$ ，不像前两个例子那样；相反，我们将依赖定理 4.1的第二个条件，即弗罗贝尼乌斯范数和迹的集中性。

请注意，从应用角度而言，本节的结果应被视为一种概念验证，而非使用行列式点过程对（随机）图进行抽样的实用方法。

[22]可访问网址：https://gitlab.irisa.fr/hjaquard/discrete-to-continuous-dpps。

事实上，文献中使用行列式点过程进行图采样，通常采用随机森林的根过程（例如，[3, 58]）。遗憾的是，本文的结果似乎并未直接涵盖这一过程。研究这一更复杂的过程是未来工作的一个重要方向。

**7.1. 潜在位置随机图与连通性核。** 在最常见的设定中，潜在位置随机图由所谓的连通性核 $\mathcal{W}$ 定义：$\mathcal{X} \times \mathcal{X} \to [0,1]$，其中紧集 $\mathcal{X} \subseteq \mathbf{R}^d$。给定这样的核、一个稀疏度 $\alpha_n \leq 1$ 以及$\mathcal{X}$上的概率分布 $\mu$，通过先根据 $\mu$独立同分布地抽取 $n$ 个点 $n$ ,..., $x_n \subseteq \mathcal{X}$，然后生成边，得到一个含有 $n$ 个节点的潜在位置随机图

$$a_{i,j} \sim \mathrm{Ber}(\alpha_n \mathcal{W}(x_i, x_j)). \tag{7.1}$$

这里，$a_{i,j} = a_{j,i}$ 是图的邻接矩阵 $A$ 的元素。需要强调的是，这个邻接矩阵通常是唯一可用的信息：与前两个例子不同，这些位置 $x_i$ 是未知的。系数 $\alpha_n$ 控制图的稀疏度，边的期望数量为$\mathcal{O}(\alpha_n n^2)$。通常会区分不同的密度状态：当 $\alpha_n \in \mathcal{O}(1)$时称为稠密图；当 $\alpha_n \in \mathcal{O}(\frac{1}{n})$时称为稀疏图；介于两者之间的任意速率对应相对稀疏图。一般而言，当 $\alpha_n \gtrsim_{\log(n)} n$ 时，可以得到非渐近收敛结果，而这正是我们考虑的状态。

---

假设核 $\mathcal{K} = \mathcal{W}$ 是一个有效的 DPP 核——也就是说，$\mathcal{W}$ 是连续且对称的，并且与 $\mathcal{W}$ 及 $\mu$ 相关联的积分算子 $T_{\mathcal{W}}$ 的特征值位于 $[0,1]$, 中，因此根据 Macchi-Soshnikov定理，DPP($\mathcal{K}$ $\mu$) 存在——我们考察以下问题：仅通过观测邻接矩阵 $A$，我们是否拥有足够的信息来近似关于 $\mathcal{K}$的 DPP? 或者，在我们的语境下，我们能否仅利用 $A$ 构造一个矩阵 $K_n$，使得 DPP($K_n, \mu_n$) 与 DPP($\mathcal{K}$ $\mu$) 弱一致? [23]

7.2. 通过通用奇异值阈值化实现弱相干性。首先需要观察到

$$\mathbf{E}\left[\frac{A}{\alpha_n} \mid X_n\right] = \mathcal{K}_{|X_n \times X_n}.$$

因此 $A$ 确实是估计格拉姆矩阵的一个良好候选。事实上，可以证明二者之间的收敛发生在算子范数 [47]意义下。然而，这不足以应用定理 1.1：显然，它并不逐项收敛，而且这种集中性在弗罗贝尼乌斯范数下也不成立（后者只能估计伯努利变量的方差之和）。

事实证明，借助矩阵估计文献中的一种强大工具——所谓的通用奇异值阈值法（USVT） [20]，通过考察一个经过修改的边加权图[24]，即可同时恢复 Frobenius 范数集中性和迹集中性。与原始 USVT 相比，完整结果仍需进行大量修改。

---

[23]请注意， $x_i$并未知。因此，对 $x_i$进行抽样，实际上是从1、.....、 $n$中抽取指标，如第2节所述。

[24]使用其他工具或许也能获得弱相干性结果。采用 *USVT* 近似有两个优点：*1/* 可以利用现有的强大集中性结果；*2/ USVT* 估计的计算效率很高（参见例如， [53] 的引言中对此话题的讨论）。

在本节其余部分，我们假设对于所有 $x \in \mathcal{K}(x,x) = c\,\mathcal{X}$， $c \leq 1$是某个常数：例如，当$\mathcal{W}$ 是径向核时，这一条件便成立；径向核是潜在位置随机图中的经典假设。为简单起见，我们还假设已知 $c$ 。

7.2.1. 基于 *USVT* 的核。 记 $A = \sum_{i=1}^{n} \lambda_i u_i u_i^t$ 为$A$ 的特征分解，并设 $\gamma_n \geq 0$ 为某个非负阈值，则 USVT 估计量定义为

$$\widetilde{A}_{\gamma_n} = \frac{1}{\alpha_n} \sum_{\lambda_i \geq \gamma_n} \lambda_i u_i u_i^t.$$

已证明，该估计量足以恢复 Frobenius 范数集中性 [39] (见补充材料G 节），但不足以确保满足迹集中条件。此外，也无法保证相应的行列式点过程存在。为了解决这些问题，我们考虑对 USVT 估计量进行修改。~第一步是对 $A_{\gamma_n}$进行对角扰动:

$$\bar{A}_{\gamma_n} = \widetilde{A}_{\gamma_n} + C(\widetilde{A}_{\gamma_n})I, \quad \text{with} \quad C(\widetilde{A}_{\gamma_n}) = \max\left(c - \frac{\operatorname{tr}\left(\widetilde{A}_{\gamma_n}\right)}{n},\ 0\right), \tag{7.2}$$

其中回顾一下， $c$ 是$\mathcal{K}(x,x)$的值，因此也是$\operatorname{tr}(\mathcal{K}_{X_n \times X_n}/n)$的值。如下文命题7.1 的证明所示，对于 $A_{\gamma_n}$，迹条件总是成立，但相应的行列式点过程可能并不存在。为满足 Macchi-Soshnikov 定理的假设，我们考虑核

(7.3)

$$K_n = C'(\bar{A}_{\gamma_n})\bar{A}_{\gamma_n}, \quad \text{where} \quad C'(\bar{A}_{\gamma_n}) = \min\left(\frac{n}{\lambda_{\max}(\bar{A}_{\gamma_n})},\ \left(1 + \frac{1}{(\alpha_n n)^{1/4}}\right)^{-1}\right),$$

~其中， $\lambda_{\max}(A)$ 是_ $A$ $C$ $A_{\gamma_n}$ $\geq 0$ $A_{\gamma_n}$ 的最大特征值。由于(), 的特征值均为非负；又由于 $0 \leq C'(A_{\gamma_n}) \leq$ $\lambda_{\max}$ ( ~~$\bar{A}_{\gamma_n}$~~)，~~我们~~得出 $K_n$ 的特征值属于 $[0,n]$。由于 $K_n$ 确实是对称的，因此 DPP($K_n, \mu_n$) 存在。接着，我们可以证明 $K_n$ 在弗罗贝尼乌斯范数和迹集中意义下趋于 $\mathcal{K}$。

命题 7.1. 假设 $\alpha_n \gtrsim \log n/n$。对于任意 $q > 0$，存在三个常数 $\rho_q$、$b_q$、$B_q \in \mathbf{R}_{>0}$，使得对于任意 $\delta \in (0,1)$， $\varepsilon > 0$，并且对于 $\gamma_n = \rho_q(\alpha_n n)^{3/4}$，

$$n \geq N_K(\delta,\varepsilon) \ \Rightarrow\ \mathbb{P}\left(\max\left(\left\|\frac{K_n - \mathcal{K}_{|X_n \times X_n}}{n}\right\|_F, \left|\operatorname{tr}\left(\frac{K_n}{n}\right) - \operatorname{tr}\left(\frac{\mathcal{K}_{|X_n \times X_n}}{n}\right)\right|\right) \geq \varepsilon\right) \leq \delta,$$

其中

$$N_K(\delta,\varepsilon) = \max\left(\frac{1}{\delta^q}, \frac{b_q}{\alpha_n \varepsilon^8}, \frac{B_q}{\alpha_n \varepsilon^4}\right). \tag{7.4}$$

在此，根据所需的保证，可以调整参数 $q$ ，以提高该界成立的概率，但代价是乘法常数变大。命题 7.1 的证明详见补充材料第G 节，是对 [39]中结果的推广。其依据是 [47] 中的一个集中性结果以及谱不等式；我们强调，该结果仅在$\alpha_n \gtrsim \log(n)\ n$ 时成立。

---

注 7.2. $\varepsilon = \varepsilon_n$ 可达到的速率直接取决于稀疏度 $\alpha_n$。对于较稀疏的图，若 $\alpha_n \gtrsim$ $\log_n{}^n$，则特别是 $\frac{b_q}{\alpha_n \varepsilon^8}$ 项将速率限制在不快于 $\mathcal{O}\left(\frac{1}{n}\log()^{1/8}\right)$。一般而言，图越稀疏，可达到的速率就越慢。

现在我们可以使用定理 4.1, $ii$)，得到如下结果。

推论 7.3。 在命题 *7.1* 的设定下，DPP$(K_n, \mu_n)$ 与 DPP$(\mathcal{K}\ \mu)$弱相干。对于给定的 $q$，弱相干速率由方程 $N_K$ 中的 (7.4)给出。

注 **7.4**。实际实现我们的估计量$K_n$存在一些障碍：首先，稀疏度 $\alpha_n$ 未知，需要进行估计；其次，超参数 $\gamma_n$ 只在乘法常数的意义下已知，因此也必须通过验证程序进行调整；第三，对角线值 $c$ 可能并非总是已知。如上所述，此处给出的估计量 $K_n$ 主要是一个理论性概念验证。在实践中，使用行列式点过程进行图采样通常通过随机森林完成，这留待未来工作研究。

**8. 讨论与展望**。 除了线性回归（例如， [26, 25]）这类结构非常规整的问题之外，要为离散行列式点过程建立严格且可量化的优于独立抽样的保证，一直是众所周知的难题。另一方面，尽管计算成本有所增加，但许多实验结果支持行列式点过程在不同任务上相较于独立同分布抽样具有实际统计优势。为了量化这种优势，我们提出了一种系统方法，可用于将连续行列式点过程的统计保证转化为离散行列式点过程的统计保证。具体而言，线性统计量的所有矩的集中性由核的集中速率控制，该速率可以基于最大范数，也可以基于弗罗贝尼乌斯范数和迹。这些准则既适用于固定大小的投影行列式点过程（对于这类过程，最大范数集中性已是强有力的工具），也适用于连续核的满秩、高噪声估计。这些统计量的逼近速率下界为$\mathcal{O}\left(\sqrt{\frac{\log(\frac{1}{\delta})}{n}}\right)$，与 $n$ 独立同分布变量的经典集中速率相匹配；在这一方面，我们预计该结果是最优的。

虽然我们的论述聚焦于定义在紧空间上的行列式点过程，但第4.1 节中对核误差的分析非常一般。另一方面，第4.2节中对测度误差的分析在很大程度上依赖于紧性假设，但可以推广到集合 $X_n$ ，即按照例如强度递增的泊松过程 $r$（$n$）$\mu$ 抽样得到的集合，其中 $\mu$ 为某个拉东测度，且 $r$（$n$）随 $n$递增；这可借助于泊松样本的集中性结果（参见例如， [6]）。

就推广而言，我们的结果值得针对第5、6 和7节中所研究的三种行列式点过程分别开展广泛的实证研究。现在让我们讨论其他几个视角。

8.0.1.**中心极限定理。**对于定义在 $\mathcal{X}$ 上的某些函数类，连续多元正交多项式系综和调和系综已获得快于独立同分布情形的中心极限定理（CLT） [10, 52], ，从而为基于行列式点过程的蒙特卡罗积分建立了强有力的保证。那么，当从 $X_n$ 而非 $\mathcal{X}$ 中抽样，且使用离散行列式点过程时，是否也能为定义在 $\boldsymbol{X}$ 上的函数建立中心极限定理? 特别地，根据 [64],中提出的判据，只要（1点）线性统计量的方差相对于其期望增长得足够快，就基本上可以为行列式点过程建立中心极限定理。当 $n$ 足够大时，弱相干性能够控制这两个量，因此很可能可以将现有的中心极限定理推广到离散行列式点过程。需要强调的是，除了速率之外，这类结果还能为具有挑战性的统计情境提供有用的保证；在这些情境中，我们关注的是定义在未观测到的潜在空间 $\mathcal{X}$上的函数（例如，第6节中所述的未知流形）。

8.0.2. *L*-系综。在实践中，研究者通常更方便将离散行列式点过程定义为（扩展）L-系综 [67], ，而不是像方程（2.7）和（2.8）中那样通过其边缘概率来定义。L-系综描述每个可能的 $\mathcal{X}$样本的概率，而不是样本 $\mathcal{S}$中给定点的共现概率，并使用辅助核 $L$。特别地，核 $L$ 是边缘核的 Lipschitz 函数 $K$，因此我们可以预期定理 4.1中的集中界可以推广到这一情形（例如，Frobenius范数界可借助矩阵分析中的结果进行转换 [41]）。尽管这一思路相当简单， $L$-系综的连续对应形式是通过过程的 Janossy 密度描述的，而这种扩展还需要额外的技术处理。

8.0.3. 随机森林的根。 正如我们在第 7 节所提到的，我们工作的一个重要动机是研究随机森林根过程的连续极限 [3]。这是一个核为 $q(L+qI)^{-1}$ 的行列式点过程，其中 $q>0$，且 $L$ 是通常的图拉普拉斯算子，即 相对于计数测度 $\sum_{x\in X_n}\delta_x$ 定义。使用这种非归一化测度会带来一个重要差异：这相当于在定理 $K_n$ 4.1 中将 (分别将 $\mathcal{K}$) 替换为 $nK_n$ (分别将 $n\mathcal{K}$)，这对核的集中提出了严格得多的条件。尤其是，对于我们在第 7 节考察的这类随机图，我们很可能无法满足该条件。

另一方面，我们针对 非紧集 空间 $\mathcal{X}$ （第 4.1 节）提出的弱相干性判据在这种情形下仍可能满足：这是因为，要使这些结果适用，我们只要求核在所有紧集 $X_n\cap C$ 上集中， $C$ 其中 $\mathcal{X}$，而核 $K_n$ 则可以由整个数据集 $X_n$ 构造。事实证明，文献中已研究了非紧潜在空间上的若干重要随机图类别，例如所谓的 graphex [71, 17, 16], ，它们最广为人知的是作为稀疏图 极限 的适当框架；此外还有双曲随机图 [42]以及其他相关的稀疏图模型（参见例如， [69]）。针对这些模型研究局部的类 USVT 集中性结果，将是未来工作的方向。

8.0.4. 单样本DPP识别。最后，让我们简要回到第 1.2 节中讨论过的基于 [60], 的单个无限样本进行DPP估计的情形。考虑定义在 $\mathrm{R}^d$ 上、核为 K 且相对于勒贝格测度 $\lambda^{\otimes r}$ 定义的一个DPP。给定来自 DPP(K $\lambda^{\otimes r}$) 的单个无限样本 S，目标是使用最大似然准则，根据参数族 $(\mathrm{K}_\theta)_\theta$ 估计核 K，其中观测窗口为 $W\subseteq \mathrm{R}^d$。在实践中，无法计算 S $\cap W$ 的似然 $l_W(\theta)\tilde{}$，因此 [60] 提出了替代似然 $l_W(\theta)$ 来近似它。不深入探讨技术细节，他们得出了一系列结果：首先，他们证明，在任意精细的网格 $\eta\mathrm{Z}^d$ $(\eta>0)$ 上定义的行列式点过程 $\tilde{l}_W(\theta)$，随着窗口大小 $W$ 增大，几乎必然收敛到 $l_W(\theta)$（此处， $\eta$ 保持不变）；其次，他们证明，在 $\eta Z^d$ 上适当定义的DPP会随着 $\eta\to 0$ 弱收敛到定义在 $\mathrm{R}^d$ 上的DPP。

这留下了一些尚待解答的问题：1/ 是否 $l_W$ 收敛到一些对于定义在 $\mathrm{R}^d$ 上的行列式点过程而言？2/ 最大似然估计量 （而不只是似然），也就是说， $\eta\mathbf{Z}^d$ 上的一个核，是否收敛到 $\mathcal{K}$？3/ 这种核估计会如何影响下游处理任务？我们预计，弱相干性（在非紧空间上）可能为回答其中一些问题提供合适的框架。

附录 A：点过程的弱收敛

弱收敛将实值随机变量的依分布收敛推广到定义在称为波兰空间的一般空间上的测度。本节的目标是

简要回顾这一概念，尤其是它作为点过程极限概念的具体形式。简言之，它要求一大类统计量的期望收敛。关于这一主题的详细论述，读者可参阅 [70, 24]。

考虑定义在波兰空间 $\Omega$ 上的一列测度 $\lambda_n$。[25] 若 $\lambda_n$ 在 $\Omega$ 上弱收敛到测度 $\lambda$，则称

$$\int_\Omega h d\lambda_n \xrightarrow{n\to\infty} \int_\Omega h d\lambda$$

对所有有界连续函数 $h: \Omega \to \mathbf{R}$，均成立。换言之，任意可用函数 $f$ 关于 $\lambda_n$ 的积分都必须收敛到其关于 $\lambda$ 的积分。

为了涵盖点过程的情形，我们需要将 $\Gamma$ 的局部有限子集与取值于 $\{0,1\}$ 的局部有限测度对应起来；这类测度称为 $\Gamma$ 中的构型：

$$\mathrm{Conf}(\Gamma) = \left\{ \mathcal{S} = \sum_{x\in B} \delta_x \ ; \ B \subseteq \Gamma, \text{ such that } \mathcal{S}(C) < \infty \text{ for all compact sets } C \subseteq \Gamma \right\}.$$

