



全局 Fréchet 回归的最优模型平均方法

张建强^{1†}, 陈泽^{2†}, 廖军^{3,4}, 许王莉^{3,4*}

1. 浙江工商大学统计与数据科学学院, 杭州 310018;
2. 山东大学中泰证券金融研究院, 济南 250100;
3. 中国人民大学应用统计科学研究中心, 北京 100872;
4. 中国人民大学统计学院, 北京 100872

† 同等贡献

E-mail: zhangjqs@163.com, chze96@163.com, junliao@ruc.edu.cn, wbxu@ruc.edu.cn

收稿日期: 2024-04-24; 接受日期: 2025-11-25; 网络出版日期: 2026-02-09; * 通信作者

国家重点研发计划 (批准号: 2023YFA1008702 和 2024YFA1013502)、国家自然科学基金 (批准号: 12571300, 12501370 和 12001534)、国家资助博士后研究人员计划 (批准号: GZC20231478)、中国博士后科学基金 (批准号: 2024M761782) 和浙江省登峰学科 (浙江工商大学统计学) 资助项目

摘要 近年来, 在众多应用领域中不断涌现大量复杂的非 Euclid 数据, 而全局 Fréchet 回归模型已成为分析非 Euclid 数据的重要工具. 然而, 在实际应用中, 全局 Fréchet 回归模型中自变量的选择往往具有较大挑战. 为了缓解这种模型选择不确定性问题, 本文采用数据驱动交叉验证权重选取准则, 并提出一种可行的模型平均方法. 在理论方面, 针对可能存在误设定的候选模型, 本文建立全局 Fréchet 回归估计的相合性, 并得到最优的逐点收敛速度与接近最优的一致收敛速度. 进一步地, 本文证明所提出的模型平均方法的渐近最优性, 并得到权重的收敛速度. 此外, 如果候选模型集至少包含一个正确模型, 则交叉验证所选择的权重渐近集中于正确模型. 基于所提出的模型平均方法, 本文发展一个可靠的变量重要性度量, 用于衡量 Fréchet 回归模型自变量的变量重要性, 并证明该方法能渐近识别出真模型中的所有变量. 模拟研究和实际数据分析结果显示, 所提方法的表现优于一些常用的模型选择和模型平均方法.

关键词 渐近最优性 Fréchet 回归 度量空间数据 模型平均 变量重要性

MSC (2020) 主题分类 62J99

1 引言

由于现代数据收集技术的进步, 取值于特定度量空间的非 Euclid 数据日益增多, 如大脑功能连接的协方差或相关矩阵^[11]和不同国家或地区的人口年龄分布^[5]. 文献^[29]指出, 分析非 Euclid 数据具有重要研究价值. 为了分析一般类型的非 Euclid 数据, 文献^[25]提出了全局 Fréchet 回归模型, 该模

英文引用格式: Zhang J Q, Chen Z, Liao J, et al. Optimal model averaging for global Fréchet regression (in Chinese). Sci Sin Math, 2026, 56: 1–36, doi: [10.1360/SCM-2024-0265](https://doi.org/10.1360/SCM-2024-0265)

型可视为经典线性回归模型的推广. 全局 Fréchet 回归模型可以处理取值于一般度量空间的响应变量, 如对称正定矩阵和概率分布, 因此已在生物学^[14] 和经济学^[26] 等多个领域获得广泛应用.

然而, 在实际中, 全局 Fréchet 回归模型的自变量选择通常具有不确定性. 而自变量的选取至关重要, 它将极大地影响预测表现. 因此, 实践中需要决定哪些变量被纳入模型中. 模型选择是一类重要的方法, 旨在选择由特定自变量构成的最终模型. 文献^[27] 提出了对全局 Fréchet 回归的模型选择方法, 该方法可视为岭选择算子 (ridge selection operator) 的推广. 然而, 模型选择的局限性在于其只选择单个模型, 从而忽略了模型不确定性可能带来的影响 (参见文献^[8, 34]). 相比之下, 模型平均方法能够有效缓解这种不确定性. 模型平均方法综合了来自不同候选模型的估计, 通常比模型选择具有更低的统计风险 (参见文献^[6, 24]).

模型平均方法可分为两大类: Bayes 模型平均和频率模型平均. 尽管 Bayes 模型平均具有良好的灵活性和适用性, 但确定适当的先验概率是一个巨大的挑战^[16]. 近年来, 基于参数模型的频率模型平均方法得到了迅速发展, 并在多个统计建模框架中取得了重要进展, 包括线性回归^[15, 39]、分位回归^[21, 30]、广义线性模型^[36, 38] 和时间序列模型^[19, 20] 等. 对于更为复杂的半参数和非参数模型, 文献中也有大量关于频率模型平均的研究, 如变系数模型^[18]、部分线性模型^[37]、变系数部分线性模型^[40] 和广义可加部分线性模型^[6] 等.

然而, 上述模型平均方法均针对取值于 Euclid 空间的响应变量. 由于一般的度量空间通常不存在代数和几何结构, 在 Fréchet 回归框架下建立最优模型平均方法面临显著挑战. 针对一维概率密度函数类型的响应变量, 文献^[32] 研究了基于全局 Fréchet 回归的一个特例—全局 Wasserstein-Fréchet 回归的最优模型平均方法. 文献^[17] 提出了全局 Fréchet 回归的最优模型平均方法但并未研究权重的渐近性质以及当候选模型集包含正确模型时该方法的表现. 更为重要的是, 已有研究并未建立关于全局 Fréchet 回归估计量在各候选模型下的相合性与收敛速度的理论结果. 本文针对上述问题, 提出了一种全局 Fréchet 回归模型的最优模型平均方法. 该方法不仅允许响应变量是概率密度函数, 而且允许响应变量在一般的度量空间取值 (如对称正定矩阵). 因此, 该方法具有广泛的适用性.

本文的主要贡献有以下 3 个方面.

第一, 建立了全局 Fréchet 回归模型估计量在各候选模型下的相合性, 并在允许候选模型误设定的条件下得到了一致收敛速度. 这些收敛结果对建立模型平均方法的理论性质具有重要作用. 据我们所知, 关于这些收敛性质的现有研究结果 (如文献^[25]) 仅适用于真模型. 当候选模型可能存在误设定时, 以上收敛结论是否仍成立目前尚不清楚. 因此, 本文的理论结果弥补了该方面的研究空缺.

