arXiv:2302.04292v1 [math.ST] 8 Feb 2023

使用无限多视图模型的样本复杂度

Robert A. Vandermeulen VANDERMEULEN@TU-BERLIN.DE
柏林工业大学：机器学习组，柏林学习与数据基础研究所

摘要

近期研究表明，当目标密度是具有利普希茨连续边缘分布的可分概率密度的凸组合（即多视图模型）时，使用低秩估计器可以显著提高非参数密度估计器的收敛速率。然而，这一假设限制性很强，目前尚不清楚这些发现能够在多大程度上推广到一般的概率密度函数。本文通过引入一种刻画概率密度函数复杂度的新方法——非负利普希茨谱（NL谱）——回答了这一问题。与光滑性性质不同，NL谱几乎可以用于刻画任何概率密度函数。文中给出了依赖于目标密度NL谱的有限样本界，并由此推导出维度无关的收敛速率，用以刻画NL谱在何种情况下能够带来快速收敛。

关键词：非参数密度估计、低秩模型、密度估计、张量分解、样本复杂度

1. 引言

非参数密度估计是一项数学性质已得到充分研究的统计任务。直方图和核密度估计器等常用非参数密度估计量的普遍一致性早已为人所知。此外，当已知目标密度属于某个光滑密度类时，已有针对非参数密度估计量的有限样本界和收敛速率（Silverman，1978；Gy¨orfi 等，1985；Devroye 和 Lugosi，2001；Tsybakov，2008；Vandermeulen 和 Scott，2013；Dasgupta 和 Kpotufe，2014；Jiang，2017）。众所周知，非参数密度估计深受维数灾难的影响。在理论研究中，这通常表现为收敛速率或界中出现某种维数指数（例如，参见 Jiang（2017）中的定理1）。近期研究已证明，将光滑性假设与低秩/多视图假设相结合，可以显著提高非参数密度估计器的收敛速率，从而避免维数灾难。具体而言，已有研究证明，只要目标密度满足多视图模型假设，就存在以速率 $O_n(1/\sqrt[3]{n})$ 收敛的普遍一致非参数密度˜估计量（Vandermeulen 和 Ledent，2021），

$$p\left(x_1, \ldots, x_d\right) = \sum_{i=1}^{k} w_i \prod_{j=1}^{d} p_{i,j}\left(x_j\right), \tag{1}$$

其中 $p_{i,\ j}$ 是利普希茨连续的概率密度函数（pdf），且 w 位于概率单纯形中。值得注意的是，该收敛速率与维度 d、分量数 k以及分量边缘密度的 Lipschitz 常数均无关。对于同一类密度，还已证明，无论如何选择速率，标准直方图估计器都以 $\omega(1/\sqrt[d]{n})$ 的速率收敛

© R.A. Vandermeulen。

关于分箱宽度（Vandermeulen 和 Ledent，2021），这是维数灾难的一个明显例证。

尽管上述结果展示了将多视图结构纳入密度估计的潜在益处，但假设（1）相当强。本研究扩展了Vandermeulen 和 Ledent （2021）的分析，使其几乎适用于任何概率密度函数。为此，本文引入了一种新的密度复杂度刻画方式，即非负利普希茨谱（NL谱）。NL谱是对（1）的无限求和（$k = \infty$）扩展，用于刻画w_i 的衰减速度以及利普希茨常数的增长情况。随后证明，NL谱适用于极其一般的一类概率密度函数：每个（勒贝格意义下）几乎处处连续的概率密度函数都有 NL 谱。这使得Vandermeulen 和 Ledent （2021）中的分析得以扩展到范围大得多的一类概率密度函数。接着，为Vandermeulen 和 Ledent （2021）中引入的低秩直方图估计量推导出一个依赖于 NL 谱的有限样本界。最后，利用这一界，根据 NL 谱的增长率和衰减率，证明了估计量的收敛速率。特别地，在（1）的无限分量版本中，若已知 $\sum_{i=k+1}^{\infty} w_i \in O_k\left(k^{-\alpha}\right)$ 且 $p_{k,j}$, L_k 的利普希茨常数满足 $L_k \in O_k\left(k^{\beta}\right)$，则存在一个普遍一致估计量，其收敛速率为 $\tilde{O}_n\left(n^{-\alpha/(3\alpha+\beta+1)}\right)$。

1.1. 相关工作

最早研究多视图模型非参数密度估计量的是 Song 等人（2014），他们提出了一种对核密度估计进行因式分解以恢复多视图模型的方法。尽管 Song 等人（2014）给出了多视图分量恢复速率的理论保证，但并未证明可以利用多视图假设来改善估计量收敛。相反，他们提出的是一种非参数混合建模方法。其他具有较强理论保证的非参数混合建模方法包括：假设数据按照混合分量分组（Vandermeulen 和 Scott，2015；Vandermeulen 和 Scott，2019；Ritchie 等人，2020；Vandermeulen 和 Saitenmacher，2022），或假设混合分量满足集中性假设（Dan 等人，2018；Aragam 等人，2020；Aragam 和 Yang，2021；Aragam 和 Tai，2022）。与 Song 等人（2014）的方法一样，这些方法得到的收敛速率并未优于典型的非参数速率。

关于低秩密度估计的其他研究主要关注可辨识性和可恢复性（Allman 等，2009；Kargas 和 Sidiropoulos，2019），而另一些研究则提出了低秩方法，并通过实证表明，这些方法能够改善非参数估计，但没有给出速率保证（Song 和 Dai，2013；Novikov 等，2021；Amiridi 等，2022）。矩阵估计中的低秩方法也已得到广泛研究。非负矩阵或张量分解是一项与多视图密度估计相似的任务，因为它可用于恢复低秩概率矩阵/张量（Lee 和 Seung，1999；Donoho 和 Stodden，2004；Kim 和 Choi，2007；Arora 等，2012）。压缩感知领域也已广泛研究低秩矩阵方法，并利用受限等距性质（Recht 等，2010）或受限强凸性（Negahban 和 Wainwright，2011，2012），提出了具有强有力理论保证的改进型矩阵估计量。尽管与前文介绍的方法略有不同，但在 Wasserstein 距离估计中对耦合测度施加低非负秩约束，也已得到研究。研究表明，这种方法能够改善统计估计并带来计算优势（Scetbon 等，2021；Scetbon 和 Cuturi，2022）。虽然距离估计与

密度估计有很大不同，但这种方法值得关注，因为它在一类低秩概率测度上进行优化，并且具有扎实的理论分析，证明了估计量收敛性的改善。

这些工作 Vandermeulen (2020)；Vandermeulen 和 Ledent (2021) 首次通过强有力的理论保证表明，多视图假设可以用于改善非参数密度估计量的收敛速率。这些工作还表明，非负 Tucker 分解同样可以达到这一效果，并发现 Tucker 分解在实践中产生了更好的估计量。与这些工作中的证明类似，本文的结果并非对为压缩感知或非负矩阵分解而开发的技术进行改编。

2. 结果

本节介绍并讨论本文的主要结果。所有结果的证明见第 3 节。

在介绍结果之前，需要先说明一些符号和术语。本文还会在行文中陆续介绍其他符号；本文符号汇总表见附录 A。对于两个集合，A 和 B，$A \times B$ 表示笛卡尔积。对于两个实值函数，$f\colon A \to \mathbb{R}$ 和 $g\colon B \to \mathbb{R}$，它们的乘积定义为 $f \times g\colon A \times B \to \mathbb{R}\colon (a, b) \mapsto f(a)g(b)$。注意，如果在此情况下 $A = B$，那么 $f \times g$ 是定义在 $A \times A$ 上的函数，而不是定义在 A上的函数；$f \cdot g\colon x \mapsto f(x)g(x)$。$\mathbb{N}$ 表示所有大于 0 的整数。对于集合 A，$\mathbb{I}^A$ 表示 A上的示性函数。对于 $n \in \mathbb{N}$，$[n] = \{1, 2, \ldots, n\}$。对于集合和函数，乘积算子 $\prod$ 和幂算子 $\cdot^{\times n}$ 始终表示上文定义的乘积 $\times$。术语几乎处处（a.e.）始终指勒贝格测度。pdf 是一个几乎处处非负且勒贝格积分等于 1 的函数。函数的等式（或不等式）始终表示几乎处处相等（或几乎处处满足不等式）。