$\mathsf{Conf}(\Gamma)$ 中的点构型与 $\Gamma$ 的局部有限子集一一对应，而点过程也可视为取值于 $\mathsf{Conf}(\Gamma)$ 的随机变量。现在固定一个第二可数局部紧豪斯多夫空间 $\Gamma$。考虑其构型空间 $\mathsf{Conf}(\Gamma)$，并赋予它模糊拓扑，即满足以下条件的拓扑：当$\mathcal{S}_n \xrightarrow{n\to\infty}\mathcal{S}$ 时，

$$\int_\Gamma f d\mathcal{S}_n \xrightarrow{n\to\infty} \int_\Gamma f d\mathcal{S}$$

对于所有具有紧支集的连续 $f: \Gamma \to \mathbf{R}$。特别地，这使得 $\mathsf{Conf}(\Gamma)$ 成为一个波兰空间，从而可以在其上定义弱收敛。

现在我们可以考虑在 $\Gamma$ 上的一列点过程 $\mathcal{P}_n$。这里，我们不把点过程视为一个随机局部有限子集，而是视为一个取值于 $\mathsf{Conf}(\Gamma)$ 的随机变量，因此其对应的分布不过是 $\mathsf{Conf}(\Gamma)$ 上的一个测度。接着，当这一列测度所对应的分布收敛到 $\mathcal{Q}$ 的分布时，我们称点过程 $\mathcal{P}$ 是 $\mathcal{P}_n$ 弱收敛到的极限，即当

$$\mathbf{E}_{\mathcal{S}_n\sim\mathcal{P}_n}[h(\mathcal{S}_n)] \xrightarrow{n\to\infty} \mathbf{E}_{\mathcal{S}\sim\mathcal{Q}}[h(\mathcal{S})]$$

对于所有关于模糊拓扑连续的有界函数 $h: \mathsf{Conf}(\Gamma) \to \mathbf{R}$。从实际角度来看，这意味着，对于任何要在随机构型上计算的可接受统计量 $h$，其在 $\mathcal{P}_n$ 下的期望应收敛到其在 $\mathcal{Q}$ 下的期望。

事实上，弱收敛满足一个所谓的连续映射定理，这是一种稳定性结果，与第3.2节的矩映射定理类似：如果 $\lambda_n$ 弱收敛到 $\lambda$，那么对于任意连续函数 $g$，$\int_\Omega (h\circ g) d\lambda_n \to \int_\Omega (h\circ g) d\lambda$。这两个结果的不同之处在于，矩映射定理涉及对随机点过程计算得到的统计量的矩，并且还额外保证了集中速率。

[25] 我们赋予该空间其 Borel $\sigma$-代数。

## 附录 B：第 4 节中缺失的证明

为便于阅读，我们仅给出适用于矩阵的命题 4.6 和 4.10的变体的证明。由于我们的证明仅依赖于三角不等式和柯西-施瓦茨不等式，因此，关于一般核的版本只需对一般核应用完全相同的论证，并将相应的求和替换为关于$\mathrm{supp}(\varphi_r)$ 相对于 $\nu_n$ 的积分（详情请参见下文的注释 B.2和 B.4 ）。

我们先回顾一个有用的恒等式。

命题 B.1。 设 $a_1,...,a_r \in \mathbf{C}$ 且 $b_1,...,b_r \in \mathbf{C}$。那么，

$$\prod_{i=1}^{r} a_i - \prod_{i=1}^{r} b_i = \sum_{j=1}^{r} \left( \prod_{k=1}^{j-1} a_k \right) (a_j - b_j) \left( \prod_{k=j+1}^{r} b_k \right). \tag{B.1}$$

**B.1. 命题 4.6 的证明。** 为了简化若干记账操作，适当地调整我们的形式体系会很有帮助。考虑一个有限集 $I_r = \{i_1,...,i_r\}$，将其视为有限指标集，并设有一个值函数$\mathtt{v} : I_r \to [n]$ ，为每个指标赋予一个值。我们的目标是证明，对于$b(\mathtt{v}) = |\det(A)_{\mathtt{v}(I_r)} - \det(B)_{\mathtt{v}(I_r)}|$,

$$b(\mathtt{v}) \le r! \sum_{j=1}^{r} \left( \max_{k,l\in[n]} |A_{k,l}| \right)^{j-1} \left( \max_{k,l\in[n]} |A_{k,l} - B_{k,l}| \right) \left( \max_{k,l\in[n]} |B_{k,l}| \right)^{r-j},$$

其中我们采用如下约定：$\det(M)_{\mathtt{v}(I_r)} = 0$ 当 $\mathtt{v}$ 不是单射时。读者可以验证，这等价于命题 4.6 的表述。记 $\mathfrak{S}_{I_r}$ 为指标 $I_r$ 上的排列集合（特别强调，这些排列作用于 $I_r$ 本身，而非 $\mathtt{v}(I_r)$）。依次应用三角不等式和方程（B.1），可得

$$\begin{aligned}
b(\mathtt{v}) &= \left| \sum_{\sigma\in\mathfrak{S}_{I_r}} \mathrm{sgn}(\sigma) \prod_{j=1}^{r} A_{\mathtt{v}(i_j),\mathtt{v}(\sigma(i_j))} - \sum_{\sigma\in\mathfrak{S}_{I_r}} \mathrm{sgn}(\sigma) \prod_{j=1}^{r} B_{\mathtt{v}(i_j),\mathtt{v}(\sigma(i_j))} \right| \\
&\le \sum_{\sigma\in\mathfrak{S}_{i_r}} \left| \prod_{j=1}^{r} A_{\mathtt{v}(i_j),\mathtt{v}(\sigma(i_j))} - \prod_{j=1}^{r} B_{\mathtt{v}(i_j),\mathtt{v}(\sigma(i_j))} \right| \\
&= \sum_{\sigma\in\mathfrak{S}_{i_r}} \left| \sum_{j=1}^{r} \left( \prod_{k=1}^{j-1} A_{\mathtt{v}(i_k),\mathtt{v}(\sigma(i_k))} \right) \left( A_{\mathtt{v}(i_j),\mathtt{v}(\sigma(i_j))} - B_{\mathtt{v}(i_j),\mathtt{v}(\sigma(i_j))} \right) \left( \prod_{k=j+1}^{r} B_{\mathtt{v}(i_k),\mathtt{v}(\sigma(i_k))} \right) \right| \\
&\le r! \sum_{j=1}^{r} \left( \max_{k,l\in[n]} |A_{k,l}| \right)^{j-1} \left( \max_{k,l\in[n]} |A_{k,l} - B_{k,l}| \right) \left( \max_{k,l\in[n]} |B_{k,l}| \right)^{r-j},
\end{aligned}$$

这就证明了命题 4.6。

注 B.2. 为证明命题 4.7，我们可以重复完全相同的论证，将矩阵 $A$ 和 $B$ 替换为核 $K_n^{\mathcal{P}}$和 $K_n^{\mathcal{Q}}$。在这种情况下，我们不再能考虑赋值 $\mathtt{v} : I_r \to [n]$，而形如 $A_{\mathtt{v}(i_j)\,\mathtt{v}(\sigma(i_j))}$ 的表达式则替换为 $K_n^{\mathcal{P}}(x_j, x_{\sigma(j)})$，其中 $[r]$的一个给定排列为 $\sigma$。

**B.2. 引理的证明 4.10.** 我们再次采用第 B.1 节中的形式体系，并考虑一个有限指标集 $I_r$ 以及值函数 $\mathrm{v}\colon I_r \to \mathbf{R}$。我们将证明

$$\sum_{\mathrm{v}:I_r\to[n]} |\det(A)_{\mathrm{v}(I_r)} - \det(B)_{\mathrm{v}(I_r)}| \le (r\times r!) M_{A,B}^{r-1} \max\left(\|A-B\|_F, |\mathrm{tr}\,(A) - \mathrm{tr}\,(B)|\right),$$

其中 $M_{A,B} = \max\left(h(A,B)\ t(A,B)\right)$，对于 $h(A,B) = \max\left(\|A\|_F, \|B\|_F\right)$ 以及 $t(A,B) = \max\left(\mathrm{tr}(A), \mathrm{tr}(B)\right)$。这不过是对引理 4.10 的重新表述。

我们首先应用三角不等式，得到

$$\sum_{\mathrm{v}:I_r\to[n]} |\det(A)_{\mathrm{v}(I_r)} - \det(B)_{\mathrm{v}(I_r)}| \le \sum_{\sigma\in\mathfrak{S}_{I_r}} \left| \sum_{\mathrm{v}:I_r\to[n]} \left( \prod_{j=1}^{r} A_{\mathrm{v}(i_j),\mathrm{v}(\sigma(i_j))} - \prod_{j=1}^{r} B_{\mathrm{v}(i_j),\mathrm{v}(\sigma(i_j))} \right) \right|$$

因此，利用方程 (B.1) 中的 $E_\sigma = \left|\sum_{\mathrm{v}:I_r\to[n]} \left(\prod_{j=1}^{r} A_{\mathrm{v}(i_j)\,\mathrm{v}(\sigma(i_j))} - \prod_{j=1}^{r} B_{\mathrm{v}(i_j)\,\mathrm{v}(\sigma(i_j))}\right)\right|$，

$$E_\sigma \le \sum_{j=1}^{r} \left| \sum_{\mathrm{v}:I_r\to[n]} \left( \prod_{k=1}^{j-1} A_{\mathrm{v}(i_k),\mathrm{v}(\sigma(i_k))} \right) \left( A_{\mathrm{v}(i_j),\mathrm{v}(\sigma(i_j))} - B_{\mathrm{v}(i_j),\mathrm{v}(\sigma(i_j))} \right) \left( \prod_{k=j+1}^{r} B_{\mathrm{v}(i_k),\mathrm{v}(\sigma(i_k))} \right) \right|.$$

为简化接下来的方程，我们引入另一种记号，并重写每一项

$$E_{\sigma,j} = \left| \sum_{\mathrm{v}:I_r\to[n]} \left( \prod_{k=1}^{j-1} A_{\mathrm{v}(i_k),\mathrm{v}(\sigma(i_k))} \right) \left( A_{\mathrm{v}(i_j),\mathrm{v}(\sigma(i_j))} - B_{\mathrm{v}(i_j),\mathrm{v}(\sigma(i_j))} \right) \left( \prod_{k=j+1}^{r} B_{\mathrm{v}(i_k),\mathrm{v}(\sigma(i_k))} \right) \right|$$

as

$$E_{\sigma,j} = \left| \sum_{\mathrm{v}:I_r\to[n]} \prod_{k=1}^{r} c(\sigma,j,\mathrm{v},k) \right|,$$

其中

$$c(\sigma,j,\mathrm{v},k) = \begin{cases} A_{\mathrm{v}(i_k),\mathrm{v}(\sigma(i_k))} & \text{if } k<j, \\ A_{\mathrm{v}(i_j),\mathrm{v}(\sigma(i_j))} - B_{\mathrm{v}(i_j),\mathrm{v}(\sigma(i_j))} & \text{if } k=j, \\ B_{\mathrm{v}(i_k),\mathrm{v}(\sigma(i_k))} & \text{if } k>j. \end{cases}$$

现在考虑将 $\sigma$ 唯一分解为 $c$ 互不相交的环的形式 $\sigma = \gamma_1 \circ \ldots \circ \gamma_c$。那么，我们可以写成

$$\begin{aligned} E_{\sigma,j} &= \left| \sum_{\mathrm{v}:I_r\to[n]} \prod_{l=1}^{c} \left( \prod_{\substack{k\in[r] \\ i_k\in\mathrm{supp}(\gamma_l)}}^{r} c(\sigma,j,\mathrm{v},k) \right) \right| \\ &= \prod_{l=1}^{c} \left| \sum_{\mathrm{v}:I_r\to[n]} \left( \prod_{\substack{k\in[r] \\ i_k\in\mathrm{supp}(\gamma_l)}}^{r} c(\sigma,j,\mathrm{v},k) \right) \right|, \end{aligned}$$

第二个等式成立，是因为所有的环彼此不相交，而且我们认出这是一个“形式为 $\sum_i \sum_j a_i b_j = \left(\sum_i a_i\right)\left(\sum_j b_j\right)$”的恒等式。

下文中，我们记作

$$E_{\sigma,j,l} = \left| \sum_{\mathrm{v}:I_r \to [n]} \left( \prod_{\substack{k\in[r] \\ i_k \in \mathrm{supp}(\gamma_l)}}^{r} c(\sigma, j, \mathrm{v}, k) \right) \right|$$

each 上述乘积中的项，并根据循环的长度进行穷尽讨论 s.

如果 $\gamma_l$ 的长度为 $l = 1$，那么它在其支集 $\mathrm{supp}(\gamma_l) = \{i_k\}$ 上等于恒等映射，并且有：

- $E_{\sigma,\ j,\ l} = |\sum_{i=1}^{n} A_{i,\ i}| = |\mathrm{tr}(A)|$ 若 $k < j$;
- $E_{\sigma,\ j,\ l} = |\sum_{i=1}^{n} B_{i,\ i}| = |\mathrm{tr}(B)|$ 若 $k > j$;
- $E_{\sigma,\ j,\ l} = |\sum_{i=1}^{n} A_{i,\ i} - B_{i,\ i}| = |\mathrm{tr}(A) - \mathrm{tr}(B)|$ 若 $k = j$。

当 $\gamma_l$ 的长度至少为 2时，我们将在下文的命题 B.3 中证明

$$E_{\sigma,j,l} \leq \prod_{\substack{k\in[r] \\ i_k \in \mathrm{supp}(\gamma_l)}} \|C(j,k)\|_F,$$

其中

$$C(j,k) = \begin{cases} A & \text{if } k < j, \\ A - B & \text{if } k = j, \\ B & \text{if } k > j, \end{cases}$$

由此我们得到所需的不等式。

由于我们每次只需处理一个循环 $\gamma_l$，不妨不失一般性地假设 $\gamma_l(i_j) = i_{j+1}$（其中，根据定义，$i_{r_l+1} = i_1$）。于是，我们有

$$E_{\sigma,j,l} = \sum_{i_1,\ldots,i_{r_l}=1}^{n} \prod_{k=1}^{r_l} (C(j,k))_{i_k,i_{k+1}},$$

接下来只需证明如下结论。

命题 B.3. 设 $M_1,\ldots,M_s$ 表示具有实系数的 $s$ $n \times n$ 矩阵。于是，我们有

$$\left| \sum_{i_1,\ldots,i_s=1}^{n} \prod_{k=1}^{s} (M_k)_{i_k,i_{k+1}} \right| \leq \prod_{k=1}^{s} \|M_k\|_F.$$

证明。在 $s = 2$的情形下，应用柯西-施瓦茨不等式即可直接得到如下结果：

$$\left| \sum_{i_1,i_2=1}^{n} (M_1)_{i_1,i_2} (M_2)_{i_2,i_1} \right| \leq \|M_1\|_F \|M_2\|_F.$$

若 $s \geq 3$，由柯西-施瓦茨不等式可得

$$\left| \sum_{i_1,\ldots,i_{r_l}=1}^{n} \prod_{k=1}^{n} (M_k)_{i_k,i_{k+1}} \right| \leq \|M_1\|_F \left\| \left[ \sum_{i_3,\ldots,i_{r_l}=1}^{n} \prod_{k=2}^{n} (M_k)_{i_k,i_{k+1}} \right]_{i_1,i_2=1}^{n} \right\|_F,$$

在这种情况下，我们将通过归纳法证明

$$\|M_1\|_F \left\| \left[ \sum_{i_3,\dots,i_{r_l}=1}^{n} \prod_{k=2}^{n} (M_k)_{i_k,i_{k+1}} \right]_{i_1,i_2=1}^{n} \right\|_F^2 \le \prod_{k=1}^{s} \|M_k\|_F^2,$$

由此得到结论。

现在我们继续进行归纳。若 $s=3$,

$$\begin{aligned}
\left\| \left[ \sum_{i_3=1}^{n} (M_2)_{i_2,i_3}(M_3)_{i_3,i_1} \right]_{i_1,i_2=1}^{n} \right\|_F^2 &= \sum_{i_1,i_2=1}^{n} \left| \sum_{i_3=1}^{n} (M_2)_{i_2,i_3}(M_3)_{i_3,i_1} \right|^2 \\
&\le \sum_{i_1,i_2=1}^{n} \|(M_2)_{i_2,:}\|_2^2 \|(M_3)_{:,i_1}\|_2^2 \\
&= \sum_{i_1,i_2=1}^{n} \left( \sum_{i_3=1}^{n} |(M_2)_{i_2,i_3}|^2 \right) \left( \sum_{i_3'=1}^{n} |(M_3)_{i_3,i_1}|^2 \right) \\
&= \|M_2\|_F^2 \|M_3\|_F^2,
\end{aligned}$$

其中第二行是通过应用柯西-施瓦茨不等式得到的。当 $s>3$ 时，我们重复一个经典的次乘性论证，并且对于

$$C = \left\| \left[ \sum_{i_3,\dots,i_{r_l}=1}^{n} \prod_{k=2}^{n} (M_k)_{i_k,i_{k+1}} \right]_{i_1,i_2=1}^{n} \right\|_F^2,$$

