arXiv:2603.05897v1 [math.ST] 6 Mar 2026

协变量偏移下非参数回归的极小极大理论

Petr Zamolodtchikov 比勒费尔德大学
pzamolod@math.uni-bielefeld.de

2026年3月9日

摘要

我们研究协变量偏移下的非参数回归问题，其中观测到来自目标分布和一个相关但不同的源分布的样本。我们引入一个新概念——迁移函数，并证明其定义域的性质决定了我们的极小极大速率。这些速率呈现出多种参数范围，包括经典速率，其由仅源域速率和仅目标域速率中较优者决定；此外，还包括收敛速率与样本量之间呈乘性交互、且快于两者择优基准的参数范围。我们证明，设计自适应估计量能够以相差至多对数因子的精度达到这些速率。与现有理论相比，我们的结果涵盖了协变量具有无界支撑的情形。

1 引言

迁移学习是一种广泛采用的方法，通过将相关数据纳入训练集来提升预测模型的性能。其核心前提是，不同数据集可能包含与当前任务相关的信息，模型能够提取并利用这些信息，因此多年来涌现出许多成功的应用。示例包括计算机视觉 [17, 32], 医疗保健 [1,11], 自然语言处理 [23], 和实验粒子物理学 [21, 8]。除了提升模型性能这一自然动机外，当相关数据稀缺时，迁移学习尤其值得关注。

在监督学习场景中，迁移学习旨在设计能够在目标分布 $Q_{X,\,Y}$ 上表现良好的算法，同时使用由源分布 $P_{X,\,Y}$ 中的 n个样本，或源分布中的 n 个样本与目标分布中的 m 个样本混合而成的数据集进行拟合。经典统计学习理论的大部分内容都是在训练集与测试集之间满足分布不变性的假设下推导出来的，也就是说 $P_{X,\,Y} = Q_{X,\,Y}$ 因此无法解释、预测并控制实际中的迁移学习。因此，迁移学习是一个活跃的研究领域 [36]。理论文献主要围绕由分布偏移引出的两种在数学上不同但互为补充的范式逐渐形成体系。第一种称为后验漂移，指协变量的边缘分布保持不变的情形，即 $P_X = Q_X$ 但输出的条件分布因任务而异，即 $P_{Y|X} \neq Q^{Y|X}$· 参见 [2, 7] 及其中的参考文献。本文讨论协变量偏移 (CS) [25, 27], ，这是第二种情形，此时只有协变量的边缘分布发生变化，即 $P_X \neq Q_X$ 而输出的条件分布保持相同，即 $P^{Y|X} = Q^{Y|X}$·

分析协变量偏移的主要理论目标有两方面。首先，人们旨在识别那个通过调节收敛速率来描述并量化可迁移性的正确对象。其次，人们

关键词：协变量偏移、极小极大、非参数回归、迁移学习、无界支撑。

旨在设计能够最优地适应分布偏移的算法。在本文中，可迁移性由迁移函数这一概念加以量化，它是一个依赖于(P_X,Q_X)的一元函数，其解析行为编码了可迁移性。更确切地说，后者在其定义域的边界处发散，而该边界的位置决定了可达到的极小极大速率。第二个目标在相差对数因子的意义下，由一个设计自适应 k-最近邻回归估计量实现。

迁移量化是协变量偏移文献中的核心议题。在线性模型中，迁移可自然地通过源协方差矩阵与目标协方差矩阵 [20, 26] 之间的相互作用来描述。针对经验风险最小化器（ERM），一种主流方法通过重新加权经验风险来处理协变量偏移，从而通过关于密度比 [19, 13, 9, 33,18, 10, 12] 的矩假设来刻画可迁移性。

我们现在扩展的这条研究脉络已不再采用密度比类型的假设，而是聚焦于源-目标对的几何型正则性。值得注意的是，在分类设定中，开创性工作 [16] 引入了迁移指数，它通过球质量比的增长来衡量源-目标对的相对奇异性；另见 [6]。同样地，在非参数回归背景下，α个族 的源-目标对由 [22] 和 [28] 引入。

在非参数设定中，典型的收敛速率形式为 $(n^{r_S} + m^{r_T})^{-1} \asymp (n^{r_S} \vee m^{r_T})^{-1}$，我们称之为楔形速率，并将其确定为源速率与目标速率中的最小值。楔形速率几乎总是对应于这样一种估计量的收敛速率：该估计量从分别仅使用 n 源样本和 m目标样本训练的两个独立估计量中选出表现最佳者。

在 α族的背景下，固定带宽的 Nadaraya-Watson 估计量是极小极大最优的。然而， [24], 表明，未惩罚的最小二乘估计量具有精细得多的迁移性质，与采用逐点最优或设计自适应带宽的 Nadaraya-Watson 估计量相当。这使幂分布到均匀分布迁移获得了更快的速率，其中 P_X 服从幂律，而 Q_X 服从均匀分布；同时也推动了对 [34], 中设计自适应估计量的研究，首个经过分析的是局部 k-NN。此外， [24]揭示了乘法速率的一个典型实例。近期预印本 [35] 针对一般光滑性条件下幂分布到均匀分布的协变量偏移，给出了积分收敛速率的下界，从而对乘法机制提供了更精细的认识。

这些观察揭示了现有极小极大理论在非参数回归中的协变量偏移方面存在的不足。首先， [22] 中的源-目标对类别包含非常不规则的分布。这些分布构成了困难情形，并导致相关理论将乘法速率等更精细的机制平滑掉。其次，一旦设计分布变得无界，迁移指数和 α族便不再适用。事实上，此时这两个量都会变为无穷大。

通过迁移函数这一方法，我们得到了一个简洁而富有表现力的协变量偏移下非参数回归极小极大理论。对于每一种出现的分布，结果都对应一个标量参数，用以决定极小极大速率。分析针对一类受限的源-目标对展开，这类源-目标对的局部质量具有足够良好的性质，从而使我们的速率能够捕捉迁移的微妙机制。特别是，我们的方法给出了丰富的参数范围图谱，并对其进行了清晰界定。如第 4.3 节所述，这种限制剥离出一种结构上稳健的清晰现象，而有原则地放宽限制只会拓展这一理论。困难情形是非常规整的帕累托分布源-目标对，这也是现实的极小极大理论应有的特点。无论协变量支撑集是否有界，迁移函数都具有明确意义。由于所有这些速率都是新的，我们的理论也为无界协变量支撑集下的非参数回归作出了贡献。

论文组织结构。 我们在第 2 节介绍统计模型、迁移函数及所采用的假设，随后在第 3 节陈述并讨论主要结果。第 4 节专门讨论我们对迁移函数的理解、估计量的设计，以及这些

所采用假设的必要性或缺失。所有证明均收录于附录。

记号。我们使用记号 $a_n \lesssim b_n$ 表示存在一个正常数 c，使得对所有 $n \in \mathbb{N}$，$a_n \le c b_n$。$a_n \gtrsim b_n$ 的定义类似；当 $a_n \lesssim b_n$ 且 $a_n \gtrsim b_n$. 时，我们记作 $a_n \asymp b_n$。对于任意 $x \in \mathbb{R}^d$ 和 $r > 0$，我们用 $\mathbb{B}(x, r)$ 表示以 x 为中心、半径为 r 的闭球，其中欧几里得范数记作 $\|\cdot\|$.。函数 f 的无穷范数记作 $\|f\|_\infty$.

2 预备知识

非参数回归模型。我们考虑标准的非参数回归模型。设 (X, Y) 是一对取值于 $\mathbb{R}^d \times \mathbb{R}$，的随机变量。我们假设存在一个函数 $f_*: \mathbb{R}^d \to \mathbb{R}$，满足

$$Y = f_*(X) + \varepsilon, \tag{1}$$

其中 ε 是一个中心化随机变量，且与 X. 独立。我们还假设 ε 服从参数为 α，$\nu > 0$. 的次指数分布。也就是说，对所有 $|\lambda| \le \alpha^{-1}$

$$\mathbb{E}[\exp(\lambda \varepsilon)] \le \exp\Big(\frac{\nu^2 \lambda^2}{2}\Big).$$

此外，我们假设 f_* 属于赫尔德球 $\mathcal{H}(L, \beta)$，其定义如下

$$\mathcal{H}(L, \beta) := \Big\{ f: \mathbb{R}^d \to \mathbb{R} : \|f\|_\infty + \sup_{\substack{x, y \in \mathbb{R}^d \\ x \ne y}} \frac{|f(x) - f(y)|}{\|x - y\|^\beta} \le L \Big\},$$

其中，$L > 0$ 和 $\beta \in (0, 1]$ 为固定常数。

协变量偏移。在协变量偏移 (CS) 的设定中，我们有来自目标分布 $\mathrm{Q}_{\mathrm{X},\,\mathrm{Y}}$ 的 m 个观测样本，以及来自源分布 $\mathrm{P}_{\mathrm{X},\,\mathrm{Y}}$ 的 n 个观测样本，并且它们在给定协变量的条件下相同，

$$\mathrm{P}_{\mathrm{Y}|\mathrm{X}} = \mathrm{Q}_{\mathrm{Y}|\mathrm{X}}. \tag{CS}$$

在 (CS) 条件下，如果 (X, Y) 在源分布（分别为目标分布）下满足模型 (1)，那么完全相同的模型也在目标分布（分别为源分布）下成立。这意味着，在 $\mathrm{P}_{\mathrm{X},\,\mathrm{Y}}$ 和 $\mathrm{Q}_{\mathrm{X},\,\mathrm{Y}}$. 下，回归函数 f_* 和噪声变量的分布相同。差异体现在边缘分布 P_{X} 和 Q_{X}.上

预测损失。给定来自 $\mathrm{P}_{\mathrm{X},\,\mathrm{Y}}$ 的独立同分布观测样本 $\mathcal{D}_{\mathrm{P}} = \{(X_i, Y_i)\}_{i=1}^n$，以及来自 $\mathrm{Q}_{\mathrm{X},\,\mathrm{Y}}$，的额外独立样本 $\mathcal{D}_{\mathrm{Q}} = \{(X'_j, Y'_j)\}_{j=1}^m$，其观测为独立同分布，预测目标是构造一个估计量 $f(\cdot, \mathcal{D}_{\mathrm{Q}}, \mathcal{D}_{\mathrm{Q}})$，使均方误差 $\ell(f) := \mathbb{E}_{\mathrm{Q}_{\mathrm{X},\,\mathrm{Y}}}[(\widehat{f}(X') - Y')^2 | \mathcal{D}_{\mathrm{Q}}, \mathcal{D}_{\mathrm{Q}}]$.最小。在模型 (1) 下，这等价于在 $L^2(\mathrm{Q}_{\mathrm{X}})$ 损失下估计 f_*

$$\ell(\widehat{f}) - \ell(f_*) = \mathbb{E}_{X' \sim \mathrm{Q}_{\mathrm{X}}}\big[(\widehat{f}(X') - f_*(X'))^2\big] = \big\|\widehat{f} - f_*\big\|^2_{L^2(\mathrm{Q}_{\mathrm{X}})}.$$

在模型 (1) 下，Y 的分布完全由协变量 X的分布, 回归函数 f_*, 以及 ε的分布, 所决定，这意味着指定 (X, Y) 的联合分布是多余的。因此，在本文其余部分，我们将 P_{X} 称为源分布，将 Q_{X} 称为目标分布。

迁移函数。在（CS）和模型（1）中，显然极小极大估计率取决于源-目标对 $(\mathrm{P}_{\mathrm{X}}, \mathrm{Q}_{\mathrm{X}})$，并且如第1节所述，关键步骤之一是识别并形式化

能够刻画这种依赖关系的量。这里，这种依赖关系通过迁移函数 及其定义域的行为来量化。首先，我们将 $\mathcal{M}$引入为 R^d 上具有勒贝格密度的博雷尔概率测度类。

定义 1（迁移函数）。 对于一对分布 P、$\mathrm{Q} \in \mathcal{M}$，以及实数 $\gamma \in [0, \infty)$，迁移函数定义为

$$\mathcal{T}(\mathrm{P}, \mathrm{Q}, \gamma) := \mathrm{E}_{X \sim \mathrm{Q}}[p(X)^{-\gamma}] \in [0, \infty],$$

其中， p 和 q 分别是P 和Q的密度。

直观地说， $\mathcal{T}$（P、Q $\cdot$）衡量了 Q 分配给 P 的低密度区域的质量。对于固定的 P、Q，迁移函数关于 γ 是一个对数凸函数，通常当其自变量达到定义域的边界时会发散至无穷大。这引出了可积性指标的定义。

定义 2（可积性指标）。 对于P、$\mathrm{Q} \in \mathcal{M}$，我们定义（P、Q）的可积性指标 as

$$\gamma^*(\mathrm{P}, \mathrm{Q}) := \sup\{\gamma \geq 0 : \mathcal{T}(\mathrm{P}, \mathrm{Q}, \gamma) < \infty\} \in [0, \infty].$$

迁移函数和可积性指标的更多性质将在第 4.1 节中讨论。

正则性。 我们的结果推导依赖于额外的正则性假设，这些假设限制了本文所研究的协变量分布类别。首先，记 $\mathcal{M}(D)$ 为 $\mathcal{M}$ 中勒贝格密度由常数 $D > 0$ 上界控制的测度类，其中该常数任意选定但固定。

定义 3。 令 $0 < \theta < \infty$. 。我们定义 $\mathcal{P}(D, \theta)$ 为分布类，其中分布 $\mathrm{P} \in \mathcal{M}(D)$ 的密度为 p，且对所有 $x \in$ P的支撑集 以及所有 $r \in (0, 1]$,

$$\theta^{-1} p(x) r^d \leq \mathrm{P}\{\mathrm{B}(x, r)\} \leq \theta p(x) r^d. \tag{2}$$

我们将 (2) 称为局部质量 假设。对于均匀分布，或密度有正下界和有限上界的分布，定义 3 均成立；这是非参数回归中最标准的设定之一。高斯分布等标准轻尾分布不满足 (2)，而帕累托分布和指数分布等重尾分布则满足该条件（见示例 7 和 8）。若 P 的密度 p 在 $x_0 \in \mathrm{R}^d$，处消失，且 $p(x)$ 在 x_0 的某个邻域内与 $\|x - x_0\|^\alpha$ 成正比，其中某个正数为 α，则 P 不满足 (2)。我们将在第 4.3 节进一步讨论定义 3 及其对结果的影响

3 主要结果

在本节中，我们先陈述上界和下界，再对所得收敛速率指数进行分析。我们在 (CS) 下的模型 (1) 设定中开展研究。我们固定 $D, \theta > 0$. 我们将始终用 $\mathrm{P_X}$ 表示源分布，用 $\mathrm{Q_X}$. 表示目标分布。因此，我们也简化记号，用 $\gamma^* := \gamma^*(\mathrm{P_X}, \mathrm{Q_X})$ 表示，并用 $s^* := \gamma^*(\mathrm{Q_X}, \mathrm{Q_X})$. 为便于阅读，我们也用 $r_\beta := \frac{2\beta}{2\beta + d}$. 表示

3.1 可学习性与协变量偏移的影响

在下文中，我们不再具体展开上界和下界中的乘法常数表达式，仅保留其对参数的关键依赖关系。更精确的陈述见附录 B，参见定理 29 和定理 32。现在陈述我们的主要结果。一方面，我们给出在类 $\mathcal{H}(L, \beta)$, 上的概率意义下上确界风险上界，该上界对任意源-目标对逐点成立。另一方面，我们给出在 $\mathcal{H}(L, \beta)$ 以及可迁移性指数有界的各类上的期望意义下上确界风险下界。

定理4（上界）。设 $\mathrm{P_X}$、$\mathrm{Q_X} \in \mathcal{P}$（$D$、$\theta$）。那么，存在一个估计量 f、一个整数$N = N$（θ、D）和一个常数 C（τ）$= C$（τ、d、β、θ、γ^*、s^*），使得对所有 $\tau > 1$、所有 $m \wedge n \geq N$、所有$\gamma \in [0, \gamma^*)$ 以及所有 $s \in [0, s^*)$，

$$\sup_{f_* \in \mathcal{H}(L,\beta)} \mathbb{P}\left\{ \left\| \widehat{f} - f_* \right\|^2_{L^2(\mathrm{Q_X})} \geq C(\tau)\mathcal{R}(n, m, \gamma, s) \right\} \leq 3(n^{1-\tau} + m^{1-\tau}),$$

其中，速率 $\mathcal{R} = \mathcal{R}$（$n$，$m$，$\gamma$，$s$）定义为

$$\mathcal{R} = \begin{cases} \mathcal{T}_{\mathrm{Q_X}}(\mathrm{P_X}, \gamma)^{\frac{r_\beta - s}{\gamma - s}} \mathcal{T}_{\mathrm{Q_X}}(\mathrm{Q_X}, s)^{\frac{\gamma - r_\beta}{\gamma - s}} \left(\frac{\log(nm)}{n} \right)^{\gamma \frac{r_\beta - s}{\gamma - s}} \left(\frac{\log(nm)}{m} \right)^{s \frac{\gamma - r_\beta}{\gamma - s}} & \text{if } (\gamma - r_\beta)(s - r_\beta) < 0 \\ & \text{and } m \in [n, n^{\gamma/s}], \\ \left[\mathcal{T}_{\mathrm{Q_X}}(\mathrm{P_X}, \gamma \wedge r_\beta) \left(\frac{\log(nm)}{n} \right)^{\gamma \wedge r_\beta} \right] \wedge \left[\mathcal{T}_{\mathrm{Q_X}}(\mathrm{Q_X}, s \wedge r_\beta) \left(\frac{\log(nm)}{m} \right)^{s \wedge r_\beta} \right] & \text{otherwise.} \end{cases}$$

这里结果的逐点性质是可取的：在满足 $\mathcal{P}(D,\ \theta)$ 所施加的正则性条件时，速率仅通过迁移函数取决于具体的源-目标对。需要注意的是，定理4中的估计量不依赖于源-目标对的具体选择，也不依赖于对可迁移性指数（$\gamma^*,\ s^*$）的任何了解。不过，它依赖于对 $2\beta/(2\beta + d)$ 的了解。事实上，我们设计了一种局部 k近邻回归器，并在第4.2节讨论其构造。定理4指出了关于（$\gamma,\ s,\ r_\beta$）的条件，在这些条件下可能出现特殊机制，使速率变为乘法形式。当（$\gamma - r_\beta$）（$s - r_\beta$）< 0时，就会出现这种情况。在此情形下，若再满足 $m \in [n, n^{\gamma/s}]$.，速率便变为乘法形式。后一个区间可理解为：若 $\gamma/s < 1$，则为 $[n^{\gamma/s}, n]$；当 $\gamma/s > 1$.时，则为 $[n, n^{\gamma/s}]$。因此，收敛速率取决于五个参数（γ^*，s^*，r_β，n，m）。

值得注意的是，在定理4中，$\mathrm{Q_X}$ 兼具数据生成过程（的一部分）和测试分布这双重角色，这一设定是任意的。仔细阅读证明可知，若改为考虑数据由来自 $\mathrm{P}^{(1)}\,\mathrm{x}$ 的 n 个样本和来自 $\mathrm{P}^{(2)}\,\mathrm{x}$ 的 m 个样本组成，同时仍将 $\mathrm{Q_X}$ 作为测试分布，同样的结果依然成立。此时，速率指数中的相关参数为 $\gamma^*(\mathrm{P}^{(1)}\,\mathrm{x}\ ,\mathrm{Q_X})$ 和 $\gamma^*(\mathrm{P}^{(2)}\,\mathrm{x}\ ,\mathrm{Q_X})$。这一观察为多任务估计 [5] 或领域泛化 [4]提供了自然的扩展方向。

现在我们陈述一个更专门化的结果，它形式上对应于定理4中的情形 $m = 0$，在该情形下，训练数据仅从源分布中抽取。

推论 5 (上界，纯迁移)。设 $\mathrm{P_X}$,$\mathrm{Q_X} \in \mathcal{P}(D, \theta)$。则存在一个估计量$\frown f$、一个整数 $N = N(\theta, D) \geq 2d \vee 3$, 以及一个常数 $C(\tau) = C(\tau, d, \beta, \theta, \gamma^*)$，使得对于所有 $\tau > 1$、所有 $n \geq N$ 以及所有 $\gamma \in [0, \gamma^*)$,

$$\sup_{f_* \in \mathcal{H}(L,\beta)} \mathbb{P}\left\{ \left\| \widehat{f} - f_* \right\|^2_{L^2(\mathrm{Q_X})} \geq C(\tau)\mathcal{T}(\mathrm{P_X}, \mathrm{Q_X}, \gamma \wedge r_\beta) \left(\frac{\log n}{n} \right)^{\gamma \wedge r_\beta} \right\} \leq 3n^{1-\tau}.$$

鉴于推论 5，可以注意到，在 $(\gamma^*,\ s^*,\ r_\beta,\ n,\ m)$ 这类速率并非乘法形式的配置中，定理4中的速率本质上是通过分别仅在源样本和仅在目标样本上训练的两个估计量中选出最佳者所得到的速率的最小值。

我们通过以下方式，为定理4的上界补充具有有界可迁移性指数的配对类别上的一致下界

$$\underline{\mathcal{P}}(D, \theta, s, \gamma) := \left\{ (\mathrm{P_X}, \mathrm{Q_X}) \in \mathcal{P}(D, \theta)^2 : \gamma^*(\mathrm{P_X}, \mathrm{Q_X}) \geq \gamma \ \text{ and } \ \gamma^*(\mathrm{Q_X}, \mathrm{Q_X}) \geq s \right\}.$$

根据此定义，我们得到类别 $\mathcal{P}(D,\ \theta,\ s,\ \gamma)$ 上极小极大风险的下界。

定理 6. 在（协变量偏移）条件下，考虑带有高斯噪声 $\varepsilon \sim \mathcal{N}(0,\ \sigma^2)$的回归模型(1)，其中 $\sigma^2 > 0$. 令 L、$D > 0$、 $\theta \geq$勒贝格测度(B(0，1)) $\beta \in (0,1]$ 且$(\gamma,\ s) \in (0,\ \infty) \times (0,\ 1)$。那么，存在一个与 n、m 无关的常数 $C_0 > 0$，使得对所有 $n \vee m > 1$,

$$\inf_{\widehat{f}} \sup_{\substack{f_* \in \mathcal{H}(L,\beta) \\ \mathrm{P_X}, \mathrm{Q_X} \in \underline{\mathcal{P}}(D,\theta,\gamma,s)}} \mathrm{E}\left[\|\widehat{f} - f_*\|^2_{L^2(\mathrm{Q_X})}\right] \geq \begin{cases} C_0 m^{-s\frac{\gamma - r_\beta}{\gamma - s}} n^{-\gamma\frac{r_\beta - s}{\gamma - s}} & if\ (\gamma - r_\beta)(s - r_\beta) < 0, \\ & and\ m \in [n, n^{\gamma/s}] \\ C_0\left[n^{-(\gamma \wedge r_\beta)} \wedge m^{-(s \wedge r_\beta)}\right] & else. \end{cases}$$

其中，下确界取遍所有由来自 $\mathrm{P_X}$ 的 个样本Y 和来自$\mathrm{Q_X}$ 的 个样本Y.得到的估计量

对于可迁移性指数固定（而非仅有下界）的源-目标对类别，同样的结论也一致成立。 $\theta \geq$ Leb(B(0,1))这一要求可以省略，因为根据勒贝格微分定理，对于 $\theta <$ Leb(B(0,1))， $\mathcal{P}(D,\theta)$ 为空。对于 $s \in$ (0,1)，使用帕累托源-目标对证明下界。诸如 $\gamma \wedge s = 0$ 或$s = 1$ 这样的边界情形，可以通过针对目标分布的替代构造来处理，但这会增加技术复杂性，故本文不作探讨。就形式化的下界构造而言， $s > 1$ 或 $\gamma = \infty$ 这两种情形不在本文讨论范围之内。

3.2 速率分析

如定理 4 和定理 6 所示，可迁移性指数决定了我们收敛速率中的指数。我们的上界和下界表述并不对称：上界中的指数针对任意 $0 \leq \gamma < \gamma^*$ 和 $0 \leq s < s^*$，而下界中的指数则用可迁移性指数表示。此外，上界还将迁移函数纳入乘法常数。这是自然的，因为下界的存在表明，对于大于 γ^* 和 s^*的指数，乘法常数必然会趋于无穷。迁移函数作为乘法常数出现，正体现了这一效应。在可迁移性指数处的奇异性类型满足某些条件时，可以将这些因素吸收到速率中，但需付出引入多对数因子的代价。我们将在第 4.1 节中对此作更详细的讨论；现在先分析第 3.1 节中得到的收敛速率。

以下各节的目标是描述上述收敛速率的结构、它们的过渡，以及由我们的主要结果产生的不同参数范围。这促使我们优先考虑可读性与清晰性。作为朝此方向迈出的第一步，我们引入术语，用以描述第 3.1 节中出现的不同情形。

配置。 当 $(\gamma - r_\beta)(s - r_\beta) < 0$. 时，我们将 (γ, s, r_β) 的配置称为超临界 。每当 $(\gamma - r_\beta)(s - r_\beta) = 0$（相应地，$(\gamma - r_\beta)(s - r_\beta) > 0$）时，我们将相同参数的配置称为（次）临界 。

参数范围。 我们将楔形区域 称为 $(\gamma, s, r_\beta, n, m)$ 的任何配置，其中速率与楔形速率 $n^{-(\gamma \wedge r_\beta)} \wedge m^{-(s \wedge r_\beta)}$成比例. 同样，加速区域 指 $(\gamma, s, r_\beta, n, m)$ 的任何配置，其中极小极大速率与加速速率 $n^{-\frac{\gamma r_\beta}{\gamma - s}} m^{-\frac{s(\gamma - r_\beta)}{\gamma - s}}$.成比例

为进一步简化后续第 3.2 节和第 3.3 节的表述，我们省略所有对数因子，并将乘法常数设为 1。特别地，我们省略迁移函数的乘法项。最后，我们将收敛速率中依赖于 n 的部分称为源速率 ，将依赖于 m 的部分称为目标速率。

3.2.1 相图

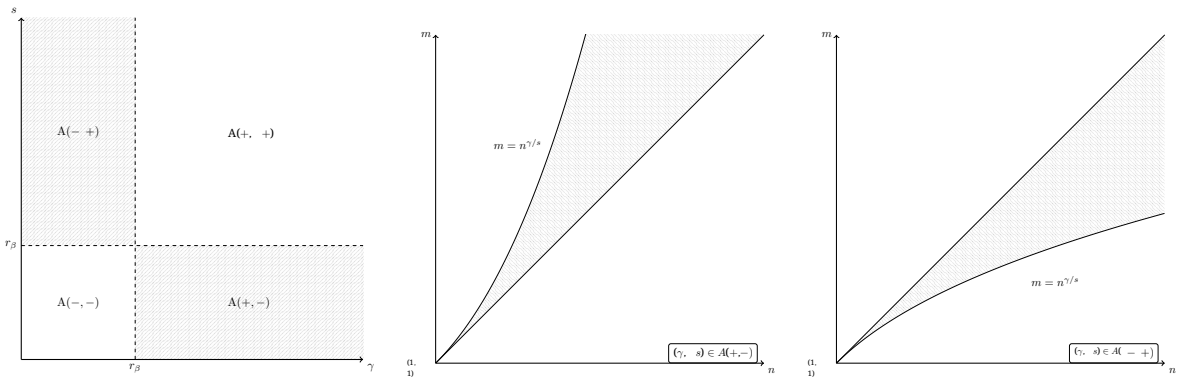

图1：**左图：**在（γ, s）坐标系中的相图。虚线区域对应超临界配置。**中图：**在（n, m）坐标系中的相图，其中（γ, s） $\in A$（$+, -$）。**右图：**在（n, m）坐标系中的相图，其中（γ, s） $\in A$（$-, +$）。虚线区域对应加速区域。

图1最左侧的面板展示了全局相图。迁移性指数空间被清晰地划分为四个区域。区域 A(–, –) ∪A(+, +) 表示亚临界配置的集合，在这些配置下，加速区域不会出现；而 A(–, +) ∪A(+, –) 表示超临界配置的集合，在这些配置下，加速区域可能出现，具体取决于 (n, m)。A(–, –) ∪A(+, +) 与 A(–, +) ∪ A(+, –) 之间的边界（虚线）表示临界配置的集合。在区域 A(–, –) 中，速率始终为 $n^{-\gamma} \wedge m^{-s}$；在 A(+, +) 中，速率始终为 $n^{-r_\beta} \wedge m^{-r_\beta}$.。固定 (γ, s) 后，可以用 (n, m) 坐标来观察相图。这便得到图1中央和最右侧的面板。它们描述的是 (γ, s) 固定且为超临界的情形。由于两个加速区域互斥，因此我们得到两张图。中央图对应于 (γ, s) ∈A(+, –)，这意味着 $\gamma/s > 1$；而最右侧的图对应于 (γ, s) ∈ A(–, +)，其中 $\gamma/s < 1$. 。在图1的中央图和最左侧图中，阴影区域对应加速区域。

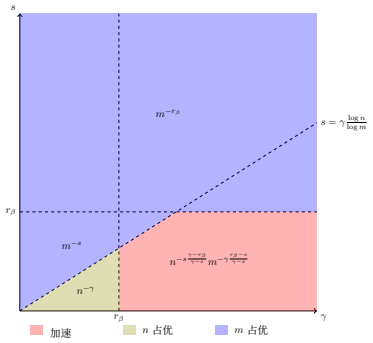

图2：固定（n, m）数对时（γ, s）相的细分。这里，$n < m$.