2.1. 非负利普希茨谱

在密度估计中，概率密度函数的光滑性常被用来衡量其复杂度，或衡量密度估计的难度。下面给出一种描述密度复杂度的方法，其形式类似于线性算子的谱，也是本文其余部分的重点。

定义 1 若概率密度函数 p 具有非负利普希茨谱 (w, L)，其中 w、$(w_i)_{i=1}^{\infty}$ 且 L、$(L_i)_{i=1}^{\infty}$，则

$$p = \sum_{i=1}^{\infty} w_i \prod_{j=1}^{d} p_{i,j}, \tag{2}$$

其中 $(w_i)_{i=1}^{\infty}$ 是一个非负序列，且 $\sum_{i=1}^{\infty} w_i = 1$ 并且 $(p_{i,\ j})_{(i,j)\in\mathbb{N}\times[d]}$ 是利普希茨连续概率密度函数，其 Lipschitz 常数为 $p_{i,\ j}$ 等于 L_i 对所有 i，j。[1] 若边缘概率密度函数 光滑 ，则称该非负利普希茨谱为 $p_{i,\ j}$ 中 (2) 的所有边缘概率密度函数均光滑。

直观而言，若 NL谱中的 w 衰减缓慢，且 L_i 较大，则说明概率密度函数更复杂。后续结果将对此作出精确描述。在本文其余部分的所有结果中，NL谱的“光滑”描述都不会发挥任何作用。之所以将其纳入，仅仅是因为它是一种额外的正则性性质，易于在证明中包含，或许也会对未来的研究有所帮助。

1. 为了技术上的方便，我们始终假设 $w_1 > 0$。

值得注意的是，概率密度函数的 NL谱并不唯一。这种不唯一性不仅源于对谱进行重新排序或重复求和项等简单修改，也可能是因为，与矩阵的奇异值分解不同，最小非负分解和张量分解不一定在缩放和重新排序意义下唯一。许多研究探讨了（非负）矩阵、张量及测度分解的唯一性或非唯一性（Kruskal, 1977; Sidiropoulos and Bro, 2000; Donoho and Stodden, 2004; Comon et al., 2008; Allman et al., 2009; Anandkumar et al., 2014; Vandermeulen 和 Scott, 2015; Tah- masebi et al., 2018; Vandermeulen 和 Scott, 2019; Vandermeulen 和 Saitenmacher, 2022）。相较于基于光滑性的复杂度刻画（如 H¨older、Sobolev 和 Nikol'ski 光滑性），NL谱的这种不唯一性或许可被视为一种潜在缺点；这些基于光滑性的刻画由一两个标量值表征，而非两列无穷级数（Tsybakov, 2008）。

下面的定理表明，NL-谱描述了一类非常丰富的概率密度函数。

定理 2 若概率密度函数几乎处处连续，则它具有平滑的非负利普希茨谱。

可以说，这类概率密度函数涵盖了所有具有实际意义的概率密度函数；很难想象现实中会有人想要估计一个在正测度集合上本质不连续的概率密度函数。相比之下，H¨older、Sobolev 和 Nikol'ski 光滑性类不适用于含有单个不连续点的概率密度函数。

在分析学和概率论的其他领域，也存在让人联想到 NL 谱的分解或近似。例如，多元黎曼和的形式与（1）类似，

$$\sum_{i=1}^{k} w_i \mathbb{I}_{(a_{i,1},b_{i,1})\times\cdots\times(a_{i,d},b_{i,d})} = \sum_{i=1}^{k} w_i \mathbb{I}_{(a_{i,1},b_{i,1})} \times \cdots \times \mathbb{I}_{(a_{i,d},b_{i,d})} = \sum_{i=1}^{k} w_i \prod_{j=1}^{d} \mathbb{I}_{(a_{i,j},b_{i,j})}.$$

有关随机过程的文献中也提到过类似的分解。以下引文出自 Kendall 和 Montana （2002）：

> .....可以证明，当状态空间是光滑流形且核具有连续密度 $p(x, y)$ 时，存在性成立；事实上，还可以证明，存在大量阶数为 1 的小集，即它们可用于构造表示 $p(x, y) = \sum_{i=1}^{\infty} f_i(x)g_i(y)$，其中 $f_i(x)$ 是支撑在小集上的非负连续函数，而 $g_i(y)$ 是概率密度函数。

定理 2与以往结果不同，因为 NL谱分解是精确的，而非近似；它是概率密度函数的混合，分量边缘分布具有很强的正则性，并且描述了一类非常丰富且在实践中很有用的概率密度函数。

2.2. 非参数估计量结果

以下定理表明，存在一些非参数密度估计量，其性能取决于目标密度的 NL谱。令 $\mathcal{D}_d$ 为所有支撑在d维单位立方体上的概率密度函数构成的集合，$[0, 1]^{\times d}$。以下定理基于 Vandermeulen 和 Ledent （2021）引入的低秩直方图估计量推导而来。

定理 3 令 $d, b, k, n \in \mathbb{N}$ 和 $0 < \delta \leq 1$。存在一个估计量 $V_n \in \mathcal{D}_d$，使得对于任意密度 $p \in \mathcal{D}_d$，其 NL谱为 (w, L) 时，以下结论成立

$$P\left(\|p - V_n\|_1 > \frac{\sqrt{3}d}{2b}\sum_{i=1}^{k} w_i L_i + 6\sum_{i=k+1}^{\infty} w_i + 7\sqrt{\frac{2bdk\log(4bdkn)}{n}} + 7\sqrt{\frac{\log(\frac{3}{\delta})}{2n}}\right) < \delta,$$

其中，V_n 是 X_1、.....、$X_n \overset{iid}{\sim} p$的函数。

虽然本文中的估计量在技术上可以实现，但计算上难以处理。引入一种略有不同的刻画 NL-谱的方法将会有所帮助。

定义 4 一个 NL-类，记为 $\mathfrak{V}\left(W, \tilde{L}\right)$，其中 W、$(W_i)_{i=1}^{\infty}$是一个非负、非递增且收敛到 0 的序列，而 $\tilde{L}$、$\left(\tilde{L}_i\right)_{i=1}^{\infty}$ 是一个非负、非递减序列，它是所有满足 $L_k \leq L_k$且对所有 k均有 $\sum_{i=k+1}^{\infty} w_i \leq W_k$ 的 NL- 谱 (w, L) 的概率密度函数的 ˜ 集合。

存在一些估计量，在 NL-类上具有如下行为。

定理 5 设 d、b、k、$n \in \mathbb{N}$ 和 $0 < \delta \leq 1$。存在一个估计量 $V_n \in \mathcal{D}_d$，使得对于任意密度 $p \in \mathcal{D}_d \bigcap \mathfrak{V}(W, L)$，以下结论成立：

$$P\left(\|p - V_n\|_1 > \frac{\sqrt{3}d}{2b}L_k + 6W_k + 7\sqrt{\frac{2bdk\log(4bdkn)}{n}} + 7\sqrt{\frac{\log(\frac{3}{\delta})}{2n}}\right) < \delta,$$

其中，V_n 是 X_1、.....、$X_n \overset{iid}{\sim} p$ 的函数。

可以利用这一点得到适用于具有多项式收敛速率的 NL-类的估计量。

命题 6 设 α、$\beta > 0$ 和 $d \in \mathbb{N}$。存在一个普遍一致的估计量序列 $(V_n)_{n=1}^{\infty}$，定义在 $\mathcal{D}_d$ 上，使得对于固定的采样密度 $p \in \mathcal{D}_d \cap \mathfrak{V}(W, L)$，满足 $W_k \in O_k(k^{-\alpha})$ 和 $L_k \in O_k\left(k^{\beta}\right)$ 时，以下结论成立：$\|V_n - p\|_1 \in \tilde{O}_n\left(n^{-\alpha/(3\alpha+\beta+1)}\right)$。