第二, 在全局 Fréchet 回归框架下, 本文提出了一种可行的模型平均方法. 在候选模型误设定假设下, 本文证明了所提方法的渐近最优性. 当候选模型集中至少包含一个正确模型时, 进一步证明了权重的过一致性, 即基于交叉验证准则选出的权重渐近地集中于正确模型. 此外, 还得到了权重的收敛速度. 值得注意的是, 与传统的回归模型相比, 上述理论性质的证明更具挑战性. 常用的工具 (如文献^[31, 定理 2] 和 Taylor 展开技术) 在此不再适用. 这种限制来自于响应变量不再是标量且一般的度量空间不具有微分结构. 为此, 本文应用经验过程理论以克服上述困难.

第三, 基于所提出的模型平均方法, 本文进一步构建了一种可靠的变量重要性度量, 并从理论上证明该度量方法可以渐近识别所有重要变量. 因此, 本文的研究正式引入并发展了一种有效的方法来评估全局 Fréchet 回归模型自变量的变量重要性, 回答了文献^[27] 所提出的如何衡量全局 Fréchet 回归变量重要性的问题.

本文余下内容的安排如下. 第 2 节引入全局 Fréchet 回归模型并介绍模型平均方法以及数据驱动的权重选择准则. 理论结果在第 3 节中呈现. 第 4 节为全局 Fréchet 回归模型构建一种可靠的变量重

要性度量. 第 5 和 6 节分别报告模拟研究和实际数据分析. 第 7 节是本文研究的总结与展望. 附录 A 提供了理论结果的证明, 附录 B 提供了条件的合理性说明和示例.

2 模型设定和模型平均估计

2.1 记号

为方便阅读, 本小节在表 1 中汇总了全文核心记号及其定义, 其余符号将在首次引入时定义.

2.2 全局 Fréchet 回归模型

考虑取值于乘积空间 $(\mathbb{S}, d)$ 的随机元 $(X, Y) \sim F$, 其中自变量 X 取值于 $\mathcal{X} \subseteq \mathbb{R}^p$, 随机元 Y 取值于 $\mathbb{S}$. 记 F 为 (X, Y) 的概率分布. 假定 X 的均值和方差存在, 记 $\mu = E(X)$ 和 $\Sigma = \text{Var}(X)$, 并假设 Σ 正定.

传统的均值和方差定义仅适用于取值于 Euclid 空间的随机变量, 将其推广到取值于度量空间的随机元时存在困难. 为解决这一问题, 文献 [12] 将 Euclid 空间中随机变量的均值和方差的定义推广到一般度量空间, 提出了 Fréchet 均值和 Fréchet 方差. 具体而言, 对于取值于度量空间的随机元 Y , 其 Fréchet 均值和 Fréchet 方差分别定义为

$$m_{\oplus} = \arg \min_{s \in \mathbb{S}} E(d^2(Y, s)), \quad V_{\oplus} = E(d^2(Y, m_{\oplus})).$$

为了研究随机元与相应自变量之间的关系, 文献 [25] 提出了在给定 $X = x = (x_1, \dots, x_p)^T$ 时 Y 的 Fréchet 回归函数的概念, 即

$$m_{\oplus}(x) = \arg \min_{s \in \mathbb{S}} M_{\oplus}(s, x),$$

其中 $M_{\oplus}(\cdot, x) = E(d^2(Y, \cdot) | X = x)$. 基于以上概念, 文献 [25] 引入全局 Fréchet 回归模型

$$m_{\oplus}(x) = \arg \min_{s \in \mathbb{S}} E\{[1 + (X - \mu)^T \Sigma^{-1}(x - \mu)]d^2(Y, s)\}, \quad x \in \mathcal{X}, \quad (2.1)$$

表 1 本文使用的核心记号

记号	定义
C	一般的正常数 (每次取值可能不同)
$(\mathbb{S}, d)$	度量空间, 其中 d 表示度量
M_k	第 k 个候选模型包含的自变量指标集
$\mathcal{M}$	候选模型集合 $\{M_k, k = 1, \dots, K\}$
$x_{(k)}$	向量 x 中由 M_k 索引的分量构成的子向量
$A_{(k)}$	方阵 A 中由 M_k 索引的行列构成的子矩阵
$w = (w_1, \dots, w_K)^T$	权重向量, 属于集合 $\mathcal{W} = \{w \in [0, 1]^K : \sum_{k=1}^K w_k = 1\}$
$W_{(k)}(s, x)$	$E\{[1 + (X_{(k)} - \mu_{(k)})^T \Sigma_{(k)}^{-1}(x_{(k)} - \mu_{(k)})]d^2(Y, s)\}$
$W_{n(k)}(s, x)$	$n^{-1} \sum_{i=1}^n [1 + (X_{(k)i} - \bar{X}_{(k)})^T \hat{\Sigma}_{(k)}^{-1}(x_{(k)} - \bar{X}_{(k)})]d^2(Y_i, s)$
$\ \cdot\ $	$\mathbb{R}^p$ 空间上的 Euclid 范数
$B(\delta; s)$	以 s 为中心、 δ 为半径的球体
$N(\epsilon, B(\delta; s), d)$	$B(\delta; s)$ 关于度量 d 的 ϵ -覆盖数 (覆盖球体尺寸为 ϵ)

该全局模型的拟合不涉及调节参数, 因此便于在实际中应用. 可以验证经典的线性回归模型是全局 Fréchet 回归模型 (2.1) 的一个特例 (推导细节参见文献 [27]). 因此, 全局 Fréchet 回归模型 (2.1) 可视为经典线性回归在一般度量空间响应变量情形下的自然推广.

假定 $\{(X_i, Y_i) : i = 1, \dots, n\}$ 是来自于分布 F 的一个随机样本. 记 $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ 为样本均值且 $\hat{\Sigma} = n^{-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$ 为样本协方差矩阵. 对于任意 $x \in \mathcal{X}$, $m_{\oplus}(x)$ 的全局 Fréchet 回归估计量定义为

$$\hat{m}_{\oplus}(x) = \arg \min_{s \in \mathbb{S}} \sum_{i=1}^n [1 + (X_i - \bar{X})^T \hat{\Sigma}^{-1} (x - \bar{X})] d^2(Y_i, s).$$

2.3 模型平均估计

在许多情况下, 真实的条件均值函数 $m_{\oplus}(x)$ 不满足等式 (2.1), 因此将全局 Fréchet 回归模型看作近似模型更为合适. 假设考虑 K 个候选模型 $\{M_k, k = 1, \dots, K\}$, 其中第 k 个候选模型可表示为

$$m_{\oplus(k)}(x) = \arg \min_{s \in \mathbb{S}} E\{[1 + (X_{(k)} - \mu_{(k)})^T \Sigma_{(k)}^{-1} (x_{(k)} - \mu_{(k)})] d^2(Y, s)\}, \quad x \in \mathcal{X}.$$