固定特定数对（n, m），并观察图1最左侧示意图由此产生的细分，也很有启发性。我们固定 $n < m$，得到图2。正如第 3.2.3 节将要看到的，速率会从一个参数范围平滑过渡到另一个参数范围。由于 $n < m$,唯一可能的加速区域是情形 $s < r_\beta < \gamma$,，对应于 A（$+, -$）。事实上，这是我们能够满足$m \in [n, n^{\gamma/s}]$. 的唯一方式。图2中，加速区域以红色区域表示，而图的其余部分对应楔形区域。在绿色区域中，楔形速率由源速率决定。在蓝色区域中，速率由目标速率决定。蓝色区域又细分为两个区域。在线 $s = r_\beta$ 下方，速率为 m^{-s}；在线上方，速率为 m^{-r_β}。可以看到，红色区域是 A（$+, -$）与直线 $s = \gamma \log n \log m$· 下方区域的交集。随着比值$\log n$ $\log m$ 增大并趋近于 1，红色区域的面积随之增大。当该比值大于 1, 时，图示会变为关于直线 $s = \gamma$.的镜像。

这是因为，如果 $n > m$，那么加速发生的唯一方式就是满足 $n^{\gamma/s} \leq m$，也就是说$\gamma < s$.因此，加速区域由 $\mathrm{A}(-, +)$ 与直线 $s = \gamma \log n \log m$· 上方区域的交集表示

至此，我们完成了对配置和参数范围的一般映射，并为接下来的章节奠定了基础。在后续章节中，我们将讨论并可视化 (γ, s) 和 (n, m) 的相变。 由于加速区间具有对称性，我们将聚焦于 $s < r_\beta < \gamma$. 这一情况

3.2.2 可迁移性指标的相变

在（log n，log m）坐标系中，加速区域的呈现效果更好。我们首先通过图 3 展示一种相变。具体过程如下：固定 γ 和 r_β，并以（log n，log m）坐标展示 s 从 $s \in$（r_β，γ）向$s \in$（0，r_β）平滑过渡的三个快照，或者等价地，（γ，s）从 A（+，+）过渡到 A（+，−）。图 3 展示了这一过渡的三个阶段。

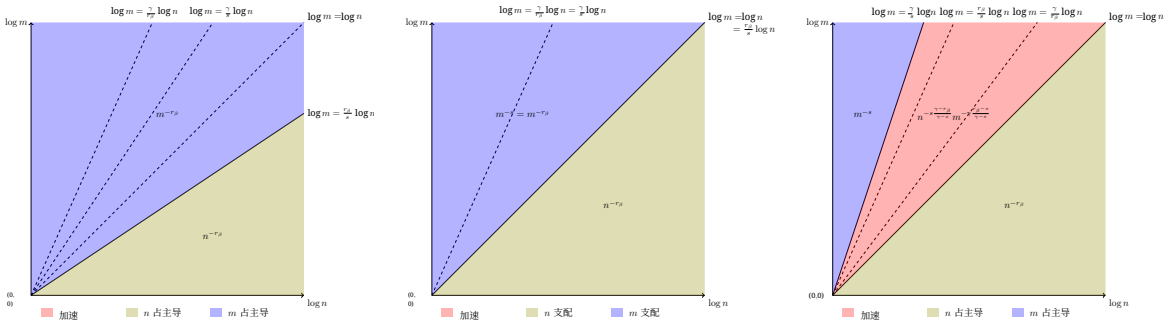

图 3：固定 γ/r_β 时，（log n，log m）坐标中的相变。**左：** $r_\beta < s < \gamma$，亚临界配置。**中：** $r_\beta = s < \gamma$，临界配置。**右：** $s < r_\beta < \gamma$，超临界配置。

图 3 最左侧的图对应于 $(\gamma,\ s) \in A(+,\ +)$。该速率为楔形速率。绿色区域表示速率由源速率驱动的数对 (log n，log m)，蓝色区域表示速率由目标速率驱动的数对。我们现在描述最左侧图上的实线与虚线。这两个区域之间的边界由直线 $s\log\ m = r_\beta \log\ n$ 给出，我们称之为 $a(s)$。直线 $s\log\ m = \gamma\log\ n$，我们称之为 $b(s)$。后者表示 $m = n^{\gamma/s}$，对应于加速区域在可能发生时的最大 m。相比之下，直线 $\log m = \log\ n$ 表示加速会发生的最小 m，并称为 (I)。直线 $r_\beta \log\ m = \gamma \log\ n$，记为 (M)，对应于使加速速率中的两个因子达到平衡的数对 $(n,\ m)$，并作为参考而标出。

当 s 逐渐减小至 r_β，$a(s)$ 会收缩至 (I)，而 $b(s)$ 会收缩至 (M)，从而得到图 3 中央的示意图。后者表示临界配置 $s = r_\beta$. 一旦$s < r_\beta$，我们便得到图 3 最右侧的示意图。在那里，配置 $(\gamma,\ s,\ r_\beta)$ 是超临界的。直线 $a(s)$ 和 (M) 位于由 (I) 与 $b(s)$ 所界定的红色锥体内。红色锥体是与加速区域相对应的 (log n，log m) 集合。

我们描述了在 $(\gamma,\ s)$。中跨越相变时加速区间的出现。另一种互补的视角是：固定处于超临界状态的 $(\gamma,\ s)$，并沿着源-目标样本量路径 $(n,\ m)$ 直观地分析速率。这引出了下一节。

3.2.3 样本规模路径

现在，我们固定图 3 最右侧图示的配置，并探索在 (log n,log m) 平面中，沿着 (n, m) 的两条路径变化时速率的演变。本节的一个主要目的，是直观地展示楔形速率与加速速率之间的过渡是平滑的，并且

在加速区域内，加速速率显著低于楔形速率。

鉴于速率的结构，我们在对数尺度上表示其演变。第一条路径称为线性路径，由满足 $\log\, n + \log\, m = \log\, nm = R.$ 的数对（n，m）给出。令$R = \log(B)$，这样的路径可由 $\lambda \in [0,1]$ 参数化为 $n = B^{1-\lambda}$ 和 $m = B^{\lambda}.$。第二条路径对应于固定预算路径，其中总样本量固定。即 $n + m = B,$ 或$\lambda \mapsto ((1-\lambda)B,\ \lambda B)$，且 $n \wedge m \geq 1.$。（γ，s）平面中的路径，以及沿这些路径对应速率的对数演变，见图4。

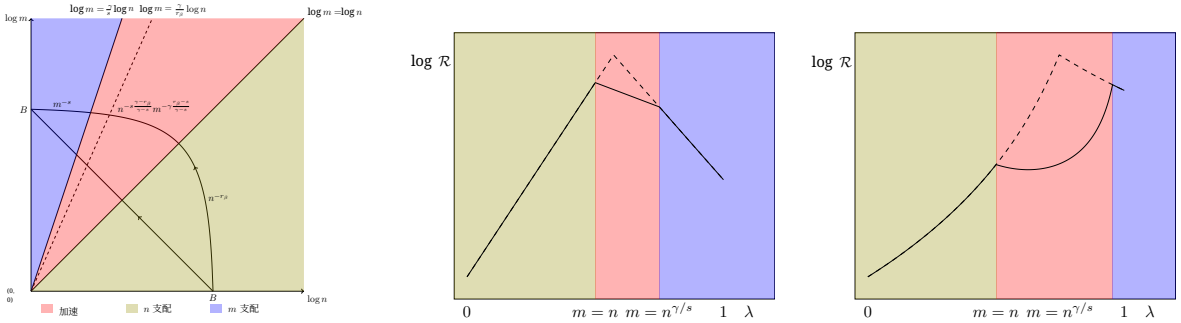

图4：**左：**（$\log\, n$，$\log\, m$）坐标中的两条路径。线性路径对应于（n，m）$= (B^{1-\lambda},\ B^{\lambda})$，而凹路径对应于（$n$，$m$）$= ((1-\lambda)B,\ \lambda B)$。**中：**沿线性路径的速率对数随 $\lambda.$ 的演变。**右：**沿固定预算路径的速率对数随 $\lambda.$ 的演变

在图4中间和最左侧的图中，点线对应楔形速率。我们可以清楚地看到，加速速率在边界点 $m = n$ 和 $m = n^{\gamma/s}$，处与楔形速率平滑衔接，而且加速速率低于楔形速率。通过将 $m^s = n^{u\gamma+(1-u)s}$ 参数化为 $u \in [0,1].$，也可以看出后一点。此时，加速速率变为$n^{-(r_\beta + u(\gamma - r_\beta))}$，对于所有 $u \in (0,\ 1)$，它都严格小于楔形速率。至此，我们对定理 4 和定理 6 的极小极大速率所呈现的相图的阐述告一段落。

3.3 启发式解释、相关工作和示例

我们首先给出一种启发式解释，说明加速速率如何以及为何出现。随后，我们将结合另一项获得了类似参数范围的研究来讨论我们的结果，并以源-目标对的具体示例所对应的速率作为结尾。

加速区域出现的启发式解释。 现在，我们以启发式的方式提取加速区域出现或未出现背后的机制。首先，我们选取一个能够优化偏差-方差权衡逐点 的估计量。 $x.$ 相应的步骤在此处并不重要，详见第4.2节。这排除了改善积分误差率的任何可能性。由此得到的逐点平方误差具有如下形式

$$(\widehat{f}(x) - f(x))^2 \leq (np(x) + mq(x))^{-r_\beta} \asymp (np(x) \vee mq(x))^{-r_\beta},$$

由此我们看到，偏差-方差机制施加了指数 r_β，并使其不可避免。特别地，它的出现并不依赖于迁移函数的任何可迁移性指数，而且即使放宽 $\mathcal{P}(D,\ \theta)$的正则性假设，这一性质仍然成立。我们现在引入一对可容许参数（γ，s）。为了推导超额风险界，下一步是将后者对$\mathrm{Q_X}.$ 积分。写出该积分，并利用对所有$\lambda \in [0,1]$ 以及 $a,\ b > 0,$ 成立的 $a \wedge b = \inf_{\lambda\in[0,1]} a^\lambda b^{1-\lambda}$，可得

$$\left\|\widehat{f} - f_*\right\|^2_{L^2(\mathrm{Q_X})} \leq \mathrm{E}_{\mathrm{Q_X}}\left[\inf_{\lambda\in[0,1]}\left\{(np(X))^{-\lambda r_\beta}(mq(X))^{-(1-\lambda)r_\beta}\right\}\right]. \tag{3}$$

使 λ 最小化的取值在 $mq(x) > np(x)$时为零，在 $mq(x) < np(x)$时为一。然而，为了使加速区域出现，我们转而引入一个参数 $\tau > 1$和一个子集 $E_\tau := \{x : \tau^{-1}mq(x) \le np(x) \le \tau mq(x)\}$，其中 $np(x)$与 $mq(x)$相当。对 $x \in E_\tau$，对任意 $\lambda \in [0,1]$成立$(np(x)) \wedge (nq(x)) \asymp (np(x))^{\lambda}(mq(x))^{1-\lambda}$，因此 λ 可被视为一个自由调节参数，而不是由 $np(x)$与 $mq(x)$之间的不平衡所施加。于是，我们可以把该积分分成三项。为此，记 $E_- := \{x : np(x) < \tau^{-1}mq(x)\}$，并记$E_+ := \{x : \tau mq(x) < np(x)\}$. 。这一点，连同此前关于 λ选取的考虑，导致

$$\begin{aligned}\left\|\widehat{f} - f_*\right\|^2_{L^2(\mathrm{Q_X})} \le \mathrm{E}_{\mathrm{Q_X}}\left[(mq(X))^{-r_\beta}\mathbb{1}_{E_-}(X)\right] + \mathrm{E}_{\mathrm{Q_X}}\left[(np(X))^{-r_\beta}\mathbb{1}_{E_+}(X)\right] \\ + \mathrm{E}_{\mathrm{Q_X}}\left[(np(X))^{-\lambda r_\beta}(mq(X))^{-(1-\lambda)r_\beta}\mathbb{1}_{E_\tau}(X)\right].\end{aligned} \tag{4}$$

因此，积分误差的量级将与（4）中的主导项相同。需要特别指出的是，第三项显然是带来加速速率的那一项；当所有样本都来自 $\mathrm{P_X}$ 或 $\mathrm{Q_X}$. 时，它不可能出现。直观而言，当 $\mathrm{Q_X}(E_\tau)$ 较大时，第三项可以 占主导地位；也就是说，在 R^d. 的一个足够大的子集上， $np(x)$ 与 $mq(x)$ 的“量级”相同。这说明，加速区域的出现取决于 n 和 m。严格来说，$\mathrm{Q_X}(E_\tau)$ 的大小也取决于 γ 和 s. 。不过，它们的作用可以更好地解释如下。由于参数 γ 和 s 与 $\mathrm{Q_X}$ 下 () 和 () 的矩是否存在有关，因此它们决定了（4）中期望内两项关于 的可积性。因此，它们完全决定了导致（或根据配置（, s）而不导致）加速速率的最优 λ 。至此，我们对加速速率及其出现方式的启发式解释就结束了。

相关工作。 我们现在将我们的结果与近期的工作 [35], 进行比较。据我们所知，该工作与 [24], 并列，是文献中少数报告协变量偏移问题乘法收敛速率的工作之一。

作者在非参数回归设置下展示了一个类似的结果，其中 $\mathrm{Q_X} = \mathcal{U}([0,1])$ 和 $\mathrm{P_X}$ 属于一个由幂律构成的单参数类，即对于 $a > 0$， $\mathrm{P_X}$ 的密度与 $x^{a-1}\mathbb{1}(x \in [0,1])$ 成正比。 尽管他们的速率以 a 和 r_β 表示，但我们可以计算出相应的可迁移性指数。我们得到 $\gamma^* = 1/(a-1)$ 和 $s^* = \infty$。用这些记号表示，他们的极小极大速率为

$$\inf_{\widehat{f}} \sup_{f_* \in \mathcal{H}(L,\beta)} \mathrm{E}\left[\left\|\widehat{f} - f_*\right\|^2_{L^2(\mathrm{Q_X})}\right] \gtrsim \begin{cases} n^{-r_\beta} \wedge m^{-r_\beta} & \text{if } \gamma \ge r_\beta \\ n^{-\frac{2\beta+1}{2\beta+a}} \wedge m^{-r_\beta} & \text{if } \gamma^* < r_\beta \text{ and } m \notin \left[n, n^{\frac{2\beta+1}{2\beta+a}}\right] \\ m^{-(r_\beta - \gamma^*)}n^{-\gamma^*} & \text{if } \gamma^* < r_\beta \text{ and } m \in \left[n, n^{\frac{2\beta+1}{2\beta+a}}\right]. \end{cases}$$

首先，我们观察到，我们对（超/次）临界配置的定义与他们的结果是一致的。事实上，由于 $s^* = \infty$，唯一的超临界配置是 $\gamma^* < r_\beta$。其次，由于局部质量假设，我们的设置不包含幂律行为。然而，启发式地将 s 取为无穷大，我们的加速速率会得到与 [35] 中相同的速率指数。这些对应关系表明，可迁移性指数具有某种普适性。超临界配置对应于 $r_\beta \in [\gamma^*, s^*]$，并且在 (n, m) 满足附加条件时，会得到形式为$m^{-a}n^{-b}$ 的加速速率，其中 $a + b = r_\beta$.

现在我们讨论结果中不一致的方面。在我们的结果中令 s^* 趋于无穷，会使加速区域取决于 $m \in [1, n]$, 而这通常远大于他们的区间 $[n^{(2\beta+1)/(2\beta+a)}, n]$. 。在（次）临界配置下，我们的楔形速率与他们一致，但在超临界非加速情形下则不一致。速率上的差异表明，

分布类别本身会同时影响速率以及发生加速时 $(n,\ m)$ 的取值范围。尽管很容易让人想这样做，但我们不会试图通过比较这两个结果来推断缺失的参数，尤其是因为针对显式密度计算的收敛速率往往会掩盖关键参数的识别，下面的示例将说明这一点。

示例。 现在我们给出使用帕累托分布和指数分布源-目标对得到的速率示例。这些示例的证明见附录 C。再次强调，我们省略对数因子，并将乘法常数设为 1。

示例 7 (帕累托分布源-目标对**)**。 我们考虑定义在 $\mathbb{R}$ 上的二参数帕累托分布族，记为 $\mathrm{Par}(\alpha,\sigma)$。对于 α, $\sigma>0$，$\mathrm{Par}(\alpha,\sigma)$ 的密度 r 定义为

$$r(x)=\frac{\alpha}{\sigma}\Big(1+\frac{x}{\sigma}\Big)^{-(\alpha+1)}\mathbb{1}(x\geq 0).$$

设 $\alpha_{\mathrm{P}},\alpha_{\mathrm{Q}},\sigma_{\mathrm{P}},\sigma_{\mathrm{Q}}>0$，并考虑 $\mathrm{P_X}=\mathrm{Par}(\alpha_{\mathrm{P}},\sigma_{\mathrm{P}})$ 和 $\mathrm{Q_X}=\mathrm{Par}(\alpha_{\mathrm{Q}},\sigma_{\mathrm{Q}})$，其密度分别为 p 和 q.。对于选取为 D,θ 的

$$D=(\alpha_{\mathrm{P}}/\sigma_{\mathrm{P}})\vee(\alpha_{\mathrm{Q}}/\sigma_{\mathrm{Q}}),\quad and\quad \theta=2\big(1+1/(\sigma_{\mathrm{P}}\wedge\sigma_{\mathrm{Q}})\big)^{\alpha_{\mathrm{P}}\wedge\alpha_{\mathrm{Q}}+1},$$

有 $\mathrm{P_X},\mathrm{Q_X}\in\mathcal{P}(D,\theta)$。此外，可迁移性指数为

$$\gamma^*=\gamma^*(\mathrm{P_X},\mathrm{Q_X})=\frac{\alpha_{\mathrm{Q}}}{\alpha_{\mathrm{P}}+1},\quad and\ s^*=\gamma^*(\mathrm{Q_X},\mathrm{Q_X})=\frac{\alpha_{\mathrm{Q}}}{\alpha_{\mathrm{Q}}+1}.$$

为确定速率中的指数，我们可以验证

$$s^*<r_\beta\Leftrightarrow\alpha_{\mathrm{Q}}<2\beta/d,\quad and\quad \gamma^*<r_\beta\Leftrightarrow\alpha_{\mathrm{P}}>\alpha_{\mathrm{Q}}/r_\beta.$$

若 $\alpha_{\mathrm{Q}}<2\beta/d$ 和 $\alpha_{\mathrm{P}}>\alpha_{\mathrm{Q}}/r_\beta$ 均成立，则楔形速率为

$$n^{-\frac{\alpha_{\mathrm{Q}}}{\alpha_{\mathrm{P}}+1}}\wedge m^{-\frac{\alpha_{\mathrm{Q}}}{\alpha_{\mathrm{Q}}+1}}.$$

加速速率出现的第一个必要条件是

$$\underbrace{\Big(\alpha_{\mathrm{Q}}<\frac{2\beta}{d}\quad and\quad \alpha_{\mathrm{P}}\leq\frac{\alpha_{\mathrm{Q}}}{r_\beta}\Big)}_{s^*<r_\beta<\gamma^*}\quad or\quad \underbrace{\Big(\alpha_{\mathrm{Q}}\geq\frac{2\beta}{d}\quad and\quad \alpha_{\mathrm{P}}>\frac{\alpha_{\mathrm{Q}}}{r_\beta}\Big)}_{s^*\geq r_\beta>\gamma^*}.\tag{5}$$

换言之，两个分布都必须具有足够重的重尾，或者都必须具有足够轻的尾部。加速速率出现的第二个必要条件是

$$\underbrace{n\leq m\leq n^{\frac{\alpha_{\mathrm{Q}}+1}{\alpha_{\mathrm{P}}+1}}}_{s^*<r_\beta\leq\gamma^*}\quad or\quad \underbrace{n^{\frac{\alpha_{\mathrm{Q}}+1}{\alpha_{\mathrm{P}}+1}}\leq m\leq n}_{s^*\geq r_\beta>\gamma^*}.$$

在这两种情况下，加速速率为

$$n^{-\frac{(\alpha_{\mathrm{Q}}+1)r_\beta-\alpha_{\mathrm{Q}}}{\alpha_{\mathrm{Q}}-\alpha_{\mathrm{P}}}}m^{-\frac{\alpha_{\mathrm{Q}}-(\alpha_{\mathrm{P}}+1)r_\beta}{\alpha_{\mathrm{Q}}-\alpha_{\mathrm{Q}}}}.$$

示例 8（指数源-目标对）。我们考虑定义在 R 上、记作 $\mathcal{E}(\lambda)$ 的指数分布族。对于 $\lambda>0$, $\mathcal{E}(\lambda)$ 的密度 r 定义为 $r(x)=\lambda e^{-\lambda x}\mathbb{1}(x\geq 0)$。令 $\lambda_{\mathrm{P}},\lambda_{\mathrm{Q}}>0$，并考虑具有相应密度 p 和 q. 的 $P_{\mathrm{X}}=\mathcal{E}(\lambda_{\mathrm{P}})$ 和 $Q_{\mathrm{X}}=\mathcal{E}(\lambda_{\mathrm{Q}})$。对于 $D=\lambda_{\mathrm{P}}\vee\lambda_{\mathrm{Q}}$ 和 $\theta=e^D$，有 $P_{\mathrm{X}},\ Q_{\mathrm{X}}\in\mathcal{P}(D,\theta)$。可迁移性指数为

$$\gamma^*=\gamma^*(\mathrm{P_X},\mathrm{Q_X})=\frac{\lambda_{\mathrm{Q}}}{\lambda_{\mathrm{P}}},\quad and\ s^*=\gamma^*(\mathrm{Q_X},\mathrm{Q_X})=1.$$

或者，我们可以说， s^* 可以取 $[0,1)$ 中的任意值。楔形速率为

$$n^{-\frac{\lambda_{\mathrm{Q}}}{\lambda_{\mathrm{P}}}\wedge r_\beta}\wedge m^{-r_\beta}.$$

加速速率唯一出现的情形是当 $\lambda_{\mathrm{Q}}\le r_\beta\lambda_{\mathrm{P}}$，且 $n^{\lambda_{\mathrm{Q}}/\lambda_{\mathrm{P}}}\le m\le n.$。特别是，$\mathrm{Q_X}$的尾部必须比$\mathrm{P_X}$. 重得足够多。此时，加速速率为

$$n^{-\lambda_{\mathrm{Q}}\frac{1-r_\beta}{\lambda_{\mathrm{P}}-\lambda_{\mathrm{Q}}}}m^{-\frac{r_\beta\lambda_{\mathrm{P}}-\lambda_{\mathrm{Q}}}{\lambda_{\mathrm{P}}-\lambda_{\mathrm{Q}}}}.$$

4 迁移函数、估计量与局部质量

本节将介绍支撑我们结果的技术细节。我们首先进一步研究第 2 节中引入的迁移函数的性质。接着，我们详细介绍在定理4和推论 5的上界中达到相应速率的估计量，最后讨论分布类 $\mathcal{P}(D,\ \theta)$。

4.1 迁移函数

本节展示第 2 节中定义的迁移函数的若干性质，以及它对定理 4 和定理 6 收敛速率的影响。我们将描述它作为 γ 的函数时的行为，并解释其衡量的内容。随后，我们聚焦于它的可迁移性指数，并将其与矩和 Rényi 熵等标准量联系起来。最后，我们讨论当 $\mathcal{T}$(P，Q γ) 接近 γ^*(P，Q) 时的爆发如何影响我们上界中的收敛速率。

对于本节 中，我们考虑两个任意分布 P 和 Q，其密度分别为 p和 q.让我们回顾迁移函数的定义如下

$$\mathcal{T}(\mathrm{P},\mathrm{Q},\gamma)=\mathrm{E}_{X\sim\mathrm{Q}}\left[p(X)^{-\gamma}\right].$$

迁移函数满足 $\mathcal{T}$(P，Q，0) $=1$，是对数凸的，并且 $\gamma\mapsto\mathcal{T}$(P，Q γ)$^\gamma$是非递减的（见附录 D 中的引理 37）。可迁移性指数 γ^*(P，Q) 是迁移函数不再有限并趋于无穷时对应的最终（可能为无穷）边界点。任何使迁移函数有限的 γ 都可以近似表示 Q 中落在 P 的低密度区域内的概率质量。 事实上，根据马尔可夫不等式，对于任意 $0\le\gamma<\gamma^*$(P，Q)，，有

$$\mathrm{Q}\{x:p(x)\le t\}\le t^\gamma\mathcal{T}(\mathrm{P},\mathrm{Q},\gamma).$$

类似地，迁移函数控制 $p(X)^{-1}$ 的截断矩。这种控制主要是技术性的，因此此处省略。不过，正是这种控制界定了第 3.3 节中式 (4) 的第三项，并得出加速速率。现在我们聚焦于迁移函数的可迁移性指数。

首先，只要 P $=$Q 具有无界支撑，我们便立即得到上界 γ^*(P，P) <1。 此外，广义矩的存在意味着可迁移性指数存在下界。

引理 9。 设 $X\sim\mathrm{P}\in\mathcal{M}$和 $\rho,\varepsilon>0.$若 $\mathrm{E}[\|X\|^{\rho+\varepsilon}]<\infty$，则 $\gamma^*(\mathrm{P},\ \mathrm{P})\ge\rho/(\rho+d)$。

当 P $\neq$ Q时，通过其他已知量来获得 γ^*(P,Q) 的下界会更加复杂。这里，我们根据 Rényi/Tsallis 散度的指标给出下界。对于 $\alpha>0,\ \ \alpha\neq1$，α-Tsallis 散度在 $\mathcal{M}$ 中 Q 相对于 P 定义为

$$D_\alpha^T(\mathrm{Q}\,\|\,\mathrm{P}):=\frac{1}{\alpha-1}\Big(1-\int q(x)^\alpha p(x)^{1-\alpha}\,dx\Big)=\frac{1}{\alpha-1}\,\mathrm{E}_{X\sim\mathrm{Q}}\left[\Big(\frac{q(X)}{p(X)}\Big)^\alpha\right].$$

Q 相对于 P 的 Rényi 散度定义为

$$D_\alpha(\mathrm{Q}\,\|\,\mathrm{P}) := \frac{1}{\alpha-1}\log\int_{\mathbb{R}^d} q(x)^\alpha p(x)^{1-\alpha}\,dx = \frac{1}{\alpha-1}\log\big((\alpha-1)(1-D_\alpha^T(\mathrm{Q}\,\|\,\mathrm{P}))\big).$$

对于相同的参数，两种散度都有限。对于某个参数而言，α. 的有限性 D_α(Q,P) 加上对 γ^*(P, P) 的了解，足以得到 γ^*(P,Q) 的下界。

引理 10。 设 P、Q $\in \mathcal{M}$ 和 $\alpha > 0, \alpha \neq 1$. 若 $D_\alpha(\mathrm{P}, \mathrm{Q}) < \infty$，则

$$\gamma^*(\mathrm{P},\mathrm{P})\Big(\frac{\alpha-1}{\alpha}\Big) \le \gamma^*(\mathrm{P},\mathrm{Q}).$$

Rényi/Tsallis 散度有限等价于密度比 q/p属于L^α(Q)，因此，引理 10 可以重新解释为密度比的可积性，并结合 p^{1-s}. 的可积性。引理 10 中的下界并不精确，因为对于某些特定的 (P,Q)，它可能会变得松弛。例如，可以取 P $=$Par(a, 1)，并令 Q $=\mathcal{E}$(1)。在这种情况下，对于所有 $\alpha > 1$， γ^*(P, Q) 都是无限的，而 D_α(Q ||P) 是有限的，并且 γ(P, P) $= a/(a+1)$。引理 10 的界并不精确，但建立了可积性条件与可迁移性指数之间的联系。

最后一个值得关注的问题是迁移函数在可迁移性指数处的行为。首先，回想一下， γ^* 表示 γ^*(P, Q)，很容易看出，取决于所考虑的概率分布对 (P， Q) $\mathcal{T}$(P， Q γ^*) 可能是有限的，也可能是无限的。

例 11。 设 p 和 q 分别是P 和Q的密度。设 a、$b > 0$ 和 $c \ge 0$，以及

$$p \propto \frac{\mathbb{1}(x \ge 2)}{x^{a+1}} \;\; and \;\; q \propto \frac{\mathbb{1}(x \ge 2)}{x^{b+1}\log(x)^c} \;\; so\ that\ \gamma^* = \frac{b}{a+1}.$$