在 $\mathcal{D}_d$上具有普遍一致性；如果采样密度 $p \in \mathcal{D}_d$ 固定，则 $\|p - V_n\|_1 \xrightarrow{p} 0$。这里，$w$ 的衰减会抵消 L 的增长对收敛速率的影响。保持 β˜不变并令 $\alpha \to \infty$，该估计量的收敛速率趋近于 $O_n(1/\sqrt[3]{n})$，这与Vandermeulen 和 Ledent (2021) 对有限秩密度得到的收敛速率一致。这近似达到了Gy¨orfi 等人(1985) 中一维直方图的最优收敛速率：“对于光滑密度，直方图估计的平均 L^1 误差至少必须以 $n^{-1/3}$的速率变化。”

3. 证明

本节分为两个小节。第一个小节逐步铺垫并证明定理2，第二个小节则证明估计量界和收敛速率。在证明本文的结果之前，还需要引入更多符号说明。

令 $\mathcal{F}$为定义在 $\mathbb{R}$上的光滑的利普希茨连续概率密度函数集合。令 $\mathcal{F}^d$、$\{f_1 \times \cdots \times f_d \mid f_i \in \mathcal{F}\}$，注意它们也都是概率密度函数。回顾一下，实向量空间中集合 S 的锥包是 cone (S) $\{\sum_{i=1}^{n} w_i s_i \mid n \in \mathbb{N}, w_i \geq 0, s_i \in S\}$。锥算子只会应用于$\mathcal{F}^d$。虽然这是标准符号，仍提醒读者，对于集合 S，$S^{\mathbb{N}}$ 是由 S中的元素组成的所有无限序列的集合：$S^{\mathbb{N}}$、$\{(s_i)_{i=1}^{\infty} \mid s_i \in S\}$。最后，$\lambda$ 表示勒贝格测度，其维数总能从上下文中明确。

3.1. 朝着定理的证明迈进 2

下面的引理指出，在多元区间上的示性函数 $\prod_{i=1}^{d} \mathbb{I}^{(a_i,b_i)}$下，可以找到 $\mathcal{F}^d$ 中的一个正比例缩放元素，使其在 L^1 距离下任意逼近该示性函数。

引理 7 设 $\varepsilon > 0$和 $I = \prod_{i=1}^{d} \mathbb{I}^{(a_i,\ b_i)}$。存在 $f \in \mathcal{F}^d$和 $w \geq 0$，使得 $wf \leq I$和 $\int I - wf d\lambda \leq \varepsilon$。

引理的证明7 证明将按维数进行归纳， d。

基础情形， $d = 1$：如果 $\varepsilon \geq b_1 - a_1$，只需任选一个 $f \in \mathcal{F}$ 并令 $w = 0$，即可得到$wf = 0 \leq I$ 和 $\int I - wf d\lambda = b_1 - a_1 - 0 \leq \varepsilon$，因此当 $\varepsilon \geq b_1 - a_1$时引理成立。本情形的其余部分将省略 $\varepsilon < b_1 - a_1$ 以及下标“1”。

基础情形将使用光滑鼓包函数来证明；关于鼓包函数的技术性介绍，参见 Tu（2010）第 13.1 节。此证明只需要一组非常简单的鼓包函数：对于任意 $c_1 < c_2 < c_3 < c_4 \in \mathbb{R}$，存在一个光滑函数 ρ，使得 ρ在 $(c_1, c_4)^C$上等于零， ρ在 (c_2, c_3)上等于 1，并且 ρ在其他地方属于 $[0, 1]$。利用这一点，令 $\tilde{f}$ 为一个类似于 ρ的光滑函数 $\tilde{}$ $\tilde{}$ $\tilde{}$，满足 $c_1 = a, c_2 = a + \varepsilon/4, c_3 = b - \varepsilon/4, c_4 = b$。令 $f = f/\int f d\lambda$和 $w = \int f d\lambda$，于是 $\tilde{}f = wf$，且 f是一个满足 $wf \leq I$的概率密度函数。

由于 f 光滑，因此其一阶导数处处存在且连续。由于 df 在紧集 $[a, b]$上连续， df 在 $[a, b]$上有界。由于 f 在 $[a, b]^C$ 上为零， df 在该集合上恒等于零。由此可知， df 有界，因此 f 是利普希茨连续的；此外，它还光滑，因而 $f \in \mathcal{F}$。下面的不等式即可完成基础情形 $d = 1$，

$$
\begin{aligned}
&\int I - wf d\lambda \\
&\quad = \int_{(a+\varepsilon/4, b-\varepsilon/4)} \mathbb{I}^{(a,b)} - wf d\lambda + \int_{(a,b)^C} \mathbb{I}^{(a,b)} - wf d\lambda + \int_{(a,a+\varepsilon/4)\cup(b-\varepsilon/4,b)} \mathbb{I}^{(a,b)} - wf d\lambda \\
&\quad = \int_{(a,a+\varepsilon/4)\cup(b-\varepsilon/4,b)} \mathbb{I}^{(a,b)} - wf d\lambda \leq \int_{(a,a+\varepsilon/4)\cup(b-\varepsilon/4,b)} 1 d\lambda = \varepsilon/2.
\end{aligned}
$$

归纳步骤：假设该引理对某个 $d \in \mathbb{N}$ $\varepsilon > 0$ $I = \mathbb{I}^{a_1 b_1} \times \cdots \times$ 成立。 $\mathbb{I}^{a_d b_d} \times \mathbb{I}^{a_{d+1} b_{d+1}}$ $J = \mathbb{I}^{a_1 b_1} \times \cdots \times \mathbb{I}^{a_d b_d}$ $J' = \mathbb{I}^{a_{d+1} b_{d+1}}$ $I = J \times J'$ 令 和 (,) ,) ,)。令 (,) ,) 以及 (,)，因此 。根据归纳假设，存在 w、 $w' \geq 0$、 $f \in \mathcal{F}^d$ 和 $f' \in \mathcal{F}$，使得 $wf \leq J$、 $w'f' \leq J'$，

$$
\int J - wf d\lambda \leq \frac{\varepsilon}{4\int J' d\lambda}, \text{ and } \int J' - w'f' d\lambda \leq \frac{\varepsilon}{4\int J d\lambda}.
$$

注意， $f \times f' \in \mathcal{F}^{d+1}$和 $ww'f \times f' \leq I$。以下内容完成证明

$$
\begin{aligned}
\int J \times J' - ww'f \times f' d\lambda &= \int J \times J' - J \times w'f' + J \times w'f' - ww'f \times f' d\lambda \\
&= \int J \times \left(J' - w'f'\right) + (J - wf) \times w'f' d\lambda \\
&= \int J d\lambda \int \left(J' - w'f'\right) d\lambda + \int (J - wf)\, d\lambda \int w'f' d\lambda \\
&\leq \int J d\lambda \frac{\varepsilon}{4\int J d\lambda} + \frac{\varepsilon}{4\int J' d\lambda} \int J' d\lambda = \varepsilon/2.
\end{aligned}
$$

■

以下引理表明，对于一类具有足够正则性的非负函数中的任意函数，都可以在 $\mathcal{F}^d$ 的锥包中找到一个元素，使其不超过该函数，并且在 L^1 距离下任意逼近该函数。

引理 8 设 $p:\ \mathbb{R}^d \to \mathbb{R}$ 为非负、紧支撑、有界且几乎处处连续，并且 $\varepsilon > 0$。存在 $k \in \mathbb{N}$、w_1、……、$w_k \geq 0$，以及 f_1、……、$f_k \in \mathcal{F}^d$，使得

$$\int p - \sum_{i=1}^{k} w_i f_i d\lambda \leq \varepsilon \quad \textit{and} \quad p \geq \sum_{i=1}^{k} w_i f_i.$$