故 $m_{\oplus(k)}(x)$ 在第 k 个候选模型下的估计量可通过

$$\hat{m}_{\oplus(k)}(x) = \arg \min_{s \in \mathbb{S}} \sum_{i=1}^n [1 + (X_{(k)i} - \bar{X}_{(k)})^T \hat{\Sigma}_{(k)}^{-1} (x_{(k)} - \bar{X}_{(k)})] d^2(Y_i, s) \quad (2.2)$$

得到, 其中 $\hat{\Sigma}_{(k)} = n^{-1} \sum_{i=1}^n (X_{(k)i} - \bar{X}_{(k)})(X_{(k)i} - \bar{X}_{(k)})^T$ 为 M_k 所对应的样本协方差矩阵. 由于 $m_{\oplus(k)}(x)$ 是 $E\{[1 + (X_{(k)} - \mu_{(k)})^T \Sigma_{(k)}^{-1} (x_{(k)} - \mu_{(k)})] d^2(Y, \cdot)\}$ 的最小化点, 因此可以将其理解为在给定自变量 $X_{(k)}$ 时, 对 $m_{\oplus}(x)$ 的最佳近似 (在其存在的前提下). 给定权重向量 $w = (w_1, \dots, w_K)^T$, $m_{\oplus}(x)$ 的模型平均估计量为 $\hat{m}_{\oplus}(x; w) = \sum_{k=1}^K w_k \hat{m}_{\oplus(k)}(x)$. 为了保证 $\hat{m}_{\oplus}(x; w)$ 属于 $\mathbb{S}$, 本文假定 $\mathbb{S}$ 是凸集, 即 $\mathbb{S}$ 包含其内任意有限个点的凸组合.

对关键的权重选择准则, 本文采用留一交叉验证法以确定关键权重. 对 $k = 1, \dots, K$, 记 $\tilde{m}_{\oplus(k)}^{-i}(x)$ 为 $m_{\oplus(k)}(x)$ 的 Jackknife 估计, 其计算公式为

$$\tilde{m}_{\oplus(k)}^{-i}(x) = \arg \min_{s \in \mathbb{S}} \sum_{j \neq i} [1 + (X_{(k)j} - \bar{X}_{(k)}^{-i})^T \hat{\Sigma}_{(k), -i}^{-1} (x_{(k)} - \bar{X}_{(k)}^{-i})] d^2(Y_j, s),$$

其中 $\bar{X}_{(k)}^{-i} = (n-1)^{-1} \sum_{j \neq i} X_{(k)j}$ 且 $\hat{\Sigma}_{(k), -i} = (n-1)^{-1} \sum_{j \neq i} (X_{(k)j} - \bar{X}_{(k)}^{-i})(X_{(k)j} - \bar{X}_{(k)}^{-i})^T$. 基于留一交叉验证的权重选择准则由下式给出:

$$CV(w) = \frac{1}{n} \sum_{i=1}^n d^2(Y_i, \tilde{m}_{\oplus}(X_i; w)),$$

其中 $\tilde{m}_{\oplus}(X_i; w) = \sum_{k=1}^K w_k \tilde{m}_{\oplus(k)}^{-i}(X_i)$.

最优的权重估计可以通过最小化留一交叉验证准则得到, 即

$$\hat{w} = \arg \min_{w \in \mathcal{W}} CV(w). \quad (2.3)$$

进而, $m_{\oplus}(x)$ 的最终模型平均估计为

$$\hat{m}_{\oplus}(x; \hat{w}) = \sum_{k=1}^K \hat{w}_k \hat{m}_{\oplus(k)}(x).$$

对于 2-Wasserstein 空间 (定义参见第 5.1 小节), 不难验证优化问题 (2.3) 是凸二次规划问题, 因此可以通过常用软件中的约束优化程序 (如 R 语言中的 `optim` 和 `CVXR` 等软件包) 实现 $\hat{w}$ 的求解. 当 $CV(w)$ 关于 w 非凸时, 可使用 R 软件包 `Rsolnp` 等求解最优权重.

3 模型平均的理论性质

本节建立 $\hat{m}_{\oplus(k)}(x)$ 的一致收敛速度, 以及所提出的模型平均估计的渐近最优性及最优权重的渐近表现. 本文的理论部分假设 $\mathcal{X}$ 是紧空间且 $\mathbb{S}$ 是全有界的. 除非特殊说明, 所有的极限性质均为当 n 趋于无穷时成立.

3.1 $\hat{m}_{\oplus(k)}(x)$ 的一致收敛速度

本小节建立 $\hat{m}_{\oplus(k)}(x)$ 的相合性和相应的一致收敛速度. 为了研究 $\hat{m}_{\oplus(k)}(x)$ 的渐近表现, 需要以下假设条件.

条件 3.1 对任意 $x \in \mathcal{X}$, $k = 1, \dots, K$, 假定 $m_{\oplus}(x)$ 和 $m_{\oplus(k)}(x)$ 存在且唯一. 对任意 $x \in \mathcal{X}$, $k = 1, \dots, K$, $i = 1, \dots, n$, 假定 $\hat{m}_{\oplus(k)}(x)$ 和 $\tilde{m}_{\oplus(k)}^{-i}(x)$ 存在且唯一几乎必然成立. 此外, 对任意 $\varepsilon > 0$ 和 $k = 1, \dots, K$, 有

$$\inf_{x \in \mathcal{X}} \inf_{d(s, m_{\oplus(k)}(x)) > \varepsilon} (W_{(k)}(s, x) - W_{(k)}(m_{\oplus(k)}(x), x)) > 0,$$

且存在 $\tau = \tau(\varepsilon) > 0$ 使得

$$P\left(\inf_{x \in \mathcal{X}} \inf_{d(s, \hat{m}_{\oplus(k)}(x)) > \varepsilon} (W_{n(k)}(s, x) - W_{n(k)}(\hat{m}_{\oplus(k)}(x), x)) \geq \tau\right) \rightarrow 1.$$

条件 3.2 对任意给定的 $x \in \mathcal{X}$, 存在 $\eta > 0$ 和 $C > 0$ (可能与 x 有关), 使得对于 $k = 1, \dots, K$ 与 $d(m_{\oplus(k)}(x), s) < \eta$, 有 $W_{(k)}(s, x) - W_{(k)}(m_{\oplus(k)}(x), x) \geq Cd^2(m_{\oplus(k)}(x), s)$. 此外, 随着 $\delta \rightarrow 0^+$, 有

$$\max_{k=1, \dots, K} \int_0^1 [1 + \log N(\delta\epsilon, B(\delta; m_{\oplus(k)}(x)), d)]^{1/2} d\epsilon = O(1).$$