可以验证，对所有 a、$b > 0$,

$$if\ c \le 1, \;\; then \;\; \mathcal{T}(\mathrm{P_X}, \mathrm{Q_X}, \gamma^*) = \infty, \;\; and \;\; \mathcal{T}(\mathrm{P_X}, \mathrm{Q_X}, \gamma^*) < \infty \;\; otherwise.$$

如果迁移函数在可迁移性指数处有限，那么下界所给速率中的指数是可达到的。如果迁移函数在可迁移性指数处为无限，那么下界中的指数能否达到，取决于 $\mathcal{T}$(P， Q ·) 在 γ^*. 处的发散速度。让我们聚焦于推论 5 的设定。假设 $\gamma^* < r_\beta$，我们可以将上界改写为：以高概率且相差一个 log n 因子，

$$\big\|\widehat{f} - f_*\big\|^2_{L^2(\mathrm{Q})} \le C\mathcal{T}(\mathrm{P},\mathrm{Q},\gamma)n^{\gamma^*-\gamma}n^{-\gamma^*} = Ce^{(\gamma^*-\gamma)\log n + \log\mathcal{T}(\mathrm{P},\mathrm{Q},\gamma)}n^{-\gamma^*}.$$

根据 $\mathcal{T}$(P， Q γ) 在 γ^*_-，处的行为，可能出现多种情形。例如，对于一个多项式极点，假设 $\mathcal{T}$(P， Q γ) $= 1/(\gamma^*-\gamma)^a$ 对某个 $a \ge 1$，那么取 $\delta_n := \gamma^* - \gamma_n = a/\log n$可得

$$\big\|\widehat{f} - f_*\big\|^2_{L^2(\mathrm{Q})} \le Ce^{a(1-\log(a))}(\log n)^a n^{-\gamma^*},$$

这表明，当迁移函数以多项式速度发散时，我们需要在收敛速率上付出一个多对数因子的代价，而 γ^* 仍可作为 n 的指数实现。

4.2 局部最近邻回归器

为获得相应的速率，我们考虑一个局部 k近邻回归器（k-NN）。令 $\mathcal{D}_\mathrm{P} := \{(X_i,\ Y_i)\}_{i=1}^n$和 $\mathcal{D}_\mathrm{Q} := \{(X_i',\ Y_i')\}_{i=1}^m$ 分别为来自 $\mathrm{P_{X,\ Y}}$ 和 $\mathrm{Q_{X,\ Y}}$.的两组 n 和 m 独立同分布观测。给定两个近邻函数 k_P 和 k_Q，，局部 k近邻估计量定义为

$$\widehat{f}(x) = \frac{1}{k_\mathrm{P}(x) + k_\mathrm{Q}(x)}\left[\sum_{i=1}^{k_\mathrm{P}(x)} Y_i(x) + \sum_{j=1}^{k_\mathrm{Q}(x)} Y_j'(x)\right], \tag{6}$$

其中， $Y_i(x)$ 和 $Y_i'(x)$ 分别表示在 $\{X_1,\dots,X_n\}$ 和$\{X_1',\dots X_m'\}$中， x 的第 i 个最近邻的标签，。函数 k_{P} 和 k_{Q} 分别针对每个数据集 $\mathcal{D}_{\mathrm{P}}$ 和 $\mathcal{D}_{\mathrm{Q}}$. 逐点选择，以平衡偏差和方差。具体而言，考虑一个仅依赖于 $\mathcal{D}_{\mathrm{P}}$. 的 k-NN回归器。随后可以证明，以高概率逐点地，平方误差可分解为

$$(\widehat{f}(x)-f_*(x))^2 \lesssim \underbrace{R_k(x)^{2\beta}}_{\text{squared-bias}} + \underbrace{\frac{\log n}{k}}_{\text{variance}},$$

其中，对数因子源于该结论具有高概率性质，而 $R_k(x) = \|x - X_k(x)\|_2^2$ 是 x 到其在 $\{X_1,\dots,X_n\}$. 中第 k 个近邻的距离。接着，将后者与其总体版本 $\zeta_k(x)$ 相关联: $=\inf\{r\ge 0 : \mathrm{P_X}\{\mathrm{B}(x,\ r)\}\ge k/n\}$，并利用分布的正则性假设，我们得到 $\zeta_k(x)\asymp (k/(np(x)))^{1/d}$，由此可得

$$(\widehat{f}(x)-f_*(x))^2 \lesssim \left(\frac{k_{\mathrm{P}}(x)}{np(x)}\right)^{2\beta} + \frac{\log n}{k_{\mathrm{P}}}.$$

通过选择 $k_{\mathrm{P}}^*(x)\asymp \log(n)^{d/(2\beta+d)}(np(x))^{2\beta/(2\beta+d)}$.，即可平衡平方偏差与方差。现在，由于分布 $\mathrm{P_X}$ 未知，我们改用插件近似。最后，我们需要对 $k_{\mathrm{P}}(x)$ 进行截断，使其保持在对数范围内 n，以保证该结论以高概率成立，且 n. 这就得到了

$$k_{\mathrm{P}}(x)\asymp \left\lceil \log(n)\vee \log(n)^{d/(2\beta+d)}(n\widehat{p}(x))^{2\beta/(2\beta+d)}\wedge n\right\rceil,$$

其中， $\widehat{p}$ 是 ℓ的 p最近邻密度估计器。其在 $x\in\mathbb{R}^d$，处的表达式为

$$\widehat{p}(x) := \frac{\mathrm{P}_n\{\mathrm{B}(x,R_\ell(x))\}}{R_\ell(x)^d} = \frac{\ell}{nR_\ell(x)^d}, \tag{7}$$

其中， P_n 是 $\{X_1,\dots,X_n\}$. 的经验分布。定理4和推论 5 中的结果是通过选取 $\ell\asymp \log((n\vee 1)(m\vee 1))$。这一选择得到的。选择 ℓ 会得到一个方差较高、偏差较低的密度估计量，这与我们的方法一致，因为我们只需使比率$p/\widehat{p}$ 在 $\mathbb{R}^d$的一个大子集内，对 x一致地保持在两个正常数之间。相同的步骤可用于从 $\mathcal{D}_{\mathrm{Q}}$ 中选择 k_{Q} ，并得到最终估计量 (6)。现在我们回到它的定义，并强调，即使近邻函数是分别选取的，也可以认为该估计量会在 $\mathcal{D}_{\mathrm{P}}$ 和 $\mathcal{D}_{\mathrm{Q}}$，之间选择信息量最大的样本。将 (6) 代入逐点误差 $(\widehat{f}(x)-f_*(x))^2$, 的界中，并重新整理可得

$$(\widehat{f}(x)-f_*(x))^2 \le \frac{k_{\mathrm{P}}(x)}{k_{\mathrm{P}}(x)+k_{\mathrm{Q}}(x)}\big(\widehat{f}_{\mathrm{P}}(x)-f_*(x)\big)^2 + \frac{k_{\mathrm{Q}}(x)}{k_{\mathrm{P}}(x)+k_{\mathrm{Q}}(x)}\big(\widehat{f}_{\mathrm{Q}}(x)-f_*(x)\big)^2, \tag{8}$$

其中， $\widehat{f}_{\mathrm{P}}$ 是仅在 $\mathcal{D}_{\mathrm{P}}$ 上训练的 k_{P}-近邻估计量，而 $\widehat{f}_{\mathrm{Q}}(x)$ 是仅在 $\mathcal{D}_{\mathrm{Q}}$. 上训练的 k_{Q}-近邻估计量。我们看到，逐点误差是 $\widehat{f}_{\mathrm{P}}$ 和 $\widehat{f}_{\mathrm{Q}}$. 的逐点误差的凸组合。注意，近邻函数 k_{P} 和 k_{Q} 关于底层密度是单调的。因此，在 p和 q， 之间不平衡程度较大的区域中，权重倾向于选择基于局部更密集样本得到的估计量。

4.3 局部质量假设

最后一部分，我们讨论 $\mathcal{P}(D,\ \theta)$。 中的分布的局部质量性质。这可以说是我们设定中限制性最强的假设。尽管如此，我们的受约束设定恰当地凸显了这一现象。

我们这一受限设定的一个特点是，它避免了极小极大速率由近奇异分布所主导。事实上，在 [16, 22, 28] 中，下界是沿着序列（在 n， m）导出的

这些测度类似于 [14, 第三部分], 中的测度，它们弱收敛到一个相对于勒贝格测度奇异的极限测度。相比之下，我们的下界是针对帕累托分布得到的。

尽管我们的主要结果以及第 4.1 节中的讨论表明，即使在更广泛的设定下，迁移函数也是值得考虑的自然对象，但局部质量条件并非如此。现在回顾第3.3节的研究结果。我们已经看到，将结果形式化地推广到幂律分布时，在一定程度上仍然有意义。特别是，我们刻画了加速区域中速率的结构。然而，我们的结果既未能再现楔形速率中的指数，也未能再现$n,\ m$ 中的阈值。这表明，可迁移性指数与其他量之间还存在进一步的相互作用。后者可能源自我们设定中未包含的分布形状。推导出对 (2) 的定量弱化条件超出了本文的范围，但这是我们这条研究路线的自然延续。另一个值得探索的方向是 $\mathrm{P_X}$ 与 $\mathrm{Q_X}$. 的支撑集之间存在维度不匹配的情形。可以在本文的同一通用框架内探讨这一问题，只需将 (2) 弱化为仅要求存在某些 $\theta > 0,$ 和 $0 < d' \le d,$

$$\theta^{-1} p(x) r^{d} \le \mathrm{P_X}\{\mathrm{B}(x,r)\} \le \theta p(x) r^{d}, \quad \text{and} \quad \theta^{-1} q(x) r^{d'} \le \mathrm{Q_X}\{\mathrm{B}(x,r)\} \le \theta q(x) r^{d'},$$

同时保持上界 $p \vee q \le D$. 原则上，自然可以预期，维度不匹配的情况能够通过对（2）作更一般的弱化来涵盖。我们将这一方向的精确处理留待未来工作。

参考文献

[1] Md Shahin Ali、Md Sipon Miah、Jahurul Haque、Md Mahbubur Rahman和Md Khairul Islam。“一种利用深度卷积神经网络与迁移学习模型进行皮肤癌分类的增强技术”。载于：*Machine Learning with Applications* 5（2021年），第100036页。issn: 2666-8270。doi: https : / / doi . org / 10 . 1016 / j . mlwa . 2021 . 100036。url: https : // www.sciencedirect.com/science/article/pii/S2666827021000177。[2] Arnab Auddy、T Tony Cai和Abhinav Chakraborty。“分布式差分隐私约束下非参数分类的极小极大与自适应迁移学习”。载于：*Journalof the Royal Statistical Society Series B: Statistical Methodology* （2025年11月），qkaf070。issn: 1369-7412。doi: 10 . 1093 /jrsssb / qkaf070。eprint: https : / / academic . oup . com / jrsssb / advance-article-pdf/doi/10.1093/jrsssb/qkaf070/65272455/qkaf070.pdf。url: https: // doi.org/10.1093/jrsssb/qkaf070。[3] Gérard Biau和Luc Devroye。近邻方法讲义。Springer数据科学丛书。Springer，Cham，2015年，第ix+290页。isbn: 978-3-319-25386-2。doi: 10.1007/978-3-319-25388-6。url: https://doi.org/10.1007/978-3-319-25388-6。[4] Gilles Blanchard、Aniket Anand Deshmukh、Ürun Dogan、Gyemin Lee和Clayton Scott。“通过边缘迁移学习实现领域泛化”。载于：*Journal of Machine Learning Research*22.1（2021年1月）。issn: 1532-4435。url: http://jmlr.org/papers/v22/17-679.html。[5] Gilles Blanchard、Jean-Baptiste Fermanian和Hannah Marienwald。“高维多个均值向量的估计”。载于：*arXiv* 预印本 *arXiv:2403.15038* （2024年）。[6] Changxiao Cai、T. Tony Cai和Hongzhe Li。“上下文多臂老虎机的迁移学习”。英语。载于：《统计年刊》 52.1（2024年），第207–232页。issn: 0090-5364。doi: 10.1214/23- AOS2341。url: projecteuclid.org/journals/annals-of-statistics/volume-52/issue-1/ Transfer-learning-for-contextual-multi-armed-bandits/10.1214/23-AOS2341.full。[7] T. Tony Cai、Dongwoo Kim和Hongming Pu。“函数均值估计的迁移学习：相变与自适应算法”。英语。载于：《统计年刊》52.2（2024年），第654–678页。issn: 0090-5364。doi: 10.1214/24-AOS2362.

[8] Benedetta Camaiani、Roberto Seidita、Lucio Anderlini、Rudy Ceccarelli、Vitaliano Ciulli、Piergiuli o Lenzi、Mattia Lizzo 和 Lorenzo Viliani。“利用领域适应进行标准模型截面的模型无关测量”。载于：《欧洲物理学杂志 *C*》82.10（2022年10月）。ISSN：1434-6052。DOI：10.1140/epjc/s10052- 022- 10871-3。网址：http://dx.doi.org/10. 1140/epjc/s10052-022-10871-3。[9] Yihang Chen、Fanghui Liu、Taiji Suzuki 和 Volkan Cevher。“协变量偏移下的高维核方法：数据依赖的隐式正则化”。载于：第*41*届国际机器学习大会论文集。ICML’24。奥地利维也纳：JMLR.org，2024年。[10] Andrea Della Vecchia、Arnaud Mavakala Watusadisi、Ernesto De Vito 和 Lorenzo Rosasco。“核岭回归在协变量偏移下的计算效率”。载于：*arXiv*预印本 *arXiv:2505.14083*（2025年）。[11] Andreas Ebbehoj、Mette Østergaard Thunbo、Ole Emil Andersen、Michala Vilstrup Glindtvad 和 Adam Hulman。“临床研究中非图像数据的迁移学习：范围综述”。载于：*PLOS Digital Health* 1（2022年2月），第1–22页。DOI：10.1371/journal.pdig.0000014。网址：https://doi.org/10.1371/journal.pdig.0000014。[12] Xingdong Feng、Xin He、Yuling Jiao、Lican Kang 和 Caixing Wang。“协变量偏移下的深度非参数分位数回归”。载于：*Journal of Machine Learning Research* 25.385（2024年），第1–50页。网址：http://jmlr.org/papers/v25/24-0906.html。[13] Davit Gogolashvili、Matteo Zecchin、Motonobu Kanagawa、Marios Kountouris 和 Maurizio Filippone。“协变量偏移适应何时需要重要性加权校正？”载于：*arXiv*预印本 *arXiv:2303.04020*（2023年）。[14] J.-P. Kahane。“关于完美线性集的三则札记”。法语。载于：*Enseign. Math. (2)* 15（1969年），第185–192页。ISSN：0013-8584。[15] Michael R. Kosorok。经验过程与半参数推断导论。Springer统计学丛书。Springer，纽约，2008年，第xiv页+483。ISBN：978-0-387-74977-8。DOI：10.1007/978-0-387-74978-5。网址：https://doi.org/10.1007/978-0-387-74978-5。[16] Samory Kpotufe 和 Guillaume Martinet。“边缘奇异性与协变量偏移中标签的益处”。载于：《统计年刊》49.6（2021年），第3299–3323页。ISSN：0090-5364,2168-8966。DOI：10.1214/21-aos2084。网址：https://doi.org/10.1214/21-aos2084。[17] Xuhong Li、Yves Grandvalet、Franck Davoine、Jingchun Cheng、Yin Cui、Hang Zhang、Serge Belongie、Yi-Hsuan Tsai 和 Ming-Hsuan Yang。“计算机视觉任务中的迁移学习：铭记来源”。载于：《图像与视觉计算》93（2020年），第103853页。ISSN：0262-8856。DOI：https://doi.org/10.1016/j.imavis.2019.103853。网址：https://www.sciencedirect.com/science/article/pii/S0262885619304469。[18] Ren-Rui Liu 和 Zheng-Chu Guo。“模型设定错误回归中的谱算法：协变量偏移下的收敛”。载于：*arXiv*预印本 *arXiv:2509.05106*（2025年）。[19] Cong Ma、Reese Pathak 和 Martin J. Wainwright。“最优应对基于再生核希尔伯特空间的非参数回归中的协变量偏移”。英语。载于：《统计年刊》51.2（2023年），第738–761页。ISSN：0090-5364。DOI：10.1214/23-AOS2268。[20] Neil Mallinar、Austin Zane、Spencer Frei 和 Bin Yu。“协变量偏移下的最小范数插值”。载于：第*41*届国际机器学习大会论文集。ICML’24。奥地利维也纳：JMLR.org，2024年。[21] Farouk Mokhtar、Joosep Pata、Dolores Garcia、Eric Wulff、Mengke Zhang、Michael Kagan 和 Javier Duarte。“为未来对撞机的新型探测器几何结构微调机器学习粒子流重建”。载于：*Phys. Rev. D* 111（2025年5月9日），第092015页。DOI：10.1103/ PhysRevD.111.092015。网址：https://link.aps.org/doi/10.1103/PhysRevD.111.092015。

[22] Reese Pathak、Cong Ma 和 Martin Wainwright。“协变量偏移的一种新相似性度量及其在非参数回归中的应用”。载于：第*39*届国际机器学习大会论文集。Kamalika Chaudhuri、Stefanie Jegelka、Le Song、Csaba Szepesvari、Gang Niu 和 Sivan Sabato 编。第162卷。机器学习研究论文集。PMLR，2022年7月，第17517–17530页。网址：https://proceedings.mlr.press/v162/pathak22a. html。[23] Sebastian Ruder、Matthew E. Peters、Swabha Swayamdipta 和 Thomas Wolf。“自然语言处理中的迁移学习”。载于：*2019*年北美计算语言学协会年会论文集：教程。Anoop Sarkar 和 Michael Strube 编。明尼阿波利斯，明尼苏达州：计算语言学协会，2019年6月，第15–18页。DOI：10.18653/v1/N19-5004。网址：https://aclanthology.org/N19-5004。[24] Johannes Schmidt-Hieber 和 Petr Zamolodtchikov。“非参数最小二乘估计量的局部收敛速率及其在迁移学习中的应用”。英语。载于：*Bernoulli*30.3（2024年），第1845–1877页。ISSN：1350-7265。DOI：10.3150/23-BEJ1655。网址：projecteuclid. org / journals / bernoulli / volume - 30 / issue - 3 / Local - convergence - rates - of - the - nonparametric-least-squares-estimator-with/10.3150/23-BEJ1655.full。[25] Hidetoshi Shimodaira。“通过加权对数似然函数改善协变量偏移下的预测推断”。载于：*J. Statist. Plann. Inference* 90.2（2000年），第227–244页。ISSN：0378- 3758,1873-1171。DOI：10.1016/S0378-3758(00)00115-4。网址：https://doi.org/10.1016/ S0378-3758(00)00115-4。[26] Yanke Song、Sohom Bhattacharya 和 Pragya Sur。“迁移学习中最小范数插值器的泛化误差”。载于：*arXiv*预印本 *arXiv:2406.13944* （2024年）。[27] Masashi Sugiyama、Matthias Krauledat 和 Klaus-Robert Müller。“通过重要性加权交叉验证实现协变量偏移适应”。载于：*Journal of Machine Learning Research* 8.35（2007年），第985–1005页。网址：http://jmlr.org/papers/v8/sugiyama07a.html。[28] Lukas Trottner。“马尔可夫设计下非参数回归中的协变量偏移”。载于：*Information and Inference: A Journal of the IMA*13.2（2024年5月），iaae011。ISSN：2049-8772。DOI：10.1093/imaiai/iaae011。电子版：https://academic.oup.com/imaiai/article-pdf/13/2/ iaae011/57420104/iaae011.pdf。网址：https://doi.org/10.1093/imaiai/iaae011。[29] Alexandre B. Tsybakov。非参数估计导论。Springer统计学丛书。根据2004年法语原版修订扩充，由Vladimir Zaiats翻译。Springer，纽约，2009年，第xii页+214。ISBN：978-0-387-79051-0。DOI：10 . 1007 / b13794。网址：https : // doi.org/10.1007/b13794。[30] Aad W. van der Vaart 和 Jon A. Wellner。弱收敛与经验过程。Springer统计学丛书。附统计学应用。Springer-Verlag，纽约，1996年，第xvi页+508。ISBN：0-387-94640-3。DOI：10.1007/978-1-4757-2545-2。网址：https://doi.org/10.1007/978-1-4757-2545-2。[31] Martin J. Wainwright。高维统计学。一种非渐近视角。2019年。DOI：10.1017/9781108627771。网址：https://doi.org/10.1017/9781108627771。[32] Jindong Wang 和 Yiqiang Chen。“计算机视觉中的迁移学习”。载于：迁移学习导论：算法与实践。新加坡：Springer Nature Singapore，2023年，第265–273页。[33] Shuntuo Xu、Zhou Yu 和 Jian Huang。“无界密度比估计：在协变量偏移下的误差控制中的应用”。载于：*arXiv*预印本 *arXiv:2504.01031* （2025年）。[34] Petr Zamolodtchikov 和 Hanyuan Hang。“协变量偏移下的迁移学习：重尾设计下的局部 k-最近邻回归”。载于：*arXiv*预印本 *arXiv:2401.11554*（2024年）。[35] Ling Zhou 和 Yuhong Yang。“关于非参数领域适应中的协同学习现象”。载于：*arXiv*预印本 *arXiv:2511.17009* （2025年）。

[36] Zhengyu Zhu、Yibo Yan、Gefei Li 和 Riquan Zhang。《统计迁移学习的最新进展》。载于：《国际统计评论》n/a.n/a（2025）。doi: https : / / doi. org/10.1111/insr.12613。电子版：https://onlinelibrary.wiley.com/doi/pdf/10.1111/ insr.12613。网址：https://onlinelibrary.wiley.com/doi/abs/10.1111/insr.12613。

A 初步结果

A.1 邻居上的一致控制

命题12。令 $\mathcal{B} := \{\mathrm{B}(x,r)\colon\ x\in\mathbb{R}^d, r>0\}$。$\mathcal{B}$ 的 *VC* 维数为 $2d+1$。此外，存在一个普适常数 K，使得对于任意 $\varepsilon\in(0,1)$，

$$\mathcal{N}(\mathbb{1}_{\mathcal{B}}, \|\cdot\|_{L_1(\mathrm{Q})}, \varepsilon) \le K(2d+1)(4e)^{2d+1}\varepsilon^{-2d}$$

对任意概率测度 Q均成立，其中 $\mathbb{1}_{\mathcal{B}} := \{\mathbb{1}(\cdot\in B)\ B\in\mathcal{B}\}$.

证明。关于 VC 维数的第一个结果可由 [30] 中的例 2.6.1 得出。关于覆盖数的第二个结果可由 [15] 中的定理 9.2 得出。 □

引理 13 (伯恩斯坦不等式**)**。设 $X_1,\dots,X_n$ 为取值于 $\mathbb{R}$ 的 n 独立同分布随机变量。若$\mathbb{P}\{|X_1|\le C\}=1$ 对某个 $C>0$ 成立，且 $\mathrm{E}[X_1]=\mu$，则对所有 $t>0$,

$$\mathbb{P}\left\{\Big|\frac{1}{n}\sum_{i=1}^n X_i-\mu\Big|\ge t\right\} \le 2\exp\left(-\frac{nt^2}{2\sigma^2+2Ct/3}\right), \tag{9}$$

其中 $\sigma^2=\mathrm{Var}(X_1)$。

定义 14. 对于$\mathrm{P}\in\mathcal{M}$和 $h>0$，我们将函数 ζ_h^{P} 定义为

$$\zeta_h^{\mathrm{P}}\colon x\mapsto \inf\Big\{r>0:\mathrm{P}\{\mathrm{B}(x,r)\}\ge h\Big\}. \tag{10}$$

当所考虑的分布P 可由上下文明确判断时，我们将在记号中省略P指数，改写为 ζ_h 。

接下来，我们证明到第 k近邻的距离的一致界。设 $\mathrm{P}\in\mathcal{M}$, $n\ge 2d$, $X_1,\dots,X_n$ 是服从分布 P 的独立同分布随机变量。记 $R_k(x)$为x 到 $\{X_1,\dots X_n\}$. 中第 k个最近邻的距离。设 $\delta,h\in(0,1)$。

命题 15. 设 δ、$h>0$和 $K_0=2K(2d+1)(4e)^{2d+1}$，其中 K 是命题*12*中定义的绝对常数。以下两个断言成立。

(i) 若 $\lambda\in(0,1)$ 且整数 n 满足

$$n\ge\frac{4d(1+\lambda/3)}{\lambda^2 h}\log\left(\Big(\frac{K_0}{\delta}\Big)^{1/(2d)}\frac{2}{(1-\lambda)h}\right), \tag{11}$$

then

$$\mathbb{P}\left\{\inf_{x\in\mathbb{R}^d}\frac{1}{nh}\sum_{i=1}^n\mathbb{1}(X_i\in\mathrm{B}(x,\zeta_h(x)))<\frac{(1-\lambda)}{2}\right\}\le\delta.$$

(ii) 若 $\lambda\in\mathbb{R}_+^*$，且整数 n 满足

$$n\ge\frac{4d(1+\lambda/3)}{\lambda^2 h}\log\left(\Big(\frac{K_0}{\delta}\Big)^{1/(2d)}\frac{1}{(1+\lambda)h}\right), \tag{12}$$

则，

$$\mathbb{P}\left\{\sup_{x\in\mathbb{R}^d}\frac{1}{nh}\sum_{i=1}^n\mathbb{1}(X_i\in\mathrm{B}(x,\zeta_h(x)))>2(1+\lambda)\right\}\le\delta.$$

证明。设 $z^m := (z_1, \ldots\ldots, z_m) \in \mathrm{R}^{d\times m}$ 为 m 个不同点的任意向量，记 $S(x, h) := \mathrm{B}(x, \zeta_h(x))$，并定义集合

$$\mathcal{I}(h) := \Big\{ \mathbb{1}(\cdot \in S(x,h)) \Big\}.$$

考虑事件

$$A_n(t, z^m) := \bigcap_{j=1}^{m} \Big\{ \Big| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\big(X_i \in S(z_j,h)\big) - h \Big| \le t \Big\}.$$

令 $X \sim \mathrm{P}$。对于所有 $x \in \mathrm{R}^d$，均有

$$(i)\ \mathrm{E}[\mathbb{1}(X \in S(x,\ h))] = \mathrm{P_X}\{S(x,\ h)\} = \mathrm{P_X}\{\mathrm{B}(x,\ \zeta_h(x))\} = h (ii)\ \mathrm{Var}(\mathbb{1}(X \in S(x,\ h))) = h(1-h) \le h. (iii)\ 0 \le \mathbb{1}(X \in S(x,\ h)) \le 1.$$

根据并集界和伯恩斯坦不等式（9），

$$\mathbb{P}\big\{A_n(t, z^m)^c\big\} \le \sum_{j=1}^{m} \mathbb{P}\Big\{ \Big| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\big(X_i \in S(z_j,h)\big) - h \Big| \ge t \Big\} \le 2m \exp\Big(- \frac{nt^2}{2h + 2t/3} \Big). \tag{13}$$

令 $\mathrm{P}_n := n^{-1} \sum_{i=1}^{n} \delta_{X_i}$ 为样本 X_1、……、X_n 的经验分布，并令 $\varepsilon \in (0, 1)$。应用命题12可确保存在一个覆盖 $(\mathcal{B}, \|\cdot\|_{L_1(\mathrm{P}_n)})$ 的 ε-网，其基数有限，为 $M = M(X^n)$，且 $\|M\|_\infty \le K_0 \varepsilon^{-2d}$。由于 $\mathcal{I}(h) \subseteq \mathcal{B}$，存在一个覆盖 $(\mathcal{I}(h), \|\cdot\|_{L_1(\mathrm{P}_n)})$ 的 ε-网，其基数有限（但为随机变量） $m = m(X^n)$，且上界为 $\|M\|_\infty$。令 $Z_1, \ldots Z_m$ 使得该 ε-网为集合 $\big\{S(Z_j, h) 1 \le j \le m\big\}$。由于 m 的上界不依赖于样本，因此将 z^m 替换为 Z^m. 时，上界（13）仍然成立。因此，利用命题12的覆盖界可得

$$\mathbb{P}\big\{A_n(t, Z^{\overline{m}})^c\big\} \le K_0 \varepsilon^{-2d} \exp\Big(- \frac{nt^2}{2h + 2t/3} \Big).$$

我们现在考虑 $A_n(t,\ Z^m)$。Z_j 的 ε-网性质保证，对所有 $x \in \mathrm{R}^d$，存在 $j \in [m]$，使得

$$\big\| \mathbb{1}(\cdot \in S(x,h)) - \mathbb{1}(\cdot \in S(Z_j,h)) \big\|_{L_1(\mathrm{P}_n)} \le \varepsilon,$$

并且，在 $A_n(t, Z^m)$，由此可得

$$\Big| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \in S(x,h)) - h \Big| \le \big\| \mathbb{1}(\cdot \in S(x,h)) - \mathbb{1}(\cdot \in S(Z_j,h)) \big\|_{L_1(\mathrm{P}_n)} + \Big| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \in S(Z_j,h)) - h \Big|,$$

其中，第一项的上界为 ε，而根据 $A_n(t,\ Z^m)$ 的定义，第二项的上界为 t.。点 $x \in \mathrm{R}^d$ 在 R^d 中任意选取，因此，

$$\sup_{x \in \mathbb{R}^d} \Big| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \in S(x,h)) - h \Big| \le \varepsilon + t. \tag{14}$$

现在我们继续证明第一个结论。特别地，由（14）可知

$$\inf_{x\in\mathbb{R}^d}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(X_i\in S(x,h))\geq h-\varepsilon-t. \tag{15}$$

令 δ，$\lambda\in(0,1)$。我们选取 $t=\lambda h$，并且

$$\varepsilon=\delta^{-1/(2d)}K_0^{1/(2d)}\exp\Big(-\frac{n\lambda^2 h}{4d(1+\lambda/3)}\Big).$$

可以验证，对于所有满足以下条件的 n

$$n\geq\frac{4d(1+\lambda/3)}{\lambda^2 h}\log\Big(\frac{2}{(1-\lambda)h}\Big(\frac{K_0}{\delta}\Big)^{1/(2d)}\Big),$$

有 $\varepsilon\leq(1-\lambda)h/2$. 最后，将这些值代入 (15) 可得，在 $A_n(t,\ Z^m)$，

$$\inf_{x\in\mathbb{R}^d}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(X_i\in S(x,h))\geq\frac{(1-\lambda)h}{2}.$$

这证明了

$$\mathbb{P}\left\{\inf_{x\in\mathbb{R}^d}\frac{1}{nh}\sum_{i=1}^{n}\mathbb{1}(X_i\in S(x,h))<\frac{(1-\lambda)}{2}\right\}<\mathbb{P}\big\{A_n(t,Z^{\overline{m}})^c\big\},$$

并且可以验证 $\mathbb{P}\big\{A_n(t,\ Z^m)^c\big\}\leq\delta$，由此完成了第一个结论的证明。现在我们证明第二个结论，它是对第一个结论证明的略微变体。(14) 的一个推论是

$$\sup_{x\in\mathbb{R}^d}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(X_i\in S(x,h))\leq h+\varepsilon+t.$$

接着，我们按之前的方式选取 $t=\lambda h$，使得 $\lambda\in(0,\ \infty)$，并选取 ε。特别地，对于所有

$$n\geq\frac{4d(1+\lambda/3)}{\lambda^2 h}\log\left(\frac{1}{(1+\lambda)h}\left(\frac{K_0}{\delta}\right)^{1/(2d)}\right),$$

我们有 $\varepsilon\leq(1+\lambda)h$. 因此，对于这样的 n,

$$\sup_{x\in\mathbb{R}^d}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(X_i\in S(x,h))\leq 2(1+\lambda)h,$$

这证明了

$$\mathbb{P}\left\{\sup_{x\in\mathbb{R}^d}\frac{1}{nh}\sum_{i=1}^{n}\mathbb{1}(X_i\in S(x,h))<2(1+\lambda)\right\}<\mathbb{P}\big\{A_n(t,Z^{\overline{m}})^c\big\}\leq\delta.$$

证明至此完成。 □

可以验证，只要满足 (11)，命题 15 的两个结论都成立。从现在起，出于技术原因，我们假设命题12中的常数 K 大于 1。这一假设至多只会使覆盖数界变差。现在我们考虑一个整数 $k\in\{1,\ldots,n\}$.