引理 8 的证明 由于 p 具有紧支撑、有界且几乎处处连续的性质，根据黎曼-勒贝格定理，它是黎曼可积的，因此也是达布可积的[2]。于是，利用达布下和，存在一个“阶梯函数” $\phi = \sum_{i=1}^{k} w_i' \prod_{j=1}^{d} \mathbb{I}_{(a_{i,j}, b_{i,j})}$，满足 $w_i' \geq 0$，使得 $\phi \leq p$ 且 $\int p - \phi d\lambda \leq \varepsilon/2$。为简洁起见，记 $I_i = \prod_{j=1}^{d} \mathbb{I}_{(a_{i,j}, b_{i,j})}$。根据引理 7，存在 f_1、……、$f_k \in \mathcal{F}^d$ 以及 w_1、……、$w_k \geq 0$，使得 $w_i f_i \leq w_i' I_i$ 且 $\int w_i' I_i - w_i f_i d\lambda \leq \varepsilon/(2n)$。下面的论述即可完成证明，

$$\begin{aligned}
\int p - \sum_{i=1}^{k} w_i f_i d\lambda &= \int p - \sum_{i=1}^{k} w_i' I_i + \sum_{i=1}^{k} w_i' I_i - \sum_{i=1}^{k} w_i f_i d\lambda \\
&= \left(\int p - \sum_{i=1}^{k} w_i' I_i d\lambda \right) + \left(\sum_{i=1}^{k} \int w_i' I_i - w_i f_i d\lambda \right) \\
&\leq \varepsilon/2 + n\frac{\varepsilon}{2n} = \varepsilon.
\end{aligned}$$

■

引理 9 设 $p:\ \mathbb{R}^d \to \mathbb{R}$ 为非负、紧支撑、有界且几乎处处连续的函数。存在一个非负序列 $(w_i)_{i=1}^{\infty}$ 和 $(f_i)_{i=1}^{\infty} \in \left(\mathcal{F}^d\right)^{\mathbb{N}}$ 使得 $p = \sum_{i=1}^{\infty} w_i f_i$。

引理 9 的证明 由引理 8 可知，存在 $q_1 \in$ 锥 $\left(\mathcal{F}^d\right)$，使得 $p \geq q_1$ 且 $\int p - q_1 d\lambda \leq 1/2$。注意，$p - q_1$ 是非负、紧支撑、有界且几乎处处连续的。通过归纳法，存在一个序列 $(q_i)_{i=1}^{\infty} \in$ 锥 $\left(\mathcal{F}^d\right)^{\mathbb{N}}$，满足 $\int p - \sum_{i=1}^{k} q_i d\lambda \leq 1/2^k$ 和 $p \geq \sum_{i=1}^{k} q_i$，对所有 k 均成立。由于 $\int p - \sum_{i=1}^{k} q_i d\lambda = \int \left| p - \sum_{i=1}^{k} q_i \right| d\lambda = \left\| p - \sum_{i=1}^{k} q_i \right\|_1$，该求和在 L^1 中是柯西序列，且 $\sum_{i=1}^{\infty} q_i = p$。由于 $q_i \in$ 锥 $\left(\mathcal{F}^d\right)$，可以写作 $q_i = \sum_{j=1}^{k_i} w_{i,j} f_{i,j}$，其中 $w_{i,j} \geq 0$、$f_{i,j} \in \mathcal{F}^d$。由此，$p = \sum_{i=1}^{\infty} \sum_{j=1}^{k_i} w_{i,j} f_{i,j}$。注意，$\|p\|_1 = \left\| \sum_{i=1}^{\infty} \sum_{j=1}^{k_i} w_{i,j} f_{i,j} \right\|_1 < \infty$，并且由于 $w_{i,j} f_{i,j}$ 是非负函数，因此可得

$$\infty > \left\| \sum_{i=1}^{\infty} \sum_{j=1}^{k_i} w_{i,j} f_{i,j} \right\|_1 = \sum_{i=1}^{\infty} \left\| \sum_{j=1}^{k_i} w_{i,j} f_{i,j} \right\|_1 = \sum_{i=1}^{\infty} \sum_{j=1}^{k_i} \left\| w_{i,j} f_{i,j} \right\|_1. \tag{3}$$

2. 这里用到的事实在单变量分析学教材中很常见。关于这些事实的多变量版本的完整论述，请参阅Edwards (1994) 第四章第2节和第3节。请注意，该著作中的术语有些非标准。

因此，所有 $w_{i,j}f_{i,j}$ 在 L^1中绝对可求和。于是，通过重新标记，存在 $w_i \geq 0$ 和$f_i \in \mathcal{F}^d$，使得 $\sum_{i=1}^{\infty} w_i f_i = p$。■

引理 9现在可以推广，以证明定理 2。

定理 2 设 p是几乎处处连续的概率密度函数。显然，可以将 p 分解为绝对可和的级数 $p = \sum_{i=1}^{\infty} p_i$，其中 p_i 均为非负、几乎处处连续且紧支撑的函数。例如，可以将 p 简单地拆分为支撑在 $\left\{\prod_{i=1}^{d} [a_i, a_i+1] \mid a_1,\dots,a_d \in \mathbb{Z}\right\}$ 上的分量，这构成一个可数集。我们只考虑该级数中的一个加项，因此固定某个 p_i。

现在证明 $p_i = \sum_{j=0}^{\infty} q_{i,j}$ 其中 $q_{i,j}$ 是有界、几乎处处连续、紧支撑的非负函数。为此，定义 $q_{i,j} = \max(\min(p_i - j, 1), 0)$。显然 $1 \geq q_{i,j} \geq 0$ 对所有 j 均成立，因此 $q_{i,j}$ 有界且非负。对于任意 x，若 $p_i(x) = 0$，则

$$q_{i,j}(x) = \max(\min(p_i(x) - j, 1), 0) = \max(\min(0 - j, 1), 0) = \max(0 - j, 0) = 0,$$

因此 $p_i(x) = 0 \Rightarrow q_{i,j}(x) = 0$，所以 $\operatorname{supp}(q_{i,j}) \subseteq \operatorname{supp}(p_i)$；又由于 p_i 的支撑是紧的，因此 $q_{i,j}$ 的支撑也是紧的。$q_{i,j}$ 几乎处处连续，这是因为最大值、最小值和减法都能保持逐点连续性。

现在证明 $\sum_{j=0}^{\infty} q_{i,j} = p_i$。任取一个 $x \in \mathbb{R}^d$。存在某个$\ell \in \mathbb{N} \cup \{0\}$ 使得 $p_i(x) \in [\ell, \ell+1]$。由此可得

$$\begin{aligned}
\sum_{j=0}^{\infty} q_{i,j}(x) =& \sum_{j=0}^{\ell-1} \max(\min(p_i(x) - j, 1), 0) + \max(\min(p_i(x) - \ell, 1), 0) \\
&\cdots + \sum_{j=\ell+1}^{\infty} \max(\min(p_i(x) - j, 1), 0) \\
=& \sum_{j=0}^{\ell-1} \max(1, 0) + \max(p_i(x) - \ell, 0) + \sum_{j=\ell+1}^{\infty} \max(p_i(x) - j, 0) \\
=& \ell + p_i(x) - \ell + \sum_{j=\ell+1}^{\infty} 0 = p_i(x).
\end{aligned}$$

由于 $q_{i,j}$ 均为非负数，沿用引理 9 的证明中式 (3) 的论证，可知和 $\sum_{i=1}^{\infty}\sum_{j=0}^{\infty} q_{i,j}$ 是 L^1 绝对可求和的。由引理 9 可知，对于所有 $q_{i,j} = \sum_{\ell=1}^{\infty} w_{i,j,\ell} f_{i,j,\ell}$，满足 $w_{i,j,\ell} \geq 0$ 以及 $f_{i,j,\ell} \in \mathcal{F}^d$ 和 p。代入后得到 $p = \sum_{i=1}^{\infty}\sum_{j=0}^{\infty}\sum_{\ell=1}^{\infty} w_{i,j,\ell} f_{i,j,\ell} = p$，它同样绝对可求和。由于该求和是在可数集上进行的，因此可以重新标记 $p = \sum_{i=1}^{\infty} w_i f_i$。由于 $f_i \in \mathcal{F}^d$、$f_i = \prod_{j=1}^{d} f_{i,j}$，其中 $f_{i,j}$ 是利普希茨连续的，因此可以构造一个序列 $(L_i)_{i=1}^{\infty}$，使得 $f_{i,j}$ 均为 L_i-利普希茨连续。最后注意到 $1 = \int \sum_{i=1}^{\infty} w_i f_i = \sum_{i=1}^{\infty} w_i$。由此可知，$p$ 的 NL谱为 (w, L)。■