我们最终得到：

$$\begin{aligned}
C &= \sum_{i_1,i_2=1}^{n} \left| \sum_{i_3=1}^{n} \left( (M_2)_{i_2,i_3} \sum_{i_4,\dots,i_{r_l}=1}^{n} \prod_{k=3}^{r_l} (M_k)_{i_k,i_{k+1}} \right) \right|^2 \\
&\le \sum_{i_1,i_2=1}^{n} \|(M_2)_{i_2,:}\|_2^2 \left\| \sum_{i_4=1}^{n} (M_3)_{:,i_4} \sum_{i_5,\dots,i_{r_l-1}=1}^{n} \prod_{k=4}^{r_l} (M_k)_{i_k,i_{k+1}} \right\|_2^2 \\
&= \sum_{i_1,i_2=1}^{n} \left( \sum_{i_3=1}^{n} |(M_3)_{i_2,i_3}|^2 \right) \left( \sum_{i_3,\dots,i_{r_l}=1}^{n} \left| \prod_{k=3}^{r_l} (M_k)_{i_k,i_{k+1}} \right|^2 \right) \\
&= \|M_3\|_F^2 \left\| \left[ \sum_{i_4,\dots,i_{r_l}=1}^{n} \prod_{k=3}^{r_l} (M_k)_{i_k,i_{k+1}} \right]_{i_3,i_1} \right\|^2 \\
&\le \prod_{k=1}^{s} \|M_k\|_F^2,
\end{aligned}$$

其中，不等式再次由柯西-施瓦茨不等式得到，而最后一行则由归纳假设推出。

□

应用命题B.3 于$E_{\sigma,\ j,\ l}$ 即可完成证明。请注意，在命题B.3的证明中，我们只用到了矩阵为实值这一事实。

备注B.4。与备注B.2类似，我们可以运用完全相同的论证来证明引理4.11。不过，不再考虑对赋值求和，而是改为对变量$(x_1, \ldots\ldots, x_r)$在$\mathrm{supp}(\varphi_r) = C^r_{\varphi_r}$上进行多重积分。特别地，对于命题B.3 ，相同的论证表明，对于实值核$K_1, \ldots\ldots, K_s : \mathcal{X} \times \mathcal{X} \to \mathbf{R}$,

$$\left| \underbrace{\int_{C_{\varphi_r}} \ldots \int_{C_{\varphi_r}}}_{k \text{ times}} \prod_{k=1}^{s} K_k(x_k, x_{k+1}) d\nu_n(x_1) \ldots d\nu_n(x_s) \right| \leq \prod_{k=1}^{s} \|K_k\|_{L^2(\nu_n^{\otimes 2})}.$$

我们强调 $C_{\varphi_r}$ 是紧集，并且在证明过程中，所有相关积分确实都可以交换次序。

**B.3. 命题的证明 4.12.** 由三角不等式，我们有

$$E_n^{\mu} \leq \left| \int_{X_n^r} f d\mu_n^{\otimes r} - \int_{\mathcal{X}^n} \int_{X_n^r} f d\mu_n^{\otimes r} d\mu^{\otimes n}(X_n) \right| + \left| \int_{\mathcal{X}^n} \int_{X_n^r} f d\mu_n^{\otimes r} d\mu^{\otimes n}(X_n) - \int_{\mathcal{X}^r} f d\mu^{\otimes r} \right|.$$

我们用

$$M_1 = \left| \int_{X_n^r} f d\mu_n^{\otimes r} - \int_{\mathcal{X}^n} \int_{X_n^r} f d\mu_n^{\otimes r} d\mu^{\otimes n}(X_n) \right|,$$

$$M_2 = \left| \int_{\mathcal{X}^n} \int_{X_n^r} f d\mu_n^{\otimes r} d\mu^{\otimes n}(X_n) - \int_{\mathcal{X}^r} f d\mu^{\otimes r} \right|$$

右端的两个被加项。

我们再次考虑一个有限指标集 $I_r = \{i_1, \ldots, i_r\}$ 以及赋值 $\mathtt{v} : I_r \to \mathbf{R}$。然后我们引入如下分拆:

$$\begin{aligned} \int_{\mathcal{X}^n} \int_{X_n^r} f d\mu_n^{\otimes r} d\mu^{\otimes n}(X_n) &= \int_{\mathcal{X}^n} \sum_{\mathtt{v}: I_r \to [n]} \frac{1}{n^r} f(x_{\mathtt{v}(i_1)}, \ldots, x_{\mathtt{v}(i_r)}) d\mu^{\otimes n}(X_n) \\ &= \int_{\mathcal{X}^n} \sum_{\substack{\mathtt{v}: I_r \to [n] \\ \mathtt{v} \text{ injective}}} \frac{1}{n^r} f(x_{\mathtt{v}(i_1)}, \ldots, x_{\mathtt{v}(i_r)}) d\mu^{\otimes n}(X_n) \\ &\quad + \int_{\mathcal{X}^n} \sum_{\substack{\mathtt{v}: I_r \to [n] \\ \mathtt{v} \text{ non injective}}} \frac{1}{n^r} f(x_{\mathtt{v}(i_1)}, \ldots, x_{\mathtt{v}(i_r)}) d\mu^{\otimes n}(X_n). \end{aligned}$$

我们从 $M_2$开始，并用 $T_{inj}$ （相应地， $T_{\neg inj}$）表示对单射（相应地，非单射）赋值之和的积分，用 $c_{min}, c_{max} \in \mathbf{R}$ 表示 $f$的下界和上界。于是，我们有

$$\sum_{\substack{\mathtt{v}: I_r \to [n] \\ \mathtt{v} \text{ non injective}}} \frac{1}{n^r} c_{min} \leq T_{\neg inj} \leq \sum_{\substack{\mathtt{v}: I_r \to [n] \\ \mathtt{v} \text{ non injective}}} \frac{1}{n^r} c_{max}.$$

对非单射值函数的数量进行计数，我们得到

$$\#\{\mathtt{v} : I_r \to [n] \text{ non injective}\} = \sum_{l=1}^{r-1} \binom{r-1}{l} n^l \leq 2^{r-1} n^{r-1},$$

因此 $|T_{\neg inj}| \le \frac{2^{r-1}n^{r-1}}{n^r}\max(-c_{min}, c_{max}) = \frac{2^{r-1}n^{r-1}}{n^r}\beta$。论证的关键在于计算 $T_{inj}$:

$$\begin{aligned} T_{inj} &= \sum_{\substack{\mathtt{v}:I_r\to[n]\\ \mathtt{v}\text{ injective}}} \frac{1}{n^r}\int_{\mathcal{X}^n} f(x_{\mathtt{v}(i_1)},...,x_{\mathtt{v}(i_r)})d\mu^{\otimes n}(X_n) \\ &= \sum_{\substack{\mathtt{v}:I_r\to[n]\\ \mathtt{v}\text{ injective}}} \frac{1}{n^r}\int_{\mathcal{X}^r} f d\mu^{\otimes r} \\ &= \frac{n^r - \sum_{l=1}^{r-1}\binom{r-1}{l}n^l}{n^r}\int_{\mathcal{X}^r} f d\mu^{\otimes r}, \end{aligned}$$

其中，第二个等号之所以成立，关键在于积分是对 $r$ 不同变量进行的（$\mathtt{v}$ 为单射）。

将两个估计代入 $M_2$，最终得到

$$\begin{aligned} M_2 &\le \frac{2^{r-1}n^{r-1}}{n^r}\left(\left|\int_{\mathcal{X}^r} f d\mu^{\otimes r}\right| + \beta\right) \\ &\le \frac{M}{n}. \end{aligned}$$

现在我们转向 $M_1$，我们将使用麦克迪尔米德不等式 [18]对其进行界定。为简化接下来的方程，记

$$g(x_1,...,x_n) = \sum_{\mathtt{v}:I_r\to[n]} \frac{1}{n^r} f(x_{\mathtt{v}(i_1)},...,x_{\mathtt{v}(i_r)})$$

对于由元素 $X_n = (x_1,...,x_j,...,x_n)$ 构成的有序集合 $\in \mathcal{X}^n$ 以及某个 $x' \in \mathcal{X}$，我们定义 $Y_n = (y_1,...,y_n) := (x_1,...,x',...,x_n)$。于是，我们有

$$|g(X_n) - g(Y_n)| = \left|\sum_{\substack{\mathtt{v}:I_r\to[n]\\ j\in\mathtt{v}(I_r)}} \frac{1}{n^r} f(x_{\mathtt{v}(i_1)},...,x_{\mathtt{v}(i_r)}) - f(y_{\mathtt{v}(i_1)},...,y_{\mathtt{v}(i_r)})\right|$$

并且，由于 $\#\{\mathtt{v}: I_r \to [n];\ j \in \mathtt{v}(I_r)\} = \sum_{l=1}^{r}\binom{r}{l}n^{l-1}$，

$$|g(X_n) - g(Y_n)| \le \max(c_{min}, c_{max})\frac{\sum_{l=1}^{r}\binom{r}{l}n^{l-1}}{n^r}.$$

结论可通过应用 McDiarmid 不等式得出。

**B.4. 推论 4.13 的证明。** 我们将证明，对于 $n \ge \max\left(\frac{2^{r+5}\beta}{\varepsilon^2}\log\left(\frac{2}{\delta}\right), \frac{C_r}{\varepsilon}\right)$ 且概率至少为 $1 - \frac{\delta}{2}$，

$$E_n^{\mu} \le \frac{\varepsilon}{2}.$$

回顾一下，使用命题 4.12 $E_n^{\mu} \le \widetilde{\varepsilon} + \frac{M}{n}$ 中的记号。因此，只需证明右侧的每一项都小于 $\frac{\varepsilon}{4}$。

对于确定性项，$\frac{M}{n} \le \frac{\varepsilon}{4}$ 只要 $n \ge \frac{4M}{\varepsilon}$。使 $M$ 显式化后，只要

$$n \ge 4 \times 2^{r-1}\left(\mu^{\otimes r}(C)\beta + \beta^2\right),$$

其中 $\beta$ 定义于方程（4.12）。

因此，$E_n^\mu \le \frac{\varepsilon}{2}$ 只要 $\widetilde{\varepsilon} \le \frac{\varepsilon}{4}$ 根据命题4.12 以至少 $1-\frac{\delta}{2}$的概率成立 $\frac{\delta}{2} \ge 2\exp\left(-\left(\frac{\frac{\varepsilon}{4}^2 n}{2b_n}\right)\right)$，其中 $b_n$定义为

$$b_n = n^2\left(\beta\frac{\sum_{l=1}^{r}\binom{r}{l}n^{l-1}}{n^r}\right)^2. \tag{B.2}$$

等价地，这意味着

$$n \ge \frac{32 b_n \log\left(\frac{4}{\delta}\right)}{\varepsilon^2}$$

并且，注意到 $b_n \le 2^r\beta^2$，只需满足

$$n \ge \frac{2^{r+5}\beta\log\left(\frac{4}{\delta}\right)}{\varepsilon^2}.$$

## 参考文献

[1] Anari, N., Liu, Y. P. 和 Vuong, T.-D.（2022）。通过平均情形熵独立性，实现生成树和行列式点过程的最优次线性抽样。载于*2022 IEEE* 第*63*届计算机科学基础年度研讨会（*FOCS*），123–134。IEEE。[2] Avena, L., Castell, F., Gaudillière, A. 和 Mélot, C.（2018）。随机森林与网络分析。《统计物理学杂志》**173**，985–1027。[3] Avena, L. 和 Gaudillière, A.（2018）。随机生成森林的两种应用。《理论概率杂志》**31**，1975–2004。[4] Baccelli, F., Błaszczyszyn, B. 和 Karray, M.（2024）。随机测度、点过程与随机几何。[5] Bachem, O., Lucic, M. 和 Krause, A. 面向机器学习的实用核心集构造。arXiv 2017。*arXiv* 预印本 *arXiv:1703.06476*。[6] Bachmann, S.（2016）。泊松泛函的集中性：随机几何图中的连通分量计数。《随机过程及其应用》**126**，1306–1330。[7] Bardenet, R., Feller, A., Bouttier, J., Degiovanni, P., Hardy, A., Rançon, A., Rous-sel, B., Schehr, G. 和 Westbrook, C.（2023）。从点过程到量子光学，再回到点过程。工作论文或预印本。[8] Bardenet, R., Ghosh, S. 和 Lin, M.（2021）。基于正交多项式的行列式点过程：用于随机梯度下降小批量抽样。《神经信息处理系统进展》**34**，16226–16237。[9] Bardenet, R., Ghosh, S., Simon-Onfroy, H. 和 Tran, H.-S.（2024）。通过负依赖构造小型核心集：行列式点过程、线性统计量与集中性。[10] Bardenet, R. 和 Hardy, A.（2020）。使用行列式点过程进行蒙特卡罗。《应用概率年刊》。[11] Barthelmé, S., Tremblay, N. 和 Amblard, P.-O.（2023）。一种更快的离散行列式点过程抽样器。载于国际人工智能与统计会议，5582–5592。PMLR。[12] Barthelmé, S.（2024）。行列式点过程及相关主题导论。草稿。[13] Belkin, M. 和 Niyogi, P.（2006）。拉普拉斯特征映射的收敛性。《神经信息处理系统进展》**19**。[14] Bi, Z. 和 de Micheaux, P. L.（2025）。流形维数估计：一项实证研究。*arXiv* 预印本 *arXiv:2509.15517*。[15] Borda, B., Grabner, P. 和 Matzke, R. W.（2024）。球面和平坦环面上行列式点过程的黎曼能量、L 2 L^2 差异度与最优传输。《*Mathematika*》**70**，e12245。[16] Borgs, C., Chayes, J. T., Cohn, H. 和 Lovász, L. M.（2020）。graphex 的可识别性与弱核度量。载于《搭建桥梁 *II*：*László Lovász* 数学》，29–157。施普林格。

[17] Borgs, C., Chayes, J. T., Cohn, H. 和 Veitch, V.（2019）。稀疏可交换图的抽样视角。《概率年刊》47，2754–2800。https://doi.org/10.1214/18-AOP1320[18] Boucheron, S., Lugosi, G. 和 Massart, P.（2013）。《集中不等式：独立性的非渐近理论》。牛津大学出版社。https://doi.org/10.1093/acprof: oso/ 9780199535255 001 0001 。[19] Brubeck, P. D., Nakatsukasa, Y. 和 Trefethen, L. N.（2021）。使用 Arnoldi 方法的范德蒙德矩阵。《SIAM Review》63，405–415。[20] Chatterjee, S.（2015）。通过通用奇异值阈值化进行矩阵估计。[21] Coifman, R. R. 和 Lafon, S.（2006）。扩散映射。《Applied and Computational Harmonic Analysis》21，5–30。[22] Crane, H.（2018）。《统计网络分析的概率基础》。Chapman and Hall/CRC。[23] Daley, D. J. 和 Vere-Jones, D.（2003）。《点过程理论导论：第一卷：基础理论与方法》。施普林格。[24] Daley, D. J. 和 Vere-Jones, D.（2008）。《点过程理论导论：第二卷：一般理论与结构》。施普林格。[25] Derezinski, M. 和 Mahoney, M. W.（2021）。随机化数值线性代数中的行列式点过程。《Notices of the American Mathematical Society》68，34–45。[26] Derezinski, M. 和 Warmuth, M. K.（2018）。用于线性回归的反向迭代体积采样。《J. Mach. Learn. Res.》19，23:1–23:39。[27] Dieudonné, J.（1970）。《分析论著》第2卷。Academic press。[28] Dunson, D. B., Wu, H.-T. 和 Wu, N.（2021）。图拉普拉斯算子与热核重建在 L∞ 中的谱收敛，基于随机样本。《Applied and Computational Harmonic Analysis》55，282–336。[29] Ehler, M., Gräf, M. 和 Oates, C. J.（2019）。闭流形上的最优蒙特卡罗积分。《Statistics and Computing》29，1203–1214。[30] Fanuel, M. 和 Bardenet, R.（2021）。使用核方法估计连续行列式点过程。《神经信息处理系统进展》34，24124–24136。[31] Gallot, S., Hulin, D. 和 Lafontaine, J.（2004）。《黎曼几何》，第3版。Universitext。施普林格。[32] Gouriéroux, C. 和 Lu, Y.（2025）。行列式点过程相关核矩阵的一个简单估计量。arXiv 预印本 arXiv:2505.14529。[33] Grigorescu, E., Juba, B., Wimmer, K. 和 Xie, N.（2022）。行列式点过程最大似然学习的困难性。载于《学习理论会议》，3800–3819。PMLR。[34] Hough, J. B., Krishnapur, M., Peres, Y. 和 Virág, B.（2006）。《行列式过程与独立性》。《Probability Surveys》3，206–229。https://doi.org/10.1214/154957806000000078[35] Huhtanen, M. 和 Larsen, R. M.（2002）。离散正交二元多项式的生成。《BIT Numerical Mathematics》42，393–407。[36] Kato, T.（2013）。《线性算子的扰动理论》132。Springer Science & Business Media。[37] Katori, M.（2020）。《行列式点过程、随机对数气体及其他》。arXiv 预印本 arXiv:2002.07760。[38] Katori, M. 和 Shirai, T.（2022）。球面上行列式点过程的标度极限。arXiv 预印本 arXiv:2203.07590。[39] Keriven, N.（2023）。随机图上的熵最优传输。《SIAM J. Math. Data Sci.》。 **5** 1028 1050 10 1137 22 1518281 –。[40] 2016 https://doi.org/。/ M Kim, J., Rinaldo, A. 和 Wasserman, L.（）。流形维数估计的极小极大速率。arXiv 预印本 arXiv:1605.01011。[41] Kittaneh, F.（1985）。正规算子的 Lipschitz 函数。《Proceedings of the American Mathematical Society》94，416–418。[42] Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A. 和 Boguná, M.（2010）。复杂网络的双曲几何。《Physical Review E—Statistical, Nonlinear, and Soft Matter Physics》82，036106。[43] Kulesza, A. 和 Taskar, B.（2012）。《面向机器学习的行列式点过程》。《Found. Trends Mach. Learn.》。 **5** 123 286 10 1561 2200000044 –。[44] 2010 $\varepsilon$ https://doi.org/。/ Langberg, M. 和 Schulman, L. J.（）。积分的通用 -近似器。载于《第二十一届 ACM-SIAM 离散算法年度研讨会论文集》，598–607。SIAM。