条件 3.3 存在 $\tau > 0$ 和 $C > 0$ 满足

$$\min_{k=1, \dots, K} \inf_{x \in \mathcal{X}} \inf_{d(s, m_{\oplus(k)}(x)) < \tau} [W_{(k)}(s, x) - W_{(k)}(m_{\oplus(k)}(x), x) - Cd^2(m_{\oplus(k)}(x), s)] \geq 0.$$

此外, 随着 $\delta \rightarrow 0^+$, 有

$$\max_{k=1, \dots, K} \int_0^1 \sup_{x \in \mathcal{X}} [1 + \log N(\delta\epsilon, B(\delta; m_{\oplus(k)}(x)), d)]^{1/2} d\epsilon = O(\log(1/\delta)).$$

条件 3.1 被用于建立 M 估计的相合性. 条件 3.1 的第一部分保证了 $W_{(k)}(\cdot, x)$ 存在唯一的极小值点, 从而当 $W_{n(k)}$ 弱收敛到 $W_{(k)}$ 时可得出 $\hat{m}_{\oplus(k)}(x)$ 的一致相合性. 条件 3.1 的第二部分用于证明 $\{d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) : x \in \mathcal{X}\}$ 依概率渐近等度连续. 当 $\mathbb{S} \subseteq \mathbb{R}$ 时, 条件 3.1 自动满足. 具体而言, 记 $W'_{n(k)}(\cdot, x)$ 和 $W''_{n(k)}(\cdot, x)$ 分别为 $W_{n(k)}(\cdot, x)$ 的一阶和二阶导数. 注意到 $W'_{n(k)}(\hat{m}_{\oplus(k)}(x), x) = 0$ 且 $W''_{n(k)}(\cdot, x) = 2n^{-1} \sum_{i=1}^n [1 + (X_{(k)i} - \bar{X}_{(k)})^T \hat{\Sigma}_{(k)}^{-1} (x_{(k)} - \bar{X}_{(k)})]$. 对 $W_{n(k)}(\cdot, x)$ 在 $\hat{m}_{\oplus(k)}(x)$ 处进行 Taylor 展开, 可以得到

$$W_{n(k)}(s, x) - W_{n(k)}(\hat{m}_{\oplus(k)}(x), x)$$

$$= n^{-1} \sum_{i=1}^n [1 + (X_{(k)i} - \bar{X}_{(k)})^T \hat{\Sigma}_{(k)}^{-1} (x_{(k)} - \bar{X}_{(k)})] (s - \hat{m}_{\oplus(k)}(x))^2.$$

进而大数定律和 $\mathcal{X}$ 的紧性保证了 $\sup_{x \in \mathcal{X}} |W''_{n(k)}(\cdot, x) - 2| = o_p(1)$. 这表明条件 3.1 的第二部分对 $\tau = \varepsilon^2$ 成立. 运用类似的方法可知条件 3.1 的第一部分在此情形下也成立. 条件 3.2 和 3.3 与 $m_{\oplus(k)}(x)$ 邻域的曲率和局部覆盖数有关. 其被用于推导 $\hat{m}_{\oplus(k)}(x)$ 的收敛速度. 条件 3.1–3.3 是 Fréchet 回归文献中的常用条件, 参见文献 [4, 9, 25]. 附录 B.1 提供了满足条件 3.1–3.3 的具体示例.

下述的定理保证了 $\hat{m}_{\oplus(k)}(x)$, $k = 1, \dots, K$ 的一致相合性.

定理 3.1 若条件 3.1 成立, 则对任意 $k \in \{1, \dots, K\}$, 有

$$\sup_{x \in \mathcal{X}} d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) = o_p(1).$$

与文献 [25] 的结论不同, 定理 3.1 允许候选模型被错误设定. 进一步地, 下述定理研究了 $\hat{m}_{\oplus(k)}(x)$ 在允许模型误设定时的收敛速度.

定理 3.2 若条件 3.1 和 3.2 成立, 则对任意 $k \in \{1, \dots, K\}$ 和任意给定的 $x \in \mathcal{X}$, 有

$$d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) = O_p(n^{-1/2}).$$

进一步地, 若条件 3.1 和 3.3 成立, 则有

$$\max_{k=1, \dots, K} \sup_{x \in \mathcal{X}} d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) = O_p((n/\log n)^{-1/2}).$$

定理 3.2 给出了逐点收敛速度和一致收敛速度. 值得注意的是, 逐点收敛速度达到 $n^{-1/2}$, 与 Euclid 空间中的最优收敛速度一致. 同时, 一致收敛速度是接近最优的, 其与最优收敛速度仅相差一个对数因子. 该结果改进了文献 [25] 的理论结果: 一致收敛速度为 n^{-c} 对任意 $c < 1/2$ 成立. 定理 3.2 给出的一致收敛速度对于建立模型平均方法的渐近最优性将发挥关键作用. 值得指出的是, 本文将一致收敛速度作为定理结论加以证明, 而在文献 [17, 32] 中为保证方法的理论性质, 类似的结果作为假设而未给出证明.

3.2 渐近最优性

本小节研究所提出的模型平均方法在所有候选模型误设定假设下的理论性质. 记 (X, Y) 为来自于 F 的一个独立样本, 并记 $\mathcal{D}_n$ 为数据 $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$. 样本外预测损失和对应的风险分别定义为

$$L_n(w) = \mathbb{E}[d^2(Y, \hat{m}_{\oplus}(X; w)) \mid \mathcal{D}_n] - \mathbb{E}[d^2(Y, m_{\oplus}(X))]$$

和

$$R_n(w) = \mathbb{E}(L_n(w)) = \mathbb{E}[d^2(Y, \hat{m}_{\oplus}(X; w))] - \mathbb{E}[d^2(Y, m_{\oplus}(X))].$$

记 $\xi_n = \inf_{w \in \mathcal{W}} R_n(w)$. 为了得到渐近最优性, 所需条件如下:

条件 3.4 存在绝对常数 $C > 0$ 使得对所有 $w \in \mathcal{W}$ 与 $s_k, \tilde{s}_k \in \mathbb{S}$, $k = 1, \dots, K$, 有

$$d\left(\sum_{k=1}^K w_k s_k, \sum_{k=1}^K w_k \tilde{s}_k\right) \leq C \sum_{k=1}^K d(s_k, \tilde{s}_k).$$