命题16。令 $\tau > 0$. 若 $k \geq \lceil$对数 $n\rceil$, $n \geq 2d$, 且 $h_+ = h_+(k) := 2k/((1-\lambda_\tau^+)n)$ 满足

$$\lambda_\tau^+ := \frac{4\big(2d+\tau+\log(K_0)/\log(2d)\big)\big)+9}{4\big(2d+\tau+\log(K_0)/\log(2d)\big)\big)+12},$$

那么，

$$\mathbb{P}\left\{\sup_{x\in\mathbb{R}^d}\frac{R_k(x)}{\zeta_{h_+}(x)} \geq 1\right\} \leq n^{-\tau}.$$

我们进一步定义

$$B_\tau := \frac{2}{1-\lambda_\tau^+} = \frac{8(2d+3+\log(K_0)/\log(2d)+\tau)}{3}, \tag{16}$$

使得 $h_+ = B_\tau k/n$.

证明。应用命题 15 (i)，其中 $\delta = n^{-\tau}$，以及 $h = 2k/((1-\lambda)n)$ 取任意的 $\lambda \in (0,\ 1)$，可得，对所有满足以下条件的 n

$$n \geq \frac{2d(1+\lambda/3)(1-\lambda)n}{\lambda^2 k}\log\Big(K_0^{1/(2d)}n^{1+\tau/(2d)}k^{-1}\Big), \tag{17}$$

有

$$\mathbb{P}\left\{\inf_{x\in\mathbb{R}^d}\frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i \in \mathrm{B}(x,\zeta_{h_+}(x))) < \frac{k}{n}\right\} \leq n^{-\tau}, \tag{18}$$

其中 K_0 在命题 15 中定义。高概率界 (18) 表明

$$\mathbb{P}\left\{\sup_{x\in\mathbb{R}^d}\frac{R_k(x)}{\zeta_{h_+}(x)} \geq 1\right\} \leq n^{-\tau}.$$

现在我们检查（17）的右侧始终小于 $2d$。（17）中的不等式等价于

$$\frac{(1+\lambda/3)(1-\lambda)}{\lambda^2 k}\log\Big(K_0 n^{2d+\tau}k^{-2d}\Big) \leq 1. \tag{19}$$

对于 $n \geq k \geq \log n$，利用 K_0 的表达式、 的假设以及 ，我们可以验证，对于 0，有 $\log(K_0 n^{2d+\tau}k^{-2d}) \geq 0$. 后一个条件对任意 $\tau > 0$ 和任意 $n \geq 1$. 均成立。因此，我们可以断言，要使（19）成立，只需满足

$$\frac{(1+\lambda/3)(1-\lambda)}{\lambda^2}\log\Big(K_0 n^{2d+\tau}\Big) \leq \log n.$$

Using $a\log(b) = \log(b^a)$，对两边取指数并重新整理各项可得

$$1 \leq n^{\frac{\lambda^2}{(1+\lambda/3)(1-\lambda)}-2d-\tau-\log(K_0)/\log(n)}.$$

要使后者成立，只需指数非负。由于 $n \geq 2d$，上述公式中指数的一个下界为 $\lambda^2/((1+\lambda/3)(1-\lambda)) - 2d-\tau-\log(K_0)/\log(2d)$。将所得下界的表达式重新整理，可得到二次不等式形式的充分条件

$$\lambda^2\Big(1+\frac{a}{3}\Big)+\frac{2a}{3}\lambda - a.$$

其中 $a := 2d + \tau + \log(K_0)/\log(2d)$。不难发现，正根 z_+ 满足

$$z_+ = \frac{a}{a+3}\left(\sqrt{4+\frac{9}{a}} - 1\right) \le \frac{4a+9}{4a+12} = \lambda_\tau^+ < 1,$$

其中，不等式由伯努利不等式得出。这表明，按照我们对 λ_τ^+, 的选择，(17) 对所有 $n \ge 2d.$ 都成立。因此，对于所有 $k \ge \log n,$

$$\mathbb{P}\left\{\sup_{x\in\mathbb{R}^d} \frac{R_k(x)}{\zeta_{h_+}(x)} \ge 1\right\} \le n^{-\tau}.$$

□

令 $\mathrm{P} \in \mathcal{P}.$ 令 $X \sim \mathrm{P},$ 且 $Y = f(X) + \varepsilon,$ 其中 ε 是一个取值于 R 的随机变量，且与 $X.$ 独立。另假设 X 是参数为 $\alpha, \nu > 0.$ 的次指数随机变量。令 $(X_1, Y_1) \ldots, (X_n, Y_n)$ 为 (X, Y) 的 n 个独立同分布副本。对于 $k \in \mathbf{N}$ 和 $x \in \mathbf{R}^d,$，我们用 $R_k^{\mathrm{P}}(x) := \|x - X_k(x)\|.$ 表示。我们定义以下事件

$$U_n^{\mathrm{P}}(k) := \left\{\sup_{x\in\mathbb{R}^d} \frac{R_k^{\mathrm{P}}(x)}{\zeta_{h_+(k)}^{\mathrm{Px}}(x)} \le 1\right\}, \text{ and } U_n^{\mathrm{P}} := \bigcap_{k=\lceil \log n\rceil}^{n} U_n^{\mathrm{P}}(k).$$

备注 17. 虽然命题*16*表明，对于每个单独选取的 $k \ge \log n,$ $R_k(x) \le \zeta_{h_+}(x)$ 以高概率一致成立，但很容易将其推广到 $k = k(x) \in \{1, \ldots, n\}$ 依赖于 $x \in \mathbb{R}^d.$ 的情形。由于 k 现在可以随 x 改变取值，因此必须付出代价：概率中的 sup 不再针对整个 $\mathbb{R}^d$，而只针对如下定义的一个子集 $\Theta_n^*(k)$

$$\Theta_n(k) := \left\{x \in \mathbb{R}^d : k(x) \ge \lceil \log n\rceil\right\}.$$

同样，根据并集界可知，我们需要在概率中付出一个因子 n 的代价

$$\begin{aligned}
\mathbb{P}\left\{\sup_{x\in\Theta_n^*} \frac{R_{k(x)}(x)}{\zeta_{h_+}} \ge 1\right\} &\le \mathbb{P}\left\{\bigcup_{\ell=\lceil \log n\rceil}^{n} \left\{\sup_{x\in\mathbb{R}^d} \frac{R_\ell(x)}{\zeta_{h_+}(x)} \ge 1\right\}\right\} \\
&\le \sum_{\ell=\lceil \log n\rceil}^{n} \mathbb{P}\left\{\sup_{x\in\mathbb{R}^d} \frac{R_\ell(x)}{\zeta_{h_+}(x)} \ge 1\right\} \\
&\le n^{1-\tau}.
\end{aligned}$$

命题18。 令 $\tau > 0$ 和 $V_\tau := K_0^{1/(2d)} \vee (8d + 4\tau)$。若 $k \ge \lceil V_\tau$对数 $n\rceil,$ $n \ge 2d,$ 且 $h_- = h_-(k) := k/(2n(1 + \lambda_\tau^-))$，其中

$$\lambda_\tau^- := 1 + \sqrt{\frac{5}{2}},$$

那么，

$$\mathbb{P}\left\{\inf_{x\in\mathbb{R}^d} \frac{R_k(x)}{\zeta_{h_-}(x)} \le 1\right\} \le n^{-\tau}.$$

我们进一步定义

$$c_1 := \frac{1}{2 + \sqrt{10}}, \tag{20}$$

从而 $h_- = c_1 k/n.$

证明。 将命题 15（ii）与 $\delta = n^{-\tau}$ 以及任意 $\lambda \in \mathrm{R}_+^*$ 下的 $h = k/(2(1+\lambda)n)$ 相结合，，可得，对于所有满足条件的 n

$$n \geq \frac{8d(1+\lambda/3)(1+\lambda)n}{\lambda^2 k} \log\left(K_0^{1/(2d)} n^{1+\tau/(2d)} k^{-1}\right), \tag{21}$$

可得

$$\mathbb{P}\left\{ \sup_{x\in\mathbb{R}^d} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i \in \mathrm{B}(x, \zeta_{h_-}(x))) \geq \frac{k}{n} \right\} \leq n^{-\tau}, \tag{22}$$

其中， K_0 在命题 15 中定义。高概率界 (22) 表明

$$\mathbb{P}\left\{ \inf_{x\in\mathbb{R}^d} \frac{R_k(x)}{\zeta_{h_-}(x)} \geq 1 \right\} \leq n^{-\tau}.$$

将 $\lambda_\tau^- = 1 + \sqrt{5/2}$ 代入 (21) 得到

$$k \geq 4 \log\left(K_0 n^{2d+\tau} k^{-2d}\right),$$

并且很容易验证，对于所有 $n \geq 2d$ 和 $k \geq \lceil V_\tau \log n \rceil$，该不等式都成立 . □

我们定义事件

$$L_n^{\mathrm{P}}(k) := \left\{ \inf_{x\in\mathbb{R}^d} \frac{R_k^{\mathrm{P}}(x)}{\zeta_{h_+(k)}^{\mathrm{P}}(x)} \geq 1 \right\}, \text{ and } L_n^{\mathrm{P}} := \bigcap_{k=\lceil V_\tau \log n\rceil}^{n} L_n^{\mathrm{P}}(k).$$

我们还定义集合

$$\Delta_n^{\mathrm{P}}(c,k) := \left\{ x \in \mathbb{R}^d : p(x) \geq \frac{ck}{n} \right\}.$$

引理19。 令 $\tau > 0$ 和 p 为P的密度。若 $k \geq \lceil$对数 $n\rceil$，则对所有 $n \geq 2d$，

$$U_n^{\mathrm{P}}(k) \subseteq \left\{ \sup_{x\in\Delta_n^{\mathrm{P}}(S_\tau,k)} p(x) R_k(x)^d \leq \frac{S_\tau k}{n} \right\},$$

其中 $S_\tau := \theta B_\tau$.

证明。 由于 $\mathrm{P} \in \mathcal{P}$，我们可以应用 (2) 得到，对所有 $x \in \mathbb{R}^d$，

$$h_+ = \mathrm{P}\{\mathrm{B}(x, \zeta_{h_+}(x))\} \geq \theta^{-1}\left(\zeta_{h_+}(x) \wedge 1\right)^d p(x).$$

通过重新整理前面显示的等式，我们得到

$$\zeta_{h_+}(x) \wedge 1 \leq \left(\frac{\theta h_+}{p(x)}\right)^{1/d}.$$

我们注意到，若 $x \in \Delta_n^{\mathrm{P}}(S_\tau,\ k)$，则 $\mathrm{P}\{\mathrm{B}(x,\ 1)\} \geq \theta^{-1}p(x) \geq B_\tau k/n = h_+$.因此， $\zeta_{h_+}(x) \leq 1$，且对于所有 $x \in \Delta_n^{\mathrm{P}}(S_\tau,\ k)$，在 $U_n^{\mathrm{P}}(k)$，

$$R_k(x)^d \leq \frac{S_\tau k}{np(x)},$$

这意味着该结果。 □

注记20。 给定一个函数 k: $\mathbb{R}^d \to \{1,\dots,n\}$ 和一个常数 $c>0$，我们定义集合

$$\Theta_n^{\mathrm{P}}(c,k) := \left\{ x \in \mathbb{R}^d : p(x) \geq \frac{ck(x)}{n} \right\}.$$

由引理*19*可直接推出，

$$\begin{aligned}
\left\{ \sup_{x\in\Theta_n^{\mathrm{P}}(c,k)} \frac{p(x)R_{k(x)}(x)^d}{k(x)} \geq \frac{S_\tau}{n} \right\} &= \bigcup_{i=\lceil \log n\rceil}^{n} \left\{ \sup_{x\in k^{-1}(i)\cap\Delta_n^{\mathrm{P}}(c,i)} p(x)R_i(x)^d \geq \frac{S_\tau i}{n} \right\} \\
&\subseteq \bigcup_{i=\lceil \log n\rceil}^{n} \left\{ \sup_{x\in\Delta_n^{\mathrm{P}}(c,i)} p(x)R_i(x)^d \geq \frac{S_\tau i}{n} \right\} \\
&= \left(\bigcap_{i=\lceil \log n\rceil}^{n} U_n^{\mathrm{P}}(i) \right)^c = U_n^{\mathrm{P}\,c}.
\end{aligned}$$

因此，

$$\mathbb{P}\left\{ \sup_{x\in\Theta_n^{\mathrm{P}}(c,k)} \frac{p(x)R_{k(x)}(x)^d}{k(x)} \leq \frac{S_\tau}{n} \right\} \geq 1 - n^{1-\tau}.$$

引理21。 设 $\tau>0$. 若 $k \geq \lceil V_\tau \log n\rceil$，且 $n \geq 2d$，则

$$\left\{ \inf_{x\in\Delta_n^{\mathrm{P}}(c_2,k)} p(x)R_k(x)^d \leq \frac{c_1 k}{\theta n} \right\} \subseteq L_n^{\mathrm{P}}(k)^c,$$

其中 $c_2 := \theta c_1$.

证明。 我们在事件 $L_n^{\mathrm{P}}(k)$。对于所有满足 $\zeta_{h_-}(x) \leq 1$， 的 $x \in \mathrm{P}$的支撑集，局部质量性质 (2) 蕴含

$$h_- = \mathrm{P}\{\mathrm{B}(x,\zeta_{h_-}(x))\} \leq \theta p(x)\zeta_{h_-}(x)^d.$$

将后者重新整理后，可得

$$\zeta_{h_-}(x) \geq \left(\frac{c_1 k}{\theta n p(x)} \right)^{1/d}.$$

还需证明，若 $x \in \Delta_n^{\mathrm{P}}(c_2,k)$，则 $\zeta_{h_-}(x) \leq 1$. 。设 $x \in \Delta_n^{\mathrm{P}}(c_2,\ k)$，并假设$\zeta_{h_-}(x) > 1$. 。局部质量性质 (2) 表明

$$h_- = \mathrm{P}\{\mathrm{B}(x,\zeta_{h_-}(x))\} \overset{(i)}{>} \mathrm{P}\{\mathrm{B}(x,1)\} \geq \theta^{-1}p(x) \geq \frac{c_1 k}{n} = h_-,$$

这就产生了矛盾。因此，对所有 $x \in \Delta_n^{\mathrm{P}}(c_2,\ k)$，都有 $\zeta_{h_-}(x) \leq 1$，并且更有

$$L_n^{\mathrm{P}}(k) \subseteq \left\{ \inf_{x\in\Delta_n^{\mathrm{P}}(c_2,k)} p(x)R_k(x)^d \geq \frac{c_1 k}{\theta n} \right\}.$$

□

A.2 一致逐点绝对偏差和标准差界

^现在我们在模型 (1) 下进行讨论，其中 $f_* \in \mathcal{H}(L,\ \beta)$，且 $L > 0$和 $\beta \in (0,\ 1]$. 设 f为一个局部k-近邻回归估计量，其近邻函数为 k: $\mathrm{R}^d \to \{1, \ldots, n\}$，且与 $Y_i|X_i$独立，适用于$1 \le i \le n$. 。也就是说，对所有 $x \in \mathrm{R}^d$,

$$\widehat{f}(x) = \frac{1}{k(x)} \sum_{i=1}^{n} Y_i(x), \tag{23}$$

其中，$Y_i(x)$ $1 \le i \le k(x)$ 是与 x 在$\{X_1, \ldots, X_n\}$. 中第 k个最近邻对应的标签。我们定义相应的逐点平均函数

$$\overline{f}\colon x \mapsto \mathrm{E}\left[\widehat{f}(x)|X_1, \ldots, X_n\right]. \tag{24}$$

上一小节得出了以下逐点偏差界。

引理22。 如果 $n \ge \|k(x)\|_\infty$，那么对于所有 $x \in \mathbb{R}^d$,

$$|\overline{f}(x) - f_*(x)| \le 2L \wedge L \sum_{i=1}^{k(x)} k(x)^{-1} R_i(x)^\beta,$$

此外，如果 $\mathrm{P} \in \mathcal{P}$，且对于所有 $x \in \mathbb{R}^d$， $k(x) \ge \lceil \log n \rceil$ ，那么

$$\bigcap_{i \in k(\mathbb{R}^d)} U_n^{\mathrm{P}}(i) \subseteq \left\{ \sup_{x \in \Theta_n^{\mathrm{P}}(S_\tau, k)} \frac{p(x)^{\beta/d} |\overline{f}(x) - f_*(x)|}{k(x)^{\beta/d}} \le \left(\frac{S_\tau}{n}\right)^{\beta/d} \right\}.$$

证明。 由于 $f_* \in \mathcal{H}(L, \beta)$，我们得到，对于所有 $x \in \mathrm{R}^d$，以及所有 $i \in \{1, \ldots, k(x)\}$,

$$\begin{aligned} \left|f(X_i(x)) - f_*(x)\right| &\le \left(f(X_i(x)) + f_*(x)\right) \wedge L\|X_i(x) - x\|^\beta \\ &\le 2L \wedge L R_i(x)^\beta. \end{aligned}$$

结合 $f(x)$ 和 $\sum_{i=1}^{k(x)} k(x)$ 的定义$^{-1}$ $=^{-}1$ 可得

$$\begin{aligned} \left|\overline{f}(x) - f_*(x)\right| &= \left| \sum_{i=1}^{k(x)} k(x)^{-1} f(X_i(x)) - f_*(x) \right| \\ &\le \sum_{i=1}^{k(x)} k(x)^{-1} \left|f_*(X_i(x)) - f_*(x)\right| \\ &\le \sum_{i=1}^{k(x)} k(x)^{-1} \left(2F \wedge L R_i(x)^\beta\right) \\ &\le 2L \wedge L \sum_{i=1}^{k(x)} k(x)^{-1} R_i(x)^\beta. \end{aligned}$$

这证明了第一个陈述。第二个陈述可通过进一步将先前推导出的不等式上界化为$|f(x) - f_*(x)| \le L R_k(x)^\beta$ 并应用引理19得到。 □

命题23。 设 $\widehat{f}$如（23）中所述，其近邻函数为 k: $\mathbb{R}^d \to \{1, \ldots, n\}$. 设 f由（24）定义。则，对于所有 $n \ge 2d$，均有

$$\mathbb{P}\left\{ \sup_{x \in \mathbb{R}^d} k(x)^{1/2} \left|\widehat{f}(x) - \overline{f}(x)\right| \ge \eta \right\} \le K_d n^{2d+1} e^{-H\eta^2},$$

其中 $K_d = 2(25/d)^d$，且 $H = \lambda_0^2/(8\sigma_0)$ 对任意满足以下条件的 $\lambda_0, \sigma_0 > 0$： $\mathrm{E}[\exp(\lambda|\varepsilon|)] \le \sigma_0$.

证明。 对于任意 $x \in \mathbb{R}^d$，我们有

$$\widehat{f}(x) - \overline{f}(x) = \sum_{i=1}^{n} k(x)^{-1}[Y_i(x) - f(X_i(x))]\mathbb{1}(i \le k(x)).$$

令 $Z_i(x) := Y_i(x) - f(X_i(x))$ 对所有 $1 \le i \le n$ 以及所有 $x \in \mathbb{R}^d$. 成立。根据 [3] 中的命题 8.1 以及回归模型 (1)，在给定 $X_1, \ldots, X_n$ 的条件下，随机变量 $\{Z_i(x)\}_{i=1}^n$ 相互独立且均值为零。根据假设，它们是次指数的，因此满足一致噪声条件；存在常数 $\lambda_0, \sigma_0 > 0$，使得

$$\sup_{x^n \in \mathbb{R}^{d\times n}} \mathrm{E}\Big[\exp\big(\lambda_0 |Z_i(x)|\big)\Big| X^n = x^n\Big] \le \sigma_0 < \infty, \tag{25}$$

其中 $x^n \in \mathbb{R}^{d\times n}$. 根据 [3], 中的引理 12.1，对任意 $x \in \mathbb{R}^d$, 以及任意 $\eta \le \min(1, 2\sigma_0)/\lambda_0$，有 t

$$\mathbb{P}\Big\{\big|\widehat{f}(x) - \overline{f}(x)\big| \ge \eta \Big| X^n = x^n\Big\} \le 2\exp\Big(-\frac{\eta^2\lambda_0^2 k(x)}{8\sigma_0}\Big). \tag{26}$$

为了导出在 $x \in \mathbb{R}^d$, 上一致的高概率界，我们引入如下置换集合

$$\mathcal{W}(x^n) := \Big\{\sigma \in \mathcal{S}_n : \exists x \in \mathbb{R}^d,\ X_i(x) = X_{\sigma(i)},\ \forall i \in \{1, \ldots, n\}\Big\},$$

其中 $\mathcal{S}_n$ 是 $\{1, \ldots, n\}$. 上的对称群。 [3] 中的定理 12.2 表明，对所有 $n \ge 2d$, 以及所有 $x^n \in \mathbb{R}^{d\times n}$, ， $\mathcal{W}(x^n)$ 的基数有上界

$$\ell(x^n) := \mathrm{Card}(\mathcal{W}(x^n)) \le \Big(\frac{25}{d}\Big)^d n^{2d}.$$

特别地，给定 $x^n \in \mathbb{R}^{d\times n}$，我们可以将 $\sim_{x^n}$ 在 $\mathbb{R}^d$ 上的等价关系定义为：对所有z_1， $z_2 \in \mathbb{R}^d$,

$$\big(z_1 \sim_{x^n} z_2\big) \iff \Big(\forall i \in \{1, \ldots, n\} : x_i(z_1) = x_i(z_2)\Big),$$

其中 $x_i(z)$ 是 z在 $\{x_1, \ldots, x_n\}$. 中的第 i 个最近邻。此外，$\mathbb{R}^d/\sim_{x^n}$ 的基数至多为 $\mathrm{Card}(\mathcal{W}(x^n))$，并且与 $\mathcal{W}(x^n)$. 一一对应。于是，设 $\{A_\sigma\}_{\sigma\in\mathcal{W}(x^n)}$ 为 $\mathbb{R}^d/\sim_{x^n}$. 的元素。现在，我们可以按照 $\sim_{x^n}$, 对 $\mathbb{R}^d$ 进行划分，利用并集界并应用 (26)，从而推导出一个高概率成立的一致上界，如下所示

$$\begin{aligned}
&\mathbb{P}\Big\{\sup_{x\in\mathbb{R}^d} k(x)^{1/2}\big|\widehat{f}(x) - \overline{f}(x)\big| \ge \eta \Big| X^n = x^n\Big\} \\
&\quad\le \mathbb{P}\Big\{\bigcup_{\sigma\in\mathcal{W}(X^n)}\Big\{\sup_{x\in A_\sigma}\Big|\sum_{i=1}^{n}\frac{k(x)^{1/2}}{k(x)}\big(Y_{\sigma(i)} - f_*(X_{\sigma(i)})\big)\Big| \ge \eta\Big| X^n = x^n\Big\}\Big\} \\
&\quad\le \sum_{\sigma\in\mathcal{W}(x^n)}\mathbb{P}\Big\{\sup_{\ell\in\{1,\ldots,n\}}\Big|\sum_{i=1}^{\ell}\ell^{-1/2}\big(Y_{\sigma(i)} - f_*(X_{\sigma(i)})\big)\Big| \ge \eta\Big| X^n = x^n\Big\} \\
&\quad\le \sum_{\sigma\in\mathcal{W}(x^n)}\sum_{\ell=1}^{n}\mathbb{P}\Big\{\Big|\sum_{i=1}^{\ell}\ell^{-1/2}\big(Y_{\sigma(i)} - f_*(X_{\sigma(i)})\big)\Big| \ge \eta\Big| X^n = x^n\Big\} \\
&\quad\le \sum_{\sigma\in\mathcal{W}(x^n)}\sum_{\ell=1}^{n} 2\exp\Big(-\frac{\eta^2\lambda_0^2}{8c_0}\Big) \\
&\quad\le 2\Big(\frac{25}{d}\Big)^d n^{2d+1}\exp\Big(-\frac{\eta^2\lambda_0^2}{8c_0}\Big).
\end{aligned}$$

我们通过对等式两边关于 $X_1, \ldots, X_n$ 取期望来完成证明。 □

推论24. 在命题*23*的条件下，对于所有 $\tau > 0$ 和所有 $n \geq 3 \vee 2d$,

$$\mathbb{P}\left\{ \sup_{x\in\mathbb{R}^d} k(x)^{1/2}\big|\widehat{f}(x) - \overline{f}(x)\big| \geq \sqrt{C_\tau \log n}\right\} \leq n^{-\tau},$$

with

$$C_\tau := \frac{\log K_d}{H}\left(1 + \frac{2d+\tau+1}{\log(K_d))}\right)$$

其中 H 和 K_d 在命题*23*中定义。

证明。 重新整理后，命题23的结论为

$$\mathbb{P}\left\{\forall x \in \mathbb{R}^d,\ \big|\widehat{f}(x) - \overline{f}(x)\big| \leq \sqrt{\frac{1}{Hk(x)}\log\Big(\frac{K_d n^{2d+1}}{\delta}\Big)}\right\} \geq 1-\delta.$$

我们现在将此结果应用于

$$\eta_- \geq \sqrt{\frac{\log K_d}{H}\left(1 + \frac{2d+\tau+1}{\log K_d}\log n\right)}$$

从而得到

$$\mathbb{P}\left\{ \sup_{x\in\mathbb{R}^d} k(x)^{1/2}|\widehat{f}(x) - \overline{f}(x)| \geq \eta_-\right\} \leq n^{-\tau}.$$