3.2. 估计量界与收敛速率的证明

该估计量的结果建立在以下研究成果之上：Vandermeulen 和 Ledent (2021) 分析了 d维单位立方体上的密度直方图式估计量。在介绍这些结果之前，我们先补充一些

符号说明。在Vandermeulen 和 Ledent (2021) 中，作者定义了低秩直方图的概念。在此语境下，“直方图”指在由等距立方体组成的划分上保持常数的概率密度函数。具体而言，他们将 $\mathcal{H}_{1,\ b}$ 定义为单位区间上所有具有 b个等距分箱的直方图的集合，即

$$\mathcal{H}_{1,b} \triangleq \left\{ \sum_{i=1}^{b} w_i \mathbb{I}_{[(i-1)/b,i/b]} \mid w_i \geq 0, \int \sum_{i=1}^{b} w_i \mathbb{I}_{[(i-1)/b,i/b]} d\lambda = 1 \right\}.$$

A lo在 d 维空间上的 w-秩直方图，每个维度有 b 个分箱，定义为 s

$$\mathcal{H}_{d,b}^{k} \triangleq \left\{ \sum_{i=1}^{k} w_i \prod_{j=1}^{d} h_{i,j} \mid 0 \leq w_i, \sum_{i=1}^{k} w_i = 1, h_{i,j} \in \mathcal{H}_{1,b} \right\}.$$

令 Lip_L 表示从 $\mathbb{R}$ 到 $\mathbb{R}$ 的 L-利普希茨连续函数集合。

以下命题由Vandermeulen 和 Ledent （2021）证明，并给出了如何从 $\mathcal{H}_{d,b}^{k}$ 中选择估计量的界。

命题 10（Vandermeulen 和 Ledent （**2021**）中的命题 2.1）令 d、b、k、$n \in \mathbb{N}$ 和 $0 < \delta \leq 1$。存在一个估计量 $V_n \in \mathcal{H}_{d,b}^{k}$，使得

$$\sup_{p \in \mathcal{D}_d} P\left(\|p - V_n\|_1 > 3 \min_{q \in \mathcal{H}_{d,b}^{k}} \|p - q\|_1 + 7\sqrt{\frac{2bdk\log(4bdkn)}{n}} + 7\sqrt{\frac{\log(\frac{3}{\delta})}{2n}} \right) < \delta, \quad (4)$$

其中，V_n 是 X_1、.....、$X_n \overset{iid}{\sim} p$ 的函数。

开始证明定理 3 时，偏差项 $\min^{q \in \mathcal{H}_d^k}$ $_{,b}$ $\|p - q\|_1$通过 NL谱进行分析。

以下是秩一偏差的一个界，它简化了Vandermeulen 和 Ledent （2021）中的分析（参见该论文中的定理 2.5）。

引理 11 设 $f_1, \ldots, f_d \in \mathrm{Lip}_L \cap \mathcal{D}_1$。则 $\min^{h \in \mathcal{H}_{d,\ b}^1} \left\| \prod_{i=1}^{d} f_i - h \right\|_1 \leq \frac{dL}{\sqrt{12b}}$。

注意，由于集合 $\mathcal{H}_{d,\ b}^1$ 的紧致性，极小值点存在（参见 Vander- meulen 和 Ledent （2021）中的附录 A.1）。

引理的证明**11** 首先，

$$\begin{aligned}
&\min_{h \in \mathcal{H}_{d,b}^1} \left\| \prod_{i=1}^{d} f_i - h \right\|_1 \\
&= \min_{h_1,\ldots,h_d \in \mathcal{H}_{1,b}} \left\| \prod_{i=1}^{d} f_i - \prod_{j=1}^{d} h_j \right\|_1 \leq \min_{h_1,\ldots,h_d \in \mathcal{H}_{1,b}} \sum_{i=1}^{d} \|f_i - h_i\|_1 \quad (5) \\
&= \min_{h_1,\ldots,h_d \in \mathcal{H}_{1,b}} \sum_{i=1}^{d} \left\| \mathbb{I}_{[0,1]} \cdot (f_i - h_i) \right\|_1 \leq \min_{h_1,\ldots,h_d \in \mathcal{H}_{1,b}} \sum_{i=1}^{d} \left\| \mathbb{I}_{[0,1]} \right\|_2 \|f_i - h_i\|_2 \quad (6) \\
&\leq \min_{h_1,\ldots,h_d \in \mathcal{H}_{1,b}} \sum_{i=1}^{d} \frac{L}{\sqrt{12b}} = \frac{dL}{\sqrt{12b}}, \quad (7)
\end{aligned}$$

其中（5）是 Reiss（1989）引理 3.3.7 的推论（见附录 B），而（6）则由 H¨older 不等式得出。为说明（7），注意 Vandermeulen 和 Ledent（2021）的引理 C.2 和 C.4 表明，对于 $f_i \in \mathcal{D}_1 \cap \text{Lip}_L$,

$$\left\| f_i - \text{Proj}_{\text{span}(\mathcal{H}_{1,b})} f_i \right\|_2 \leq \frac{L}{\sqrt{12b}} \text{ and } \left\| f_i - \text{Proj}_{\text{span}(\mathcal{H}_{1,b})} f_i \right\|_2 = \min_{h \in \mathcal{H}_{1,b}} \|f_i - h\|_2 ,$$

其中投影采用 L^2 距离。这就完成了证明。 ■

在 Vandermeulen 和 Ledent （2021）中，作者研究了定义域为 $[0,1]^{\times d}$ 而非 $\mathbb{R}^d$ 的概率密度函数。这是一个相当微妙的问题，但这意味着密度 $p(x) = 2x_{\mathbb{I}[0,1]}(x)$ 在该论文中会是一个有效的 2-利普希茨连续密度，但在本文中则不是。该工作的结果适用于 $\mathcal{D}_d \cap \text{Lip}_L$ 中的密度。本文的结果可以调整为在定义域 $[0,1]^{\times d}$ 上稍微更一般的情形，但为简洁起见未作此处理。

下面的引理用 NL谱刻画偏差。

引理 12 设 $p \in \mathcal{D}_d$ 具有 NL谱（w, L），则

$$\min_{\mathcal{H}^k_{d,b}} \|p - h\|_1 \leq \frac{d}{\sqrt{12b}} \sum_{i=1}^{k} w_i L_i + 2 \sum_{i=k+1}^{\infty} w_i .$$

引理 **12** 的证明：由于 p 具有 NL谱（w, L），$p = \sum_{i=1}^{\infty} w_i \prod_{j=1}^{d} p_{i,j}$，其中 $p_{i,j} \in \mathcal{D}_d \cap \text{Lip}_{L_i}$，并且对所有 i均有 $w_i \geq 0$。令 $\tilde{p}_i = \arg\min_{h \in \mathcal{H}^1_{d,b}} \left\| \prod_{j=1}^{d} p_{i,j} - h \right\|_1$，且 $\tilde{w}_i = w_i / \left(\sum_{j=1}^{k} w_j \right)$（$\tilde{p}_i$ 存在，因为引理 11 中的最小值而非下确界——回忆定义 1 中的 $w_1 > 0$，因此$\tilde{w}_i$ 均存在）。所以 $\sum_{i=1}^{k} \tilde{w}_i \tilde{p}_i \in \mathcal{H}^k_{d,b}$。现在可以证明以下界:

$$\begin{aligned}
\left\| \sum_{i=1}^{\infty} w_i p_i - \sum_{j=1}^{k} \tilde{w}_j \tilde{p}_j \right\|_1 &\leq \left\| \sum_{i=1}^{k} w_i p_i - \sum_{j=1}^{k} \tilde{w}_j \tilde{p}_j \right\|_1 + \left\| \sum_{i=k+1}^{\infty} w_i p_i \right\|_1 \\
&= \left\| \sum_{i=1}^{k} w_i p_i - w_i \tilde{p}_i + w_i \tilde{p}_i - \tilde{w}_i \tilde{p}_i \right\|_1 + \sum_{i=k+1}^{\infty} w_i \\
&\leq \left\| \sum_{i=1}^{k} w_i p_i - w_i \tilde{p}_i \right\|_1 + \left\| \sum_{i=1}^{k} w_i \tilde{p}_i - \tilde{w}_i \tilde{p}_i \right\|_1 + \sum_{i=k+1}^{\infty} w_i . \qquad (8)
\end{aligned}$$