[45] Lasserre, J. B.、Pauwels, E.与Putinar, M.（2022）。数据分析的Christoffel–Darboux核 38。剑桥大学出版社。[46] Lavancier, F.、Møller, J.与Rubak, E.（2015）。行列式点过程模型与统计推断。英国皇家统计学会杂志B辑：统计方法学 77 853–877。[47] Lei, J.与Rinaldo, A.（2015）。随机块模型中谱聚类的一致性。《统计年鉴》

**43** 215 237 10 1214 14 1274 – 。

[48] 2024 https://doi.org/ . / -AOS Lemoine, T.与Bardenet, R.（）。紧复流形上使用Bergman核的蒙特卡罗方法。arXiv预印本 arXiv:2405.09203。[49] Lenard, A.（1973）。经典统计力学中的相关函数与状态的唯一性。数学物理通讯 30 35–44。[50] Lenard, A.（1975）。无穷多个粒子的经典统计力学系统的状态。I。《理性力学与分析档案》59 219–239。[51] Lenard, A.（1975）。无穷多个粒子的经典统计力学系统的状态。II。相关测度的刻画。《理性力学与分析档案》59 241–256。[52] Levi, M.、Marzo, J.与Ortega-Cerdà, J.（2024）。行列式点过程的线性统计量与范数表示。国际数学研究通报 2024 12869 12903– 。[53] Luo, Y.与Gao, C.（2024）。通过低次多项式进行图子估计的计算下界。《统计年鉴》52 2318–2348。[54] Macchi, O.（1975）。随机点过程的重合方法。应用概率进展 7 83–122。[55] Munteanu, A.与Schwiegelshohn, C.（2018）。核心集方法与历史：面向近似与流式算法的一种理论家设计模式。KI—人工智能 32 37 53– 。[56] Owen, A. B.（2013）。蒙特卡罗理论、方法与实例。https://artowen.su.domains/ mc/。[57] Pelletier, B.（2005）。黎曼流形上的核密度估计。《统计与概率快报》73 297–304。[58] Pilavcı, Y. Y.、Amblard, P.、Barthelmé, S.与Tremblay, N.（2021）。经由随机生成森林的图Tikhonov正则化与插值。IEEE Trans. Signal Inf. Process. over Networks

**7** 359 374 10 1109 2021 3084879 – 。

[59] 2024 https://doi.org/ . /TSIPN. 。Poinas, A.（）。关于具有非对称核的行列式点过程。arXiv预印本 arXiv:2406.03360。[60] Poinas, A.与Lavancier, F.（2023）。平稳行列式点过程似然的渐近近似。《斯堪的纳维亚统计学杂志》50 842–874。[61] Rosasco, L.、Belkin, M.与De Vito, E.（2010）。论基于积分算子的学习。机器学习研究杂志 11。[62] Shirai, T.与Takahashi, Y.（2003）。与某些Fredholm行列式相关的随机点场 I：费米子、泊松与玻色子点过程。《泛函分析杂志》205 414–463。[63] Soshnikov, A.（2000 55 923）。行列式随机点场。《俄罗斯数学调查》。[64] Soshnikov, A.（2002）。行列式随机点场的高斯极限。《概率年刊》171–187。[65] Tremblay, N.、Amblard, P.-O.与Barthelmé, S.（2017）。基于行列式过程的图采样。见2017年第25届欧洲信号处理会议（EUSIPCO）1674–1678。IEEE。[66] Tremblay, N.、Barthelmé, S.与Amblard, P.-O.（2019）。用于核心集的行列式点过程。机器学习研究杂志 20 1–70。[67] Tremblay, N.、Barthelmé, S.、Usevich, K.与Amblard, P.-O.（2023）。扩展L-系综：行列式点过程的一种新表示。应用概率年刊 33 613–640。[68] Urschel, J.、Brunel, V.-E.、Moitra, A.与Rigollet, P.（2017）。利用矩与环进行行列式点过程学习。见国际机器学习会议 3511–3520。PMLR。[69] van der Hoorn, P.、Stepanyants, H.与Krioukov, D.（2025）。概率对称性的射影极限及其在随机图极限中的应用。arXiv预印本 arXiv:2512.02208。

[70] Van Der Vaart, A. W. 与 Wellner, J. A. （1996）。弱收敛。见 弱收敛与经验过程：及其在统计学中的应用 16–28。施普林格。[71] Veitch, V. 与 Roy, D. M. （2019）。（稀疏）可交换图的抽样与估计。《统计年鉴》 **47** 3274 3299 10 1214 18 1778 – 。

[72] 2022 https://doi.org/./-AOSWu, H.-T.与Wu, N.（）。流形上一般核的核密度估计量的带速率强一致一致性。《信息与推断：*IMA*期刊》 **11** 781–799。

# 对“行列式点过程的离散到连续极限的统计一致性”的补充

C. 定理 3.3 的证明。我们将证明定理 3.3 的一个更为显式的版本。为便于阅读，我们引入一些记号。对于固定的紧支撑、有界且可测的 $\varphi_r : \mathrm{X} \to \mathrm{R}$，我们将与点过程 $\mathrm{P}_n$（相应地，$\mathrm{Q}_n$）相关联的线性统计量记为 $\Lambda_{\mathcal{P}} := \Lambda^{(\varphi_r)}(\mathrm{S}^{\mathcal{P}_n})$（相应地，$\Lambda_{\mathcal{Q}} := \Lambda^{(\varphi_r)}(\mathrm{S}^{\mathcal{Q}_n})$），并在上下文清楚时进一步省略期望的下标，例如 $\mathrm{E}[\Lambda_{\mathcal{P}}] := \mathrm{E}_{\mathcal{S}_{\mathcal{P}} \sim \mathrm{P}_n}[\Lambda^{(\varphi_r)}(\mathrm{S}_{\mathcal{P}})]$（相应地，$\mathrm{E}[\Lambda_{\mathcal{Q}}] := \mathrm{E}_{\mathcal{S}_{\mathcal{Q}} \sim \mathrm{Q}_n}\left[\Lambda^{(\varphi_r)}(\mathrm{S}^{\mathcal{Q}})\right]$）。我们强调 $\Lambda_{\mathcal{P}}$ 依赖于 $\mathrm{S}^{\mathcal{P}_n}$ 而非 P，但该项仅在下文取形如 $\mathrm{E}[\Lambda_{\mathcal{P}}]$ 或 $\mathrm{E}[\Lambda^j \mathrm{P}]$（对某个指数 $j$）的期望时出现，且从不被含糊地使用。利用这些记号，原始矩与中心矩的表达式为：

$$m_k^{(\varphi_r)}(\mathcal{P}_n) = \mathbf{E}\left[\Lambda_{\mathcal{P}}^k\right], \quad \overline{m}_k^{(\varphi_r)}(\mathcal{P}_n) = \mathbf{E}\left[\left(\Lambda_{\mathcal{P}} - \mathbf{E}\left[\Lambda_P\right]\right)^k\right],$$

$$m_k^{(\varphi_r)}(\mathcal{Q}_n) = \mathbf{E}\left[\Lambda_{\mathcal{Q}}^k\right], \quad \overline{m}_k^{(\varphi_r)}(\mathcal{Q}_n) = \mathbf{E}\left[\left(\Lambda_{\mathcal{Q}} - \mathbf{E}\left[\Lambda_Q\right]\right)^k\right].$$

回顾一下，我们假设矩满足一致有界性假设：以概率 1，存在一些常数 $c_j \in \mathbf{R}_{>0}$ 使得

$$\left|\mathbf{E}\left[\Lambda_{\mathcal{Q}}^j\right]\right| \leq c_j$$

对所有可容许的 $\varphi_r$。例如，在定理 4.1的情形下，这一条件总是满足的。我们将证明如下结论。

定理 C.1。 假设对于所有 $r$，紧支撑、有界且可测的函数 $\widetilde{\varphi_r} : \boldsymbol{X}^r \to \mathbf{R}$，存在某个 $\varepsilon^{(\widetilde{\varphi_r})} > 0$ 和 $\widetilde{\delta} \in (0,1)$，使得以~至少 $1-\delta$ 的概率，

$$\left|\Phi^{(\widetilde{\varphi_r})}(\mathcal{P}_n) - \Phi^{(\widetilde{\varphi_r})}(\mathcal{Q}_n)\right| \leq \varepsilon^{(\widetilde{\varphi_r})}. \tag{C.1}$$

~那么，对于给定的 $\varphi_r$，以至少 $1-\delta$ 的概率，

$$\left|m_k^{(\varphi_r)}(\mathcal{P}_n) - m_k^{(\varphi_r)}(\mathcal{Q}_n)\right| \leq \sum_{l=r}^{kr} \sum_{\substack{m_1,\ldots,m_{\lceil \frac{l}{r} \rceil} \geq 1 \\ m_1+\ldots+m_{\lceil \frac{l}{r} \rceil} = k}} \varepsilon_n^{\left(\varphi_r^{\left(m_1,\ldots,m_{\lceil \frac{l}{r} \rceil}\right)}\right)} \tag{C.2}$$

~并且，再次以至少 $1-\delta$ 的概率，

$$\left|\overline{m}_k^{(\varphi_r)}(\mathcal{P}_n) - \overline{m}_k^{(\varphi_r)}(\mathcal{Q}_n)\right| \leq \sum_{q=0}^{k} \binom{k}{q} \sum_{l=r}^{kr} \sum_{\substack{m_1,\ldots,m_{\lceil \frac{l}{r} \rceil} \geq 1 \\ m_1+\ldots+m_{\lceil \frac{l}{r} \rceil} = k}} c_{k-q}\left(1 + c_1(k-q)\left(c_1 + \varepsilon_n^{(\varphi_r)}\right)\varepsilon_n^{(\varphi_r)},\right)\varepsilon_n^{\left(\varphi_r^{\left(m_1,\ldots,m_{\lceil \frac{l}{r} \rceil}\right)}\right)}$$

(C.3)

$$+\sum_{q=0}^{k}\binom{k}{q}(k-q)c_1c_{k-q}\left(c_1+\varepsilon_n^{(\varphi_r)}\right)\varepsilon_n^{(\varphi_r)},$$

其中

$$\varphi_r^{\left(m_1,\ldots,m_{\lceil\frac{l}{r}\rceil}\right)}(x_1,\ldots,x_l)=\left(\prod_{i=1}^{\lfloor\frac{l}{r}\rfloor}\varphi_r(x_{(i-1)r+1},\ldots,x_{(i-1)r+r})^{m_i}\right)\varphi_r(x_{\lfloor\frac{r}{l}\rfloor+1,\ldots,x_l,x_1,\ldots,x_{r-(l \mod r)}})^{m_{\lceil\frac{l}{r}\rceil}}.$$

函数 $\varphi_r^{\left(m_1,\ldots,m_{\lceil\frac{l}{r}\rceil}\right)}$，连同 $\varphi_r$，构成族 $(\varphi^i)$，见定理 3.3。因此，只需首先注意到以下几点，即可完成定理3.3的证明：

1. 对于弱相干过程，方程（C.1）中的条件在 $\varepsilon^{(\widetilde{\varphi_r})}=\widetilde{\varepsilon}$ 当 $n>n(\delta,\widetilde{\varepsilon},\widetilde{\varphi_r})$ 时成立；

2. 方程（C.2）并且取为（C.3）的右端系数之和，取多项式（分别为 $\widetilde{\varepsilon}=\frac{\varepsilon}{s}$），则当 $\varepsilon$（分别为 $s$）足够小，使得 $\frac{\varepsilon}{s}<1$（分别为 $\frac{\varepsilon}{s}<1$）时，它有界于 $\varepsilon$。

取 $\widetilde{\delta}=\frac{\delta}{L}$ 并应用并集界，即可在 $\varepsilon/s<1$（分别为 $\varepsilon/s<1$）时，得到所需的条件（式 3.3）。根据下文的注释，最终可以去掉这一限制。

注 C.2。对于任意正实值变量 $X$ 和 $\varepsilon<\eta$，

$$\mathbb{P}(X>\eta)<\mathbb{P}(X>\varepsilon).$$

现在我们继续证明定理 C.1，其证明依赖于以下组合展开。对于 $\mathcal{X}$ 的任意局部有限子集 $\mathcal{S}$，

$$\left(\sum_{\substack{x_1,\ldots,x_r\in\mathcal{S}\\ x_i\neq x_j}}\varphi_r(x_1,\ldots,x_r)\right)^k=\sum_{l=r}^{kr}\sum_{\substack{m_1,\ldots,m_{\lceil\frac{l}{r}\rceil}\geq 1\\ m_1+\ldots+m_{\lceil\frac{l}{r}\rceil}=k}}\sum_{\substack{x_1,\ldots,x_l\in\mathcal{S}\\ x_i\neq x_j}}\varphi_r^{\left(m_1,\ldots,m_{\lceil\frac{l}{r}\rceil}\right)}(x_1,\ldots,x_l),$$

这是通过考察 $\left(\sum_{x_i,\ \ldots,\ x_i\neq x_j}\varphi_r(x_1,\ \ldots,\ x_r)\right)^k$ 的展开得到的。两个最外层求和以及函数 $\varphi_r^{\left(m_1,\ \ldots\ m_{\lceil\frac{l}{r}\rceil}\right)}$，的精确形式在此并不重要，但至关重要的是，最内层求和遍历的 $x_1,\ \ldots,\ x_l$ 两两不同。

这样，分别取期望可得

$$\begin{cases}\mathbf{E}\left[\Lambda_{\mathcal{P}}^k\right]=\sum_{l=r}^{kr}\sum_{\substack{m_1,\ldots,m_{\lceil\frac{l}{r}\rceil}\geq 1\\ m_1+\ldots+m_{\lceil\frac{l}{r}\rceil}=k}}\mathbf{E}_{\mathcal{S}_{\mathcal{P}}\sim\mathcal{P}}\left[\sum_{\substack{x_1,\ldots,x_l\in\mathcal{S}_{\mathcal{P}}\\ x_i\neq x_j}}\varphi_r^{\left(m_1,\ldots,m_{\lceil\frac{l}{r}\rceil}\right)}(x_1,\ldots,x_l)\right],\\ \mathbf{E}\left[\Lambda_{\mathcal{Q}}^k\right]=\sum_{l=r}^{kr}\sum_{\substack{m_1,\ldots,m_{\lceil\frac{l}{r}\rceil}\geq 1\\ m_1+\ldots+m_{\lceil\frac{l}{r}\rceil}=k}}\mathbf{E}_{\mathcal{S}_{\mathcal{Q}_n}\sim\mathcal{Q}_n}\left[\sum_{\substack{x_1,\ldots,x_l\in\mathcal{S}_{\mathcal{Q}_n}\\ x_i\neq x_j}}\varphi_r^{\left(m_1,\ldots,m_{\lceil\frac{l}{r}\rceil}\right)}(x_1,\ldots,x_l)\right],\end{cases}$$

因此，根据三角不等式，

$$\left|m_k^{(\varphi_r)}(\mathcal{P}_n)-m_k^{(\varphi_r)}(\mathcal{Q}_n)\right|\leq\sum_{l=r}^{kr}\sum_{\substack{m_1,\ldots,m_{\lceil\frac{l}{r}\rceil}\geq 1\\ m_1+\ldots+m_{\lceil\frac{l}{r}\rceil}=k}}\left|\mathbf{E}_{\mathcal{S}_\mathcal{P}\sim\mathcal{P}_n}\left[\Lambda^{\varphi_r\binom{m_1,\ldots,m_{\lceil\frac{l}{r}\rceil}}{}}(\mathcal{S}_\mathcal{P})\right]-\mathbf{E}_{\mathcal{S}_{\mathcal{Q}_n}\sim\mathcal{Q}_n}\left[\Lambda^{\varphi_r\binom{m_1,\ldots,m_{\lceil\frac{l}{r}\rceil}}{}}(\mathcal{S}_{\mathcal{Q}_n})\right]\right|.$$

将方程（C.1）中的集中速率明确写出后，立即可得到不等式（C.2）。

为了获得关于中心矩的界，我们利用方程（B.1）中的恒等式。首先应用二项式定理，得到:

$$(\Lambda-\mathbf{E}[\Lambda])^k=\sum_{q=0}^{k}\binom{k}{q}\Lambda^q(-\mathbf{E}[\Lambda_\mathcal{Q}])^{k-q},$$

$$(\Lambda_n-\mathbf{E}[\Lambda_n])^k=\sum_{q=0}^{k}\binom{k}{q}\Lambda_n^q(-\mathbf{E}[\Lambda_\mathcal{P}])^{k-q}.$$

对这两个表达式的期望取差，然后应用三角不等式，可得

$$\begin{aligned}\left|\overline{m}_k^{(\varphi_r)}(\mathcal{P}_n)-\overline{m}_k^{(\varphi_r)}(\mathcal{Q}_n)\right|&=\left|\sum_{q=0}^{k}\binom{k}{q}\left(\mathbf{E}[\Lambda_\mathcal{Q}^q](-\mathbf{E}[\Lambda_\mathcal{Q}])^{k-q}-\mathbf{E}[\Lambda_\mathcal{P}^q](-\mathbf{E}[\Lambda_\mathcal{P}])^{k-q}\right)\right|\\&\leq\sum_{q=0}^{k}\binom{k}{q}\left|\mathbf{E}[\Lambda_\mathcal{Q}^q](-\mathbf{E}[\Lambda_\mathcal{Q}])^{k-q}-\mathbf{E}[\Lambda_\mathcal{P}^q](-\mathbf{E}[\Lambda_\mathcal{P}])^{k-q}\right|,\end{aligned}$$