此外, 存在 $C > 0$ 与 $\alpha > 0$ 使得对所有 $w_1 = (w_{11}, \dots, w_{1K})^T$, $w_2 = (w_{21}, \dots, w_{2K})^T \in \mathcal{W}$ 与 $s_k \in \mathbb{S}$, $k = 1, \dots, K$, 有

$$\sup_{\substack{s_k \in \mathbb{S} \\ k=1, \dots, K}} d\left(\sum_{k=1}^K w_{1k} s_k, \sum_{k=1}^K w_{2k} s_k\right) \leq C \|w_1 - w_2\|^\alpha.$$

条件 3.5 假定 $\{\xi_n^{-1} \max_{k=1, \dots, K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))\}_{n \geq n_0}$ 与 $\{\xi_{n+1}^{-1} \max_{k=1, \dots, K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))\}_{n \geq n_0}$ 对某个正整数 n_0 一致可积.

条件 3.6 随着 $n \rightarrow \infty$, 有 $(n/\log n)^{-1/2}/\xi_n \rightarrow 0$, 其中 $\xi_n = \inf_{w \in \mathcal{W}} R_n(w)$.

条件 3.4 较弱, 任何有界赋范线性空间皆满足此条件. 条件 3.5 用于保证

$$\mathbb{E}\left[\xi_n^{-1} \max_{k=1, \dots, K} d(\hat{m}_{\oplus(k)}^{-n}(X_n), m_{\oplus(k)}(X_n))\right] = o(1), \quad \mathbb{E}\left[\xi_n^{-1} \max_{k=1, \dots, K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))\right] = o(1),$$

从而得到 $\sup_{w \in \mathcal{W}} |\text{CV}(w) - R_n(w)| = o_p(\xi_n)$. 条件 3.6 要求所有候选模型都被错误设定, 并且 ξ_n 的速度慢于 $(n/\log n)^{-1/2}$. 条件 3.5 和 3.6 是模型平均文献中保证最优性的常用条件 (参见文献 [2, 36, 39]). 条件 3.4–3.6 的示例在附录 B.2 中给出.

定理 3.3 若条件 3.1, 3.3–3.6 成立, 则

$$\frac{L_n(\hat{w})}{\inf_{w \in \mathcal{W}} L_n(w)} \xrightarrow{P} 1.$$

定理 3.3 表明基于最小化 $\text{CV}(w)$ 准则得到的模型平均估计在样本外预测损失 $L_n(w)$ 下具有渐近最优性.

3.3 权重的渐近性质

本小节研究权重 $\hat{w}$ 的渐近表现, 包括权重的收敛速度以及过一致性. 定义 $w^0 = \arg\min_{w \in \mathcal{W}} R_n(w)$ 为最小化样本外预测风险的总体最优权重. 为了推导权重估计量 $\hat{w}$ 的收敛速度, 需要如下条件:

条件 3.7 存在 $\tau > 0$, $\beta > 0$ 和 $C > 0$ 满足

$$\inf_{\|w - w^0\| < \tau} [R_n(w) - R_n(w^0) - C\|w - w^0\|^\beta] \geq 0.$$

条件 3.8 存在 $\gamma \in (0, 1/2)$ 使得 $\{n^{(1/2)-\gamma} \max_{k=1, \dots, K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))\}_{n \geq n_0}$ 对某个正整数 n_0 一致可积.

为解释条件 3.7 的合理性, 假定 $R_n(w)$ 关于 w 二次可微且在 w^0 处局部凸使得

$$\liminf_{n \rightarrow \infty} \inf_{\|w - w^0\| < \tau} \lambda_{\min}(\nabla^2 R_n(w)) \geq C > 0,$$

其中 $\lambda_{\min}(\cdot)$ 表示最小特征值, $\nabla^2 R_n(\cdot)$ 表示 $R_n(\cdot)$ 的 Hessian 矩阵. 例如, 当 Y 取值于配备 Euclid 范数的 Euclid 空间与配备 Frobenius 范数的相关矩阵空间时, $R_n(w)$ 二次可微. 对于满足 $\|w - w^0\| < \tau$ 的 $w \in \mathcal{W}$, 由 Taylor 展开可知

$$R_n(w) - R_n(w^0) = \nabla R_n(w^0)^T(w - w^0) + \frac{1}{2}(w - w^0)^T \nabla^2 R_n(\tilde{w})(w - w^0)$$

对某个介于 w 与 w^0 之间的 $\tilde{w}$ 成立, 这里 $\nabla R_n(\cdot)$ 表示梯度向量. 因为 w^0 是 $R_n(w)$ 极小值点, 故有 $\nabla R_n(w^0) = 0$. 这表明条件 3.7 对 $\beta = 2$ 成立. 条件 3.8 用于保证 $\sup_{w \in \mathcal{W}} |\text{CV}(w) - R_n(w)|$ 收敛到 0 的速度快于 $n^{-(1/2)+\gamma}$, 类似的条件也被用于文献 [36, 39], 其合理性说明参见附录 B.3.

定理 3.4 若条件 3.1, 3.3, 3.4, 3.7 和 3.8 成立, 则存在一个局部极小值点 $\hat{w}$ 满足

$$\|\hat{w} - w^0\| = O_p(n^{-\frac{1-2\gamma}{2\beta}}),$$

其中 β 和 γ 分别在条件 3.7 和 3.8 中给出.

定理 3.4 保证了估计的权重 $\hat{w}$ 以 $n^{-(1-2\gamma)/(2\beta)}$ 收敛到未知的总体最优权重. 值得指出的是, 本定理不需要条件 3.6, 而条件 3.6 要求所有的候选模型误设定. 此外, 本定理对 $\hat{w}$ 的渐近表现提供了重要补充. 不难看出, γ 和 β 越小, $\hat{w}$ 收敛到最优权重 w^0 的速度就越快.

接下来研究当候选模型集至少包含一个正确模型时权重的过一致性, 正确模型的定义将在下文中给出. 对于指标集 $\mathcal{I} \subseteq \{1, \dots, p\}$, 用 $\mathcal{I}^c = \{1, \dots, p\} \setminus \mathcal{I}$ 表示 $\mathcal{I}$ 的补集. 文献 [27] 给出了全局 Fréchet 回归的重要自变量的定义: 如果 $\mathcal{I}$ 是满足 $Y \perp X_{\mathcal{I}^c} \mid X_{\mathcal{I}}$ 的最小子集, 即 Y 和 $X_{\mathcal{I}^c}$ 在给定 $X_{\mathcal{I}}$ 时条件独立, 则 $\mathcal{I}$ 被称为全局 Fréchet 回归的重要变量. 进一步地, 如果 $\mathcal{I} \subseteq M_k$, 则称第 k 个候选模型是正确模型.