利用引理11，可得（8）左侧项的界

$$\left\| \sum_{i=1}^{k} w_i p_i - w_i \tilde{p}_i \right\|_1 \leq \sum_{i=1}^{k} w_i \|p_i - \tilde{p}_i\|_1 \leq \sum_{i=1}^{k} w_i \frac{d L_i}{\sqrt{12b}} = \frac{d}{\sqrt{12b}} \sum_{i=1}^{k} w_i L_i .$$

以下给出了式（8）中间项的界，从而完成证明，

$$\begin{aligned}
\left\| \sum_{i=1}^{k} w_i \tilde{p}_i - \tilde{w}_i \tilde{p}_i \right\|_1 &\leq \sum_{i=1}^{k} |w_i - \tilde{w}_i| \, \|\tilde{p}_i\|_1 = \sum_{i=1}^{k} \left| w_i - w_i / \left(\sum_{j=1}^{k} w_j \right) \right| \qquad (9) \\
&= \left(\sum_{i=1}^{k} w_i \right) \left| 1 - \left(\sum_{j=1}^{k} w_j \right)^{-1} \right| = \left| \left(\sum_{i=1}^{k} w_i \right) - 1 \right| = \sum_{i=k+1}^{\infty} w_i .
\end{aligned}$$

■

定理 3 的证明 根据命题 10，对于所有 $d, b, k, n \in \mathbb{N}$ 和 $0 < \delta \leq 1$，存在一个估计量 $V_n \in \mathcal{H}^k_{d,\ b}$ 使得

$$\sup_{\tilde{p} \in \mathcal{D}_d} P\left(\|\tilde{p} - V_n\|_1 > 3 \min_{q \in \mathcal{H}^k_{d,b}} \|\tilde{p} - q\|_1 + 7\sqrt{\frac{2bdk\log(4bdkn)}{n}} + 7\sqrt{\frac{\log(\frac{3}{\delta})}{2n}} \right) < \delta,$$

其中 V_n 是 X_1、.....、$X_n \overset{iid}{\sim} \tilde{p}$ 的函数。因此，对于这个 V_n，可得

$$P\left(\|p - V_n\|_1 > 3 \min_{q \in \mathcal{H}^k_{d,b}} \|p - q\|_1 + 7\sqrt{\frac{2bdk\log(4bdkn)}{n}} + 7\sqrt{\frac{\log(\frac{3}{\delta})}{2n}} \right) < \delta. \tag{10}$$

根据引理 12，$\min_{\mathcal{H}^k_{d,b}} \|p - h\|_1 \leq \frac{d}{\sqrt{12b}} \sum_{i=1}^{k} w_i L_i + 2\sum_{i=k+1}^{\infty} w_i$。因此

$$3 \min_{q \in \mathcal{H}^k_{d,b}} \|p - q\|_1 \leq 3\left(\frac{d}{\sqrt{12b}} \sum_{i=1}^{k} w_i L_i + 2 \sum_{i=k+1}^{\infty} w_i \right) \leq \frac{\sqrt{3}d}{2b} \sum_{i=1}^{k} w_i L_i + 6 \sum_{i=k+1}^{\infty} w_i.$$

将其代回式（10），即可得出该定理。 ■

引理 13 设 $p \in \mathcal{D}_d \cap \mathfrak{V}(W, \tilde{L})$，则 $\min_{\mathcal{H}^k_{d,b}} \|p - h\|_1 \leq \frac{d}{\sqrt{12b}} L_k + 2W_k$。

引理 **13** 的证明 由于 $p \in \mathfrak{V}(W, \tilde{L})$，$p$ 可写为 $p = \sum_{i=1}^{\infty} w_i \prod_{j=1}^{d} p_{i,j}$，其中 $w_i \geq 0$，$\sum_{i=k+1}^{\infty} w_i \leq W_k$ 对所有 k均成立，且 $p_{i,j} \in \mathcal{D}_d \cap \mathrm{Lip}^{L_i}$，其中 $L_i \leq L_k$ 对所有 $i \leq k$均成立。利用这个 NL谱，证明过程与引理12 证明式（9）的过程完全相同，从而得到

$$\min_{h \in \mathcal{H}^k_{d,b}} \|p - h\|_1 \leq \sum_{i=1}^{k} w_i \frac{dL_i}{\sqrt{12b}} + \sum_{i=1}^{k} \left| w_i - w_i / \left(\sum_{j=1}^{k} w_j \right) \right| + \sum_{i=k+1}^{\infty} w_i.$$

由于 $\tilde{L}$ 是非递减的，且 $w_i \geq 0$ 满足 $\sum_{i=1}^{k} w_i \leq 1$，因此可得

$$\sum_{i=1}^{k} w_i \frac{dL_i}{\sqrt{12b}} \leq \sum_{i=1}^{k} w_i \frac{d\tilde{L}_i}{\sqrt{12b}} \leq \sum_{i=1}^{k} w_i \frac{d\tilde{L}_k}{\sqrt{12b}} \leq \frac{d\tilde{L}_k}{\sqrt{12b}}.$$

使用与引理12 的证明中相同的论证可得

$$\sum_{i=1}^{k} \left| w_i - w_i / \left(\sum_{j=1}^{k} w_j \right) \right| + \sum_{i=k+1}^{\infty} w_i = 2 \sum_{i=k+1}^{\infty} w_i \leq 2W_k.$$

■

定理**5** 的证明与定理3的证明相同，其中使用引理13 来证明引理12。 ■

命题6 令 V_n 为定理5 中的估计量，其中 $k = \lceil n^{1/(3\alpha+\beta+1)} \rceil$ 且$b = \lceil n^{(\alpha+\beta)/(3\alpha+\beta+1)} \rceil$。注意，4式对 V_n也成立。令$p \in \mathcal{D}_d \cap \mathfrak{V}(W,\ L)$，其中 $W_k \in O\left(k^{-\alpha}\right)$ 且 $L_k \in O\left(k^{\beta}\right)$。根据定理5，可得以下界：

$$P\left(\|p - V_n\|_1 > \frac{\sqrt{3}d}{2b} L_k + 6W_k + 7\sqrt{\frac{2bdk\log(4bdkn)}{n}} + 7\sqrt{\frac{\log(\frac{3}{\delta})}{2n}}\right) < \delta.$$

将证明，不等式右侧的四项以 $\tilde{O}\left(n^{-\alpha/(3\alpha+\beta+1)}\right)$ 的速率趋于零。首先注意到 $k \in \Theta_n\left(n^{1/(3\alpha+\beta+1)}\right)$ 和 $b \in \Theta_n\left(n^{(\alpha+\beta)/(3\alpha+\beta+1)}\right)$。对于前两项，当 n 足够大时，

$$\begin{aligned}
\frac{\sqrt{3}d}{2b} L_k + 6W_k &\leq C\frac{k^{\beta}}{n^{(\alpha+\beta)/(3\alpha+\beta+1)}} + Ck^{-\alpha} && (C \text{ chosen sufficiently large}) \\
&\leq C'\frac{n^{\beta/(3\alpha+\beta+1)}}{n^{(\alpha+\beta)/(3\alpha+\beta+1)}} + C'n^{-\alpha/(3\alpha+\beta+1)} && (\text{substitute in } n;\ C' \text{ suf. large}) \\
&= C'n^{-\alpha/(3\alpha+\beta+1)} + C'n^{-\alpha/(3\alpha+\beta+1)}.
\end{aligned}$$