接下来对其进行界定。由方程（B.1）可知

$$\begin{aligned}\left|\mathbf{E}[\Lambda_\mathcal{Q}^q](-\mathbf{E}[\Lambda_\mathcal{Q}])^{k-q}-\mathbf{E}[\Lambda_\mathcal{P}^q](-\mathbf{E}[\Lambda_\mathcal{P}])^{k-q}\right|&\leq\left|\mathbf{E}[\Lambda_\mathcal{Q}^q]-\mathbf{E}[\Lambda_\mathcal{P}^q]\right|\left|\mathbf{E}[\Lambda_\mathcal{P}]^{k-q}\right|\\&\quad+\left|\mathbf{E}[\Lambda_\mathcal{Q}^q]\right|\left|\mathbf{E}[\Lambda_\mathcal{Q}]^{k-q}-\mathbf{E}[\Lambda_\mathcal{P}]^{k-q}\right|\\&\leq\left|\mathbf{E}[\Lambda_\mathcal{Q}^q]-\mathbf{E}[\Lambda_\mathcal{Q}^q]\right|\left|\mathbf{E}[\Lambda_\mathcal{P}]^{k-q}\right|\\&\quad+\left|\mathbf{E}[\Lambda_\mathcal{Q}^q]\right|(k-q)\left|\mathbf{E}[\Lambda_\mathcal{Q}]-\mathbf{E}[\Lambda_\mathcal{P}]\right|\left|\mathbf{E}[\Lambda_\mathcal{Q}]\right|\left|\mathbf{E}[\Lambda_\mathcal{P}]\right|,\end{aligned}$$

其中，第二个不等式由方程（B.1）得到，因为

$$\begin{aligned}\left|\mathbf{E}[\Lambda_\mathcal{Q}]^{k-q}-\mathbf{E}[\Lambda_\mathcal{P}]^{k-q}\right|&\leq\sum_{j=1}^{k-q}\left|\mathbf{E}[\Lambda_\mathcal{Q}]-\mathbf{E}[\Lambda_\mathcal{P}]\right|\left|\mathbf{E}[\Lambda_\mathcal{Q}]\right|\left|\mathbf{E}[\Lambda_\mathcal{P}]\right|\\&=(k-q)\left|\mathbf{E}[\Lambda_\mathcal{Q}]-\mathbf{E}[\Lambda_\mathcal{P}]\right|\left|\mathbf{E}[\Lambda_\mathcal{Q}]\right|\left|\mathbf{E}[\Lambda_\mathcal{P}]\right|.\end{aligned}$$

同理，我们得到

$$\left|\mathbf{E}[\Lambda_\mathcal{P}]^{k-q}\right|\leq\left|\mathbf{E}[\Lambda_\mathcal{Q}]^{k-q}\right|+(k-q)\left|\mathbf{E}[\Lambda_\mathcal{Q}]-\mathbf{E}[\Lambda_\mathcal{P}]\right|\left|\mathbf{E}[\Lambda_\mathcal{Q}]\right|\left|\mathbf{E}[\Lambda_\mathcal{P}]\right|.$$

综合以上各项，我们得到

$$\begin{aligned}\left|\overline{m}_k^{(\varphi_r)}(\mathcal{P}_n)-\overline{m}_k^{(\varphi_r)}(\mathcal{Q}_n)\right|&\leq\sum_{q=0}^{k}\binom{k}{q}\left|\mathbf{E}[\Lambda_\mathcal{Q}^q]-\mathbf{E}[\Lambda_\mathcal{P}^q]\right|\left(\left|\mathbf{E}[\Lambda_\mathcal{Q}]^{k-q}\right|+(k-q)\left|\mathbf{E}[\Lambda_\mathcal{Q}]-\mathbf{E}[\Lambda_\mathcal{P}]\right|\left|\mathbf{E}[\Lambda_\mathcal{Q}]\right|\left|\mathbf{E}[\Lambda_\mathcal{P}]\right|\right)\\&\quad+\sum_{q=0}^{k}\binom{k}{q}\left|\mathbf{E}[\Lambda_\mathcal{Q}^q]\right|(k-q)\left|\mathbf{E}[\Lambda_\mathcal{P}]-\mathbf{E}[\Lambda_\mathcal{Q}]\right|\left|\mathbf{E}[\Lambda_\mathcal{Q}]\right|\left|\mathbf{E}[\Lambda_\mathcal{P}]\right|,\end{aligned}$$

以及不等式（C.3）最终通过代入原始矩的集中性（方程（C.2））、应用一致界 $c_j$，并使集中速率 $\varepsilon^{(\widetilde{\varphi}_r)}$ 显式化而得到。

**D. 第5节中缺失的证明。**

D.1. 命题5.1的证明。我们将证明，在命题5.1的设定下，存在 $A$、$B$、$C>0$，使得对任意 $\widetilde{\varepsilon}>0$，只要其足够小，从而$\widetilde{\varepsilon}\le\frac{1}{2}\min\left(\|\mathcal{M}_1\|_\mu,\ \|\mathcal{P}_1\|_\mu,\ \ldots,\ \|\mathcal{P}_m\|_\mu\right)$，以至少 $1-C\exp\left(\frac{-2\widetilde{\varepsilon}^2 n}{B^2}\right)$的概率，

$$\max_{x\in X_n}|\mathcal{P}_k(x)-P_k(x)|\le A\widetilde{\varepsilon}\quad\forall k\in[m].$$

特别地，这对任意 $\widetilde{\delta}\ge C\exp\left(\frac{-2\widetilde{\varepsilon}^2 n}{B^2}\right)$ 都成立；其中，用 $\widetilde{\varepsilon}$替换 $A\widetilde{\varepsilon}$后，等价于方程（5.7）中的条件。最后指出，根据注记C.2，条件 $\widetilde{\varepsilon}\le\frac{1}{2}\min\left(\|\mathcal{M}_1\|_\mu,\ \|\mathcal{P}_1\|_\mu,\ \ldots,\ \|\mathcal{P}_m\|_\mu\right)$可以舍弃。

我们对 $k\in[m]$进行归纳推理，并将反复用到以下事实：若对于某个 $0<c\le\frac{|a|}{2}$有$|a_n-a|\le c$，则 $|\frac{1}{a_n}-\frac{1}{a}|\le\frac{2}{a^2}c$。我们还将使用以下引理。

引理 D.1. 设 $f,g:X\to\mathbf{R}$ 是定义在 $X$ 上的两个有界函数。那么，存在某个常数 $c=3\left(\max_{x\in\mathcal{X}}|f(x)|+\max_{x\in\mathcal{X}}|g(x)|\right)$，使得对所有 $e>0$，以至少 $1-2\exp\left(\frac{-2e^2 n}{c^2}\right)$ 的概率，

$$|\langle f,g\rangle_{\mu_n}-\langle f,g\rangle_\mu|\le e.$$

由此立即可得，当 $\|f\|_\mu>0$ 时，

$$|\|f\|_{\mu_n}-\|f\|_\mu|\le\frac{e}{\|f\|_{\mu_n}+\|f\|_\mu}\le\frac{e}{\|f\|_\mu}. \tag{D.1}$$

证明。这是对 McDiarmid 不等式的简单应用。

首先，我们证明 $\mathbf{E}\left[\langle f,g\rangle_{\mu_n}\right]=\langle f,g\rangle_\mu$，其中期望是对 $X_n=\{x_1,\ldots,x_n\}$ 取的：

$$\begin{aligned}\mathbf{E}\left[\langle f,g\rangle_{\mu_n}\right]&=\mathbf{E}\left[\sum_{i=1}^n\frac{1}{n}f(x_i)g(x_i)\right]\\&=\sum_{i=1}^n\frac{1}{n}\mathbf{E}[f(x_i)g(x_i)]\\&=\mathbf{E}_{x\sim\mu}[f(x)g(x)]\\&=\langle f,g\rangle_\mu.\end{aligned}$$

令 $h(x_1,\ldots,x_n)=\sum_{i=1}^n\frac{1}{n}f(x_i)g(x_i)$，则有

$$\begin{aligned}|g(x_1,\ldots,x_j,\ldots,x_n)-g(x_1,\ldots,x_j',\ldots,x_n)|&=\frac{1}{n}|f(x_j)g(x_j)-f(x_j')g(x_j')|\\&\le\frac{1}{n}c\end{aligned}$$

从而我们得到有界差分性质，其中 $c$ 的值由方程（B.1）和三角不等式得到。当 $\|f\|_\mu>0$时，通过取 $g=f$ 并使用恒等式 $(a^2-b^2)=(a+b)(a-b)$ 可得方程（D.1）。

□

让我们开始归纳法。当 $k=1$ 时，我们有

$$\left\| \frac{M_1}{\|M_1\|_{\mu_n}} - \frac{\mathcal{M}_1}{\|\mathcal{M}_1\|_{\mu}} \right\|_{\infty} = \|M_1\|_{\infty} \left| \frac{1}{\|M_1\|_{\mu_n}} - \frac{1}{\|\mathcal{M}_1\|_{\mu}} \right|,$$

其中我们采用记号 $\|\cdot\|_{\infty} = \|\cdot\|_{L^{\infty}(X_n;\mu_n)}$ 以求简洁。将引理 D.1 应用于 $f = g = \mathcal{M}_1$，可得，对于足够小的 $\widetilde{\varepsilon}$，使得 $\widetilde{\varepsilon} \leq \frac{\|\mathcal{M}_1\|_{\mu}}{2}$，

$$\begin{aligned}
\left| \frac{1}{\|M_1\|_{\mu_n}} - \frac{1}{\|\mathcal{M}_1\|_{\mu}} \right| &= \frac{1}{\|M_1\|_{\mu_n}\|\mathcal{M}_1\|_{\mu}} \left| \|M_1\|_{\mu_n} - \|\mathcal{M}_1\|_{\mu} \right| \\
&\leq \frac{1}{\|M_1\|_{\mu_n}\|\mathcal{M}_1\|_{\mu}} \left( \left| \|M_1\|_{\mu_n} - \|\mathcal{M}_1\|_{\mu_n} \right| + \left| \|\mathcal{M}_1\|_{\mu_n} - \|\mathcal{M}_1\|_{\mu} \right| \right) \\
&= \frac{1}{\|M_1\|_{\mu_n}\|\mathcal{M}_1\|_{\mu}} \left| \|\mathcal{M}_1\|_{\mu_n} - \|\mathcal{M}_1\|_{\mu} \right| \\
&\leq \frac{2}{\|\mathcal{M}_1\|_{\mu}^2} \left| \|\mathcal{M}_1\|_{\mu_n} - \|\mathcal{M}_1\|_{\mu} \right| \\
&\leq \frac{2}{\|\mathcal{M}_1\|_{\mu}^3} \widetilde{\varepsilon}
\end{aligned}$$

以至少 $1 - 2\exp\left(\frac{-2\widetilde{\varepsilon}^2 n}{c_1^2}\right)$ 的概率，其中 $c_1 = 6\max_{x\in\mathcal{X}} |\mathcal{M}_1(x)|$。因此，以至少 $1 - 2\exp\left(\frac{-2\widetilde{\varepsilon}^2 n}{c_1^2}\right)$ 的概率，

$$\max_{x\in X_n} |\mathcal{P}_1(x) - P_1(x)| \leq \frac{2\max_{x\in\mathcal{X}} |\mathcal{M}_1(x)|}{\|\mathcal{M}_1\|_{\mu}^3} \widetilde{\varepsilon}.$$

若 $k > 1$，我们在归纳假设下继续，并从如下界开始

$$\begin{aligned}
\|P_{k+1} - \mathcal{P}_{k+1}\|_{\infty} &= \left\| \frac{P'_{k+1}}{\|P'_{k+1}\|_{\mu_n}} - \frac{\mathcal{P}'_{k+1}}{\|\mathcal{P}'_{k+1}\|_{\mu}} \right\|_{\infty} \\
&= \left\| (P'_{k+1} - \mathcal{P}'_{k+1}) \frac{1}{\|\mathcal{P}'_{k+1}\|_{\mu}} + P'_{k+1} \left( \frac{1}{\|P'_{k+1}\|_{\mu_n}} - \frac{1}{\|\mathcal{P}'_{k+1}\|_{\mu}} \right) \right\|_{\infty} \\
&\leq \frac{1}{\|\mathcal{P}_{k+1}\|_{\mu}} \|P'_{k+1} - \mathcal{P}'_{k+1}\|_{\infty} + \|P'_{k+1}\|_{\infty} \left| \frac{1}{\|P'_{k+1}\|_{\mu_n}} - \frac{1}{\|\mathcal{P}'_{k+1}\|_{\mu}} \right|.
\end{aligned}$$

现在我们指出，

1. $\|\mathcal{P}'_{k+1}\|_{\mu}$ 是有限的，$\frac{1}{\|\mathcal{P}'_{k+1}\|_{\mu}}$ 也有限；2. $\|P'_{k+1}\|_{\infty} \leq \|\mathcal{P}'_{k+1}\|_{\infty} + \|P_{k+1} - \mathcal{P}'_{k+1}\|_{\infty}$，且 $\|\mathcal{P}'_{k+1}\|_{\infty}$ 有限；3. 根据引理D.1 最后，如果 $\widetilde{\varepsilon}$，并且足够小，使得 $\widetilde{\varepsilon} \leq \frac{\|\mathcal{P}_{k+1}\|_{\mu}}{2}$，

$$\begin{aligned}
\left| \frac{1}{\|P'_{k+1}\|_{\mu_n}} - \frac{1}{\|\mathcal{P}'_{k+1}\|_{\mu}} \right| &= \frac{1}{\|P'_{k+1}\|_{\mu_n}\|\mathcal{P}'_{k+1}\|_{\mu}} \left| \|P'_{k+1}\|_{\mu_n} - \|\mathcal{P}'_{k+1}\|_{\mu} \right| \\
&\leq \frac{1}{\|P'_{k+1}\|_{\mu_n}\|\mathcal{P}'_{k+1}\|_{\mu}} \left( \left| \|P'_{k+1}\|_{\mu_n} - \|\mathcal{P}'_{k+1}\|_{\mu_n} \right| + \left| \|\mathcal{P}'_{k+1}\|_{\mu_n} - \|\mathcal{P}'_{k+1}\|_{\mu} \right| \right) \\
&\leq \frac{1}{\|P'_{k+1}\|_{\mu_n}\|\mathcal{P}'_{k+1}\|_{\mu}} \left( \|P'_{k+1} - \mathcal{P}'_{k+1}\|_{\mu_n} + \left| \|\mathcal{P}'_{k+1}\|_{\mu_n} - \|\mathcal{P}'_{k+1}\|_{\mu} \right| \right)
\end{aligned}$$

$$\leq \frac{1}{\|P'_{k+1}\|_{\mu_n}\|\mathcal{P}'_{k+1}\|_{\mu}}\left(\|P'_{k+1}-\mathcal{P}'_{k+1}\|_{\infty}+\left|\|\mathcal{P}'_{k+1}\|_{\mu_n}-\|\mathcal{P}'_{k+1}\|_{\mu}\right|\right)$$

$$\leq \frac{2}{\|\mathcal{P}'_{k+1}\|_{\mu}^{3}}\left(\|P'_{k+1}-\mathcal{P}'_{k+1}\|_{\infty}+\widetilde{\varepsilon}\right)$$

以至少 $1-2\exp\left(\frac{-2\widetilde{\varepsilon}^2 n}{c_{k+1}^2}\right)$ 的概率，其中 $c_{k+1}=6\max_{x\in\mathcal{X}}|\mathcal{P}'_{k+1}(x)|$。

因此，我们只需证明 $\|P'_{k+1}-\mathcal{P}'_{k+1}\|_{\infty}$ 集中趋近于 0。根据定义，我们有

$$\|P'_{k+1}-\mathcal{P}'_{k+1}\|_{\infty}=\left\|\left(M_1-\frac{\langle M_{k+1},P_k\rangle_{\mu_n}}{\|P_k\|_{\mu_n}^2}P_k\right)-\left(\mathcal{M}_1-\frac{\langle \mathcal{M}_{k+1},\mathcal{P}_k\rangle_{\mu}}{\|\mathcal{P}_k\|_{\mu}^2}\mathcal{P}_k\right)\right\|_{\infty}$$

$$\leq \|M_{k+1}-\mathcal{M}_{k+1}\|_{\infty}+\left\|\frac{\langle M_{k+1},P_k\rangle_{\mu_n}}{\|P_k\|_{\mu_n}^2}P_k-\frac{\langle \mathcal{M}_{k+1},\mathcal{P}_k\rangle_{\mu}}{\|\mathcal{P}_k\|_{\mu}^2}\mathcal{P}_k\right\|_{\infty},$$

其中第二行由三角不等式得出。由于 $\|M_{k+1}-\mathcal{M}_{k+1}\|_{\infty}=0$，我们继续对剩余项进行界定。

利用方程（B.1）和三角不等式，我们得到

$$\left\|\frac{\langle M_{k+1},P_k\rangle_{\mu_n}}{\|P_k\|_{\mu_n}^2}P_k-\frac{\langle \mathcal{M}_{k+1},\mathcal{P}_k\rangle_{\mu}}{\|\mathcal{P}_k\|_{\mu}^2}\mathcal{P}_k\right\|_{\infty}\leq\left|\frac{1}{\|P_k\|_{\mu_n}^2}-\frac{1}{\|\mathcal{P}_k\|_{\mu}^2}\right|\left|\langle \mathcal{M}_{k+1},\mathcal{P}_k\rangle_{\mu}\right|\|\mathcal{P}_k\|_{\infty}$$

$$+\frac{1}{\|P_k\|_{\mu_n}^2}\left|\langle M_{k+1},P_k\rangle_{\mu_n}-\langle \mathcal{M}_{k+1},\mathcal{P}_k\rangle_{\mu}\right|\|\mathcal{P}_k\|_{\infty}$$

$$+\frac{1}{\|P_k\|_{\mu_n}^2}\left|\langle M_{k+1},P_k\rangle_{\mu_n}\right|\|P_k-\mathcal{P}_k\|_{\infty}.$$