用 $\mathcal{C} \subseteq \{1, \dots, K\}$ 表示正确模型对应的指标集. 相应地, 令 $w_{\mathcal{C}} = \sum_{k \in \mathcal{C}} w_k$ 表示正确模型的权重之和, 并记 $\hat{w}_{\mathcal{C}} = \sum_{k \in \mathcal{C}} \hat{w}_k$. 对给定的 $0 < \epsilon \leq 1$, 记 $\mathcal{W}_{\epsilon} = \{w \in \mathcal{W} : w_{\mathcal{C}} \leq 1 - \epsilon\}$ 为 $\mathcal{W}$ 的给正确模型分配至多 $1 - \epsilon$ 权重的子集. 记 $\tilde{\xi}_n(\epsilon) = \inf_{w \in \mathcal{W}_{\epsilon}} R_n(w)$. 为了建立权重的过一致性, 需要如下条件:

条件 3.9 对任意满足 $\mathcal{I} \subseteq \mathcal{S} \subseteq \{1, \dots, p\}$ 的 $\mathcal{S}$, 有

$$\{E[\Sigma_{\mathcal{S}}^{-1}(X_{\mathcal{S}} - \mu_{\mathcal{S}}) \mid X_{\mathcal{I}}]\}_{\mathcal{I}} = \Sigma_{\mathcal{I}}^{-1}(X_{\mathcal{I}} - \mu_{\mathcal{I}}), \quad \{E[\Sigma_{\mathcal{S}}^{-1}(X_{\mathcal{S}} - \mu_{\mathcal{S}}) \mid X_{\mathcal{I}}]\}_{\mathcal{S} \setminus \mathcal{I}} = 0,$$

其中 $\mathcal{S} \setminus \mathcal{I} = \{j : j \in \mathcal{S}, j \notin \mathcal{I}\}$.

条件 3.10 对任意给定的 $0 < \epsilon \leq 1$, 有 $(n/\log n)^{-1/2}/\tilde{\xi}_n(\epsilon) = o(1)$. 此外, 假定 $\{\tilde{\xi}_n^{-1}(\epsilon) \max_{k=1, \dots, K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))\}_{n \geq n_0}$ 与 $\{\tilde{\xi}_{n+1}^{-1}(\epsilon) \max_{k=1, \dots, K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))\}_{n \geq n_0}$ 对某个正整数 n_0 一致可积.

如果 X 服从多元椭球分布, 则条件 3.9 自然成立 (参见文献 [10, 定理 2.18]). 根据文献 [27, 引理 1] 可得, 条件 3.9 保证了 $m_{\oplus(k)}(x) = m_{\oplus}(x)$ 对所有的 $x \in \mathcal{X}$ 和 $k \in \mathcal{C}$ 成立. 条件 3.10 与文献 [6, 条件 10 和 11] 类似, 其被用于证明 $E[\tilde{\xi}_n^{-1}(\epsilon) \max_{1 \leq k \leq K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))] = o(1)$ 和 $\sup_{w \in \mathcal{W}} |\text{CV}(w) - R_n(w)| = o_p(\tilde{\xi}_n(\epsilon))$ 对任意给定的 $0 < \epsilon \leq 1$ 成立. 为了更好地理解条件 3.10, 附录 B.3 提供了一个具体示例.

定理 3.5 假定条件 3.1, 3.3, 3.4, 3.9 和 3.10 成立. 若 $\mathcal{C}$ 非空, 则 $\hat{w}_{\mathcal{C}} \xrightarrow{P} 1$.

定理 3.5 表明, 如果候选模型集中至少存在一个正确模型, 则交叉验证准则渐近地将全部权重集中于正确模型. 作为定理 3.5 的特例, 若候选模型集中只有一个正确模型, 则交叉验证准则渐近地将全部权重分配给该正确模型. 以上结果补充了模型平均方法在候选模型误设定假设下的渐近最优性. 下面的推论保证了当候选模型集中存在正确模型时, 所提模型平均方法的预测表现.

推论 3.1 若定理 3.5 的条件成立, 对任意给定的 $0 < \epsilon \leq 1$, 有

$$\frac{L_n(\hat{w})}{\inf_{w \in \mathcal{W}_{\epsilon}} L_n(w)} \xrightarrow{P} 0.$$

推论 3.1 说明, 当候选模型集中存在正确模型时, 相对于给正确模型至多分配 $1 - \epsilon$ 权重的任何模型平均估计, $\hat{w}$ 的样本外预测损失是渐近可忽略的.

4 变量重要性度量

本节给出一个可靠的变量重要性度量方法, 该方法给 Fréchet 回归中的每个自变量分配一个重要性度量值. 与文献 [33] 的方法类似, 下文将给出基于候选模型集 $\mathcal{M} = \{M_k, k = 1, \dots, K\}$ 和估计权重 $\hat{w} = (\hat{w}_1, \dots, \hat{w}_K)^T$ 的变量重要性定义.

定义 4.1 第 j 个变量 X_j ($j \in \{1, \dots, p\}$) 的重要性定义为

$$S_j = S(j; \hat{w}, \mathcal{M}) = \sum_{k=1}^K \hat{w}_k I(j \in M_k),$$

其中 $I(\cdot)$ 表示示性函数.

可以清楚地看到, 第 j 个变量的变量重要性值为包含 X_j 的所有候选模型所对应的权重之和, 且其取值在 0 到 1 之间. 如果 $\hat{w}$ 充分集中在正确模型上, 则该变量重要性度量方法具有良好的性质. 下述定理保证了在保证 $\hat{w}_{\mathcal{C}} \xrightarrow{P} 1$ 的条件下, 所提出的变量重要性度量方法具有过一致性.

定理 4.1 若定理 3.5 的条件成立, 则

$$\min_{j \in \mathcal{I}} S_j \xrightarrow{P} 1, \quad \max_{j \in \mathcal{I}_C^-} S_j \xrightarrow{P} 0,$$

其中 $\mathcal{I}_C^- = \{1, \dots, p\} \setminus (\bigcup_{k \in \mathcal{C}} M_k)$.