当 C 足够大时，第三项可界定为 $7\sqrt{\frac{2bdk}{n}\log(4bdkn)} \leq C\sqrt{\frac{bk}{n}}\log(4bdkn)$。对于第一项，当 n 足够大且 C 取值足够大时，

$$\begin{aligned}
\sqrt{\frac{bk}{n}} &\leq C\sqrt{\frac{n^{(\alpha+\beta)/(3\alpha+\beta+1)}n^{1/(3\alpha+\beta+1)}}{n}} = C\sqrt{\frac{n^{(\alpha+\beta)/(3\alpha+\beta+1)}n^{1/(3\alpha+\beta+1)}}{n^{(3\alpha+\beta+1)/(3\alpha+\beta+1)}}} \\
&= \sqrt{n^{-2\alpha/(3\alpha+\beta+1)}} = n^{-\alpha/(3\alpha+\beta+1)}.
\end{aligned}$$

由于 $4bdkn \in O_n(\text{poly}(n))$，可得 $\sqrt{\log(4bdkn)} \in O_n\left(\sqrt{\log(n)}\right)$。结合前述两个收敛速率可得

$$7\sqrt{\frac{2bdk\log(4bdkn)}{n}} \in O_n\left(n^{-\alpha/(3\alpha+\beta+1)}\sqrt{\log(n)}\right) \subset \tilde{O}_n\left(n^{-\alpha/(3\alpha+\beta+1)}\right).$$

最后 $7\sqrt{\frac{\log(\frac{3}{\delta})}{2n}} \in O_n(1/\sqrt{n})$。由于 $\alpha, \beta > 0$因此可得 $\alpha/(3\alpha+\beta+1) < 1/3$，并且

$$7\sqrt{\frac{\log(\frac{3}{\delta})}{2n}} \in O_n\left(n^{-1/2}\right) \subset O_n\left(n^{-\alpha/(3\alpha+\beta+1)}\right),$$

这就完成了本证明中关于收敛速率的部分。对于普遍一致性，将使用（4）。我们已经证明（4）中右侧的两个加数趋于零。以下引理同样出自 Vandermeulen 和 Ledent (2021)，并证明对于所有属于 $\mathcal{D}_d$ 的密度，偏差项都趋于零，从而完成证明。

引理 14（Vandermeulen 和 Ledent（2021）中的引理 2.1）设 $p \in \mathcal{D}_d$。若 $k \to \infty$ 且 $b \to \infty$，则 $\min_{q \in \mathcal{H}^{k,\ b}_{d,\ }} \|p - q\|_1 \to 0$。

■

致谢

本研究由德国联邦教育与研究部（BMBF）资助，资助对象为柏林学习与数据基础研究所（BIFOLD）（01IS18037A）。

参考文献

Elizabeth S. Allman、Catherine Matias 和 John A. Rhodes。含有大量观测变量的潜在结构模型中的参数可识别性。《统计年刊》，37(6A):3099–3132，2009 年 12 月。doi: 10.1214/09-AOS689。URLhttp://dx.doi.org/10.1214/09-AOS689。

Magda Amiridi、Nikos Kargas 和 Nicholas D. Sidiropoulos。低秩特征张量密度估计（第一部分）：基础。《IEEE信号处理汇刊》，70:2654–2668，2022。doi: 10.1109/TSP.2022.3175608。

Animashree Anandkumar、Rong Ge、Daniel Hsu、Sham M. Kakade 和 Matus Telgarsky。用于学习潜变量模型的张量分解。《机器学习研究杂志》，15: 2773–2832，2014。URLhttp://jmlr.org/papers/v15/anandkumar14b.html。

Bryon Aragam 和 Wai Ming Tai。《学习简单非参数混合模型难度的紧界》。arXiv 电子印本，文章编号 arXiv:2203.15150，2022 年 3 月。

Bryon Aragam 和 Ruiyi Yang。《非参数混合模型中的一致一致性》。arXiv 电子印本，文章编号 arXiv:2108.14003，2021 年 8 月。doi：10.48550/arXiv.2108.14003。

Bryon Aragam、Chen Dan、Eric P. Xing 和 Pradeep Ravikumar。非参数混合模型的可辨识性与贝叶斯最优聚类。《统计年刊》，48(4)：2277 – 2302，2020。doi：10.1214/19-AOS1887。URLhttps://doi.org/10.1214/19-AOS1887。

Sanjeev Arora、Rong Ge、Ravindran Kannan 和 Ankur Moitra。计算非负矩阵分解——具有可证明保证。在第四十四届 ACM计算理论研讨会论文集，STOC ’12，第 145–162 页，美国纽约州纽约，2012 年。ACM。ISBN 978-1-4503-1245-5。doi: 10.1145/2213977.2213994。URLhttp://doi.acm.org/10.1145/2213977.2213994。

Pierre Comon、Gene Golub、Lek-Heng Lim 和 Bernard Mourrain。《对称张量与对称张量秩》。SIAM 矩阵分析与应用期刊，30(3):1254–1279，2008。doi：10.1137/060661569。网址：http://dx.doi.org/10.1137/060661569。

Chen Dan、Liu Leqi、Bryon Aragam、Pradeep K Ravikumar 和 Eric P Xing。非参数混合模型下半监督学习的样本复杂度。载于 S. Bengio、H. Wallach、H. Larochelle、K. Grauman、N. Cesa-Bianchi 和 R. Garnett（编），《神经信息处理系统进展》第 31 卷。Curran Associates, Inc.，2018。

Sanjoy Dasgupta 和 Samory Kpotufe。k 近邻密度与众数估计的最优收敛速率。载于 Z. Ghahramani、M. Welling、C. Cortes、N. Lawrence 和 K.Q.Weinberger（编），《神经信息处理系统进展》第 27 卷。Curran Associates, Inc.，2014。

L. Devroye 和 G. Lugosi。《密度估计的组合方法》。施普林格，纽约，2001。

David Donoho 和 Victoria Stodden。非负矩阵分解何时能够正确地分解为各个部分？载于 S. Thrun、L. K. Saul 和 B. Sch¨olkopf 编，《神经信息处理系统进展16》，第 1141–1148 页。麻省理工学院出版社，2004。

C. Henry Edwards。 多元高级微积分。 Dover出版社，纽约，1994年。网址 http://www.loc.gov/catdir/description/dover032/94024204.html。原版出版信息：纽约：学术出版社，1973年。

L`aszl`o Gy¨orfi、Luc Devroye 和 Laszlo Gyorfi。《非参数密度估计：L^1视角》。约翰·威利父子公司，纽约；奇切斯特，1985年。

Heinrich Jiang。核密度估计的一致收敛速率。载于 Doina Pre- cup 和 Yee Whye Teh（编），第34届机器学习国际会议论文集，第70卷，机器学习研究论文集，第1694–1703页，澳大利亚悉尼国际会议中心，2017年8月6日至11日。PMLR。网址 http://proceedings.mlr.press/v70/jiang17b.html。

Nikos Kargas 和 Nicholas D. Sidiropoulos。《光滑乘积分布混合：可辨识性与算法》。载于 Kamalika Chaudhuri 和 Masashi Sugiyama 编，第二十二届国际人工智能与统计会议论文集，机器学习研究论文集第 89 卷，机器学习研究论文集，第 388–396 页。PMLR，2019年4月16日至18日。网址：https://proceedings.mlr.press/v89/kargas19a.html。

Wilfrid S. Kendall 和 Giovanni Montana。《小集与马尔可夫转移密度》。 随机过程及其应用，99(2):177–194，2002。ISSN 0304-4149。doi：https://doi.org/10.1016/S0304-4149(02)00090-X。网址：http://www.sciencedirect.com/science/article/pii/S030441490200090X。

Yong-Deok Kim 和 Seungjin Choi。非负Tucker分解。2007年 IEEE计算机视觉与模式识别会议，页码 1–8，2007年。

Joseph B. Kruskal。三阶数组：秩与三线性分解的唯一性，以及在算术复杂度和统计学中的应用。《线性代数及其应用》，18(2):95 – 138，1977年。ISSN 0024-3795。

Daniel D. Lee 和 H. Sebastian Seung。通过非负矩阵分解学习物体部件。《自然》，401(6755):788–791，1999年10月。ISSN 1476-4687。doi: 10.1038/44565。URLhttps://doi.org/10.1038/44565。