应用归纳假设可知，右端的第一项和最后一项之和分别以 $A\widetilde{\varepsilon}$ 的概率小于 $1-2\exp\left(\frac{-2\widetilde{\varepsilon}n}{B^2}\right)$，因此只需证明

$$E_k=\left|\langle M_{k+1},P_k\rangle_{\mu_n}-\langle \mathcal{M}_{k+1},\mathcal{P}_k\rangle_{\mu}\right|$$

向 0 集中。

利用界定双线性型之差的经典论证，我们得到

$$\begin{aligned}E_k&\leq\left|\langle M_{k+1},P_k\rangle_{\mu_n}-\langle \mathcal{M}_{k+1},P_k\rangle_{\mu_n}\right|+\left|\langle \mathcal{M}_{k+1},P_k\rangle_{\mu_n}-\langle \mathcal{M}_{k+1},\mathcal{P}_k\rangle_{\mu}\right|\\&=\left|\langle \mathcal{M}_{k+1},P_k\rangle_{\mu_n}-\langle \mathcal{M}_{k+1},\mathcal{P}_k\rangle_{\mu}\right|\\&\leq\left|\langle \mathcal{M}_{k+1},P_k-\mathcal{P}_k\rangle_{\mu_n}\right|+\left|\langle \mathcal{M}_{k+1},\mathcal{P}_k\rangle_{\mu_n}-\langle \mathcal{M}_{k+1},\mathcal{P}_k\rangle_{\mu}\right|\\&\leq\|\mathcal{M}_{k+1}\|_{\mu_n}\|P_k-\mathcal{P}_k\|_{\mu_n}+\left|\langle \mathcal{M}_{k+1},\mathcal{P}_k\rangle_{\mu_n}-\langle \mathcal{M}_{k+1},\mathcal{P}_k\rangle_{\mu}\right|,\end{aligned}$$

其中，最后一行由柯西-施瓦茨不等式得出。鉴于此，利用归纳假设，$\|P_k-\mathcal{P}_k\|_{\mu_n}\leq\|P_k-\mathcal{P}_k\|_{\infty}\leq A\widetilde{\varepsilon}$ 以至少 $1-C\exp\left(\frac{-2\widetilde{\varepsilon}n}{B^2}\right)$ 的概率成立。最后一步是证明 $\left|\mathcal{M}_{k+1},\overline{\mathcal{P}_k^{\mu_n}}-\mathcal{M}_{k+1},\mathcal{P}_k\mu\right|$ 也以同样的概率成立，这可通过对 $f=\mathcal{M}_{k+1}$ 和 $g=\mathcal{P}_k$ 应用引理 D.1 得出。综合以上结果，并应用并集界，使所有概率不等式同时成立，我们得到：除常数 $A$、$B$ 和 $C$ 的变化外，以至少 $1-C\exp\left(\frac{-2\widetilde{\varepsilon}n}{B^2}\right)$ 的概率，

$$\max_{x\in X_n}\left|\mathcal{P}_{k+1}(x)-P_{k+1}(x)\right|\leq A\widetilde{\varepsilon}.$$

D.2. 定理 5.2 的速率推导。 首先注意到，多元正交多项式系综的核具有如下形式

$$K(x,y)=\sum_{i=1}^{m}\phi_i(x)\phi_i(y) \tag{D.2}$$

对于某个有限的 $p\in\mathbf{N}_{>0}$ 以及函数 $\phi_i:\ \mathcal{X}\to\mathbf{R}$，它们关于 $\langle .,.\rangle_\nu$ 正交（相对于某个测度 $\nu$）。具有这种形式的核的行列式点过程称为投影行列式点过程，它们构成了行列式点过程的一个重要子类，也是行列式点过程现代理论很大一部分的基础 [34]。对于这些投影行列式点过程，有如下通用的弱相干性引理。

引理 D.2。设 $\mu$ 是第二可数紧豪斯多夫空间 $\boldsymbol{X}$ 上的一个概率测度，且 $X_n\subseteq\boldsymbol{X}$ 是依据 $\mu$独立同分布抽取的 $n$ 个点构成的集合。记 $\mu_n$ 为相应的经验测度，并考虑函数序列 $\phi_i:\boldsymbol{X}\to\mathbf{C}$和 $\phi_i^{[n]}:\ X_n\to\mathbf{C}$，使得 $\phi_i$ 有界。假设对于任意 $\widetilde{\delta}\in(0,1)$ 和 $\widetilde{\varepsilon}>0$，存在 $N_\phi(\delta,\widetilde{\varepsilon})$ 使得

$$n\geq N_\phi(\widetilde{\delta},\widetilde{\varepsilon})\ \Rightarrow\ \mathbb{P}\left(\max_{i\in[m]}\max_{x\in X_n}|\phi_i(x)-\phi_i^{[n]}(x)|\geq\widetilde{\varepsilon}\right)\leq\widetilde{\delta}, \tag{D.3}$$

并进一步定义投影核：

$$\mathcal{K}(x,y)=\sum_{i=1}^{m}\phi_i(x)\phi_i(y)\quad\forall x,y\in\mathcal{X}, \tag{D.4}$$

$$K_n(x,y)=\sum_{i=1}^{m}\phi_i^{[n]}(x)\phi_i^{[n]}(y)\quad\forall x,y\in X_n. \tag{D.5}$$

那么，对于 $\widetilde{M}=$最大值$_{i\in[p]}\left(\text{最大值}_{x\in X_n}\left|\phi_i^{[n]}(x)\right|+\text{最大值}_{x\in\mathcal{X}}|\phi_i(x)|\right)$，我们有

$$n\geq N_\phi\left(\widetilde{\delta},\frac{\widetilde{\varepsilon}}{m\widetilde{M}}\right)\ \Rightarrow\ \mathbb{P}\left(\max_{x,y\in X_n}|\mathcal{K}(x,y)-K_n(x,y)|\geq\varepsilon\right)\leq\delta. \tag{D.6}$$

特别地，在紧集 $\mathcal{X}$ 上，这些 $\mathcal{K}$ 和 $K_n$ 满足定理 4.1的假设，$ii$)并且由此得到弱相干性（前提是这些核确实定义了~ ~行列式点过程）。此外，任何已确定的速率 $\delta=\delta_n$ 和 $\widetilde{\varepsilon}=\widetilde{\varepsilon}_n$ 都可以从 $\phi_i$传递到核上，至多相差一个乘法常数。

引理 D.2 的证明。 应用三角不等式，我们有

$$\begin{aligned}\max_{x,y\in X_n}|\mathcal{K}(x,y)-K_n(x,y)|&\leq m\max_{i\in[m]}\max_{x,y\in X_n}|\phi_i^{[n]}(x)\phi_i^{[n]}(y)-\phi_i(x)\phi_i(y)|\\&\leq m\max_{i\in[m]}\max_{x,y\in X_n}\left(|\phi_i^{[n]}(x)-\phi_i(x)||\phi_i(y)|+|\phi_i^{[n]}(x)||\phi_i^{[n]}(y)-\phi_i(y)|\right),\end{aligned}$$

由此即可得出结论。 □

代入命题 5.1中的估计，可得

$$n\geq\frac{A^2B^2\log\left(\frac{2}{\delta}\right)}{\widetilde{\varepsilon}^2}\ \Rightarrow\ \mathbb{P}\left(\max_{x,y\in X_n}|\mathcal{K}(x,y)-K_n(x,y)|\geq\varepsilon\right)\leq\delta,$$

~ ~对于任意 $\delta \in (0,\ 1)$ 以及 $\varepsilon \geq m\widetilde{\varepsilon}M$，其中 $M = \max^{i\in[m]}(\max_{x\in X_n}|P_i(x)| + \max_{x\in X_n}|\mathcal{P}_i(x)|)$。注意到 $\max_{x\in X_n}|P_i(x)| \leq \widetilde{\varepsilon}_n + \max_{x\in\mathcal{X}}|\mathcal{P}_i(x)|$，因此只要 $\widetilde{\varepsilon} \leq \frac{\varepsilon}{mM}$对 $M = 2\max^{i\in[m]}\max_{x\in\mathcal{X}}|\mathcal{P}_i(x)|$ 成立即可。也就是说，当

$$n \geq \frac{A^2B^2m^2M^2\log\left(\frac{2}{\delta}\right)}{\varepsilon^2} \ \Rightarrow\ \mathbb{P}\left(\max_{x,y\in X_n}|\mathcal{K}(x,y) - K_n(x,y)| \geq \varepsilon\right) \leq \delta.$$

**E. 与 [8] 中的行列式点过程比较**。我们首先回顾~ 中行列式点过程的构造 [8], ，它是相对于测度 $\mu_n$ 定义的，其核 $K_n$构造如下。

1. 对于 $\mathcal{X}$ 上的目标 Nevai 类概率分布 $\nu$，其密度为 $q$，计算前 ~ ~$m$ 个正交多项式 $\mathcal{P}_i$，使其相对于 $\nu$ 正交，并构造核 $\mathcal{K}(x,y) = \sim$ $\sim\sum_{i=1}^m \mathcal{P}_i(x)\mathcal{P}_i(y)$。2. 根据未知测度 $\mu$的密度计算核密度估计量 $\tilde{\gamma}$，其中 $X_n$ 是从该测度中抽取的。3. 定义重新加权核

$$\overline{\mathcal{K}}(x,y) = \sqrt{\frac{q(x)}{\tilde{\gamma}(x)}}\widetilde{\mathcal{K}}(x,y)\sqrt{\frac{q(y)}{\tilde{\gamma}(y)}}$$

并计算其在 $X_n$上的限制 $\left(\overline{\mathcal{K}}\right)_{|X_n\times X_n}$。

4. 最后，计算 $\left(\mathcal{K}\right)_{|X_n\times X_n}$ 的前 $m$ 个特征向量 $\phi_1,\ldots,\phi_m$ ,—并构造投影核 $K_n(x,y) = \sum_{i=1}^m \phi_i(x)\phi_i(y)$，它是相对于 $\mu_n$ 的一个（投影）行列式点过程，并满足Macchi-Soshnikov定理的假设。

~接着，文中论证了 [8] 相对于函数1点线性统计量的方差，$\mathrm{DPP}(K_n,\mu_n)$与函数 $\varphi$ 对应的方差相比，DPP( ~$\mathcal{K}$ $\nu$) 相差一个$\mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right)$项。这一保证类似于推论5.4中得到的结论。

这两种方法有两个重要的比较方面。

- 作者承认，[8] 中行列式点过程的谱舍入步骤理想情况下应当省略，因为它会使分析尤为复杂。相比之下，第5$K_n$节 $\mu_n$中的 DPP(5.1 节中构造) 要直接得多：无需选择目标密度 $q$，无需进行密度估计（这些方法依赖超参数），也无需进行谱舍入。另一方面，尽管这两个过程达到相似的渐近速率，但 $q$ 在~构造中，$\mathrm{DPP}(K_n,\mu_n)$ 的目标密度理论上可以选择得更合适，以进一步降低非渐近方差。改进效果可能达到的程度尚不明确；要进行理论比较，首先必须为两个行列式点过程的方差近似建立最优估计，而这两种情况下都尚无此类估计。就实际表现而言，两种方法非常相似 [9]。

- 或许最重要的是，精确的DPP抽样算法需要对核进行特征分解，即计算 $m$ 特征向量，精确算法的计算复杂度为$\mathcal{O}(nm^2)$。[26] 这几乎是所有基于DPP的算法的瓶颈。对于我们提出的过程DPP($K_n$，$\mu_n$)，这些特征向量恰好就是离散正交多项式，而对此已有更高效的算法。尤其是在低维情况下（这正是.....的情形），

[26]需要强调的是，此处涉及的矩阵通常并非稀疏矩阵，因此无法从高效的稀疏线性代数运算算法和实现中获益。

定理5.3中的方差缩减最具吸引力，可采用以下算法：对$O(nm)$， $d=1$（参见[19]以了解相关讨论）；对$O(nm^{3/2})$， $d=2$（参见[35]作为早期示例）。

**F. 第6节缺失的证明。**

F.1. 命题*6.3*的证明。 行列式点过程的同分布等价性由以下一般断言得到。

命题 F.1。 设 $\Gamma$ 为第二可数的局部紧豪斯多夫空间，$\widetilde{\gamma}$ 为 $\Gamma$ 上的拉东测度，且 $\gamma$ 为 $\Gamma$ 上关于 $\widetilde{\gamma}$ 绝对连续的测度，其密度为 $p$，使得 $\gamma = p\widetilde{\gamma}$。考虑两个核 $K\colon \Gamma\times\Gamma\to\mathbf{C}$ 与 $\widetilde{K}\colon \Gamma\times\Gamma\to\mathbf{C}$，使得

$$K(x,y)=\sqrt{\frac{1}{p(x)}}\widetilde{K}(x,y)\sqrt{\frac{1}{p(y)}}. \tag{F.1}$$

于是，在它们存在的前提下，$(K,\widetilde{\gamma})$ 与 $(K,\gamma)$ 定义同一个行列式点过程。

证明。设 $\varphi_r\colon \mathcal{X}^r\to\mathbf{R}$ 为有界可测函数。则我们有

$$\begin{aligned}\int_{\mathcal{X}^r}\varphi_r\rho_r[K]d\gamma^{\otimes r} &= \int_{\mathcal{X}^r}\varphi_r(x_1,\dots,x_r)\det\left([K(x_i,x_j)]_{i,j=1}^r\right)d\gamma^{\otimes r}(\{x_1,\dots,x_r\})\\ &= \int_{\mathcal{X}^r}\varphi_r(x_1,\dots,x_r)\det\left(\left[\widetilde{K}(x_i,x_j)\right]_{i,j=1}^r\right)\prod_{i=1}^r\frac{1}{p(x_i)}d\gamma^{\otimes r}(\{x_1,\dots,x_r\})\\ &= \int_{\mathcal{X}^r}\varphi_r\rho_r[\widetilde{K}]d\widetilde{\gamma}^{\otimes r}.\end{aligned}$$

□

由此，这也保证了行列式点过程$(K_n,\mu_n)$存在。

现在我们转而讨论核的集中性。首先，我们要证明 $v_i$趋向于 $\phi_i$。更准确地说。

命题 F.2. 在命题 *6.3*的假设下，存在常数 $A_1$， $A_2$， $A_3$， $B$， $C>0$ 使得，对于任意 $\widetilde{\varepsilon}>0$，以至少 $1-C\left(\frac{1}{n^2}+\exp\left(\frac{-2\widetilde{\varepsilon}^2 n}{B^2}\right)\right)$的概率，

$$\max_{x\in X_n}|v_i(x)-\phi_i(x)|\le A_1\widetilde{\varepsilon}+A_2\overline{\alpha}+A_3\overline{\beta}\quad\forall i\in[m],$$

其中 $\alpha\propto\overline{\alpha}<1$ 且 $\beta\propto\overline{\beta}<1$。

我们指出，与命题 5.1的证明类似，我们的证明是在假设 $\widetilde{\varepsilon}$ 足够小的条件下进行的，然后按照备注 C.2舍弃这一条件。关于 $\widetilde{\varepsilon}$ 的精确上界可以从证明中推导出来，但过程相当繁琐。此外，假设 $\alpha$ （分别为 $\beta$）与某个 $\alpha$ 成正比，这主要是形式上的处理，目的是隐藏命题 6.3中的一些常数。

我们将用到以下不等式：

$$\begin{aligned}|\langle a_n,b_n\rangle_{\omega_n}-\langle a,b\rangle_\omega| \le{}& |\langle a,b\rangle_{\omega_n}-\langle a,b\rangle_\omega|\\ &+\|a\|_{\omega_n}\|b_n-b\|_{\omega_n}+\|b\|_{\omega_n}\|a_n-a\|_{\omega_n}\\ &+\|a_n-a\|_{\omega_n}\|b_n-b\|_{\omega_n},\end{aligned} \tag{F.2}$$

其中， $\|.\|_{\omega_n}$ 和 $\|.\|_{\omega}$（分别为 $\langle .,.\rangle_{\omega_n}$ 和 $\langle .,.\rangle_{\omega}$）表示与测度 $\omega_n$ 和 $\omega$ 相关的 $L^2$ 范数（分别为内积）。这是因为

$$|\langle a_n, b_n\rangle_{\omega_n} - \langle a, b\rangle_{\omega}| = |\langle a_n, b_n\rangle_{\omega_n} - \langle a, b\rangle_{\omega_n}| + |\langle a, b\rangle_{\omega_n} - \langle a, b\rangle_{\omega}|,$$

而方程（F.2）可由恒等式

$$|\langle a_n, b_n\rangle_{\omega_n} - \langle a, b\rangle_{\omega_n}| = |\langle a_n - a, b\rangle_{\omega_n} + \langle a, b_n - b\rangle_{\omega_n} + \langle a_n - a, b_n - b\rangle_{\omega_n}|$$

对于向量 $a_n$， $b_n$， $a$， $b \in \mathbf{R}^n$，再应用三角不等式和柯西-施瓦茨不等式即可得到。

根据三角不等式，我们需要证明

$$\|v_i - \phi_i\|_\infty \le \|v_i - u_i\|_\infty + \|u_i - \phi_i\|_\infty \tag{F.3}$$

当 $i = 1,\dots,m$ 时趋于 0，其中我们采用记号 $\|a\|_\infty = \max_{x\in X_n} |a(x)|$，并回顾一下， $v_i$ 是通过 Gram-Schmidt 正交化过程定义的：

$$\begin{cases} v_1 = \frac{u_1}{\|u_i\|_{\omega_n}} \\ v'_{k+1} = u_{k+1} - \frac{u_{k+1}, v_k}{\|v_k\|_{\omega_n}} v_k \\ v_{k+1} = \frac{v'_{k+1}}{\|v'_{k+1}\|_{\omega_n}}. \end{cases} \tag{F.4}$$