根据定义, 可知指标集 $\mathcal{I}_C^-$ 中不包含任何属于正确模型的指标. 定理 4.1 保证了所提出的变量重要性度量的过一致性, 即 $\mathcal{I}$ 中每个重要变量的重要性值依概率收敛到 1, 同时每个不包含在任何正确模型中的变量重要性值依概率收敛到 0. 特别地, 如果指标集 $\mathcal{I}_C = \bigcup_{k \in \mathcal{C}} M_k$ 中仅包含对应于真模型的指标, 则不属于真模型的变量所对应的变量重要性值全部收敛到 0. 在此情况下, 该变量重要性度量方法具有一致性而非过一致性.

此外, 变量重要性度量的过一致性不能保证所有不重要变量的重要性值收敛到 0, 除非有额外的假设条件, 如权重的一致性 (参见文献 [33, 定理 1]). 当最大的候选模型 $\{1, \dots, p\}$ 属于候选模型集时, $\mathcal{I}_C = \{1, \dots, p\}$ 和 $\mathcal{I}_C^- = \emptyset$. 在此情况下, 定理 4.1 的第二部分是平凡的. 然而, 采用与定理 4.1 类似的证明方法可以证明如下更一般的结论. 记 $\mathcal{D}$ 为 $\{1, \dots, K\}$ 的满足 $\sum_{k \in \mathcal{D}} \hat{w}_k \xrightarrow{P} 1$ 的任一子集并定义 $\mathcal{I}_D^- = \{1, \dots, p\} \setminus (\bigcup_{k \in \mathcal{D}} M_k)$, 则当 $n \rightarrow \infty$ 时, $\max_{j \in \mathcal{I}_D^-} S_j \xrightarrow{P} 0$.

5 模拟研究

本节研究所提模型平均方法的有限样本预测表现、权重的过一致性以及变量重要性度量的过一致性.

5.1 预测表现

本小节选用两种不同类型的响应变量取值空间 $\mathbb{S}$ 以评估所提模型平均方法的有限样本预测表现, 包括配备了 2-Wasserstein 度量的一维概率分布空间以及基于 Cholesky 分解度量的对称正定矩阵空间. 将本文所提出的基于全局 Fréchet 回归的模型平均方法简记为 MAGFR, 并考虑以下 3 种方法进行比较: (1) 文献 [25] 提出的基于全部自变量的全局 Fréchet 回归 (简记为 GFR); (2) 文献 [27] 提出的 Fréchet 岭选择算子 (简记为 FRiSO); (3) 等权重模型平均估计 (简记为 Equal).

候选模型集为自变量的全体子集模型, 故候选模型的总数为 $K = 2^p - 1$. 为了评估不同方法的预测表现, 本小节考虑两种误差衡量准则:

$$D_1 = n^{-1} \sum_{i=1}^n d^2(m_{\oplus}(x_i), \hat{m}_{\oplus}(x_i)) \quad \text{和} \quad D_2 = n^{-1} \sum_{i=1}^n d^2(m_{\oplus}(\tilde{x}_i), \hat{m}_{\oplus}(\tilde{x}_i)),$$

其中 $\hat{m}_{\oplus}(x_i)$ 表示基于 GFR, FRiSO, Equal 和 MAGFR 得到的对 $m_{\oplus}(x_i)$ 的预测, $\{\tilde{x}_i\}_{i=1}^n$ 为一组独立的测试集. 重复实验 500 次并计算 D_1 和 D_2 的平均值.

首先给出响应变量为一维概率分布的例子. 取 $\mathbb{S}$ 为由满足 $\int_{\mathbb{R}} x^2 dF(x) < \infty$ 的分布函数 $F(\cdot)$ 全体所构成的度量空间. 对 $F(\cdot), G(\cdot) \in \mathbb{S}$, 2-Wasserstein 度量的定义为

$$W_2(F, G) = \left(\int_0^1 (F^{-1}(t) - G^{-1}(t))^2 dt \right)^{1/2},$$

其中 $F^{-1}(\cdot)$ 和 $G^{-1}(\cdot)$ 分别表示对应于 $F(\cdot)$ 和 $G(\cdot)$ 的分位函数. 为了使 2-Wasserstein 度量在计算上便于实施, 引入一组 $[0, 1]$ 上的等距节点 $\{t_1, \dots, t_m\}$, 并计算 $(m^{-1} \sum_{i=1}^m (F^{-1}(t_i) - G^{-1}(t_i))^2)^{1/2}$ 作为 $W_2(F, G)$ 的离散近似. 根据 Tucker 等 [27] 的建议, 本小节取 $m = 20$.

本小节采用与文献 [25, 27] 类似的例子. 具体而言, 为了产生相关的自变量 $X_j \sim U(-1, 1)$, $j = 1, \dots, p$, 首先生成 $(V_1, \dots, V_p) \sim N_p(0, \Sigma)$, 其中 $\Sigma = (\rho^{|i-j|})_{p \times p}$, 并计算 $X_j = 2\Phi(V_j) - 1$, $j = 1, \dots, p$, 其中 Φ 表示标准正态分布的分布函数.

例 5.1 给定 X , 响应变量 Y 按照如下方式产生: $Y^{-1}(\cdot) = \mu + \sigma\Phi^{-1}(\cdot)$, 其中 $\mu \mid X \sim N(\mu_0 + \beta_1 x_1 + \beta_2 x_2, \nu_1)$ 与 $\sigma \mid X \sim \text{Gamma}((\sigma_0 + \beta_3 x_4 + \dots + \beta_7 x_7)^2 / \nu_2, \nu_2 / (\sigma_0 + \beta_3 x_4 + \dots + \beta_7 x_7))$ 独立产生. Fréchet 回归函数为

$$m_{\oplus}^{-1}(x) = E(Y^{-1}(\cdot) \mid X = x) = \mu_0 + \beta_1 x_1 + \beta_2 x_2 + (\sigma_0 + \beta_3 x_4 + \dots + \beta_7 x_7) \Phi^{-1}(\cdot).$$

真实回归系数为 $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7) = (1, 0, 0, 0, 1, 0, -0.5)$. 其他参数设置为 $\mu_0 = 0$, $\sigma_0 = 3$, $\nu_1 = 0.5$, $\nu_2 = 0.5$, $\rho = 0.5$.

例 5.2 真实回归系数为 $\beta_i = 1.5^{-i}$, $i = 1, \dots, 6$, $\beta_7 = 1$, 其余设置与例 5.1 相同.

例 5.3 真实回归系数为 $\beta_i = i^{-1.5}$, $i = 1, \dots, 6$, $\beta_7 = 1$, 其余设置与例 5.1 相同.