Sahand Negahban 和 Martin J. Wainwright。含噪声情况下（近似）低秩矩阵的估计与高维尺度。《统计年刊》，39(2):1069–1097，2011年4月。doi: 10.1214/10-AOS850。URL https://doi.org/10.1214/10-AOS850。

Sahand Negahban 和 Martin J. Wainwright。受限强凸性与加权矩阵补全：含噪声情形下的最优界。《机器学习研究杂志》，13:1665–1697，2012年5月。ISSN 1532-4435。URLhttp://dl.acm.org/citation.cfm?id=2188385.2343697。

Georgii S. Novikov、Maxim E. Panov 和 Ivan V. Oseledets。张量列车密度估计。载于 Cassio P. de Campos、Marloes H. Maathuis 和 Erik Quaeghe- beur（编），《第三十七届不确定性人工智能会议论文集》，UAI 2021，线上活动，2021年7月27日至30日，机器学习研究论文集第161卷，第1321–1331页。AUAI出版社，2021。URL https://proceedings.mlr.press/v161/novikov21a.html。Benjamin Recht、Maryam Fazel 和 Pablo A. Parrilo。通过核范数最小化求解线性矩阵方程的有保证最小秩解。《SIAM评论》，52(3):471–501，2010年8月。ISSN 0036-1445。doi: 10.1137/070697835。URL http://dx.doi.org/10.1137/070697835。R.D. Reiss。次序统计量的近似分布：及其在非参数统计学中的应用。Springer统计学丛书。施普林格，1989。ISBN 9783540968511。URL https://books.google.de/books?id=DxzvAAAAMAAJ。Alexander Ritchie、Robert A Vandermeulen 和 Clayton Scott。基于分组观测对可识别非参数混合模型进行一致估计。载于 H. Larochelle、M. Ranzato、R. Hadsell、M. F. Balcan 和 H. Lin（编），《神经信息处理系统进展》第33卷，第11676–11686页。Curran Associates, Inc.，2020。Meyer Scetbon 和 Marco Cuturi。低秩最优传输：近似、统计与去偏。载于 Alice H. Oh、Alekh Agarwal、Danielle Belgrave 和 Kyunghyun Cho（编），《神经信息处理系统进展》，2022。URL https://openreview.net/forum?id=4btNeXKFAQ。Meyer Scetbon、Marco Cuturi 和 Gabriel Peyr´e。低秩 Sinkhorn 分解。载于 Marina Meila 和 Tong Zhang（编），《第38届国际机器学习会议论文集》，机器学习研究论文集第139卷，第9344–9354页。PMLR，2021年7月18–24日。URL https://proceedings.mlr.press/v139/scetbon21a.html。Nicholas D. Sidiropoulos 和 Rasmus Bro。关于n维数组多线性分解的唯一性。《化学计量学杂志》，14(3):229–239，2000。B W Silverman。密度核估计及其导数的弱一致性与强一致性。《统计年刊》，6(1)，1978。Le Song 和 Bo Dai。多元分布的鲁棒低秩核嵌入。载于 C. J. C. Burges、L. Bottou、M. Welling、Z. Ghahramani 和 K. Q. Weinberger（编），《神经信息处理系统进展》第26卷。Curran Associates, Inc.，2013。Le Song、Animashree Anandkumar、Bo Dai 和 Bo Xie。多视图潜变量模型的非参数估计。载于 Eric P. Xing 和 Tony Jebara（编），《第31届国际机器学习会议论文集》，机器学习研究论文集第32卷，第640–648页，中国北京，2014年6月22日至24日。PMLR。URL https://proceedings.mlr.press/v32/songa14.html。Behrooz Tahmasebi、Seyed Abolfazl Motahari 和 Mohammad Ali Maddah-Ali。有限乘积测度有限混合的可识别性。arXiv电子预印本，文章编号 arXiv:1807.05444，2018年7月。doi: 10.48550/arXiv.1807.05444。

Alexandre B. Tsybakov。《非参数估计导论》。Springer 出版公司，第一版，2008。

L.W.Tu. 流形导论。Universitext。Springer 纽约，2010。ISBN 9781441973993。网址 https://books.google.de/books?id=br1KngEACAAJ。

Robert Vandermeulen 和 Clayton Scott。稳健核密度估计的一致性。载于 Shai Shalev-Shwartz 和 Ingo Steinwart 编，第 26 届学习理论年会论文集，第 30 卷，机器学习研究论文集，第 568–591 页，美国新泽西州普林斯顿，2013 年 6 月 12–14 日。PMLR。网址https://proceedings.mlr.press/v30/Vandermeulen13.html。

Robert A. Vandermeulen。利用张量分解改进非参数密度估计。arXiv 电子预印本，文章编号 arXiv:2010.02425，2020 年 10 月。doi：10.48550/arXiv.2010.02425。

Robert A Vandermeulen 和 Antoine Ledent。超越光滑性：将低秩分析纳入非参数密度估计。载于 M. Ranzato、A. Beygelzimer、Y. Dauphin、P.S. Liang 和 J. Wortman Vaughan（编），《神经信息处理系统进展》，第 34 卷，第 12180–12193 页。Curran Associates, Inc.，2021。

Robert A. Vandermeulen 和 Ren´e Saitenmacher。分组样本混合模型的广义可辨识性界。arXiv 电子预印本，文章编号 arXiv:2207.11164，2022 年 7 月。

Robert A. Vandermeulen 和 Clayton D. Scott。分组样本混合模型的可辨识性。arXiv 电子预印本，文章编号 arXiv:1502.06644，2015 年 2 月。doi：10.48550/arXiv. 1502.06644。

Robert A. Vandermeulen 和 Clayton D. Scott。《非参数混合模型的算子理论方法》。《统计年刊》，47(5):2704–2733，2019 年 10 月。doi: 10.1214/18-AOS1762。URLhttps://doi.org/10.1214/18-AOS1762。

附录 A。符号说明

$A \times B$	笛卡尔积
$f \times g$	$(f \times g)(a,b) = f(a)g(b)$
$f \cdot g$	$(f \cdot g)(a) = f(a)g(a)$
pdf	积分为一的非负函数
$\mathbb{N}$	$\{1,2,3,\ldots\}$
$\mathbb{I}^A$	A上的示性函数
$[n]$	$\{1,2,\ldots,n\}$
$\mathcal{D}_d$	支撑包含于 $[0,1]^{\times d}$ 的所有概率密度函数的集合
$\mathcal{F}$	$\mathbb{R}$ 上所有光滑的利普希茨连续概率密度函数的集合。
$\mathcal{F}^d$	$\{f_1 \times \cdots \times f_d \mid f_i \in \mathcal{F}\}$
锥（S）	$\{\sum_{i=1}^n w_i s_i \mid n \in \mathbb{N}, w_i \geq 0, s_i \in S\}$
$S^{\mathbb{N}}$	S 的所有无限序列的集合，即 $\{(s_i)_{i=1}^\infty \mid s_i \in S\}$
λ	勒贝格测度，维数未明确给出
Lip_L	定义域为 $[0,1]$ 且 L-利普希茨连续的实值函数集合
$\mathcal{H}_{1,b}$	$\left\{\sum_{i=1}^k w_i \mathbb{I}_{[(i-1)/b,i/b)} \mid w_i \geq 0, \int \sum_{i=1}^k w_i \mathbb{I}_{[(i-1)/b,i/b)} d\lambda = 1\right\}$
$\mathcal{H}_{d,b}^k$	$\left\{\sum_{i=1}^k w_i \prod_{j=1}^d h_{i,j} \mid w_i \in \Delta_k,\ h_{i,j} \in \mathcal{H}_{1,b}\right\}$

附录 B. 外部结果

Lem引理 15（Reiss（**Reiss (1989)**））对于概率测度 $\mu_i,\ \nu_i,\ i = 1,\ \ldots\ldots,\ d$,

$$\left\| \prod_{i=1}^d \mu_i - \prod_{i=1}^d \nu_i \right\| \leq \sum_{i=1}^d \|\mu_i - \nu_i\|,$$

其中范数为全变差范数。（注意，全变差范数与概率密度函数的 L^1 等价。）