最后，由于 $n \geq 3$, $\log n \geq 1$ ，我们可以选取

$$\eta = \sqrt{C_\tau \log n} = \sqrt{\frac{\log K_d}{H}\left(1 + \frac{2d+\tau+1}{\log K_d}\right)\log n} \geq \eta_-,$$

并得到所述结论。 □

我们定义事件

$$V_n^{\mathrm{P}}(\widehat{f}) := \left\{ \sup_{x\in\mathbb{R}^d} k(x)^{1/2}\big|\widehat{f}(x) - \overline{f}(x)\big| \leq \sqrt{C_\tau \log n}\right\}.$$

命题 25 (偏差-方差分解**)**。 在模型(1)的假设下，若 f 是一个局部*k-NN*，其近邻函数为 $k\colon \mathbb{R}^d \to \{\lceil \log n\rceil, \dots, n\}$，且$\mathrm{P} \in \mathcal{P}$，则对所有 $n \geq 2d \vee 3$ 及所有$\Delta \subseteq \Theta_n^{\mathrm{P}}(S_\tau, k)$,

$$\left\{\forall x \in \mathbb{R}^d,\ \big(\widehat{f}(x) - f(x)\big)^2 \leq 2C_\tau\frac{\log n}{k(x)} + 8L^2\left(\frac{S_\tau k(x)}{np(x)}\right)^{2\beta/d}\mathbb{1}(x\in\Delta) + 8L^2\mathbb{1}(x\in\Delta^c)\right\}$$
$$\supseteq \left(\bigcap_{i\in k(\mathbb{R}^d)} U_n^{\mathrm{P}}(i)\right)\bigcap V_n^{\mathrm{P}}(\widehat{f}).$$

证明。 由引理22与引理24可知，在事件

$$\left(\bigcap_{i\in k(\mathbb{R}^d)} U_n^{\mathrm{P}}(i)\right)\bigcap V_n^{\mathrm{P}}(\widehat{f}),$$

对所有 $x \in \mathrm{R}^d$,

$$\begin{aligned}\big(\widehat{f}(x) - f(x)\big)^2 &\leq 2\big|\widehat{f}(x) - \overline{f}(x)\big|^2 + 2|\overline{f}(x) - f(x)\big|^2 \\ &\leq \frac{2C_\tau\log n}{k(x)} + 8L^2\left(\frac{S_\tau k(x)}{np(x)}\right)^{2\beta/d}\mathbb{1}(x\in\Delta) + 8L^2\mathbb{1}(x\in\Delta^c).\end{aligned}$$

证明到此结束。 □

A.3 密度估计量的证明

在证明我们的上界时，我们选择一个近邻函数 $k(x)$，它依赖于相应协变量密度的一个 ℓ-NN 密度估计量。 $\widehat{p}$ 的定义见式 (7)，它是一个 ℓ-最近邻密度估计量。我们首先给出一个初步结果，表明在高密度区域中 $\widehat{p} \asymp p$。

引理 26. 设 $\mathrm{P} \in \mathcal{P}(D,\theta)$ 满足 $D,\theta > 0.$。记 p 为P的密度，并令 $\widehat{p}$ 为基于来自P的 n 个独立同分布样本得到的 ℓ-*NN* 密度估计量。考虑 $n \geq 3 \vee 2d.$

(i) 若 $\ell \geq \lceil V_\tau \log n \rceil$，则

$$L_n^{\mathrm{P}}(\ell) \subseteq \left\{ \sup_{x \in \Delta_n^{\mathrm{P}}(c_2,\ell)} \frac{\widehat{p}(x)}{p(x)} \leq \frac{\theta}{c_1} \right\}.$$

(ii) 若 $\ell \geq \lceil \log n \rceil$，则

$$U_n^{\mathrm{P}}(\ell) \subseteq \left\{ \inf_{x \in \Delta_n^{\mathrm{P}}(S_\tau,\ell)} \frac{\widehat{p}(x)}{p(x)} \geq \frac{1}{S_\tau} \right\}$$

证明。 (i) 我们考虑事件 $L_n^{\mathrm{P}}(\ell)$。对所有 $x \in \Delta_n^{\mathrm{P}}(c_2,\ell)$，均有

$$\frac{\widehat{p}(x)}{p(x)} = \frac{\ell}{np(x)R_\ell(x)^d} \leq \frac{\theta}{c_1},$$

其中，最后一个不等式由引理21得出。这表明

$$L_n^{\mathrm{P}}(\ell) \subseteq \left\{ \sup_{x \in \Delta_n^{\mathrm{P}}(c_2,\ell)} \frac{\widehat{p}(x)}{p(x)} \leq \frac{\theta}{c_1} \right\}.$$

(ii) 我们在事件 $U_n^{\mathrm{P}}(\ell)$。对于所有 $x \in \Delta_n^{\mathrm{P}}(S_\tau,\ \ell)$，我们有

$$\frac{\widehat{p}(x)}{p(x)} = \frac{\ell}{np(x)R_\ell(x)^d} \geq \frac{1}{S_\tau},$$

其中，最后一个不等式来自引理19。这表明

$$U_n^{\mathrm{P}}(\ell) \subseteq \left\{ \inf_{x \in \Delta_n^{\mathrm{P}}(c_2,\ell)} \frac{\widehat{p}(x)}{p(x)} \geq \frac{1}{S_\tau} \right\}.$$

□

注释27。这表明，如果 $\ell \geq \lceil V_\tau \log\, n \rceil$，且 $n \geq 3 \vee 2d$，那么，在 $L_n^{\mathrm{P}}(\ell) \cap U_n^{\mathrm{P}}(\ell)$ 上，对于所有$x \in \Delta_n^{\mathrm{P}}(S_\tau \vee c_2, \ell)$，

$$\frac{1}{S_\tau} \leq \frac{\widehat{p}(x)}{p(x)} \leq \frac{\theta}{c_1}.$$

引理 28。令 $\mathrm{P} \in \mathcal{P}$，$\tau > 0$ 以及 $\kappa > 0.$ 考虑函数

$$k(x) = n \wedge \left\lceil \kappa \log(n)^{d/(2\beta+d)} \big(n\widehat{p}(x)\big)^{2\beta/(2\beta+d)} \right\rceil \vee \lceil \log n \rceil,$$

其中 $\widehat{p}$是 针对P的 ℓ-近邻密度估计量，其中 $\ell \geq \lceil V_\tau$对数 $n \rceil.$令 $N = N(\kappa,\theta,D) \geq 3 \vee 2d$满足 fy

$$\frac{N}{\log N} \geq \kappa^{(2\beta+d)/d} \left(\frac{\theta D}{c_1} \right)^{2\beta/d},$$

并定义常数

$$C := S_\tau \vee c_2 \vee \frac{S_\tau}{\kappa^{(2\beta+d)/(2\beta)} V_\tau} \vee \frac{(2S_\tau\kappa)^{(2\beta+d)/d}}{V_\tau} \left(\frac{\theta}{c_1}\right)^{2\beta/d}.$$

那么，以下断言成立

(i) 对于所有 $n \geq N$,

$$L_n^{\mathrm{P}}(\ell) \cap U_n^{\mathrm{P}}(\ell) \subseteq \left\{ \forall x \in \Delta_n^{\mathrm{P}}(C,\ell),\ k(x) = \left\lceil \kappa \log(n)^{d/(2\beta+d)} \big(n\widehat{p}(x)\big)^{2\beta/(2\beta+d)} \right\rceil \right\}.$$

(ii) 对于所有 $n \geq N$,

$$L_n^{\mathrm{P}}(\ell) \cap U_n^{\mathrm{P}}(\ell) \subseteq \left\{ \Delta_n^{\mathrm{P}}(C,\ell) \subseteq \Theta_n^{\mathrm{P}}(S_\tau, k) \right\}.$$

证明。 对 $L_n^{\mathrm{P}}(\ell) \cap U_n^{\mathrm{P}}(\ell)$. 通过证明对于所有 $x \in \Delta_n^{\mathrm{P}}(C,\ell)$,, 可得 (i)

$$\log n \leq \kappa \log(n)^{d/(2\beta+d)} \big(n\widehat{p}(x)\big)^{2\beta/(2\beta+d)} \leq n.$$

令 $x \in \Delta_n^{\mathrm{P}}(C,\ell)$, 且 $u(x) := \kappa \log(n)^{d/(2\beta+d)} (n\widehat{p}(x))^{2\beta/(2\beta+d)}$. 我们首先应用引理 26，并注意到，由于 $\Delta_n^{\mathrm{P}}(C,\ell) \subseteq \Delta_n^{\mathrm{P}}(S_\tau \vee c_2, \ell)$,

$$\kappa \log(n)^{d/(2\beta+d)} \left(\frac{np(x)}{S_\tau}\right)^{2\beta/(2\beta+d)} \leq u(x) \leq \kappa \log(n)^{d/(2\beta+d)} \left(\frac{\theta n D}{c_1}\right)^{2\beta/(2\beta+d)}. \tag{27}$$

容易验证，如果满足以下条件，则式 (27) 中最左侧的项以下界 $\log(n)$ 为界

$$p(x) \geq \frac{S_\tau}{\kappa^{(2\beta+d)/(2\beta)}} \frac{\log n}{n},$$

显然，对于所有 $x \in \Delta_n^{\mathrm{P}}(C,\ \ell)$，这也成立，因为

$$p(x) \geq \frac{C\ell}{n} \geq \frac{S_\tau}{\kappa^{(2\beta+d)/(2\beta)} V_\tau} \frac{\ell}{n} \geq \frac{S_\tau}{\kappa^{(2\beta+d)/(2\beta)}} \frac{\log n}{n}.$$

式 (27) 中最右侧的一项在 n 的条件下有上界

$$\frac{n}{\log n} \geq \kappa^{(2\beta+d)/d} \left(\frac{\theta D}{c_1}\right)^{2\beta/d},$$

这由假设可知。我们已证明前述两个结论对任意 $x \in \Delta_n^{\mathrm{P}}(C,\ \ell)$，均成立，因此,

$$L_n^{\mathrm{P}}(\ell) \cap U_n^{\mathrm{P}}(\ell) \subseteq \left\{ \forall x \in \Delta_n^{\mathrm{P}}(C,\ell),\ k(x) = \left\lceil \kappa \log(n)^{d/(2\beta+d)} \big(n\widehat{p}(x)\big)^{2\beta/(2\beta+d)} \right\rceil \right\}.$$

现在我们 证明 (ii)。 令 $x \in \Delta_n^{\mathrm{P}}(C,\ell)$。 对于 $n \geq e \vee N$， (i) 可推出 $k(x) \geq \lceil \log\ n \rceil \geq 1$. 因此 e,

$$\begin{aligned} \frac{S_\tau k(x)}{n} &\leq \frac{2S_\tau \kappa \log(n)^{d/(2\beta+d)} (n\widehat{p}(x))^{2\beta/(2\beta+d)}}{n} \\ &\leq 2S_\tau \kappa \left(\frac{\log n}{n}\right)^{d/(2\beta+d)} \left(\frac{\theta p(x)}{c_1}\right)^{2\beta/(2\beta+d)}, \end{aligned}$$

且后者由 $p(x)$ 上界控制，因为

$$p(x) \geq \frac{C\ell}{n} \geq \frac{(2S_\tau\kappa)^{(2\beta+d)/d}}{V_\tau}\left(\frac{\theta}{c_1}\right)^{2\beta/d}\frac{\ell}{n} \geq (2S_\tau\kappa)^{(2\beta+d)/d}\left(\frac{\theta}{c_1}\right)^{2\beta/d}\frac{\log n}{n}.$$

我们已经证明

$$p(x) \geq \frac{C\ell}{n} \Rightarrow p(x) \geq \frac{S_\tau k(x)}{n},$$

并且，由于 $x \in \Delta_n^{\mathrm{P}}(C,\ \ell)$ 是任意的，

$$L_n^{\mathrm{P}}(\ell) \cap U_n^{\mathrm{P}}(\ell) \subseteq \left\{\Delta_n^{\mathrm{P}}(C,\ell) \subseteq \Theta_n^{\mathrm{P}}(S_\tau,k)\right\}.$$

证明至此完成。 □

B 主要结果的证明

B.1 推论5的证明

设 $\mathrm{P_X}$、$\mathrm{Q_X} \in \mathcal{P}(D,\theta)$。设 $X \sim \mathrm{P_X}$，$X' \sim \mathrm{Q_X}$，$Y = f_*(X) + \varepsilon$，且 $Y' \sim f_*(X') + \varepsilon'$，其中$\varepsilon, \varepsilon'$ 是两个与 X 和 X' 独立、参数为 $\alpha, \nu > 0$. 的次指数独立同分布随机变量。令 $(X_1,Y_1)\ldots,(X_n,Y_n)$ 为 (X,Y) 的 n 个独立同分布副本，令 $(X_1',Y_1')\ldots,(X_m',Y_m')$ 为 (X',Y') 的 m 个独立同分布副本。对于 $k \in \mathbb{N}$ 和 $x \in \mathbb{R}^d$，只要这些量有意义，我们记 $R_k^{\mathrm{P}}(x) := \|x - X_k(x)\|$，并记 $R_k^{\mathrm{Q}}(x) := \|x - X_k'(x)\|$. 。考虑局部 k-NN 估计量

$$\widehat{f}(x) = \frac{1}{k(x)}\sum_{i=1}^{n} Y_i(x), \tag{28}$$

其中 $k(x)\colon \mathbb{R}^d \to \{1,\ldots,n\}$，且 $Y_i(x)$ 是 x 在$\{X_1,\ldots,X_n\}$. 中第 i 个最近邻的标签。接下来的结果涉及局部 k-NN 估计量。我们选择近邻函数k 时包含一个密度估计步骤。

定理 29. 设 $\gamma > 0$，$\kappa > 0$. 令 C 为引理 *28* 中定义的常数，$r := \gamma \wedge 2\beta/(2\beta+d)$，且 $\mathcal{T}(r) := \int q(x)/p(x)^r\,dx$. 若 $\widehat{p}$ 是针对 p 的 ℓ-*NN* 密度估计量，且满足 $\ell \geq \lceil V_\tau$ 对数 $n\rceil$，并且 $\widehat{f}$ 是定义于(28) 且使用近邻函数 k 的局部 k-*NN* 估计量，其中 k 给定为

$$k(x) := n \wedge \left\lceil \kappa \log(n)^{d/(2\beta+d)}(n\widehat{p}(x))^{2\beta/(2\beta+d)} \right\rceil \vee \lceil \log n \rceil.$$

那么，存在一个 $N > 2d \vee e$，使得对所有 $n \geq N$，以至少 $1 - 3n^{1-\tau}$ 的概率，

$$\|\widehat{f} - f_*\|_{L^2(\mathrm{Q_X})}^2 \leq \left(\frac{V_\tau}{C}\right)^{r-2\beta/(2\beta+d)}\left[\frac{2C_\tau}{\kappa}(S_\tau)^{2\beta/(2\beta+d)} + A\right]\mathcal{T}_{\mathrm{P_X}}(r)\left(\frac{\log n}{n}\right)^r + \left[2C_\tau + 8L^2\right]\mathcal{T}_{\mathrm{P_X}}(r)\left(\frac{C\ell}{n}\right)^r.$$

with

$$A := 8L^2(2\kappa S_\tau)^{2\beta/d}\left(\frac{\theta}{c_1}\right)^{4\beta^2/(d(2\beta+d))}.$$

注记 30。 特别地，在定理 *29* 中选择 $\ell = \lceil V_\tau$ 对数 $n \rceil$ ，可得

$$\big\|\widehat{f} - f_*\big\|^2_{L^2(\mathrm{Q_X})} \le V_\tau^r \mathcal{T}_{\mathrm{P_X}}(r) \left[\frac{2C_\tau}{\kappa} \left(\frac{S_\tau C}{V_\tau} \right)^{r_\beta} + A \left(\frac{C}{C_\tau} \right)^{r_\beta} + 2^r \Big[2C_\tau + 8L^2 \Big] \right] \left(\frac{\log n}{n} \right)^r,$$

以至少 $1 - 3n^{1-\tau}$. 的概率。此外，在选择 $\kappa = 1.$ 后即可得到推论 *5* 的陈述

证明。 令 $X \sim \mathrm{Q_X}$, C 为引理 28 中定义的常数，并记 $\Delta = \Delta_n^{\mathrm{P_X}}(C, \ell)$。 我们在事件

$$L_n^{\mathrm{P_X}} \cap U_n^{\mathrm{P_X}}(\ell) \cap V_n^{\mathrm{P_X}}(\widehat{f}). \tag{29}$$

根据引理 28 (ii) $\Delta \subseteq \Theta_n^{\mathrm{P_X}}(S_\tau, k)$，并由定义可知， $k(x) \ge \lceil \log\, n \rceil$。因此，我们可以应用命题 25 得到

$$\big\|\widehat{f} - f_*\big\|^2_{L^2(\mathrm{Q_X})} \le \mathrm{E} \left[\frac{2C_\tau \log n}{k(X)} + 16L^2 \left(\frac{S_\tau k(X)}{np(X)} \right)^{2\beta/d} \mathbb{1}(X \in \Delta) + 8L^2 \mathbb{1}(X \in \Delta^c) \right]. \tag{30}$$

我们首先对式 (30) 中期望算子内出现的三个量逐点给出界。对于第一项，根据 k、引理 28 和引理 26 (ii) 的定义，我们得到

$$\begin{aligned} \frac{2C_\tau \log n}{k(x)} &\le \frac{2C_\tau \log n}{\kappa \log(n)^{d/(2\beta+d)} (n\widehat{p}(x))^{2\beta/(2\beta+d)}} \mathbb{1}(x \in \Delta) + \frac{2C_\tau \log n}{\lceil \log n \rceil} \mathbb{1}(x \in \Delta^c) \\ &\le \frac{2C_\tau}{\kappa} \left(\frac{S_\tau \log n}{np(x)} \right)^{2\beta/(2\beta+d)} \mathbb{1}(x \in \Delta) + 2C_\tau \mathbb{1}(x \in \Delta^c). \end{aligned} \tag{31}$$

根据引理 28 (i) 和条件 $n \ge N \vee e$, 我们有 $k(x) \ge \lceil \log n \rceil \ge 1.$。这使我们可以使用 $\lceil a \rceil \le 2a$ ，其中 a 是 $k(x)$ 的表达式中上取整算子内的表达式。 此外，我们使用引理 26 (i) 对式 (30) 中的第二项作如下界定

$$\begin{aligned} 8L^2 \left(\frac{S_\tau k(x)}{np(x)} \right)^{2\beta/d} &\le 8L^2 \big(S_\tau\big)^{2\beta/d} \left(\frac{2\kappa \log(n)^{d/(2\beta+d)} (n\widehat{p}(x))^{2\beta/(2\beta+d)}}{np(x)} \right)^{2\beta/d} \\ &\le 8L^2 \big(S_\tau\big)^{2\beta/d} \left(2\kappa \left(\frac{\theta}{c_1} \right)^{2\beta/(2\beta+d)} \left(\frac{\log(n)}{np(x)} \right)^{d/(2\beta+d)} \right)^{2\beta/d} \\ &\le \underbrace{8L^2 \big(2\kappa S_\tau\big)^{2\beta/d} \left(\frac{\theta}{c_1} \right)^{4\beta^2/(d(2\beta+d))}}_{=A} \left(\frac{\log(n)}{np(x)} \right)^{2\beta/(2\beta+d)}. \end{aligned} \tag{32}$$

现在，我们对先前推导出的界关于 $\mathrm{Q_X}$. 进行积分。我们先对 (31) 的第一项进行积分，得到:

$$\begin{aligned} \int_\Delta \frac{2C_\tau \log n}{k(X)} \, d\,\mathrm{Q_X}(x) &\le \frac{2C_\tau}{\kappa} \left(\frac{S_\tau \log n}{n} \right)^{2\beta/(2\beta+d)} \int_\Delta \frac{q(x)}{p(x)^r} p(x)^{r - 2\beta/(2\beta+d)} \, dx \\ &\le \frac{2C_\tau}{\kappa} \big(S_\tau\big)^{2\beta/(2\beta+d)} \mathcal{T}_{\mathrm{P_X}}(r) \left(\frac{\log n}{n} \right)^{2\beta/(2\beta+d)} \left(\frac{C\ell}{n} \right)^{r - 2\beta/(2\beta+d)} \\ &\overset{(i)}{\le} \frac{2C_\tau}{\kappa} \big(S_\tau\big)^{2\beta/(2\beta+d)} \mathcal{T}_{\mathrm{P_X}}(r) \left(\frac{V_\tau}{C} \right)^{r - 2\beta/(2\beta+d)} \left(\frac{\log n}{n} \right)^r, \end{aligned}$$

其中，步骤（i）由以下事实得出： $r - 2\beta/\left(2\beta + d\right) \le 0.$。对于（31）中的第二项，我们得到

$$\int_{\Delta^c} \frac{2C_\tau \log n}{k(x)} \, d\,\mathrm{Q_X}(x) \le 2C_\tau \mathcal{T}_{\mathrm{P_X}}(r) \left(\frac{C\ell}{n} \right)^r.$$

对 $\mathrm{Q_X}$ 积分（32），得到

$$\begin{aligned}\int_\Delta 8L^2\Big(\frac{S_\tau k(x)}{np(x)}\Big)^{2\beta/d} d\,\mathrm{Q_X}(x) &\le A\Big(\frac{\log n}{n}\Big)^{2\beta/(2\beta+d)}\int_\Delta \frac{q(x)}{p(x)^r}p(x)^{r-2\beta/(2\beta+d)}\,dx\\ &\le A\mathcal{T}_{\mathrm{P_X}}(r)\Big(\frac{\log n}{n}\Big)^{2\beta/(2\beta+d)}\Big(\frac{C\ell}{n}\Big)^{r-2\beta/(2\beta+d)}\\ &\le A\mathcal{T}_{\mathrm{P_X}}(r)\Big(\frac{V_\tau}{C}\Big)^{r-2\beta/(2\beta+d)}\Big(\frac{\log n}{n}\Big)^{r}\end{aligned}$$

我们现在对(30)中的最后一项给出界。我们利用在 $\Delta^c,\ p(x) < C\ell/n$ 上的事实，从而得到

$$8L^2\,\mathrm{Q_X}\{\Delta^c\} = 8L^2\int_{\Delta^c}\frac{q(x)}{p(x)^r}p(x)^r\,dx \le 8L^2\Big(\frac{C\ell}{n}\Big)^r\mathcal{T}_{\mathrm{P_X}}(r). \tag{33}$$

综合以上各项，可得

$$\begin{aligned}\|\widehat{f}-f_*\|^2_{L^2(\mathrm{Q_X})} \le \Big(\frac{V_\tau}{C}\Big)^{r-2\beta/(2\beta+d)}&\left[\frac{2C_\tau}{\kappa}\big(S_\tau\big)^{2\beta/(2\beta+d)}+A\right]\mathcal{T}_{\mathrm{P_X}}(r)\Big(\frac{\log n}{n}\Big)^r\\ &+\left[2C_\tau+8L^2\right]\mathcal{T}_{\mathrm{P_X}}(r)\Big(\frac{C\ell}{n}\Big)^r.\end{aligned}$$

最后，命题16、命题18、推论24以及备注 17连同并集界共同确保

$$\mathbb{P}\left\{L_n^{\mathrm{P_X}}(\ell)\cap U_n^{\mathrm{P_X}}(\ell)\cap V_n^{\mathrm{P_X}}(\widehat{f})\right\} \ge 1-2n^{-\tau}-n^{1-\tau} \ge 1-3n^{1-\tau}.$$

□

B.2 定理4的证明

为得到我们的主要结果，我们还需要一个引理，使我们能够对低密度区域中的 x给出 $\zeta_{h_}(x)$的下界。

引理 31. 设$\mathrm{P}\in\mathcal{P}(D,\theta)$，其中 D、 $\theta>0$和 $k\in\{1,\dots,n\}$.。若 $x\in$P的支撑集 满足 $p(x)\le\alpha k/n$，且 $\alpha>0$，则

$$\zeta_{h_}(x) \ge \left(\frac{c_1}{\alpha\theta}\right)^{1/d}\wedge 1.$$

证明。 若 $\zeta_{h_}(x)>1$，则结论显然。下面通过反证法证明在$\zeta_{h_}(x)\le 1$的情况下该结论成立。令 $t\in(0,1]$ 任意，并假设 $\zeta_{h_}(x)\le t^{1/d}\le 1$. 那么，由 (2) 可得$h_-=c_1k/n\le\theta tp(x)$。对后者进行整理并利用该假设，可得

$$\frac{c_1k}{\theta tn}\le p(x)\le\frac{\alpha k}{n},$$

这意味着 $t\ge c_1/(\alpha\theta)$。 因此

$$\zeta_{h_}(x) \ge \left(\frac{c_1}{\theta\alpha}\right)^{1/d}\wedge 1.$$

□

在证明主要结果之前，我们先回顾一下问题设定。设 $D,\ \theta>0$. 对于 $\mathrm{P_X},\ \mathrm{Q_X}\in\mathcal{P}(D,\ \theta)$，在协变量偏移 (CS) 下，我们假设模型 (1) 成立，其中未知函数 f_* 属于 $\mathcal{H}(L,\ \beta)$。我们可以访问

1. $\{(X_1,Y_1)\ldots,(X_n,Y_n)\}$, 来自模型 (1)、协变量分布为 $\mathrm{P_X}$ 的独立同分布 n-样本, 以及
2. $\{(X_1',Y_1')\ldots,(X_n',Y_n')\}$, 来自模型 (1)、协变量分布为 $\mathrm{Q_X}$. 的独立同分布 m-样本

下文中，我们使用如下估计量

$$\widehat{f}(x)=\frac{1}{k_{\mathrm{P}}(x)+k_{\mathrm{Q}}(x)}\Big(\sum_{i=1}^{k_{\mathrm{P}}(x)}Y_i(x)+\sum_{j=1}^{k_{\mathrm{Q}}(x)}Y_j'(x)\Big),\tag{34}$$

其中 $k_{\mathrm{P}}\colon\mathrm{R}^d\to\{1,\ldots n\},\ k_{\mathrm{Q}}\colon\mathrm{R}^d\to\{1,\ldots,m\}$, 且 $Y_i(x)$（相应地，$Y_j'(x)$）表示 x 在 $\{X_1,\ldots,X_n\}$（相应地，$\{X_1',\ldots,X_m'\}$）中的第 i 个（相应地，第 j 个）最近邻所关联的标签。

定理 32（双样本局部最近邻的速率）**.** 设 $\tau>0,\kappa_{\mathrm{P}},\kappa_{\mathrm{Q}}>0$, 并定义

$$r_S:=2\beta/(2\beta+d)\wedge\gamma,\ \textit{and}\ r_T:=s\wedge2\beta/(2\beta+d).$$

设 $\widehat{p}$ 和 $\widehat{q}$ 分别为 p 和 q 的 ℓ_{P} 和 ℓ_{Q}-近邻密度估计量，其中

$$\ell:=\ell_{\mathrm{P}}=\ell_{\mathrm{Q}}=\lceil V_\tau\log nm\rceil.$$

最后，考虑 f 为局部（$k_{\mathrm{P}},k_{\mathrm{Q}}$）-近邻估计量，其近邻函数定义为

$$k_{\mathrm{P}}(x):=n\wedge\Big\lceil\log(nm)^{d/(2\beta+d)}(n\widehat{p}(x))^{2\beta/(2\beta+d)}\Big\rceil\vee\lceil\log(nm)\rceil$$
$$k_{\mathrm{Q}}(x):=m\wedge\Big\lceil\log(nm)^{d/(2\beta+d)}(m\widehat{q}(x))^{2\beta/(2\beta+d)}\Big\rceil\vee\lceil\log(nm)\rceil.$$

那么，存在 $N\geq2d$ 和 $M\geq2d$，使得对所有 $n\geq N,m\geq M$，以至少 $1-3n^{1-\tau}-3m^{1-\tau}$的概率，有

$$\sup_{f\in\mathcal{H}(L,\beta)}\mathbb{P}\Big\{\mathcal{E}_{\mathrm{Q}}(\widehat{f},f)\geq C(\tau)R(n,m,\gamma,s)\Big\}\leq3(n^{1-\tau}+m^{1-\tau})$$

若 $(\gamma-r_\beta)\ \ (s-r_\beta)\ <0$ 且 $m\in[n,n^{\gamma/s}]$，则

$$R(n,m,\gamma,s)=\tilde{C}_0\Big(\frac{\log(nm)}{n}\Big)^{\gamma a}\Big(\frac{\log(nm)}{m}\Big)^{s(1-a)}\mathcal{T}_{\mathrm{P}}(\gamma)^a\mathcal{T}_{\mathrm{Q}}(s)^{1-a},$$

其中 C_0 是一个与 n,m,γ,s,τ无关的常数；若 $(\gamma-r_\beta)\ \ (s-r_\beta)\ \geq0$ 或 $m\notin[n,n^{\gamma/s}]$，则

$$R=\overline{C}_0\Big[\mathcal{T}_{\mathrm{P}}(r_S)\Big(\frac{\log(nm)}{n}\Big)^{r_S}\Big]\wedge\Big[\mathcal{T}_{\mathrm{Q}}(r_T)\Big(\frac{\log(nm)}{m}\Big)^{r_T}\Big],$$