对于方程（F.3）右端最右侧的项，其上界

$$\|u_i - \phi_i\|_\infty \le \overline{\alpha}$$

由以下结果 [28]得到。

定理 F.3 ([28]). 设 $m \in \mathbf{N}_{>0}$，并定义 $\Gamma_m = \min^{i\in[m]} \mathrm{dist}(\lambda_i, \sigma(-L \setminus \{\lambda_i\})$（其中 $(\sigma(-L) \subseteq \mathbf{R}_{\ge 0}$ 表示 $\sigma(-L)$ 的谱，而dist 表示点集欧几里得距离），并假设

$$h_1(n) \le C_1 \min\left( \left( \frac{\min(\Gamma_m, 1)}{C_2 + \lambda_p^{d_X/2+5}} \right)^2, \frac{1}{\left(C_3 + \lambda_p^{(5d_X+7)/4}\right)^2} \right), \tag{F.5}$$

其中 $C_1$、 $C_2$、 $C_3$ 是依赖于 $X$ 的几何以及密度 $p$ 的常数。然后，假设 $h_1(n) \ge \left(\frac{\log(n)}{n}\right)^{4d} x^{+8}$，并且以至少 $1 - \frac{1}{n^2}$ 的概率，存在 $a_i \in \{-1, 1\}$ 使得

$$\max_{i\in[m]} \max_{x\in X_n} |a_i u_i(x) - \phi_i(x)| \le C_4 h_1(n)^{1/2}, \tag{F.6}$$

其中 $C_4$ 是依赖于 $X$ 的几何以及 $p$ 的常数。

我们令 $\overline{\alpha} = C_4 h_1(n)^{1/2}$，并为简单起见进一步假设所有 $a_i = 1$（这对核 $\mathcal{K}$ 的定义没有影响）。我们将证明，对于下文定义的 $\beta \propto \beta$，以适当的概率有 $\|v_i - u_i\|_\infty \le A_1\widetilde{\varepsilon} + A_2\alpha + A_3\beta$。

我们对 $i$ 进行归纳。当 $i = 1$ 时，有

$$\begin{aligned} \|v_1 - u_1\|_\infty &= \|u_1\|_\infty \left| 1 - \frac{1}{\|u_1\|_{\omega_n}} \right| \\ &\le (\max_{x\in\mathcal{X}} |\phi_1(x)| + \overline{\alpha}) \left| 1 - \frac{1}{\|u_1\|_{\omega_n}} \right|, \end{aligned}$$

其中第二行以至少 $1-\frac{1}{n^2}$ 的概率成立。此时，我们注意到$\|\phi_1\|_\omega = 1$，因此只需证明 $\left|\frac{1}{\|\phi_1\|_\omega} - \frac{1}{\|u_1\|_{\omega_n}}\right|$集中于 0。

假设 $|\|\phi_1\|_\omega - \|u_1\|_{\omega_n}| \le \eta \le \frac{1}{2}$，那么

$$\left|\frac{1}{\|\phi_1\|_\omega} - \frac{1}{\|u_1\|_{\omega_n}}\right| \le 2\eta,$$

并继续求得这样一个 $\eta$。根据方程 (F.2)，我们有

$$\begin{aligned}
|\|\phi_1\|_\omega - \|u_1\|_{\omega_n}| &\le \frac{|\|\phi_1\|_\omega^2 - \|u_1\|_{\omega_n}^2|}{\|\phi_1\|_\omega + \|u_1\|_{\omega_n}} \\
&\le |\|\phi_1\|_\omega^2 - \|u_1\|_{\omega_n}^2| \\
&\le |\|\phi_1\|_\omega^2 - \|\phi_1\|_{\omega_n}^2| + 2\|\phi_1\|_{\omega_n}\|u_1 - \phi_1\|_{\omega_n} + \|u_1 - \phi_1\|_{\omega_n}^2 \\
&\le |\|\phi_1\|_\omega^2 - \|\phi_1\|_{\omega_n}^2| + 2\left(\|\phi_1\|_\omega + \frac{|\|\phi_1\|_\omega^2 - \|\phi_1\|_{\omega_n}^2|}{\|\phi_1\|_\omega}\right)\overline{\alpha} + \overline{\alpha}^2 \\
&\le |\|\phi_1\|_\omega^2 - \|\phi_1\|_{\omega_n}^2| + 2\left(\|\phi_1\|_\omega + |\|\phi_1\|_\omega^2 - \|\phi_1\|_{\omega_n}^2|\right)\overline{\alpha} + \overline{\alpha}^2,
\end{aligned}$$

我们仍需对 $|\|\phi_1\|_\omega^2 - \|\phi_1\|_{\omega_n}^2|$进行界定。为此，首先注意到

$$\begin{cases} \|\phi_1\|_\omega^2 = \langle \phi_1^2, \frac{1}{p}\rangle_\mu \\ \|\phi_1\|_{\omega_n}^2 = \langle \phi_1^2, \frac{1}{e[p]}\rangle_{\mu_n}. \end{cases}$$

因此，再次应用方程 (F.2，我们有

$$|\|\phi_1\|_\omega^2 - \|\phi_1\|_{\omega_n}^2| \le \|\phi_1^2\|_{\mu_n}\left\|\frac{1}{p} - \frac{1}{e[p]}\right\|_{\mu_n} + \left|\langle \phi_1^2, \frac{1}{p}\rangle_\mu - \langle \phi_1^2, \frac{1}{p}\rangle_{\mu_n}\right|$$

并且，根据引理 D.1，以至少 $1 - 2\exp\left(\frac{-2\varepsilon^2 n}{b^2}\right)$的概率，有

$$\left|\langle \phi_1^2, \frac{1}{p}\rangle_\mu - \langle \phi_1^2, \frac{1}{p}\rangle_{\mu_n}\right| \le \widetilde{\varepsilon},$$

对某个常数 $b > 0$ 成立。剩下只需界定 $\|\phi_1^2\|_{\mu_n}\left\|\frac{1}{p} - \frac{1}{e[p]}\right\|_{\mu_n} \le (\max_{x\in\mathcal{X}}|\phi_1(x)^2|)\left\|\frac{1}{p} - \frac{1}{e[p]}\right\|_\infty$项，为此我们依赖 [72] 的以下结果。

定理 F.4 ([72]). 若 $h_2(n) \le C_1'$ 对某个 $C_1'$成立，则其概率至少为 $1 - \frac{1}{n^2}$，

$$\max_{x\in\mathcal{X}}|e[p](x) - p(x)| \le C_2'\left(\frac{\log(n)}{nh_2(n)^{d_\mathcal{X}}}\right)^{\kappa/2}, \tag{F.7}$$

其中 $C_1'$ 和 $C_2'$ 是两个常数，取决于 $\mathcal{X}$ 的几何性质以及概率密度 $p$。

令 $\beta = C_2'\left(\frac{\log(n)}{nh_2(n)^{d_\mathcal{X}}}\right)^{\kappa/2}$，并指出该结果还意味着

$$\|e[p] - p\|_\infty \le \overline{\beta}$$

从而，当 $\beta < \min_{x\in\mathcal{X}}\frac{|p(x)|}{2} = \frac{p_{min}}{2}$时，可得

$$\begin{aligned}
\left|\frac{1}{e[p](x)} - \frac{1}{p(x)}\right| &= \frac{1}{|e[p](x)||p(x)|}|e[p](x) - p(x)| \\
&\le \frac{2}{p_{min}^2}\overline{\beta}
\end{aligned}$$

对所有 $x \in X_n$，且以至少 $1 - \frac{1}{n^2}$的概率。

综合以上结果，应用并集界使所有概率界同时成立，并利用 $\alpha$，$\beta < 1$，最终可得，以至少$1 - \left(\frac{1}{n^2} + 2\exp\left(\frac{-2\widetilde{\varepsilon}^2 n}{b^2}\right)\right)$的概率，

$$\|v_1 - \phi_1\|_\infty \le A_1\widetilde{\varepsilon} + A_2\overline{\alpha} + A_3\overline{\beta},$$

其中常数为 $A_1$、$A_2$ 和 $A_3$。

现在我们转而讨论 $i > 1$ 的情形，并继续进行归纳法。我们将略去一些与 $i = 1$ 情形类似的细微计算。此处，将方程 (B.1) 和三角不等式应用于方程 (F.4)，可得

$$\begin{aligned}
\|v_{i+1} - u_{i+1}\|_\infty &\le \|v'_{i+1} - u_{i+1}\|_\infty \frac{1}{\|v'_{i+1}\|_{\omega_n}} + \left|1 - \frac{1}{\|v'_{i+1}\|_{\omega_n}}\right| \|u_{i+1}\|_\infty \\
&\le \|v'_{i+1} - u_{i+1}\|_\infty \frac{1}{\|v'_{+1}\|_{\omega_n}} + \left|\frac{1}{\|\phi_{i+1}\|_\omega} - \frac{1}{\|v'_{i+1}\|_{\omega_n}}\right| \|u_{i+1}\|_\infty \\
&\le \|v'_{i+1} - u_{i+1}\|_\infty \frac{1}{\|v'_{i+1}\|_{\omega_n}} + \left|\frac{1}{\|\phi_{i+1}\|_\omega} - \frac{1}{\|v'_{i+1}\|_{\omega_n}}\right| \left(\max_{x\in\mathcal{X}} |\phi_{i+1}(x)| + \overline{\alpha}\right),
\end{aligned}$$

这里需要注意，

$$\left|\frac{1}{\|\phi_{i+1}\|_\omega} - \frac{1}{\|v'_{i+1}\|_{\omega_n}}\right| \le \left|\frac{1}{\|\phi_{i+1}\|_\omega} - \frac{1}{\|u_{i+1}\|_{\omega_n}}\right| + \left|\frac{1}{\|u_{i+1}\|_{\omega_n}} - \frac{1}{\|v'_{i+1}\|_{\omega_n}}\right|,$$

当 $\|u_{i+1} - v'_{i+1}\|_\infty$ 足够小时，$\left|\frac{1}{\|u_{i+1}\|_{\omega_n}} - \frac{1}{\|v'_{i+1}\|_{\omega_n}}\right| \le 2\|u_{i+1} - v'_{i+1}\|_\infty$ 成立；此外，通过重复用于 $i = 1$ 情形的论证，我们可以对 $\left|\frac{1}{\|\phi_{i+1}\|_\omega} - \frac{1}{\|u_{i+1}\|_{\omega_n}}\right|$ 给出上界。

现在我们来估计

$$\begin{aligned}
\|v'_{i+1} - u_{i+1}\|_\infty &= \left\| \frac{\langle u_{i+1}, v_i\rangle_{\omega_n}}{\|v_i\|^2_{\omega_n}} v_i \right\|_\infty \\
&= |\langle u_{i+1}, v_i\rangle_{\omega_n}| \frac{\|v_i\|_\infty}{\|v_i\|^2_{\omega_n}}.
\end{aligned}$$

应用三角不等式和不等式（F.2）于 $|\langle u_{i+1}, v_i\rangle_{\omega_n}|$ 项，可知其概率至少为 $1 - C\left(\frac{1}{n^2} + \exp\left(\frac{-2\widetilde{\varepsilon}^2 n}{B^2}\right)\right)$，其中某个 $B > 0$，

$$\begin{aligned}
|\langle u_{i+1}, v_i\rangle_{\omega_n}| &\le \underbrace{|\langle \phi_{i+1}, \phi\rangle_\omega|}_{0 \text{ by orthogonality}} + |\langle u_{i+1}, v_i\rangle_{\omega_n} - \langle \phi_{i+1}, \phi_i\rangle_\omega| \\
&\le \|u_{i+1}\|_{\omega_n}\|v_i - \phi_i\|_{\omega_n} + \|v_i\|_{\omega_n}\|u_{i+1} - \phi_{i+1}\|_{\omega_n} \\
&\quad + \|v_i - \phi_i\|_{\omega_n}\|u_{i+1} - \phi_{i+1}\|_{\omega_n} \\
&\quad + |\langle \phi_{i+1}, \phi_i\rangle_{\omega_n} - \langle \phi_{i+1}, \phi_i\rangle_\omega| \\
&\le \left(\|\phi_{i+1}\|_{\omega_n} + \|u_{i+1} - \phi_{i+1}\|_{\omega_n}\right)\widetilde{e}_i + \left(\|u_i\|_{\omega_n} + e_i\right)\|u_{i+1} - \phi_{i+1}\|_{\omega_n} \\
&\quad + \widetilde{e}_i\|u_{i+1} - \phi_{i+1}\|_{\omega_n} \\
&\quad + |\langle \phi_{i+1}, \phi_i\rangle_{\omega_n} - \langle \phi_{i+1}, \phi_i\rangle_\omega|,
\end{aligned}$$

其中 $\widetilde{e}_i, e_i \le A_1\widetilde{\varepsilon} + A_2\overline{\alpha} + A_3\overline{\beta}$由归纳假设得到。应用定理F.3后，我们得到

$$|\langle u_{i+1}, v_i\rangle_{\omega_n}| \le (\|\phi_{i+1}\|_{\omega_n} + \overline{\alpha})\,\widetilde{e}_i + (\|\phi_i\|_{\omega_n} + \overline{\alpha} + e_i)\overline{\alpha} + \widetilde{e}_i\overline{\alpha} + |\langle \phi_{i+1}, \phi_i\rangle_{\omega_n} - \langle \phi_{i+1}, \phi_i\rangle_{\omega}|,$$

并将不等式（F.2）应用于 $|\langle \phi_{i+1}, \phi_i\rangle_{\omega_n} - \langle \phi_{i+1}, \phi_i\rangle_{\omega}| = \left|\langle \phi_{i+1}\phi_i, \frac{1}{e[p]}\rangle_{\mu_n} - \langle \phi_{i+1}\phi_i, \frac{1}{p}\rangle_{\mu}\right|$得到

$$|\langle \phi_{i+1}, \phi_i\rangle_{\omega_n} - \langle \phi_{i+1}, \phi_i\rangle_{\omega}| \le \left|\langle \phi_{i+1}\phi_i, \frac{1}{p}\rangle_{\mu_n} - \langle \phi_{i+1}\phi_i, \frac{1}{p}\rangle_{\mu_n}\right| + \|\phi_{i+1}\phi_i\|_{\mu_n}\left\|\frac{1}{p} - \frac{1}{e[p]}\right\|_{\mu_n},$$

其中我们采用记号 $(\phi_{i+1}\phi_i)(x) = \phi_{i+1}(x)\phi_i(x)$。与情形 $i = 1$ 类似，我们有 $\|\phi_{i+1}\phi_i\|_{\mu_n}\left\|\frac{1}{p} - \frac{1}{e[p]}\right\|_{\mu_n} \le (\max_{x\in\mathcal{X}}|\phi_{i+1}(x)\phi_i(x)|)\frac{2}{p_{min}^2}\beta$的概率至少为$1 - \frac{1}{n^2}$，且 $\left|\langle \phi_{i+1}\phi_i, \frac{1}{p}\rangle_{\mu_n} - \langle \phi_{i+1}\phi_i, \frac{1}{p}\rangle_{\mu_n}\right| < \widetilde{\varepsilon}$的概率至少为 $1 - 2\exp\left(\frac{-2\widetilde{\varepsilon}^2 n}{b^2}\right)$，这由引理 D.1 得到。为完成我们的归纳法，只需界定 $\frac{\|v_i\|_\infty}{\|v_i\|_{\omega_n}^2}$。注意到 $\|\phi_i\|_\omega = 1$，我们有

$$\frac{\|v_i\|_\infty}{\|v_i\|_{\omega_n}^2} \le \|\phi_i\|_\infty + \left|\frac{\|v_i\|_\infty}{\|v_i\|_{\omega_n}^2} - \frac{\|\phi_i\|_\infty}{\|\phi_i\|_\omega^2}\right|,$$

其中，应用方程（B.1）和三角不等式，

$$\begin{aligned}\left|\frac{\|v_i\|_\infty}{\|v_i\|_{\omega_n}^2} - \frac{\|\phi_i\|_\infty}{\|\phi_i\|_\omega^2}\right| &\le |\|v_i\|_\infty - \|\phi_i\|_\infty|\frac{1}{\|v_i\|_{\omega_n}^2} + \left|\frac{1}{\|v_i\|_{\omega_n}^2} - \frac{1}{\|\phi_i\|_\omega^2}\right|\|\phi_i\|_\infty \\ &\le \widetilde{e}_i\left(\frac{1}{\|\phi_i\|_{\omega_n}^2} + \left|\frac{1}{\|v_i\|_{\omega_n}^2} - \frac{1}{\|\phi_i\|_\omega^2}\right|\right) + \left|\frac{1}{\|v_i\|_{\omega_n}^2} - \frac{1}{\|\phi_i\|_\omega^2}\right|\max_{x\in\mathcal{X}}|\phi_i(x)|,\end{aligned}$$

因此，应用归纳假设和并集界即可完成命题 F.2的证明。

重现引理 D.2 的证明，并且由于 $\max_{x\in X_n}|\phi_i(x)| \le \max_{x\in\mathcal{X}}|\phi_i(x)| + \sim A(\widetilde{\varepsilon} + \alpha + \beta)$，其中 $\overline{A} = \max(C_4A_1, C_2'A_2, \overline{A_3})$，取 $A \propto A$ 使得 $\alpha, \beta < 1$，我们 $\widetilde{\ }$ $\widetilde{\ }$便可对 $B = B$ 和 $C = C$ 得到命题 6.3。