其中，$\overline{C}_0$ 是一个与 n、m、γ、s、τ和 C_p 无关的常数：$=C(\kappa_{\mathrm{P}})\ \ C_q\colon=C(\kappa_{\mathrm{Q}})$，其中

$$C(t):=S_\tau\vee c_2\vee\frac{S_\tau}{t^{(2\beta+d)/(2\beta)}V_\tau}\vee\frac{(2S_\tau t)^{(2\beta+d)/d}}{V_\tau}\Big(\frac{\theta}{c_1}\Big)^{2\beta/d}.$$

证明。令

$$C(t):=S_\tau\vee c_2\vee\frac{S_\tau}{t^{(2\beta+d)/(2\beta)}V_\tau}\vee\frac{(2S_\tau t)^{(2\beta+d)/d}}{V_\tau}\Big(\frac{\theta}{c_1}\Big)^{2\beta/d}.$$

并定义 $C_p:=C(\kappa_{\mathrm{P}})\ \ C_q:=C(\kappa_{\mathrm{Q}})$。对于 $R\in\{\mathrm{P},\ \mathrm{Q}\}$，定义事件

$$E_n^{R_{\mathrm{X}}}(\ell,\widehat{f}_R):=L_n^{R_{\mathrm{X}}}\cap\mathrm{U}_n^{R_{\mathrm{X}}}(\ell)\cap V_n^{R_{\mathrm{X}}}(\widehat{f}_R).$$

我们在该事件上进行研究

$$E_n^{\mathrm{P}_X}(\ell_{\mathrm{P}}, \widehat{f}_{\mathrm{P}}) \cap E_m^{\mathrm{Q}_X}(\ell_{\mathrm{Q}}, \widehat{f}_{\mathrm{Q}}). \tag{35}$$

设 $X \sim \mathrm{Q}_X$ 与样本相互独立。我们首先将风险分解如下。

$$\begin{aligned}\big\|\widehat{f} - f_*\big\|_{L^2(\mathrm{Q}_X)}^2 &= \mathrm{E}_X\left[\left(\frac{1}{k_{\mathrm{P}}(X)+k_{\mathrm{Q}}(X)}\Big(\sum_{i=1}^{k_{\mathrm{P}}(X)}\big(Y_i(X)-f_*(X)\big)+\sum_{j=1}^{k_{\mathrm{Q}}(X)}\big(Y'_j(X)-f_*(X)\big)\Big)\right)^2\right]\\ &= \mathrm{E}\left[\left(\frac{k_{\mathrm{P}}(X)}{k_{\mathrm{P}}(X)+k_{\mathrm{Q}}(X)}\big(\widehat{f}_{\mathrm{P}}(X)-f_*(X)\big)+\frac{k_{\mathrm{Q}}}{k_{\mathrm{P}}(X)+k_{\mathrm{Q}}(X)}\big(\widehat{f}_{\mathrm{Q}}(X)-f_*(X)\big)\right)^2\right],\end{aligned}$$

其中 $\widehat{f}_{\mathrm{P}}(k_{\mathrm{P}}\ x\ /\ k_{\mathrm{P}}\ x\ +k_{\mathrm{Q}}\ x \qquad w_{\mathrm{Q}}\ x\ = k_{\mathrm{Q}}\ x\ /\ k_{\mathrm{P}}\ x\ +k_{\mathrm{Q}}\ x\ .$ $)\coloneqq\ k_{\mathrm{P}}$ $()^{-1}\sum^{k_{\mathrm{P}}(x)}{}_{i=1}\ \ Y_i()$，且 $\widehat{f}_{\mathrm{Q}}()\coloneqq\ k_{\mathrm{Q}}()^{-1}\sum^{k_{\mathrm{Q}}(x)}{}_{j=1}\ \ Y'_j()$。记 $()\colon()\,(\,()\,()$ 和 $()\colon()\,(\,()\,()$。再应用詹森不等式可得

$$\big\|\widehat{f} - f_*\big\|_{L^2(\mathrm{Q}_X)}^2 \le \mathrm{E}_X\left[w_{\mathrm{P}}(X)\big(\widehat{f}_{\mathrm{P}}(X)-f_*(X)\big)^2 + w_{\mathrm{Q}}(X)\big(\widehat{f}_{\mathrm{Q}}(X)-f_*(X)\big)^2\right]. \tag{36}$$

设 $\Delta_{\mathrm{P}}\coloneqq\Delta_n^{\mathrm{P}_X}(C_p,\ell_{\mathrm{P}})$ 和 $\Delta_{\mathrm{Q}}\coloneqq\Delta_m^{\mathrm{Q}_X}(C_q,\ell_{\mathrm{Q}})$。我们两次应用命题 25，得到对 $(\widehat{f}_{\mathrm{P}}(X)-f_*(X))^2$ 和 $(\widehat{f}_{\mathrm{Q}}(X)-f_*(X))^2$. 的界。在事件 (35) 上，根据所处区域的不同，我们可以用四种不同的方式分解 (36) 中期望里的误差。结果如下

$$\begin{aligned}\big(\widehat{f}_{\mathrm{P}}(x)-f_*(x)\big)^2 &\le \begin{cases} b_{\mathrm{P}}^2(x)+V_{\mathrm{P}}(x) & \text{if } x\in\Delta_{\mathrm{P}}\\ 8L^2+V_{\mathrm{P}}(x) & \text{if } x\in\Delta_{\mathrm{P}}^c\end{cases}\\ \big(\widehat{f}_{\mathrm{Q}}(x)-f_*(x)\big)^2 &\le \begin{cases} b_{\mathrm{Q}}^2(x)+V_{\mathrm{Q}}(x) & \text{if } x\in\Delta_{\mathrm{Q}}\\ 8L^2+V_{\mathrm{Q}}(x) & \text{if } x\in\Delta_{\mathrm{Q}}^c,\end{cases}\end{aligned} \tag{37}$$

其中

$$\begin{aligned} b_{\mathrm{P}}^2(x) &\coloneqq 8L^2\left(\frac{S_\tau k_{\mathrm{P}}(x)}{np(x)}\right)^{2\beta/d},\ V_{\mathrm{P}}(x)\coloneqq 2C_\tau\frac{\log n}{k_{\mathrm{P}}(x)},\\ b_{\mathrm{Q}}^2(x) &\coloneqq 8L^2\left(\frac{S_\tau k_{\mathrm{Q}}(x)}{mq(x)}\right)^{2\beta/d},\ V_{\mathrm{Q}}(x)\coloneqq 2C_\tau\frac{\log m}{k_{\mathrm{Q}}(x)},\end{aligned}$$

为了推导超额风险的界，我们进一步将 R^d 划分为四个区域。令

$$\Delta\coloneqq\Delta_{\mathrm{P}}\cap\Delta_{\mathrm{Q}},\ \Gamma_{\mathrm{P}}\coloneqq\Delta_{\mathrm{P}}\cap\Delta_{\mathrm{Q}}^c,\ \Gamma_{\mathrm{Q}}\coloneqq\Delta_{\mathrm{P}}^c\cap\Delta_{\mathrm{Q}},\ \text{and}\ \Theta\coloneqq\Delta_{\mathrm{P}}^c\cap\Delta_{\mathrm{Q}}^c.$$

现在，我们分别对这些区域上的 f 逐点风险进行界定。

区域 Δ。对于 $x\in\Delta$，总逐点风险为

$$\big(\widehat{f}(x)-f_*(x)\big)^2 \le w_{\mathrm{P}}(x)\big(b_{\mathrm{P}}(x)^2+V_{\mathrm{P}}(x)\big)+w_{\mathrm{Q}}(x)\big(b_{\mathrm{Q}}(x)^2+V_{\mathrm{Q}}(x)\big). \tag{38}$$

现在记 $r_\beta\coloneqq 2\beta/(2\beta+d)$ $\mathcal{T}_{\mathrm{P}}(u)\coloneqq\int q(x)/p(x)^u\,dx$ 以及 $\mathcal{T}_{\mathrm{Q}}(v)\coloneqq\int q(x)^{1-v}\,dx.$ 根据 k_{P}，k_{Q}，引理 28 和引理 26 (ii) 的定义，我们得到

$$\begin{aligned} w_{\mathrm{P}}(x)V_{\mathrm{P}}(x)+w_{\mathrm{Q}}(x)V_{\mathrm{Q}}(x) &= 2C_\tau\Big(\frac{\log n}{k_{\mathrm{P}}(x)+k_{\mathrm{Q}}(x)}+\frac{\log m}{k_{\mathrm{P}}(x)+k_{\mathrm{Q}}(x)}\Big)\\ &= 2C_\tau\frac{\log(nm)}{k_{\mathrm{P}}(x)+k_{\mathrm{Q}}(x)}\\ &= 2C_\tau\Big(\frac{k_{\mathrm{P}}(x)}{\log(nm)}+\frac{k_{\mathrm{Q}}(x)}{\log(nm)}\Big)^{-1}\end{aligned} \tag{39}$$

$$
\begin{aligned}
&= 2C_\tau \left(\kappa_{\mathrm{P}} \frac{\log(nm)^{1-r_\beta} (n\widehat{p}(x))^{r_\beta}}{\log(nm)} + \kappa_{\mathrm{Q}} \frac{\log(nm)^{1-r_\beta} (m\widehat{q}(x))^{r_\beta}}{\log(nm)} \right)^{-1} \\
&\leq 2C_\tau \left(\kappa_{\mathrm{P}} \left(\frac{np(x)}{S_\tau \log(nm)} \right)^{r_\beta} + \kappa_{\mathrm{Q}} \left(\frac{mq(x)}{S_\tau \log(nm)} \right)^{r_\beta} \right)^{-1} \\
&= 2C_\tau S_\tau^{r_\beta} \left(\kappa_{\mathrm{P}} \left(\frac{np(x)}{\log(nm)} \right)^{r_\beta} + \kappa_{\mathrm{Q}} \left(\frac{mq(x)}{\log(nm)} \right)^{r_\beta} \right)^{-1} \qquad (40)
\end{aligned}
$$

我们现在 界定式（38）中的其余项。根据 k_{P}，的定义以及引理 26，我们得到

$$
\begin{aligned}
k_{\mathrm{P}}(x) b_{\mathrm{P}}^2(x) = k_{\mathrm{P}}(x) \left(\frac{S_\tau k_{\mathrm{P}}(x)}{np(x)} \right)^{2\beta/d} &= S_\tau^{2\beta/d} \frac{k_{\mathrm{P}}(x)^{(2\beta+d)/d}}{(np(x))^{2\beta/d}} \\
&= S_\tau^{2\beta/d} \kappa_{\mathrm{P}}^{(2\beta+d)/d} \log(nm) \left(\frac{\widehat{p}(x)}{p(x)} \right)^{2\beta/d} \\
&\leq \kappa_{\mathrm{P}}^{(2\beta+d)/d} \left(\frac{\theta S_\tau}{c_1} \right)^{2\beta/d} \log(nm) := K_{\mathrm{P}} \log(nm).
\end{aligned}
$$

对 $b^2(x)$ 表达式中的第二项进行同样的分析，可得

$$
k_{\mathrm{Q}}(x) b_{\mathrm{Q}}^2(x) = k_{\mathrm{Q}}(x) \left(\frac{S_\tau k_{\mathrm{Q}}(x)}{mq(x)} \right)^{2\beta/d} \leq \kappa_{\mathrm{Q}}^{(2\beta+d)/d} \left(\frac{\theta S_\tau}{c_1} \right)^{2\beta/d} \log(nm) =: K_{\mathrm{Q}} \log(nm).
$$

因此，

$$
8L^2 (w_{\mathrm{P}}(x) b_{\mathrm{P}}^2(x) + w_{\mathrm{Q}}(x) b_{\mathrm{Q}}^2(x)) \leq 8L^2 (K_{\mathrm{P}} + K_{\mathrm{Q}}) \frac{\log(nm)}{k_{\mathrm{P}}(x) + k_{\mathrm{Q}}(x)} =: \overline{K} \frac{\log(nm)}{k_{\mathrm{P}}(x) + k_{\mathrm{Q}}(x)},
$$

其中 $\overline{K} := K_{\mathrm{P}} + K_{\mathrm{Q}}$. 后一量可以与 (39) 中完全相同的方式加以界定，这表明，对于 $x \in \Delta$,

$$
\begin{aligned}
\big(\widehat{f}(x) - f_*(x) \big)^2 &\leq (2C_\tau + 8L^2 \overline{K}) S_\tau^{r_\beta} \left(\kappa_{\mathrm{P}} \left(\frac{np(x)}{\log(nm)} \right)^{r_\beta} + \kappa_{\mathrm{Q}} \left(\frac{mq(x)}{\log(nm)} \right)^{r_\beta} \right)^{-1} \\
&:= C_\Delta \left(\kappa_{\mathrm{P}} \left(\frac{np(x)}{\log(nm)} \right)^{r_\beta} + \kappa_{\mathrm{Q}} \left(\frac{mq(x)}{\log(nm)} \right)^{r_\beta} \right)^{-1} \\
&\leq C_\Delta \left(\kappa_{\mathrm{P}}^{-1} \Big(\frac{\log(nm)}{np(x)} \Big)^{r_\beta} \wedge \kappa_{\mathrm{Q}}^{-1} \Big(\frac{\log(nm)}{mq(x)} \Big)^{r_\beta} \right)
\end{aligned}
$$

区域 Γ_{P} **和** Γ_{Q}. 对于 $x \in \Gamma_{\mathrm{P}}$，我们有

$$
\big(\widehat{f}(x) - f_*(x) \big)^2 \leq w_{\mathrm{P}} \big(b_{\mathrm{P}}^2(x) + V_{\mathrm{P}}(x) \big)^2 + w_{\mathrm{Q}}(x) \big(8L^2 + V_{\mathrm{Q}}(x) \big).
$$

由于 $x \in \Delta_{\mathrm{P}}$, 前一种情形中关于 $w_{\mathrm{P}}(x)(\widehat{f}_{\mathrm{P}}(x) - f_*(x))^2$ 的界仍然成立，因此我们得到

$$
\big(\widehat{f}(x) - f_*(x) \big)^2 \leq \frac{(K_{\mathrm{P}} + 2C_\tau) \log(nm) + 8L^2 k_{\mathrm{Q}}(x)}{k_{\mathrm{P}}(x) + k_{\mathrm{Q}}(x)}.
$$

回顾 t hat $x \in \Delta_{\mathrm{Q}}^{\mathrm{c}}$，这意味着 $q(x) \leq C_q \ell_{\mathrm{Q}} / m$，因此，根据引理 31，有 t

$$
\widehat{q}(x) = \frac{\ell_{\mathrm{Q}}}{m R_{\ell_{\mathrm{Q}},\mathrm{Q}}(x)^d} \leq \frac{\ell_{\mathrm{Q}}}{m \zeta_{h_-,\mathrm{Q}}(x)^d} \leq \frac{a\theta}{c_1} \frac{\ell_{\mathrm{Q}}}{m}. \qquad (41)
$$

因此，回顾 $\ell_{\mathrm{P}} = \ell_{\mathrm{Q}} = \lceil V_\tau \log(nm) \rceil$,

$$\begin{aligned}
\big(\widehat{f}(x) - f_*(x)\big)^2 &\le \frac{(K_{\mathrm{P}} + 2C_\tau)\log(nm) + 16L^2\kappa_{\mathrm{Q}}(a\theta/c_1)^{r_\beta}\log(nm)^{1-r_\beta}\ell_{\mathrm{Q}}^{r_\beta}}{k_{\mathrm{P}}(x) + k_{\mathrm{Q}}(x)} \\
&\le (K_{\mathrm{P}} + 2C_\tau + 16L^2V_\tau^{r_\beta}\kappa_{\mathrm{Q}}(a\theta/c_1)^{r_\beta})\frac{\log(nm)}{k_{\mathrm{P}}(x) + k_{\mathrm{Q}}(x)} \\
&\le (K_{\mathrm{P}} + 2C_\tau + 16L^2V_\tau^{r_\beta}\kappa_{\mathrm{Q}}(a\theta/c_1)^{r_\beta})\frac{\log(nm)}{k_{\mathrm{P}}(x)} \\
&\le \frac{K_{\mathrm{P}} + 2C_\tau + 16L^2V_\tau^{r_\beta}\kappa_{\mathrm{Q}}(a\theta/c_1)^{r_\beta}}{\kappa_{\mathrm{P}}}\left(\frac{\log(nm)}{n\widehat{p}(x)}\right)^{r_\beta} \\
&\le \frac{(K_{\mathrm{P}} + 2C_\tau + 16L^2V_\tau^{r_\beta}\kappa_{\mathrm{Q}}(a\theta/c_1)^{r_\beta})S_\tau^{r_\beta}}{\kappa_{\mathrm{P}}}\left(\frac{\log(nm)}{np(x)}\right)^{r_\beta} =: C_{\Gamma_{\mathrm{P}}}\left(\frac{\log(nm)}{np(x)}\right)^{r_\beta}.
\end{aligned}$$

在区域 Γ_{Q} 上的分析是对称的，因此对于 $x \in \Gamma_{\mathrm{Q}}$，我们得到

$$\big(\widehat{f}(x) - f_*(x)\big)^2 \le C_{\Gamma_{\mathrm{Q}}}\left(\frac{\log(nm)}{mq(x)}\right)^{r_\beta}, \quad \text{with } C_{\Gamma_{\mathrm{Q}}} := \frac{(K_{\mathrm{Q}} + 2C_\tau + 16L^2V_\tau^{r_\beta}\kappa_{\mathrm{P}}(a\theta/c_1)^{r_\beta})S_\tau^{r_\beta}}{\kappa_{\mathrm{Q}}}.$$

区域 Θ. 对于 $x \in \Theta$，我们利用定义可知这一事实： $k_{\mathrm{P}}(x) \wedge k_{\mathrm{Q}}(x) \ge \log(nm)$，从而得到

$$\begin{aligned}
\big(\widehat{f}(x) - f_*(x)\big)^2 &\le 8L^2 + w_{\mathrm{P}}(x)V_{\mathrm{P}}(x) + w_{\mathrm{Q}}(x)V_{\mathrm{Q}}(x) \\
&\le 8L^2 + 2C_\tau\frac{\log(nm)}{k_{\mathrm{P}}(x) + k_{\mathrm{Q}}(x)} \\
&\le 8L^2 + C_\tau =: C_\theta
\end{aligned}$$

这就完成了对 R^d.划分上的逐点界。接下来只需将后者对 Q_{X}. 积分。我们将分别在各个区域内进行积分。对于每个区域，我们将区分**楔形区域**和**加速区域**两种情况。定义

$$\begin{aligned}
I_\Delta &:= \int_\Delta \left(\kappa_{\mathrm{P}}^{-1}\Big(\frac{1}{np(x)}\Big)^{r_\beta} \wedge \kappa_{\mathrm{Q}}^{-1}\Big(\frac{1}{mq(x)}\Big)^{r_\beta}\right) d\,\mathrm{Q}_{\mathrm{X}}(x) \\
I_{\Gamma_{\mathrm{P}}} &:= \int_{\Gamma_{\mathrm{P}}} \left(\frac{1}{np(x)}\right)^{r_\beta} d\,\mathrm{Q}_{\mathrm{X}} \\
I_{\Gamma_{\mathrm{Q}}} &:= \int_{\Gamma_{\mathrm{Q}}} \left(\frac{1}{mq(x)}\right)^{r_\beta} d\,\mathrm{Q}_{\mathrm{X}}(x) \\
I_\Theta &:= \int_\Theta d\,\mathrm{Q}_{\mathrm{X}}(x).
\end{aligned}$$

采用这些记号，超额风险的最终界为

$$\big\|\widehat{f} - f_*\big\|_{L^2(\mathrm{Q}_{\mathrm{X}})}^2 \le \log(nm)^{r_\beta}\big(C_\Delta I_\Delta + C_{\Gamma_{\mathrm{P}}}I_{\Gamma_{\mathrm{P}}} + C_{\Gamma_{\mathrm{Q}}}I_{\Gamma_{\mathrm{Q}}}\big) + C_\Theta I_\Theta.$$

关于 Δ.我们写作

$$\begin{aligned}
I_\Delta &\le \inf_{\theta\in[0,1]} \int_\Delta (\kappa_{\mathrm{P}}(np(x))^{r_\beta})^{-\theta}(\kappa_{\mathrm{Q}}(mq(x))^{r_\beta})^{1-\theta}\, d\,\mathrm{Q}_{\mathrm{X}}(x) \\
&\le \inf_{\theta\in[0,1]} (\kappa_{\mathrm{P}}n^{r_\beta})^{-\theta}(\kappa_{\mathrm{Q}}m^{r_\beta})^{-(1-\theta)} \int_\Delta p(x)^{-r_\beta\theta}q(x)^{-r_\beta(1-\theta)}\, d\,\mathrm{Q}_{\mathrm{X}} \\
&\le \inf_{\theta\in[0,1]} (\kappa_{\mathrm{P}}n^{r_\beta})^{-\theta}(\kappa_{\mathrm{Q}}m^{r_\beta})^{-(1-\theta)}\, \mathrm{E}_{X\sim\mathrm{Q}_{\mathrm{X}}}\left[p(X)^{-r_\beta\theta}q(X)^{-r_\beta(1-\theta)}\right]
\end{aligned}$$

在**楔形区域内,** 只需利用以下事实： $r_S - r_\beta \le 0$, $r_M - r_\beta \le 0$ 并且，在Δ, $p(x) \ge C_p\ell_{\mathrm{P}}/n$ 以及 $q(x) \ge C_q\ell_{\mathrm{Q}}/m$ 和

$$\begin{aligned} I_\Delta &\le \left[\frac{1}{\kappa_{\mathrm{P}} n^{r_\beta}} \int_\Delta \frac{q(x)}{p(x)^{r_\beta}}\, dx\right] \wedge \left[\frac{1}{\kappa_{\mathrm{Q}} m^{r_\beta}} \int_\Delta q(x)^{1-r_\beta}\, dx\right] \\ &\le \left[\frac{1}{\kappa_{\mathrm{P}} n^{r_\beta}} \int_\Delta \frac{q(x)}{p(x)^{r_S}} p(x)^{r_S - r_\beta}\, dx\right] \wedge \left[\frac{1}{\kappa_{\mathrm{Q}} m^{r_\beta}} \int_\Delta q(x)^{1-r_\beta}\, dx\right] \\ &\le \left[\frac{1}{\kappa_{\mathrm{P}} n^{r_S}} (C_p\ell_{\mathrm{P}})^{r_S - r_\beta} \mathcal{T}_{\mathrm{P}}(r_S)\right] \wedge \left[\frac{1}{\kappa_{\mathrm{Q}} m^{r_T}} (C_q\ell_{\mathrm{Q}})^{r_T - r_\beta} \mathcal{T}_{\mathrm{Q}}(r_T)\right]. \end{aligned}$$

在**加速区域**中，我们可以改取 $\theta = \gamma(s-r)/(r_\beta(s-\gamma))$，从而得到

$$\begin{aligned} I_\Delta &\le (\kappa_{\mathrm{P}} n^{r_\beta})^{-\theta} (\kappa_{\mathrm{Q}} m^{r_\beta})^{-(1-\theta)}\, \mathrm{E}_{X\sim \mathrm{Q_X}}\left[p(X)^{-r_\beta\theta} q(X)^{-r_\beta(1-\theta)}\right] \\ &\le \kappa_{\mathrm{P}}^{-\theta} \kappa_{\mathrm{Q}}^{-(1-\theta)} n^{-\frac{\gamma(r_\beta - s)}{\gamma - s}} m^{-\frac{s(\gamma - r)}{\gamma - s}}\, \mathrm{E}_{X\sim \mathrm{Q_X}}\left[p(X)^{-r_\beta\theta} q(X)^{-r_\beta(1-\theta)}\right]. \end{aligned}$$

还需验证上述期望是有限的。我们对共轭指数 $a := \gamma/(\theta r_\beta)$，和 $b := s/((1-\theta)r_\beta)$，应用霍尔德不等式，得到

$$\mathrm{E}_{X\sim \mathrm{Q_X}}\left[p(X)^{-r_\beta\theta} q(X)^{-r_\beta(1-\theta)}\right] \le \left(\mathrm{E}_{X\sim \mathrm{Q_X}}\left[p(X)^{-r_\beta\theta a}\right]\right)^{1/a} \left(\mathrm{E}_{X\sim \mathrm{Q_X}}\left[q(X)^{-r_\beta(1-\theta)b}\right]\right)^{1/b},$$

并且，由于 $r_\beta\theta a = \gamma$ 和 $r_\beta(1-\theta)b = s$，我们得出

$$I_\Delta \le \kappa_{\mathrm{P}}^{-\theta} \kappa_{\mathrm{Q}}^{-(1-\theta)} n^{-\frac{\gamma(r_\beta - s)}{\gamma - s}} m^{-\frac{s(\gamma - r)}{\gamma - s}} \mathcal{T}_{\mathrm{P}}(\gamma)^{\theta r_\beta/\gamma} \mathcal{T}_{\mathrm{Q}}(s)^{(1-\theta)r_\beta/s}.$$

关于 Γ_{P}.关于 Γ_{P}，根据定义，我们有

$$np(x) \ge C_p\ell_{\mathrm{P}}.$$

因此，在**楔形区域**中，有

$$I_{\Gamma_{\mathrm{P}}} = \int_{\Gamma_{\mathrm{P}}} (np(x))^{-r_\beta}\, d\,\mathrm{Q_X} \le \int_{\Gamma_{\mathrm{P}}} \left[(C_p\ell_{\mathrm{P}})^{-r_\beta} \wedge (np(x))^{-r_\beta}\right] d\,\mathrm{Q_X}.$$

由于对于 $x \in \Gamma_{\mathrm{P}}$, $q(x) \le C_q\ell_{\mathrm{Q}}/m$，且 $p(x) \ge C_p\ell_{\mathrm{P}}/n$，有

$$\begin{aligned} I_{\Gamma_{\mathrm{P}}} &\le \left[\frac{(C_p\ell_{\mathrm{P}})^{r_S - r_\beta}}{n^{r_S}} \int_{\Gamma_{\mathrm{P}}} p(x)^{-r_S}\, d\,\mathrm{Q_X}(x)\right] \wedge \left[(C_p\ell_{\mathrm{P}})^{-r_\beta} \int_{\Gamma_{\mathrm{P}}} \frac{q(x)}{q(x)^{r_T}} q(x)^{r_T}\, dx\right] \\ &\le \left[\frac{(C_p\ell_{\mathrm{P}})^{r_S - r_\beta}}{n^{r_S}} \mathcal{T}_{\mathrm{P}}(r_S)\right] \wedge \left[(C_p\ell_{\mathrm{P}})^{-r_\beta} \int_{\Gamma_{\mathrm{P}}} \frac{q(x)}{q(x)^{r_T}} q(x)^{r_T}\, dx\right] \\ &\le \left[\frac{(C_p\ell_{\mathrm{P}})^{r_S - r_\beta}}{n^{r_S}} \mathcal{T}_{\mathrm{P}}(r_S)\right] \wedge \left[(C_p\ell_{\mathrm{P}})^{-r_\beta} \Big(\frac{C_q\ell_{\mathrm{Q}}}{m}\Big)^{r_T} \mathcal{T}_{\mathrm{Q}}(r_T)\right], \end{aligned}$$

这给出了**楔形区域**中的界。接下来，在**加速区域** 中，若 $s < r_\beta < \gamma$，或 $\gamma < r_\beta < s$，则我们改为写作

$$\begin{aligned} I_{\Gamma_{\mathrm{P}}} &= n^{-r_\beta} \int_{\Gamma_{\mathrm{P}}} \frac{q(x)}{p(x)^{r_\beta}}\, dx \\ &= n^{-r_\beta} \int_{\Gamma_{\mathrm{P}}} \Big(\frac{q(x)}{p(x)^\gamma}\Big)^a q(x)^{(1-s)(1-a)} p(x)^{-(r_\beta - \gamma a)} q(x)^{s(1-a)}\, dx, \end{aligned}$$

with

$$a := \tfrac{r_\beta - s}{\gamma - s} \in (0,1). \tag{42}$$

我们验证

$$r_\beta - \gamma a = \frac{r_\beta(\gamma - s) - \gamma(r_\beta - s)}{\gamma - s} = s\frac{\gamma - r_\beta}{\gamma - s} = s(1-a) > 0.$$

因此，根据 Γ_{P} 的定义，在应用共轭指数为 $1/a$ 和 $1/$的霍尔德不等式（$1-a$）之前，我们得到

$$\begin{aligned} I_{\Gamma_{\mathrm{P}}} &\le (\ell_{\mathrm{P}} C_{\mathrm{P}})^{\gamma a - r_\beta} (\ell_{\mathrm{Q}} C_q)^{s(1-a)} n^{-\gamma a} m^{-s(1-a)} \int_{\Gamma_{\mathrm{P}}} \Big(\frac{q(x)}{p(x)^\gamma}\Big)^a q(x)^{(1-s)(1-a)}\, dx \\ &\le (\ell_{\mathrm{P}} C_p)^{\gamma a - r_\beta} (\ell_{\mathrm{Q}} C_q)^{s(1-a)} n^{-\gamma a} m^{-s(1-a)} \mathcal{T}_{\mathrm{P}}(\gamma)^a \mathcal{T}_{\mathrm{Q}}(s)^{1-a}, \end{aligned}$$

由此得到加速区域中的界。

关于 Γ_{Q}的界。 此情形与关于 Γ_{P}. 的界对称。按照完全相同的步骤，并交换 P_{X} 和 Q_{X} 的角色，可得如下结果。在**楔形区域**中，我们得到界

$$I_{\Gamma_{\mathrm{Q}}} \le \Big[\frac{(C_q \ell_{\mathrm{Q}})^{r_T - r_\beta}}{m^{r_T}} \mathcal{T}_{\mathrm{Q}}(r_T)\Big] \wedge \Big[(C_q \ell_{\mathrm{Q}})^{-r_\beta} \Big(\frac{C_p \ell_{\mathrm{P}}}{n}\Big)^{r_S} \mathcal{T}_{\mathrm{P}}(r_S)\Big],$$

而在**加速区域**中，我们则改写为

$$\begin{aligned} I_{\Gamma_{\mathrm{Q}}} &= m^{-r_\beta} \int_{\Gamma_{\mathrm{Q}}} \frac{q(x)}{q(x)^{r\beta}}\, dx \\ &= m^{-r_\beta} \int_{\Gamma_{\mathrm{Q}}} \Big(\frac{q(x)}{p(x)^\gamma}\Big)^a q(x)^{(1-s)(1-a)} p(x)^{\gamma a} q(x)^{-(r_\beta - s(1-a))}\, dx. \end{aligned}$$

那么，我们已经知道 $r_\beta - s(1-a) = \gamma a > 0$，因此，在 Γ_{Q}上，

$$q(x)^{-(r_\beta - s(1-a))} \le (m/(C_q \ell_{\mathrm{Q}}))^{r_\beta - s(1-a)},$$

由此即可得出结论

$$I_{\Gamma_{\mathrm{Q}}} \le (C_q \ell_{\mathrm{Q}})^{s(1-a) - r_\beta} (C_p \ell_{\mathrm{P}})^{\gamma a} n^{-\gamma a} m^{-s(1-a)} \mathcal{T}_{\mathrm{P}}(\gamma)^a \mathcal{T}_{\mathrm{Q}}(s)^{1-a}.$$

上界 Θ. 我们旨在给出上界

$$I_\Theta = \int_\Theta q(x)\, dx.$$

我们再次区分这两个参数范围。在**楔形区域**中，利用如下事实：如果 $\gamma > 0$，那么 $\mathrm{P}_{\mathrm{X}} \ll \mathrm{Q}_{\mathrm{X}}$，则有

$$\begin{aligned} I_\Theta = \int_\Theta q(x)\, dx &= \int_\Theta \Big(p(x)^{r_S} \frac{q(x)}{p(x)^{r_S}}\Big) \wedge \Big(q(x)^{r_T} \frac{q(x)}{q(x)^{r_T}}\Big)\, dx \\ &\le \Big(\int_\Theta p(x)^{r_S} \frac{q(x)}{p(x)^{r_S}}\, dx\Big) \wedge \Big(q(x)^{r_T} \frac{q(x)}{q(x)^{r_T}}\, dx\Big) \\ &\le \Big[\Big(\frac{C_p \ell_{\mathrm{P}}}{n}\Big)^{r_S} \mathcal{T}_{\mathrm{P}}(r_S)\Big] \wedge \Big[\Big(\frac{C_q \ell_{\mathrm{Q}}}{m}\Big)^{r_T} \mathcal{T}_{\mathrm{Q}}(r_T)\Big], \end{aligned}$$

其中我们还用到了如下定义：若 $x \in \Theta$，则 $p(x) < C_p \ell_{\mathrm{P}}/n$且 $q(x) < C_q \ell_{\mathrm{Q}}/m$. 在**加速区域**中，我们则改为按如下方式进行

$$I_\Theta = \int_\Theta q(x)\, dx \le \int_\Theta \Big(\frac{q(x)}{p(x)^\gamma}\Big)^a \Big(\frac{q(x)}{q(x)^s}\Big)^{1-a} p(x)^{\gamma a} q(x)^{s(1-a)}\, dx$$

$$\leq (C_p\ell_{\mathrm{P}})^{\gamma a}(C_q\ell_{\mathrm{Q}})^{s(1-a)}n^{-\gamma a}m^{-s(1-a)}\mathcal{T}_{\mathrm{P}}(\gamma)^a\mathcal{T}_{\mathrm{Q}}(s)^{1-a},$$

同时选择如式（42）中的 a。

结论。 接下来需要汇总各项，从而在**楔形区域**中得到：

$$\begin{aligned}
\big\|\widehat{f}-f_*\big\|_{L^2(\mathrm{Q_X})}^2 \leq\ & C_\Delta \log(nm)^{r_\beta}\left[\frac{1}{\kappa_{\mathrm{P}}n^{r_S}}(C_p\ell_{\mathrm{P}})^{r_S-r_\beta}\mathcal{T}_{\mathrm{P}}(r_S)\right]\wedge\left[\frac{1}{\kappa_{\mathrm{Q}}m^{r_T}}(C_q\ell_{\mathrm{Q}})^{r_T-r_\beta}\mathcal{T}_{\mathrm{Q}}(r_T)\right]\\
&+C_{\Gamma_{\mathrm{P}}}\log(nm)^{r_\beta}\left[\frac{(C_p\ell_{\mathrm{P}})^{r_S-r_\beta}}{n^{r_S}}\mathcal{T}_{\mathrm{P}}(r_S)\right]\wedge\left[(C_p\ell_{\mathrm{P}})^{-r_\beta}\Big(\frac{C_q\ell_{\mathrm{Q}}}{m}\Big)^{r_T}\mathcal{T}_{\mathrm{Q}}(r_T)\right]\\
&+C_{\Gamma_{\mathrm{Q}}}\log(nm)^{r_\beta}\left[(C_q\ell_{\mathrm{Q}})^{-r_\beta}\Big(\frac{C_p\ell_{\mathrm{P}}}{n}\Big)^{r_S}\mathcal{T}_{\mathrm{P}}(r_S)\right]\wedge\left[\frac{(C_q\ell_{\mathrm{Q}})^{r_T-r_\beta}}{m^{r_T}}\mathcal{T}_{\mathrm{Q}}(r_T)\right]\\
&+C_\Theta\Big[\Big(\frac{C_p\ell_{\mathrm{P}}}{n}\Big)^{r_S}\mathcal{T}_{\mathrm{P}}(r_S)\Big]\wedge\Big[\Big(\frac{C_q\ell_{\mathrm{Q}}}{m}\Big)^{r_T}\mathcal{T}_{\mathrm{Q}}(r_T)\Big].
\end{aligned}$$

取 $\kappa_{\mathrm{P}}=\kappa_{\mathrm{Q}}=1$，并回忆 $\ell_{\mathrm{P}}=\ell_{\mathrm{Q}}=\lceil V_\tau\log(nm)\rceil$，首先可得 $C_p=C_q=:C_0$，以及

$$\begin{aligned}
\big\|\widehat{f}-f_*\big\|_{L^2(\mathrm{Q_X})}^2 \leq\ & C_\Delta\left[(C_0V_\tau)^{r_S-r_\beta}\mathcal{T}_{\mathrm{P}}(r_S)\Big(\frac{\log(nm)}{n}\Big)^{r_S}\right]\wedge\left[(C_0V_\tau)^{r_T-r_\beta}\mathcal{T}_{\mathrm{Q}}(r_T)\Big(\frac{\log nm}{m}\Big)^{r_T}\right]\\
&+C_{\Gamma_{\mathrm{P}}}\left[(C_0V_\tau)^{r_S-r_\beta}\mathcal{T}_{\mathrm{P}}(r_S)\Big(\frac{\log(nm)}{n}\Big)^{r_S}\right]\wedge\left[(C_0V_\tau)^{r_T-r_\beta}\mathcal{T}_{\mathrm{Q}}(r_T)\Big(\frac{\log(nm)}{m}\Big)^{r_T}\right]\\
&+C_{\Gamma_{\mathrm{Q}}}\left[(C_0V_\tau)^{r_S-r_\beta}\mathcal{T}_{\mathrm{P}}(r_S)\Big(\frac{\log(nm)}{n}\Big)^{r_S}\right]\wedge\left[(C_0V_\tau)^{r_T-r_\beta}\mathcal{T}_{\mathrm{Q}}(r_T)\Big(\frac{\log(nm)}{m}\Big)^{r_T}\right]\\
&+C_\Theta\Big[(C_0V_\tau)^{r_S}\mathcal{T}_{\mathrm{P}}(r_S)\Big(\frac{\log(nm)}{n}\Big)^{r_S}\Big]\wedge\Big[(C_0V_\tau)^{r_T}\mathcal{T}_{\mathrm{Q}}(r_T)\Big(\frac{\log(nm)}{m}\Big)^{r_T}\Big]\\
\leq\ & \overline{C}_0\Big[\mathcal{T}_{\mathrm{P}}(r_S)\Big(\frac{\log(nm)}{n}\Big)^{r_S}\Big]\wedge\Big[\mathcal{T}_{\mathrm{Q}}(r_T)\Big(\frac{\log(nm)}{m}\Big)^{r_T}\Big],
\end{aligned}$$

其中 $\overline{C}_0$ 为上述所有常数之和。在**加速区域**中，汇总各项可得

$$\begin{aligned}
\big\|\widehat{f}-f_*\big\|_{L^2(\mathrm{Q_X})}^2 \leq\ & C_\Delta\log(nm)^{r_\beta}\kappa_{\mathrm{P}}^{-\gamma a/r_\beta}\kappa_{\mathrm{Q}}^{-s(1-a)/r_\beta}n^{-\gamma a}m^{-s(1-a)}\mathcal{T}_{\mathrm{P}}(\gamma)^a\mathcal{T}_{\mathrm{Q}}(s)^{1-a}\\
&+C_{\Gamma_{\mathrm{P}}}\log(nm)^{r_\beta}(\ell_{\mathrm{P}}C_p)^{\gamma a-r_\beta}(\ell_{\mathrm{Q}}C_q)^{s(1-a)}n^{-\gamma a}m^{-s(1-a)}\mathcal{T}_{\mathrm{P}}(\gamma)^a\mathcal{T}_{\mathrm{Q}}(s)^{1-a}\\
&+C_{\Gamma_{\mathrm{Q}}}\log(nm)^{r_\beta}(C_q\ell_{\mathrm{Q}})^{s(1-a)-r_\beta}(C_p\ell_{\mathrm{P}})^{\gamma a}n^{-\gamma a}m^{-s(1-a)}\mathcal{T}_{\mathrm{P}}(\gamma)^a\mathcal{T}_{\mathrm{Q}}(s)^{1-a}\\
&+C_\Theta(C_p\ell_{\mathrm{P}})^{\gamma a}(C_q\ell_{\mathrm{Q}})^{s(1-a)}n^{-\gamma a}m^{-s(1-a)}\mathcal{T}_{\mathrm{P}}(\gamma)^a\mathcal{T}_{\mathrm{Q}}(s)^{1-a}.
\end{aligned}$$

取 $\kappa_{\mathrm{P}}=\kappa_{\mathrm{Q}}=1$，并回忆 $\ell_{\mathrm{P}}=\ell_{\mathrm{Q}}=\lceil V_\tau\log(nm)\rceil$，可得 $C_p=C_q$，以及

$$\begin{aligned}
\big\|\widehat{f}-f_*\big\|_{L^2(\mathrm{Q_X})}^2 \leq\ & C_\Delta\Big(\frac{\log(nm)}{n}\Big)^{\gamma a}\Big(\frac{\log(nm)}{m}\Big)^{s(1-a)}\mathcal{T}_{\mathrm{P}}(\gamma)^a\mathcal{T}_{\mathrm{Q}}(s)^{1-a}\\
&+C_{\Gamma_{\mathrm{P}}}\Big(\frac{\log(nm)}{n}\Big)^{\gamma a}\Big(\frac{\log(nm)}{m}\Big)^{s(1-a)}\mathcal{T}_{\mathrm{P}}(\gamma)^a\mathcal{T}_{\mathrm{Q}}(s)^{1-a}\\
&+C_{\Gamma_{\mathrm{Q}}}\Big(\frac{\log(nm)}{n}\Big)^{\gamma a}\Big(\frac{\log(nm)}{m}\Big)^{s(1-a)}\mathcal{T}_{\mathrm{P}}(\gamma)^a\mathcal{T}_{\mathrm{Q}}(s)^{1-a}\\
&+C_\Theta(C_0V_\tau)^{r_\beta}\Big(\frac{\log(nm)}{n}\Big)^{\gamma a}\Big(\frac{\log(nm)}{m}\Big)^{s(1-a)}\mathcal{T}_{\mathrm{P}}(\gamma)^a\mathcal{T}_{\mathrm{Q}}(s)^{1-a}\\
=\ & \tilde{C}_0\Big(\frac{\log(nm)}{n}\Big)^{\gamma a}\Big(\frac{\log(nm)}{m}\Big)^{s(1-a)}\mathcal{T}_{\mathrm{P}}(\gamma)^a\mathcal{T}_{\mathrm{Q}}(s)^{1-a},
\end{aligned}$$

将 $a=\frac{r_\beta-s}{\gamma-s}$. 替换后，即可得到所声称的上界。回想一下，我们讨论的是事件 $E_n^{\mathrm{P_X}}(\ell_{\mathrm{P}},\ \widehat{f}_{\mathrm{P}})\cap E_m^{\mathrm{Q_X}}(\ell_{\mathrm{Q}},\ \widehat{f}_{\mathrm{Q}})$。因此，接下来只需利用命题16和命题18、推论24，

以及备注 17，并结合并集界，即可得到

$$\mathbb{P}\left\{E_n^{\mathrm{P_X}}(\ell_{\mathrm{P}},\widehat{f}_{\mathrm{P}})\cap E_m^{\mathrm{Q_X}}(\ell_{\mathrm{Q}},\widehat{f}_{\mathrm{Q}})\right\}\geq 1-2n^{-\tau}-n^{1-\tau}-2m^{-\tau}-m^{1-\tau}\geq 1-3n^{1-\tau}-3m^{1-\tau}.$$

□

B.3 定理 6的证明

定理 33. 在(CS) 下并在回归模型(1) 中，噪声为高斯噪声 $\varepsilon\sim\mathcal{N}(0,\sigma_e^2)$，并有 $\sigma_e^2>0$.。设 L、 $D>0$、 $\theta\geq$勒贝格测度(B(0, 1)) $\beta\in(0,1]$ 以及$(\gamma,s)\in(0,\infty)\times(0,1]$。则存在常数 $C_0>0$，与 n、 m 无关，使得对所有 $n\vee m>1$,

$$\inf_{\widehat{f}}\sup_{\substack{f_*\in\mathcal{H}(L,\beta)\\ \mathrm{P_X},\mathrm{Q_X}\in\underline{\mathcal{P}}(D,\theta,\gamma,s)}}\mathrm{E}\left[\|\widehat{f}-f_*\|^2_{L^2(\mathrm{Q_X})}\right]\geq\begin{cases}C_0 m^{-s\frac{\gamma-r_\beta}{\gamma-s}}n^{-\gamma\frac{r_\beta-s}{\gamma-s}} & \text{if } (\gamma-r_\beta)(s-r_\beta)<0,\\ & \text{and } m\in[n,n^{\gamma/s}]\\ C_0\left[n^{-(\gamma\wedge r_\beta)}\wedge m^{-(s\wedge r_\beta)}\right] & \text{else.}\end{cases}$$

其中下确界 取遍由 n 个来自$\mathrm{P_{X,Y}}$ 以及 m 个来自$\mathrm{Q_{X,Y}}$.的样本所得到的所有估计量

定理 *6*的证明。设 α_{P}、 $\alpha_{\mathrm{Q}}>0$ 并定义

$$\gamma=\frac{\alpha_{\mathrm{Q}}}{\alpha_{\mathrm{P}}+1},\quad\text{and}\quad s=\frac{\alpha_{\mathrm{Q}}}{\alpha_{\mathrm{Q}}+1}.$$

考虑分布$\mathrm{P_X}$ 与$\mathrm{Q_X}$，其密度分别为

$$p(x):=\left(\frac{\alpha_{\mathrm{P}}}{\sigma}\right)^d\prod_{i=1}^d\left(1+\frac{x_i}{\sigma}\right)^{-(\alpha_{\mathrm{P}}+1)}\mathbb{1}(x_i\geq 0)$$
$$q(x):=\left(\frac{\alpha_{\mathrm{Q}}}{\sigma}\right)^d\prod_{i=1}^d\left(1+\frac{x_i}{\sigma}\right)^{-(\alpha_{\mathrm{Q}}+1)}\mathbb{1}(x_i\geq 0).$$

如引理 36所示，有 $\gamma=\gamma^*(\mathrm{P_X},\mathrm{Q_X})$以及 $s=\gamma^*(\mathrm{Q_X},\mathrm{Q_X})$。此外，引理 36保证，取足够大的 σ可得$(\mathrm{P_X},\mathrm{Q_X})\in\mathcal{P}(D,\theta)$。设 $a\geq 0$与 z_1，……， z_M为 $[a,2a]^d$ 在无穷范数下的一个（极大） $2h$-装箱的中心。则有

$$\left(\frac{a}{2h}\right)^d\leq M\leq\left(\frac{3a}{2h}\right)^d.$$

考虑将 $\Phi\in\mathcal{H}(L,\beta)$ 视为一个支撑在 $[-1,1]^d$上的核，使得 $\|\Phi\|_\infty<\infty$.。令 $\phi_j(x):=Lh^\beta\Phi(|x-z_j|/h)$。Varshamov-Gilbert 界 [29, 引理 2.9] 保证，只要 $M\geq 8$, 就存在 $\{0,1\}^M$ 的一个子集 B，使得

(*i*) Card(B) $\geq 2^{M/8}$ 且

(*ii*) 对所有 b, $b'\in B$ 以及 $b\neq b'$.

对于 $b\in B$，定义 $f_b:=\sum_{i=1}^M b_i\phi_i$，并令 $\mathrm{P}_b=\mathrm{P}^{f_b}$, 为 n 个独立同分布（i.i.d.）样本对 $(X_i,\ Y_i)$ 的联合分布

其中 $X_i \sim \mathrm{P_X}$，且 Y_i 遵循模型 (1) 的分布。接下来，

$$\begin{aligned}
\|f_b - f_{b'}\|^2_{L^2(\mathrm{Q_X})} &= \int \sum_{i=1}^{M} (b_i - b_i')^2 \phi_i(x)^2 \, d\,\mathrm{Q}(x) \\
&= \sum_{i=1}^{M} |b_i - b_i'| \int_{\mathrm{B}(z_i,h)} \phi_i(x)^2 \, d\,\mathrm{Q}(x) \\
&\geq \Big(\frac{\alpha_\mathrm{Q}}{\sigma}\Big)^d \frac{L^2 h^{2\beta}}{(1+2a/\sigma)^{d(\alpha_\mathrm{Q}+1)}} \sum_{i=1}^{M} |b_i - b_i'| \int_{\mathrm{B}(z_i,h)} \Phi\Big(\frac{x-z_i}{h}\Big)^2 dx \qquad (43) \\
&= \Big(\frac{\alpha_\mathrm{Q}}{\sigma}\Big)^d \frac{L^2 h^{2\beta+d} \|\Phi\|_2^2}{(1+2a/\sigma)^{d(\alpha_\mathrm{Q}+1)}} \sum_{i=1}^{M} |b_i - b_i'| \\
&\geq \Big(\frac{\alpha_\mathrm{Q}}{\sigma}\Big)^d \frac{L^2 h^{2\beta+d} M \|\Phi\|_2^2}{8(1+2a/\sigma)^{d(\alpha_\mathrm{Q}+1)}} =: \frac{C_{q,2} h^{2\beta+d} M}{(1+2a/\sigma)^{d(\alpha_\mathrm{Q}+1)}}.
\end{aligned}$$

类似地，我们可以推导出以下上界。

$$\begin{aligned}
\|f_b - f_{b'}\|^2_{L^2(\mathrm{Q_X})} &= \int \sum_{i=1}^{M} (b_i - b_i')^2 \phi_i(x)^2 \, d\,\mathrm{Q}(x) \\
&\leq \Big(\frac{\alpha_\mathrm{Q}}{\sigma}\Big)^d \frac{L^2 h^{2\beta}}{(1+a/\sigma)^{d(\alpha_\mathrm{Q}+1)}} \sum_{i=1}^{M} |b_i - b_i'| \int_{\mathrm{B}(z_i,h)} \Phi\Big(\frac{x-z_i}{h}\Big)^2 dx \\
&\leq \Big(\frac{\alpha_\mathrm{Q}}{\sigma}\Big)^d \frac{L^2 h^{2\beta+d} M \|\Phi\|_2^2}{8(1+a/\sigma)^{d(\alpha_\mathrm{Q}+1)}} =: \frac{C_{q,1} h^{2\beta+d} M}{(1+a/\sigma)^{d(\alpha_\mathrm{Q}+1)}},
\end{aligned}$$

and

$$\|f_b - f_{b'}\|^2_{L^2(\mathrm{P_X})} \leq \Big(\frac{\alpha_\mathrm{P}}{\sigma}\Big)^d \frac{L^2 h^{2\beta+d} M \|\Phi\|_2^2}{8(1+a/\sigma)^{d(\alpha_\mathrm{P}+1)}} =: \frac{C_{p,1} h^{2\beta+d} M}{(1+a/\sigma)^{d(\alpha_\mathrm{P}+1)}}.$$

Fano不等式 [31, 命题 15.2]与互信息的一个基本上界相结合，可得

$$\begin{aligned}
&\inf_{\widehat{f}} \sup_{f \in \mathcal{H}(L,\beta)} \mathrm{E}_f\big[\|\widehat{f} - f\|^2_{L^2(\mathrm{Q_X})}\big] \\
&\qquad \geq \inf_{b \neq b'} \|f_b - f_{b'}\|_{L^2(\mathrm{Q_X})} \bigg(1 - \frac{\log(2) + \max_{b,b'} \mathrm{KL}(\mathrm{P}_b, \mathrm{P}_{b'})}{\log(\mathrm{Card}(B))}\bigg).
\end{aligned} \qquad (44)$$

简记为 $r_\beta := 2\beta/(2\beta+d)$。我们首先假设 $M > 8$，这一点我们将在后续所有情形中加以验证。于是我们有

$$\frac{\log(2)}{\log(\mathrm{Card}(B))} \leq \frac{8\log(2)}{M} \leq \frac{1}{4}, \qquad (45)$$

只要 $M \geq 32\log(2)$，我们将在后文验证这一点。在分情况讨论之前，我们总结一下

在我们的设定中由 Fano 方法推导出的约束。KL 散度中的装箱要求

$$\begin{aligned}\frac{\max_{b,b'}\mathrm{KL}(\mathrm{P}_b,\mathrm{P}_{b'})}{\log(\mathrm{Card}(B))} &\le \frac{8\max_{b,b'}\mathrm{KL}(\mathrm{P}_b,\mathrm{P}_{b'})}{M}\\ &\le \frac{8}{M\sigma_e^2}\max_{b,b'}\Big[n\|f_b-f_{b'}\|^2_{L^2(\mathrm{P_X})}+m\|f_b-f_{b'}\|^2_{L^2(\mathrm{Q_X})}\Big]\\ &\le \frac{16}{M\sigma_e^2}\left(\frac{C_{p,1}nMh^{2\beta+d}}{(1+a/\sigma)^{d(\alpha_\mathrm{P}+1)}}\vee\frac{C_{q,1}mMh^{2\beta+d}}{(1+a/\sigma)^{d(\alpha_\mathrm{Q}+1)}}\right)\\ &= \frac{16}{M\sigma_e^2}h^{2\beta+d}\left(\frac{C_{p,1}n}{(1+a/\sigma)^{d(\alpha_\mathrm{P}+1)}}\vee\frac{C_{q,1}m}{(1+a/\sigma)^{d(\alpha_\mathrm{Q}+1)}}\right)\\ &\le \frac{16}{\sigma_e^2}h^{2\beta+d}\left(\frac{C_{p,1}n}{(a/\sigma)^{d(\alpha_\mathrm{P}+1)}}\vee\frac{C_{q,1}m}{(a/\sigma)^{d(\alpha_\mathrm{Q}+1)}}\right)\\ &\le h^{2\beta+d}\tilde{C}\left(\frac{n}{(a/\sigma)^{d(\alpha_\mathrm{P}+1)}}\vee\frac{m}{(a/\sigma)^{d(\alpha_\mathrm{Q}+1)}}\right),\end{aligned} \tag{46}$$

其中 $\tilde{C}:=16\big(C_{p,1}\vee C_{q,1}\big)/\sigma_e^2$, 且后一上界小于 $1/4$，如果 $\mu\ge 1$，并且 a 选为

$$a=4\sigma\mu\left[\left(\tilde{C}^{\frac{1}{d(\alpha_\mathrm{P}+1)}}n^{\frac{1}{d(\alpha_\mathrm{P}+1)}}h^{\frac{2\beta+d}{d(\alpha_\mathrm{P}+1)}}\right)\vee\left(\tilde{C}^{\frac{1}{d(\alpha_\mathrm{Q}+1)}}m^{\frac{1}{d(\alpha_\mathrm{Q}+1)}}h^{\frac{2\beta+d}{d(\alpha_\mathrm{Q}+1)}}\right)\right]. \tag{47}$$

取 a 后，应用 Fano 方法，并假设 $a\ge\sigma$, （稍后将对此进行验证），我们得到

$$\begin{aligned}\inf_{\widehat{f}}\sup_{f\in\mathcal{H}(L,\beta)}\mathrm{E}_f\left[\|\widehat{f}-f\|^2_{L^2(\mathrm{Q_X})}\right] &\ge \frac{1}{2}\inf_{b\neq b'}\|f_b-f_{b'}\|_{L^2(\mathrm{Q_X})}\\ &\ge \frac{C_{q,2}}{2}\frac{h^{2\beta+d}M}{(1+2a/\sigma)^{d(\alpha_\mathrm{Q}+1)}}\\ &\ge \frac{C_{q,2}}{2}\frac{h^{2\beta}a^d}{(4a/\sigma)^{d(\alpha_\mathrm{Q}+1)}}\\ &=: C_1\frac{h^{2\beta}}{a^{d\alpha_\mathrm{Q}}}\\ &= \frac{C_1}{(4\sigma\mu)^{d\alpha_\mathrm{Q}}}\left(\frac{h^{2\beta-(2\beta+d)\gamma}}{\tilde{C}^\gamma n^\gamma}\wedge\frac{h^{2\beta-(2\beta+d)s}}{\tilde{C}^s m^s}\right)\\ &= \frac{C_1}{(4\sigma\mu)^{d\alpha_\mathrm{Q}}}\left(\frac{h^{(2\beta+d)(r_\beta-\gamma)}}{\tilde{C}^\gamma n^\gamma}\wedge\frac{h^{(2\beta+d)(r_\beta-s)}}{\tilde{C}^s m^s}\right).\end{aligned} \tag{48}$$

现在，我们根据 (γ,s)。的不同取值情况进行讨论。

情形 1： $\gamma,s\le r_\beta$。于是，我们可以取 $h=1$, 从而得到

$$a=4\sigma\mu\left[\left(\tilde{C}^{\frac{1}{d(\alpha_\mathrm{P}+1)}}n^{\frac{1}{d(\alpha_\mathrm{P}+1)}}\right)\vee\left(\tilde{C}^{\frac{1}{d(\alpha_\mathrm{Q}+1)}}m^{\frac{1}{d(\alpha_\mathrm{Q}+1)}}\right)\right].$$

注意到 $a\to\infty$，因为它要么是 $n\to\infty$，要么是 $m\to\infty$.。因此，取足够大的 μ，可得$M\ge 8\vee 32$ log(2)，且对所有 $n\vee m\ge 1.$，都有 $a\ge\sigma$。因此，Fano不等式 (48) 给出

$$\inf_{\widehat{f}}\sup_{f\in\mathcal{H}(L,\beta)}\mathrm{E}_f\left[\|\widehat{f}-f\|^2_{L^2(\mathrm{Q_X})}\right]\ge\frac{C_1}{(4\sigma\mu)^{d\alpha_\mathrm{Q}}}\left(\frac{h^{(2\beta+d)(r_\beta-\gamma)}}{\tilde{C}^\gamma n^\gamma}\wedge\frac{h^{(2\beta+d)(r_\beta-s)}}{\tilde{C}^s m^s}\right)\ge\frac{C}{n^\gamma\vee m^s}.$$