F.2. 定理 *6.4* 的证明。 使用方程（B.1）即可得到如下界

$$\begin{aligned}|[K_n]_{a,b} - \mathcal{K}(x_a, x_b)| \le &\left|\sqrt{\frac{1}{e[p](x_a)}} - \sqrt{\frac{1}{p(x_a)}}\right||\mathcal{K}(x_a, y_a)|\left|\sqrt{\frac{1}{p(x_b)}}\right| \\ &+ \left|\sqrt{\frac{1}{e[p](x_a)}}\right|\left|\widetilde{[K_n]}_{a,b} - \widetilde{\mathcal{K}}(x_a, x_b)\right|\left|\sqrt{\frac{1}{p(x_b)}}\right| \\ &+ \left|\sqrt{\frac{1}{e[p](x_a)}}\right|\left|\widetilde{[K_n]}_{a,b}\right|\left|\sqrt{\frac{1}{e[p](x_b)}} - \sqrt{\frac{1}{p(x_b)}}\right|\end{aligned}$$

并且，如前所述，我们以至少 $1 - \frac{1}{n^2}$ 的概率有

$$\begin{aligned}\left|\frac{1}{e[p](x)} - \frac{1}{p(x)}\right| &= \frac{1}{|e[p](x)||p(x)|}|e[p](x) - p(x)| \\ &\le \frac{2}{p_{min}^2}C_2'\left(\frac{\log(n)}{nh_2(n)^{d_\mathcal{X}}}\right)^{\kappa/2}\end{aligned}$$

只要 $\beta$ 足够小，使得 $\beta < \frac{p_{min}}{2}$。由此可得

$$\left|\sqrt{\frac{1}{e[p](x)}} - \sqrt{\frac{1}{p(x)}}\right| = \frac{1}{e[p](x)+p(x)}\left|\frac{1}{e[p](x)} - \frac{1}{p(x)}\right|$$
$$\leq \frac{2}{p_{min}^3}\beta,$$

由命题 6.3 可得，以至少 $1-\widetilde{C}\left(\frac{1}{n^2}+\exp\left(\frac{-2\widetilde{\varepsilon}^2 n}{\widetilde{B}^2}\right)\widetilde{\delta}_n\right)$ 的概率，

$$\begin{aligned}|[K_n]_{a,b} - \mathcal{K}(x_a,x_b)| &\leq \frac{2}{p_{min}^{3+1/2}}\beta M \\ &\quad + \left(\frac{1}{p_{min}} + \frac{2}{p_{min}^{3+1/2}}\beta\right) m\widetilde{A}e\left(\widetilde{M}+\widetilde{A}e\right) \\ &\quad + \left(\frac{1}{p_{min}^{1/2}} + \frac{2}{p_{min}^{3}}\beta\right)\frac{2}{p_{min}^3}\beta\left(M + m\widetilde{A}e\left(\widetilde{M}+\widetilde{A}e\right)\right) \\ &\leq mAe,\end{aligned}$$

其中 $M = \text{最大值}_{x,\ y\in\mathcal{X}}|\mathcal{K}(x,\ y)|$；最后一个粗略界是在某个较大的常数 $A>0$ 下，并假设 $\widetilde{\varepsilon} < \frac{1}{3}$，从而 $e = \widetilde{\varepsilon}+\alpha+\beta < \frac{1}{3}+\frac{1}{3}+\frac{1}{3} = 1$ 得益于 $h_1(n)$ 和 $h_2(n)$ 的可容许性。根据注释C.2，我们在定理 6.4 的陈述中忽略最后这个条件。

为了确保以至少 $1-\delta$ 的概率有 $\max_{a,\ b\in[n]}|[K_n]_{a,\ b} - \mathcal{K}(x_a,\ x_b)| \leq \varepsilon$，只需找到 $\delta$，使得以至少 $1-\delta$ 的概率，

$$\begin{cases} mA\widetilde{\varepsilon} \leq \frac{\varepsilon}{3}, \\ mA\alpha \leq \frac{\varepsilon}{3}, \\ mA\beta \leq \frac{\varepsilon}{3}. \end{cases}$$

最后两个界实际上是确定性的，并且满足于 $\frac{\varepsilon}{3mA} \geq \min\left(h_1(n)^{1/2},\ \left(\log(n)\, n h_2(n)^{d_{\mathcal{X}}}\right)^{\kappa/2}\right)$。要满足第一个条件，只需

$$\delta \geq \widetilde{C}\left(\frac{1}{n^2} + \exp\left(\frac{-2\left(\frac{\varepsilon}{3mA}\right)^2 n}{\widetilde{B}^2}\right)\right),$$

若 $\delta$ 满足以下条件，则该条件成立：

$$\begin{cases} \delta \geq 2\frac{\widetilde{C}}{n^2}, \\ \delta \geq 2\widetilde{C}\exp\left(\frac{-2\left(\frac{\varepsilon}{3mA}\right)^2 n}{\widetilde{B}^2}\right). \end{cases}$$

特别地，当以下条件成立时，情况就是如此：

$$n \geq \max\left(\sqrt{\frac{2\widetilde{C}^2}{\delta}}, \frac{9A^2\widetilde{B}^2 m^2 \log\left(\frac{2\widetilde{C}}{\delta}\right)}{\varepsilon^2}\right).$$

**命题 7.1的证明。** 我们首先证明以下命题。

命题 G.1。 对于任意 $q>0$，存在三个常数 $\rho_q$、$c_q$、$C_q\in\mathbf{R}_{>0}$，使得取 $\gamma_n=\rho_q(\alpha_n n)^{3/4}$ 时，以至少 $1-n^{-q}$ 的概率，

$$\begin{cases}\frac{1}{n}\|K_n-\mathcal{K}_{|X_n\times X_n}\|_F\le\frac{c_q}{(\alpha_n n)^{1/8}},\\ \left|\frac{\mathrm{tr}(K_n)}{n}-c\right|\le\frac{C_q}{(\alpha_n n)^{1/4}}.\end{cases}\tag{G.1}$$

特别地，当 $\varepsilon$ 时，这两个上界都小于

$$n\ge\frac{c_q^8}{\alpha_n\varepsilon^8}\quad\text{and}\quad n\ge\frac{C_q^4}{\alpha_n\varepsilon^4},$$

并且只要满足条件，概率至少为 $\delta$

$$\delta\ge n^{-q}\ \Leftrightarrow\ n\ge\frac{1}{\delta^q}.$$

取 $b_q=c_q^8$ 和 $B_q=C_q^4$ 即可得到命题 7.1。

我们首先回顾 [47], 中的对称矩阵伯努利条目集中性结果，并将其应用于我们的情形。

定理 G.2 ([47])。 假设我们处于第7节的设定下，因此特别地，$\alpha_n\gtrsim\log(n)\,n$。那么，对所有 $r>0$，存在 $c_r\in\mathbf{R}_{>0}$，使得概率至少为 $1-n^{-r}$

$$\|A-\alpha_n\mathcal{W}_{|X_n\times X_n}\|\le c_r\sqrt{\alpha_n n},\tag{G.2}$$

*whe*其中 ***re*** 表示 $\|\cdot\|$ 是关于$\mathbf{R}$上欧几里得范数的算子范数 $^n$.

这是我们证明的起点，以下我们始终假设方程 (G.2) 成立。

现在来考虑 $W=\mathcal{W}_{|X_n\times X_n}$ 的特征分解：

$$W=\sum_{i=1}^n\tau_i w_i w_i^t.$$

记 $S\subseteq[n]$ 为满足 $\lambda_i\ge\gamma_n$ 的指标子集，我们进一步定义

$$G=\sum_{i\in S}\tau_i w_i w_i^t,\qquad\hat{A}=\sum_{i\in S}\lambda_i u_i u_i^t,\tag{G.3}$$

其中我们回顾一下，$\lambda_i$ 和 $u_i$ 分别是邻接矩阵 $A$ 的特征值和特征向量。特别地，$\widetilde{A}_{\gamma_n}=\hat{A}_{\alpha_n}$。

现在我们要对 $\left\|\widetilde{A}_{\gamma_n}-\mathcal{W}_{|X_n\times X_n}\right\|_F$ 进行界定。根据三角不等式，有

$$\left\|\frac{\hat{A}}{\alpha_n}-W\right\|_F\le\left\|\frac{\hat{A}}{\alpha_n}-G\right\|_F+\|G-W\|_F.$$

为了界定右侧的第一项，注意到 $\hat{A}$ 和 $G$ 的秩均为 $|S|$，因此

$$\left\|\frac{\hat{A}}{\alpha_n}-G\right\|_F\le\sqrt{2|S|}\left\|\frac{\hat{A}}{\alpha_n}-G\right\|,$$

其中右侧的范数是算子范数。根据三角不等式，这个算子范数之差可分解为

$$\left\|\frac{\hat{A}}{\alpha_n}-G\right\| \le \frac{1}{\alpha_n}\left\|\hat{A}-A\right\| + \left\|\frac{\hat{A}}{\alpha_n}-W\right\| + \|W-G\|$$

根据 $S$和 $\hat{A}$, $\left\|\hat{A}-A\right\| \le \gamma_n$的定义。进一步，根据定理 G.2，以至少 $1-n^{-\gamma}$的概率， $\|A-\alpha_n W\| \le c_r\sqrt{\alpha_n n}$。由加藤不等式， [36, 61],可得

$$\max_{i\in[n]}|\lambda_i-\alpha_n\tau_i| \le \|A-\alpha_n W\| \le c_r\sqrt{\alpha_n n}, \tag{G.4}$$

因此，对所有 $i\in[n]\setminus S$,

$$0 \le \alpha_n\tau_i \le \lambda_i + c_r\sqrt{\alpha_n n} \le \gamma_n + c_r\sqrt{\alpha_n n}. \tag{G.5}$$

由此可得 $\|W-G\| = \left\|\sum_{i\in[n]\setminus S}\tau_i w_i w_i^t\right\| \le \frac{\gamma_n}{\alpha_n} + c_r\sqrt{\frac{n}{\alpha_n}}$，并得到

$$\left\|\frac{\hat{A}}{\alpha_n}-G\right\| \le 2\left(\frac{\gamma_n}{\alpha_n} + c_r\sqrt{\frac{n}{\alpha_n}}\right).$$

方程（G）右侧第一项的界可通过对 $S$的基数进行估计得到。为此，令 $i\in S$并注意到

$$\gamma_n - c_r\sqrt{\alpha_n n} \le \lambda_i - \|A-\alpha_n W\| \le \alpha_n\tau_i,$$

因此，对所有此类 $i$求和可得

$$|S|(\gamma_n - c_r\sqrt{\alpha_n n}) \le \alpha_n\sum_{i\in S}\tau_i \le \alpha_n \mathrm{tr}(W),$$

其中，注意到根据我们对$\mathcal{W}$的假设，$\mathrm{tr}(W) \le nc \le n$。因此我们得到$|S| \le \frac{\alpha_n}{\gamma_n - c_r\sqrt{\alpha_n n}}\mathrm{tr}(W)$，从而得到

$$\left\|\frac{\hat{A}}{\alpha_n}-G\right\|_F \le 2\sqrt{2\frac{\alpha_n\mathrm{tr}(W)}{\gamma_n - c_r\sqrt{\alpha_n n}}}\left(\frac{\gamma_n}{\alpha_n} + c_r\sqrt{\frac{n}{\alpha_n}}\right).$$

现在我们继续对方程（G）右端的第二项进行估计。由方程（G.5），可直接得到

$$\begin{aligned}\|G-W\|_F^2 &= \sum_{i\in[n]\setminus S}\tau_i^2 \\ &\le \left(\frac{\gamma_n}{\alpha_n} + c_r\sqrt{\frac{n}{\alpha_n}}\right)\sum_{i\in[n]\setminus S}\tau_i \\ &\le \left(\frac{\gamma_n}{\alpha_n} + c_r\sqrt{\frac{n}{\alpha_n}}\right)\mathrm{tr}(W).\end{aligned}$$

综合以上各项，可得

$$\left\|\frac{\hat{A}}{\alpha_n}-W\right\|_F \le 2\sqrt{2\frac{\alpha_n\mathrm{tr}(W)}{\gamma_n - c_r\sqrt{\alpha_n n}}}\left(\frac{\gamma_n}{\alpha_n} + c_r\sqrt{\frac{n}{\alpha_n}}\right) + \sqrt{\frac{\gamma_n}{\alpha_n} + c_r\sqrt{\frac{n}{\alpha_n}}\mathrm{tr}(W)}$$

并取 $\gamma_n \sim (\alpha_n n)^{3/4}$，我们确实得到

$$\left\|\frac{\widetilde{A}_{\gamma_n} - \mathcal{W}_{|X_n\times X_n}}{n}\right\|_F \lesssim \frac{1}{(\alpha_n n)^{1/8}}.$$

为得到核的弗罗贝尼乌斯范数集中性，我们注意到，由于 $\operatorname{tr}\left(\widetilde{A}_{\gamma_n}\right) \geq 0$, $\frac{1}{n}\left\|\max\left(c - \operatorname{tr}\left(\frac{\widetilde{A}_{\gamma_n}}{n}\right), 0\right) I\right\|_F \leq \frac{c}{\sqrt{n}}$ 因此应用三角不等式后，集中速率不会受到影响，我们得到

$$\left\|\frac{\bar{A}_{\gamma_n} - \mathcal{K}_{|X_n \times X_n}}{n}\right\|_F \lesssim \frac{1}{(\alpha_n n)^{1/8}}.$$

接下来，我们证明 $\operatorname{tr}\left(\frac{\bar{A}_{\gamma_n}}{n}\right)$ 集中于 $c$。首先注意到

$$\begin{aligned}
\operatorname{tr}\left(\widetilde{A}_{\gamma_n}\right) &= \sum_{i \in S} \frac{\lambda_i}{\alpha_n} \\
&\leq \sum_{i \in S} \tau_i + c_r |S| \sqrt{\frac{n}{\alpha_n}} \\
&\leq \operatorname{tr}(W) + c_r \sqrt{\frac{n}{\alpha_n}} \frac{\alpha_n n}{\gamma_n - c_r \sqrt{\alpha_n n}},
\end{aligned}$$

其中，第一个和第二个不等式分别由方程 (G.5) 得出。因此，对于 $\gamma_n \sim (\alpha_n n)^{3/4}$,

$$0 \leq \frac{\operatorname{tr}\left(\widetilde{A}_{\gamma_n}\right)}{n} \leq c + \frac{C_q}{(\alpha_n n)^{1/4}}$$

~对于某个常数 $C_q \in \mathbf{R}_{>0}$。因此，将 $C(A_{\gamma_n})I$ 相加后，我们确实得到

$$\left|\frac{\operatorname{tr}\left(\bar{A}_{\gamma_n}\right)}{n} - c\right| \leq \frac{C_q}{(\alpha_n n)^{1/4}}.$$

那么，我们有

$$\lambda_{\max}(\bar{A}_{\gamma_n}) \leq \operatorname{tr}(\bar{A}_{\gamma}) \leq n\left(1 + \frac{C_q}{(\alpha_n n)^{1/4}}\right)$$

利用 $c \leq 1$，由此可得

$$\begin{aligned}
\left(1 + \frac{C_q}{(\alpha_n n)^{1/4}}\right)^{-1} \leq C'(\bar{A}_{\gamma_n}) \leq \left(1 + \frac{1}{(\alpha_n n)^{1/4}}\right)^{-1} \\
1 - \left(1 + \frac{C_q}{(\alpha_n n)^{1/4}}\right)^{-1} \geq 1 - C'(\bar{A}_{\gamma_n}) \geq 1 - \left(1 + \frac{1}{(\alpha_n n)^{1/4}}\right)^{-1} \\
\frac{C_q}{(\alpha_n n)^{1/4}} \gtrsim 1 - C'(\bar{A}_{\gamma_n}) \gtrsim \frac{1}{(\alpha_n n)^{1/4}}
\end{aligned}$$

使得

$$\left|1 - C'(\bar{A}_{\gamma_n})\right| \lesssim \frac{1}{(\alpha_n n)^{1/4}}.$$

接下来，我们可以检验 $K_n$ 上的界：

$$\begin{aligned}
\left\|\frac{K_n - \mathcal{K}_{|X_n \times X_n}}{n}\right\|_F &\leq \left\|\frac{\bar{A}_{\gamma_n} - \mathcal{K}_{|X_n \times X_n}}{n}\right\|_F + \left\|\frac{K_n - \bar{A}_{\gamma_n}}{n}\right\|_F \\
&\lesssim \frac{1}{(\alpha_n n)^{1/8}} + \left|\frac{1}{n} - \frac{C'(\bar{A}_{\gamma_n})}{n}\right| \left\|\bar{A}_{\gamma_n}\right\|_F
\end{aligned}$$

$$\lesssim \frac{1}{(\alpha_n n)^{1/8}} + \frac{1}{(\alpha_n n)^{1/4}} \lesssim \frac{1}{(\alpha_n n)^{1/8}}$$

利用 $\left\|\bar{A}_{\gamma_n}\right\|_F \lesssim \|W\|_F + o(n) \lesssim n$。

类似地，对于迹，

$$\left|\mathrm{tr}\left(\frac{K_n - \mathcal{K}_{|X_n \times X_n}}{n}\right)\right| \leq \left|\mathrm{tr}\left(\frac{\bar{A}_{\gamma_n} - \mathcal{K}_{|X_n \times X_n}}{n}\right)\right| + \left|\mathrm{tr}\left(\frac{K_n - \bar{A}_{\gamma_n}}{n}\right)\right|$$

$$\lesssim \frac{1}{(\alpha_n n)^{1/4}} + \left|\frac{1}{n} - \frac{C'(\bar{A}_{\gamma_n})}{n}\right| \left|\mathrm{tr}\left(\bar{A}_{\gamma_n}\right)\right| \lesssim \frac{1}{(\alpha_n n)^{1/4}}$$

因为 $\left|\mathrm{tr}\left(\bar{A}_{\gamma_n}\right)\right| \lesssim cn + o(n) \lesssim n$，由此完成证明。