这就完成了 $\gamma,s\le r_\beta$. 情形下的证明

情形 2： γ, $s\ge r_\beta$. 在这种情况下，我们取如前所述定义的 a，并选择

$$h=(n\vee m)^{-1/(2\beta+d)}.$$

由此可得

$$a = 4\sigma\mu\Big(\frac{\tilde{C}^{1/(d(\alpha_{\mathrm{P}}+1))}n^{1/(d(\alpha_{\mathrm{P}}+1))}}{(n\vee m)^{1/(d(\alpha_{\mathrm{P}}+1))}} \vee \frac{\tilde{C}^{1/(d(\alpha_{\mathrm{Q}}+1))}m^{1/(d(\alpha_{\mathrm{Q}}+1))}}{(n\vee m)^{1/(d(\alpha_{\mathrm{Q}}+1))}}\Big) =: \mu C,$$

这是一个常数，而 h 是 $n\vee m$. 中的零序列。因此，通过选取足够大的 μ，我们有$M \asymp (a/h)^d \geq 8\vee$(32 log(2))，并且对所有 $n\vee m \geq 1$. 有 $a \geq \sigma$。因此，下界(48)变为

$$\begin{aligned}\inf_{\widehat{f}}\sup_{f\in\mathcal{H}(L,\beta)} \mathrm{E}_f\left[\|\widehat{f}-f\|^2_{L^2(\mathrm{Q_X})}\right] &\geq C_1\frac{h^{2\beta}}{a^{d\alpha_{\mathrm{Q}}}}\\ &= \frac{C_1}{C^{d\alpha_{\mathrm{Q}}}}(n\vee m)^{2\beta/(2\beta+d)},\end{aligned}$$

如结论所述。这表明，对正配置与零配置中的所有$(\gamma, s) \in (0,\infty) \times (0,1)$，下界成立。接下来，我们证明负配置的下界。

情形 3： $s < r_\beta < \gamma$. 该情形进一步细分为三个子情形。

情形 3.1： $m^s \leq n^s$. 取

$$h = n^{-1/(2\beta+d)}$$

得到

$$a = 4\sigma\mu\Big(\tilde{C}^{1/(d(\alpha_{\mathrm{P}}+1))} \vee \tilde{C}^{1/(d(\alpha_{\mathrm{Q}}+1))}m^{s/(d\alpha_{\mathrm{Q}})}n^{-s/(d\alpha_{\mathrm{Q}})}\Big) \in \big[4\sigma\mu\tilde{C}^{1/(d(\alpha_{\mathrm{P}}+1))}, 4\sigma\mu\tilde{C}^{1/(d(\alpha_{\mathrm{Q}}+1))}\big],$$

其阶为常数，而当 n （或 m, 因为 $m^s \leq n^s$）增大时， h 趋于0。因此，选取足够大的 μ 可得 $M \geq 8 \vee$ (32 log(2))，并且对所有 $0 \leq m \leq n,\ m\vee n \geq 1$ 有 $a \geq \sigma$，而下界变为

$$\begin{aligned}\inf_{\widehat{f}}\sup_{f\in\mathcal{H}(L,\beta)} \mathrm{E}_f\left[\|\widehat{f}-f\|^2_{L^2(\mathrm{Q_X})}\right] &\geq C_1\frac{h^{2\beta}}{a^{d\alpha_{\mathrm{Q}}}}\\ &\geq Cn^{-2\beta/(2\beta+d)} = C(n^{r_\beta}\vee m^s)^{-1},\end{aligned}$$

如结论所述。

情形 3.2： $m^s \geq n^\gamma$. 在此情形下，我们取 $h = 1$ 并得到

$$a = 4\sigma\mu\left[\left(\tilde{C}^{\frac{1}{d(\alpha_{\mathrm{P}}+1)}}n^{\frac{1}{d\alpha_{\mathrm{P}}+1}}\right) \vee \left(\tilde{C}^{\frac{1}{d(\alpha_{\mathrm{Q}}+1)}}m^{\frac{1}{d(\alpha_{\mathrm{Q}}+1)}}\right)\right],$$

它随 n和/或 m，增长，因此，选取足够大的 μ可确保$M \geq 8\vee 32$log(2)以及 $a \geq \sigma$对所有 $0 \leq n^\gamma \leq m^s$, $n\vee m \geq 1$ 从而,

$$\begin{aligned}\inf_{\widehat{f}}\sup_{f\in\mathcal{H}(L,\beta)} \mathrm{E}_f\left[\|\widehat{f}-f\|^2_{L^2(\mathrm{Q_X})}\right] &\geq C_1\frac{h^{2\beta}}{a^{d\alpha_{\mathrm{Q}}}}\\ &\geq C\big(n^\gamma\vee m^s\big)^{-1}.\end{aligned}$$

此情形得证。

情形 3.3： $n^s < m^s < n^\gamma$. 在这种情形下，假设 $M \geq 8\vee 32$log(2) 且 $a > \sigma$，我们注意到，Fano 方法得到的下界本质上具有以下形式

$$\inf_{\widehat{f}}\sup_{f\in\mathcal{H}(L,\beta)} \mathrm{E}_f\left[\|\widehat{f}-f\|^2_{L^2(\mathrm{Q_X})}\right] \geq \frac{C_1}{(4\sigma\mu)^{d\alpha_{\mathrm{Q}}}}\left(\frac{h^{(2\beta+d)(r_\beta-\gamma)}}{\tilde{C}^\gamma n^\gamma} \wedge \frac{h^{(2\beta+d)(r_\beta-s)}}{\tilde{C}^s m^s}\right), \tag{49}$$

并且由于 $r_\beta - \gamma < 0 < r_\beta - s$，上面最小值中的第一项是关于 h 的递减函数，而第二项是关于 h. 的递增函数。我们发现，选择以下值可以使这两项（近似）达到平衡

$$h := \Big(\frac{m^s}{n^\gamma}\Big)^{\frac{1}{(\gamma-s)(2\beta+d)}},$$

下界变为

$$\begin{aligned}\inf_{\widehat{f}} \sup_{f\in\mathcal{H}(L,\beta)} \mathrm{E}_f\left[\|\widehat{f}-f\|^2_{L^2(\mathrm{Q_X})}\right] &\geq \frac{C_1 C(\gamma,s)}{(4\sigma\mu)^{d\alpha_\mathrm{Q}}} m^{-s}\Big(\frac{m^s}{n^\gamma}\Big)^{\frac{r_\beta - s}{\gamma - s}} \\ &= \frac{C_1 C(\gamma,s)}{(4\sigma\mu)^{d\alpha_\mathrm{Q}}} m^{-s\frac{\gamma-r_\beta}{\gamma-s}} n^{-\gamma\frac{r_\beta-s}{\gamma-s}}.\end{aligned}$$

还需检查在此选择下， $M \geq 8 \vee$ (32 log(2)。这里一个有用的重新参数化方法是令 $m^s = n^{\lambda\gamma+(1-\lambda)s}$ 对应于 $\lambda \in$ (0,1)。 由此可将其改写为

$$h = n^{\frac{(\gamma-s)(\lambda-1)}{(\gamma-s)(2\beta+d)}} = n^{\frac{\lambda-1}{2\beta+d}},$$

and

$$\begin{aligned}M^{1/d} \asymp \frac{a}{h} &\asymp h^{-1}\Big(n^{\gamma/(d\alpha_\mathrm{Q})} h^{(2\beta+d)\gamma/(d\alpha_\mathrm{Q})} \vee m^{s/(d\alpha_\mathrm{Q})} h^{(2\beta+d)s/(d\alpha_\mathrm{Q})}\Big) \\ &= \frac{1}{h}\Big(n^\gamma h^{(2\beta+d)\gamma} \vee m^s h^{(2\beta+d)s}\Big)^{1/(d\alpha_\mathrm{Q})} \\ &= \frac{1}{h}\big(n^{\gamma\lambda} \vee n^{\gamma\lambda}\big)^{1/(d\alpha_\mathrm{Q})} = n^{\frac{\gamma\lambda}{d\alpha_\mathrm{Q}} + \frac{1-\lambda}{2\beta+d}},\end{aligned}$$

并且上面的 n 指数严格为正。此外， n 必然大于 1，否则不存在参数范围 $n^s < m^s < n^\gamma$. 。结合取足够大的 μ ，可得$M \geq 8 \vee$ (32 log(2))，且对所有 $n^s < m^s < n^\gamma$. ，都有 $a \geq \sigma$ 。因此，我们得到下界

$$\inf_{\widehat{f}} \sup_{f\in\mathcal{H}(L,\beta)} \mathrm{E}_f\left[\|\widehat{f}-f\|^2_{L^2(\mathrm{Q_X})}\right] \geq \frac{C_1 C(\gamma,s)}{(4\sigma\mu)^{d\alpha_\mathrm{Q}}} m^{-s\frac{\gamma-r_\beta}{\gamma-s}} n^{-\gamma\frac{r_\beta-s}{\gamma-s}}.$$

这就完成了情形 3.3 和情形 3。

情形 4： $\gamma < r_\beta < s$. 同样，这种情形分为三种情况。

情形 4.1： $n^\gamma \leq m^\gamma$. 这种情形与情形 3.1 对称。取$h = m^{-1/(2\beta+d)}$ 并按照类似步骤可得

$$\inf_{\widehat{f}} \sup_{f\in\mathcal{H}(L,\beta)} \mathrm{E}_f\left[\|\widehat{f}-f\|^2_{L^2(\mathrm{Q_X})}\right] \geq C m^{-2\beta/(2\beta+d)} = C(n^\gamma \vee m^{r_\beta})^{-1}.$$

因此，情形 4.1 中的结论成立。

情形 4.2: $n^\gamma \geq m^s$. 此情形与情形 3.2 对称。取 $h = 1$，并按照类似步骤即可得到下界

$$\begin{aligned}\inf_{\widehat{f}} \sup_{f\in\mathcal{H}(L,\beta)} \mathrm{E}_f\left[\|\widehat{f}-f\|^2_{L^2(\mathrm{Q_X})}\right] &\geq C_1 \frac{h^{2\beta}}{a^{d\alpha_\mathrm{Q}}} \\ &\geq C\big(n^\gamma \vee m^s\big)^{-1} = Cn^{-\gamma}.\end{aligned}$$

因此，该结论在情形 4.2 中成立。

情况 4.3: $m^\gamma < n^\gamma < m^s$. 此情形与情形 3.3 再次呈对称关系。为完整起见，我们详细说明推导步骤。按照情形 3.3 中的相同步骤，可以看到这次下界为

$$\inf_{\widehat{f}} \sup_{f\in\mathcal{H}(L,\beta)} \mathrm{E}_f\left[\|\widehat{f}-f\|^2_{L^2(\mathrm{Q_X})}\right] = \frac{C_1}{(4\sigma\mu)^{d\alpha_\mathrm{Q}}}\left(\frac{h^{(2\beta+d)(r_\beta-\gamma)}}{\tilde{C}^\gamma n^\gamma} \wedge \frac{h^{(2\beta+d)(r_\beta-s)}}{\tilde{C}^s m^s}\right),$$

并且由于 $r_\beta - s < 0 < r_\beta - \gamma$，上述最小值中的第一项是关于 h 的递增函数，而第二项是关于 h. 的递减函数。我们发现，选择.....可以使这两项（近似）平衡

$$h := \left(\frac{n^\gamma}{m^s}\right)^{\frac{1}{(s-\gamma)(2\beta+d)}},$$

下界变为

$$\begin{aligned}\inf_{\widehat{f}} \sup_{f\in\mathcal{H}(L,\beta)} \mathrm{E}_f\left[\|\widehat{f}-f\|^2_{L^2(\mathrm{Q_X})}\right] &\geq \frac{C_1 C(\gamma,s)}{(4\sigma\mu)^{d\alpha_\mathrm{Q}}} n^{-\gamma}\left(\frac{n^\gamma}{m^s}\right)^{\frac{r_\beta-\gamma}{s-\gamma}} \\ &= \frac{C_1 C(\gamma,s)}{(4\sigma\mu)^{d\alpha_\mathrm{Q}}} m^{-s\frac{r_\beta-\gamma}{s-\gamma}} n^{-\gamma\frac{s-r_\beta}{s-\gamma}}.\end{aligned}$$

我们再次可以对 $n^\gamma = m^{\lambda s+(1-\lambda)\gamma}$ 进行参数化，并得到

$$\begin{aligned}\frac{a}{h} &= h^{-1}\left(n^{\gamma/(d\alpha_\mathrm{Q})} h^{(2\beta+d)\gamma/(d\alpha_\mathrm{Q})} \vee m^{s/(d\alpha_\mathrm{Q})} h^{(2\beta+d)s/(d\alpha_\mathrm{Q})}\right) \\ &= \frac{1}{h}\left(n^\gamma h^{(2\beta+d)\gamma} \vee m^s h^{(2\beta+d)s}\right)^{1/(d\alpha_\mathrm{Q})} \\ &= \frac{1}{h}\left(m^{s\lambda} \vee m^{s\lambda}\right)^{1/(d\alpha_\mathrm{Q})} \\ &= m^{\frac{s\lambda}{d\alpha_\mathrm{Q}}+\frac{(1-\lambda)}{2\beta+d}},\end{aligned}$$

而后者被提升到 m 的严格正次幂，这表明 M 发散。由此，并选择足够大的 μ，可确保 $M \geq 8 \vee$ (32 log(2))，且对于所有 $m^\gamma < n^\gamma < m^s.$，$a \geq \sigma$。这就证明了下界

$$\begin{aligned}\inf_{\widehat{f}} \sup_{f\in\mathcal{H}(L,\beta)} \mathrm{E}_f\left[\|\widehat{f}-f\|^2_{L^2(\mathrm{Q_X})}\right] &\geq \frac{C_1 C(\gamma,s)}{(4\sigma\mu)^{d\alpha_\mathrm{Q}}} n^{-\gamma}\left(\frac{n^\gamma}{m^s}\right)^{\frac{r_\beta-\gamma}{s-\gamma}} \\ &= \frac{C_1 C(\gamma,s)}{(4\sigma\mu)^{d\alpha_\mathrm{Q}}} m^{-s\frac{r_\beta-\gamma}{s-\gamma}} n^{-\gamma\frac{s-r_\beta}{s-\gamma}}.\end{aligned}$$

至此，情况 4.3、情形 4 以及整个证明均告完成。 □

C 示例的证明

引理 *34.* 设 $P_\mathrm{X} = \mathcal{E}(\lambda)$，则，$P_\mathrm{X} \in \mathcal{P}(\lambda, e^\lambda)$ 且 $\gamma^*(P_\mathrm{X}, P_\mathrm{X}) = 1$. 此外，若 $P_\mathrm{X} = \mathcal{E}(\lambda_\mathrm{P})$、$Q_\mathrm{X} = \mathcal{E}(\lambda_\mathrm{Q})$，且 $\lambda = \lambda_\mathrm{P} \vee \lambda_\mathrm{Q}$，则 P_X 和 Q_X 均属于 $\mathcal{P}(\lambda, exp(\lambda))$。此外，$\gamma^*(P_\mathrm{X}, Q_\mathrm{X}) = \lambda_\mathrm{Q}/\lambda_\mathrm{P}$.

证明。我们首先验证定义 3 中的假设。根据定义，$\mathrm{P_X}$ 的密度上界为 λ. 接下来，对于任意 $x > 0$ 和 $0 < t \leq 1$,

$$\mathrm{P_X}\{\mathrm{B}(x,t)\} = \int_{x-t}^{x+t} \lambda\exp(-\lambda u)\mathbb{1}(u \geq 0)\, du \geq \int_x^{x+t} \lambda\exp(-\lambda(x+t))\, du$$

$$= t\exp(-\lambda t)p(x) \geq e^{-\lambda} tp(x),$$

这证明最小质量性质在 $\theta = e^{\lambda}$. 下成立。接下来，我们证明最大质量性质。令 $t \in (0, 1]$. 如果 $0 \leq x \leq t$, 那么 $\mathrm{P_X}\{\mathrm{B}(x, t)\} = 1 - \exp(-\lambda(x+t))$, 且

$$\begin{aligned}\frac{\mathrm{P_X}\{\mathrm{B}(x,t)\}}{tp(x)} = \frac{1-\exp(-\lambda(x+t))}{t\lambda\exp(-\lambda x)} &= \frac{\exp(\lambda x)-\exp(-\lambda t)}{t\lambda} \\ &\leq \frac{\exp(\lambda t)-\exp(-\lambda t)}{t\lambda} = \frac{2\sinh(\lambda t)}{\lambda t} \leq \exp(\lambda t).\end{aligned}$$

函数 $g\colon t \mapsto 2\sinh(\lambda t)/(\lambda t)$ 在 R_+^*. 上连续递增。此外， g 在$t \to 0$, 处存在极限，其值等于 2.我们已证明，对于 $0 < t \leq 1$和 $0 \leq x \leq t$, ，有 $\mathrm{P_X}\{\mathrm{B}(x,t)\} \leq \theta p(x)t$. 现在考虑 $t < x$, 此时有

$$\frac{\mathrm{P_X}\{\mathrm{B}(x,t)\}}{tp(x)} = \frac{\exp(-\lambda(x-t))-\exp(-\lambda(x+t))}{\lambda t\exp(-\lambda x)} = \frac{2\sinh(\lambda t)}{\lambda t} \leq \exp(\lambda t),$$

与此前一样， 对所有 $0 < t \leq 1$.。我们已证明 $\mathrm{P_X} \in \mathcal{P}(\lambda, \exp(\lambda))$. 接下来，对于 $s \in [0, 1)$, ，有

$$\mathcal{T}(\mathrm{P_X}, \mathrm{P_X}) = \int p(x)^{1-s}\,dx = \int \lambda^{1-s}\exp(-\lambda x(1-s))\,dx < \infty,$$

而对于所有 $s \geq 1, \mathcal{T}(\mathrm{P_X}, \mathrm{P_X}) = \infty$. 因此， $\gamma^*(\mathrm{P_X}, \mathrm{P_X}) = 1$. 最后，我们证明第二个结论。令$\gamma \in [0, \lambda_\mathrm{Q}/\lambda_\mathrm{P})$. 于是,

$$\int q(x)p(x)^{-\gamma}\,dx = \frac{\lambda_\mathrm{Q}}{\lambda_\mathrm{P}^{\gamma}}\int \exp((\gamma\lambda_\mathrm{P} - \lambda_\mathrm{Q})x)\,dx < \infty,$$

而 $\mathcal{T}(\mathrm{P_X}, \mathrm{Q_X}, \gamma) = \infty$ 对所有 $\gamma \geq \lambda_\mathrm{Q}/\lambda_\mathrm{P}$.。因此 $\gamma^*(\mathrm{P_X}, \mathrm{Q_X}) = \lambda_\mathrm{Q}/\lambda_\mathrm{P}$. □

定义 35 （帕累托分布）。 对于 α、 $\sigma > 0$, 我们称 *Pareto* 分布 $\mathrm{Par}(\alpha, \sigma)$ 为定义在 $\mathbb{R}$ 上的分布，其密度为

$$p(x) = \frac{\alpha}{\sigma}(1 + x/\sigma)^{-(\alpha+1)}\mathbb{1}(x \geq 0).$$

引理 36. 设 α、 $\sigma > 0$, 且 $P_\mathrm{X} = Par(\alpha, \sigma)$, 则 $P_\mathrm{X} \in \mathcal{P}(\alpha/\sigma, 2(1+\sigma^{-1})^{\alpha+1})$ 且 $\gamma^*(P_\mathrm{X}, P_\mathrm{X}) = \alpha/(\alpha+1)$。此外，若 $P_\mathrm{X} = Par(\alpha_\mathrm{P}, \sigma_\mathrm{P})$ 且 $Q_\mathrm{X} = Par(\alpha_\mathrm{Q}, \sigma_\mathrm{Q})$, 则 P_X 和 Q_X 均属于$\mathcal{P}(D, \theta)$, 其中

$$D = \frac{\alpha_\mathrm{P}}{\sigma_\mathrm{P}} \vee \frac{\alpha_\mathrm{Q}}{\sigma_\mathrm{Q}}, \quad and \quad \theta = 2\Big(1 + \frac{1}{\sigma_\mathrm{P} \wedge \sigma_\mathrm{Q}}\Big)^{\alpha_\mathrm{P}\wedge\alpha_\mathrm{Q}+1}.$$

此外， $\gamma^*(\mathrm{P_X}, \mathrm{Q_X}) = \alpha_\mathrm{Q}/(\alpha_\mathrm{P}+1)$。

证明。 设 p 为 $\mathrm{P_X} = \mathrm{Par}(\alpha, \sigma)$。 对于所有 $0 < r \leq 1$,, 我们有

$$\begin{aligned}(rp(x))^{-1}\int_{x-r}^{x+r} p(t)\,dt \leq \frac{2r\sup_{|t-x|\leq r} p(t)}{rp(x)} &= \frac{2p((x-r)\vee 0)}{p(x)} \\ &= 2\Big(1+\frac{x}{\sigma}\Big)^{\alpha+1}\mathbb{1}(x \leq r) + 2\Big(\frac{\sigma+x}{\sigma+x-r}\Big)^{\alpha+1}\mathbb{1}(x > r) \\ &\leq 2(1+r/\sigma)^{\alpha+1} \leq 2\Big(1+\frac{1}{\sigma}\Big)^{\alpha+1}.\end{aligned}$$

这意味着最大质量性质，因为

$$\mathrm{P_X}\{\mathrm{B}(x,r)\} = \int_{(x-r)}^{x+r} p(x)\,dx \leq 2rp((x-r)\vee 0) \leq 2\Big(1+\frac{x}{\sigma}\Big)^{\alpha+1}p(x)r.$$

另一方面，对于所有 $0 < r \le 1$，，我们有

$$\frac{p(x+r)}{p(x)} = \Big(\frac{\sigma + x}{\sigma + x + r}\Big)^{\alpha+1} \ge \Big(\frac{\sigma}{\sigma+1}\Big)^{\alpha+1},$$

这表明最小质量性质在 $\theta = 2(1+1/\sigma)^{\alpha+1}$. 下成立。事实上，

$$\mathrm{P_X}\{\mathrm{B}(x,r)\} = \int_{(x-r)}^{x+r} p(x)\,dx \ge rp(x+r) \ge \frac{1}{2}\Big(\frac{\theta}{\theta+1}\Big)^{\alpha+1} rp(x).$$

显然， $\|p\|_\infty = \alpha/\sigma$. 因此 $\mathrm{P_X} \in \mathcal{P}(\alpha/\sigma,\ 2(1+1/\sigma)^{\alpha+1})$。
接下来，我们可以验证

$$\mathcal{T}(\mathrm{P_X}, \mathrm{P_X}, s) = \big(\alpha/\sigma\big)^{1-s} \int \Big(1 + \frac{x}{\sigma}\Big)^{-(\alpha+1)(1-s)} \mathbb{1}(x \ge 0)\,dx$$

是有限的，当且 仅当 $s < \frac{\alpha}{\alpha+1}$. 最后，如果 $\mathrm{P_X} = \mathrm{Par}(\alpha_\mathrm{P})$ 且 $\mathrm{Q_X} \in \mathrm{Par}(\alpha_\mathrm{Q})$，那么二者都包含于 $s < \frac{\alpha}{\alpha+1}$. ($\mathrm{X} =$ f 两者都进入、其中 D，θ 在结论中定义，显然成立，我们可以看到

$$\mathcal{T}(\mathrm{P_X}, \mathrm{Q_X}, \gamma) = \frac{\alpha_\mathrm{Q}\sigma_\mathrm{P}^\gamma}{\alpha_\mathrm{P}^\gamma\sigma_\mathrm{Q}} \int \frac{(1+\frac{x}{\sigma_\mathrm{P}})^{\gamma(\alpha_\mathrm{P}+1)}}{(1+\frac{x}{\sigma_\mathrm{Q}})^{\alpha_\mathrm{Q}+1}} \mathbb{1}(x \ge 0)\,dx$$

当且仅当 $\gamma < \alpha_\mathrm{Q}/(\alpha_\mathrm{P}+1)$ 有限。证明至此完成。 □

D 第 4.1 节的证明

引理 37。 考虑 $\mathrm{P,Q} \in \mathcal{M}$. 那么，

(i) $\mathcal{T}(\mathrm{P,\ Q},0) = 1.(ii)$ $\gamma \mapsto \mathcal{T}(\mathrm{P,Q}\ \gamma)^\gamma$ 在其定义域上单调不减。(iii) $\gamma \mapsto \mathcal{T}(\mathrm{P,\ Q}\ \gamma)$ 在其定义域上是对数-凸的。

证明。 第一个结论由下式立即得到

$$\mathcal{T}(\mathrm{P,Q},0) = \int q(x)\,dx = 1.$$

若对所有 $\gamma > 0$， $\mathcal{T}(\mathrm{P},\ \mathrm{Q}\ \gamma) = \infty$ 均成立，则 (ii) 和 (iii) 显然成立。否则，取 $X \sim \mathrm{Q}$ 和 $s > 0$，使得 $\mathcal{T}(\mathrm{P},\ \mathrm{Q}\ s) < \infty$.。对于所有 $\gamma \in [0,s]$，函数 $t \mapsto t^{s/\gamma}$ 在 $\mathrm{R}_+\cdot$ 上是凸函数。由詹森不等式可得

$$\mathcal{T}(\mathrm{P,Q},\gamma) = \big(\mathrm{E}[p(X)^{-\gamma}]^{s/\gamma}\big)^{\gamma/s} \le \mathrm{E}[p(X)^{-s}]^{\gamma/s} = \mathcal{T}(\mathrm{P,Q},s)^{\gamma/s},$$

由此证明了 (ii)。 最后一个结论 (iii) 可由霍尔德不等式推出。令 γ_1， γ_2 属于 $\mathcal{T}(\mathrm{P},\ \mathrm{Q}\ \cdot)$ 的定义域，且 $\lambda \in [0,1]$.。于是，

$$\mathcal{T}(\mathrm{P,Q},\lambda\gamma_1 + (1-\lambda)\gamma_2) = \int \left(\frac{q(x)}{p(x)^{\gamma_1}}\right)^\lambda \left(\frac{q(x)}{p(x)^{\gamma_2}}\right)^{1-\lambda} dx \le \big(\mathcal{T}(\mathrm{P,Q},\gamma_1)^\lambda \big(\mathcal{T}(\mathrm{P,Q},\gamma_2)\big)^{1-\lambda},$$

由此得到该结论。 □

引理38 ([34]中的引理15)。 设$\mathrm{Q} \in \mathcal{M}$ 的密度为 q，且 $\rho,\varepsilon > 0$.。若 $X \sim \mathrm{Q}$ 具有有限的 $\rho+\varepsilon$阶广义矩，则 $\gamma^*(\mathrm{Q},\ \mathrm{Q}) \ge \rho/(\rho+d)$。

证明。 设 $\tau = d/(\rho + d)$。 由霍尔德不等式可得

$$\begin{aligned}
&\int q(x)^\tau \, dx = \int (1 + \|x\|^{\rho+\varepsilon})^\tau q(x)^\tau (1 + \|x\|^{\rho+\varepsilon})^{-\tau} \, dx \\
&\leq \left(\int \left((1 + \|x\|^{\rho+\varepsilon})^\tau q(x)^\tau \right)^{1/\tau} dx \right)^\tau \cdot \left(\int (1 + \|x\|^{\rho+\varepsilon})^{-\tau/(1-\tau)} \, dx \right)^{1-\tau} \\
&\leq (1 + M)^\tau \left(\int (1 + \|x\|^{\rho+\varepsilon})^{-\tau/(1-\tau)} \, dx \right)^{1-\tau},
\end{aligned}$$

其中 $M := \int \|x\|^{\rho+\varepsilon} \mathrm{Q}(dx)$。 由于 $\tau/(1-\tau) = d/\rho$，我们有 $\tau(\rho+\varepsilon)/(1-\tau) > d$，因此

$$\int (1 + \|x\|^{\rho+\varepsilon})^{-\tau/(1-\tau)} \, dx < \infty,$$

这证明了该结论。 □

引理 ***10*** 的证明。 令 $r' = (r-1)/r$ 为 r 的共轭，且 $0 \leq s < \gamma^*(\mathrm{P},\mathrm{P})$。 对于 $\gamma \geq 0$，由霍尔德不等式可得

$$\begin{aligned}
\int \frac{q(x)}{p(x)^\gamma} \, dx &= \int \frac{q(x)}{p(x)} p(x)^{1-\gamma} \, dx \\
&\leq \left(\int \left(\frac{q(x)}{p(x)} \right)^r d\,\mathrm{P}(x) \right)^{1/r} \left(\int p(x)^{-\gamma r'} \, d\,\mathrm{P}(x) \right)^{1/r'} \\
&= \exp\left(\frac{r-1}{r} D_r(\mathrm{Q} \,\|\, \mathrm{P}) \right) \left(\int p(x)^{1 - \frac{\gamma r}{r-1}} \, dx \right)^{(r-1)/r},
\end{aligned}$$

且对于任意 $\gamma < \gamma^*(\mathrm{P},\mathrm{P})(r-1)/r$，后者均为有限值，这表明 $\gamma^*(\mathrm{P},\mathrm{Q}) \geq \gamma^*(\mathrm{P},\mathrm{P})(r-1)/r$. □