接下来给出响应变量为对称正定矩阵 (SPD) 的例子. 取 $\mathbb{S}$ 为由 $r \times r$ 对称正定矩阵构成的度量空间. 根据 Cholesky 分解, 对 $A, B \in \mathbb{S}$, 有 $A = (A^{1/2})^T A^{1/2}$ 与 $B = (B^{1/2})^T B^{1/2}$. 定义 A 和 B 的 Cholesky 分解度量为

$$(\text{trace}((A^{1/2} - B^{1/2})^T (A^{1/2} - B^{1/2})))^{1/2},$$

其中 $\text{trace}(A)$ 表示矩阵 A 的迹. 在例 5.4–5.6 中, 自变量 $(X_1, \dots, X_p) \sim N_p(0, \Sigma)$, $\Sigma = (\rho^{|i-j|})_{p \times p}$. 记 I_r 为 $r \times r$ 单位矩阵, $U = (U_{ij})_{r \times r}$ 且 $U_{ij} = I$ ($i < j$). 令 T 表示 $r \times r$ Toeplitz 矩阵, 其元素为 $T_{ij} = \rho^{|i-j|}$.

例 5.4 Fréchet 回归函数为

$$m_{\oplus}(x) = E(Y \mid X = x) = A(x)^T A(x),$$

其中 $A(x) = (\mu_0 + \beta_1 x_1 + \beta_4 x_4 + \beta_7 x_7) I_r + (\beta_2 x_2 + \beta_5 x_5) U + (\beta_3 x_3 + \beta_6 x_6) T$. 给定 X , 响应变量 $Y = A(x)^T (N^{-1} \sum_{i=1}^N Z_i Z_i^T) A(x)$, 这里 Z_i 从 $N_r(0, I_r)$ 中独立产生. 真实回归系数为 $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7) = (1, 0, 0, 0, 1, 0, -0.5)$. 其他参数设置为 $\mu_0 = 0.5$, $\rho = 0.5$, $r = 5$, $N = 30$.

例 5.5 真实回归系数为 $\beta_i = 1.5^{-i}$, $i = 1, \dots, 6$, $\beta_7 = 1$ 且 $N = 50$, 其余设置均与例 5.4 相同.

例 5.6 真实回归系数为 $\beta_i = i^{-1.5}$, $i = 1, \dots, 6$, $\beta_7 = 1$ 且 $N = 50$, 其余设置均与例 5.4 相同.

表 2 报告了例 5.1–5.6 中每个方法的 D_1 和 D_2 的均值和标准误差 (在括号中). 对于概率分布类

表 2 例 5.1–5.6 中 4 种方法的 D_1 和 D_2 及其所对应的标准误差 ($\times 10^{-1}$)

		D_1				D_2			
		GFR	FRiSO	Equal	MAGFR	GFR	FRiSO	Equal	MAGFR
例 5.1	$n = 30$	2.366 (0.039)	2.146 (0.044)	1.890 (0.027)	1.591 (0.032)	3.309 (0.072)	2.732 (0.070)	2.340 (0.039)	2.012 (0.050)
	$n = 50$	1.393 (0.024)	1.274 (0.032)	1.607 (0.020)	0.940 (0.021)	1.694 (0.034)	1.471 (0.040)	1.834 (0.025)	1.073 (0.026)
	$n = 75$	0.926 (0.015)	0.857 (0.020)	1.497 (0.014)	0.628 (0.013)	1.036 (0.019)	0.932 (0.023)	1.628 (0.017)	0.672 (0.015)
	$n = 100$	0.699 (0.011)	0.622 (0.014)	1.410 (0.012)	0.461 (0.010)	0.769 (0.014)	0.664 (0.016)	1.524 (0.014)	0.490 (0.011)
	例 5.2	$n = 30$	2.377 (0.039)	2.395 (0.046)	1.684 (0.026)	1.701 (0.031)	3.297 (0.072)	3.099 (0.076)	2.059 (0.037)
例 5.2	$n = 50$	1.381 (0.022)	1.471 (0.030)	1.338 (0.019)	1.047 (0.019)	1.681 (0.033)	1.730 (0.040)	1.503 (0.023)	1.188 (0.024)
	$n = 75$	0.945 (0.015)	1.060 (0.021)	1.211 (0.014)	0.767 (0.014)	1.058 (0.019)	1.147 (0.024)	1.300 (0.017)	0.820 (0.015)
	$n = 100$	0.706 (0.012)	0.795 (0.014)	1.149 (0.013)	0.596 (0.011)	0.771 (0.014)	0.863 (0.017)	1.237 (0.014)	0.638 (0.013)
	例 5.3	$n = 30$	2.376 (0.039)	2.158 (0.051)	1.924 (0.028)	1.596 (0.032)	3.312 (0.073)	2.764 (0.081)	2.389 (0.040)
	$n = 50$	1.379 (0.023)	1.197 (0.028)	1.596 (0.021)	0.945 (0.017)	1.676 (0.033)	1.365 (0.035)	1.803 (0.026)	1.064 (0.022)
例 5.3	$n = 75$	0.941 (0.015)	0.885 (0.017)	1.485 (0.016)	0.687 (0.013)	1.053 (0.019)	0.954 (0.020)	1.603 (0.018)	0.732 (0.014)
	$n = 100$	0.704 (0.012)	0.695 (0.012)	1.430 (0.014)	0.538 (0.010)	0.766 (0.014)	0.746 (0.014)	1.542 (0.016)	0.574 (0.011)
	例 5.4	$n = 30$	1.325 (0.055)	1.352 (0.056)	3.219 (0.133)	0.925 (0.038)	1.802 (0.075)	1.679 (0.070)	3.998 (0.166)
	$n = 50$	0.869 (0.037)	1.046 (0.044)	3.171 (0.134)	0.643 (0.027)	1.041 (0.044)	1.178 (0.050)	3.613 (0.153)	0.741 (0.031)
	$n = 75$	0.630 (1.000)	1.001 (1.000)	3.161 (1.000)	0.485 (1.000)	0.705 (1.000)	1.082 (1.000)	3.459 (1.000)	0.542 (1.000)
例 5.4	$n = 100$	0.521 (0.024)	0.981 (0.045)	3.122 (0.142)	0.414 (0.019)	0.546 (0.025)	1.052 (0.048)	3.344 (0.152)	0.448 (0.020)
	例 5.5	$n = 30$	1.993 (0.015)	2.353 (0.039)	2.386 (0.015)	1.402 (0.011)	2.687 (0.029)	3.132 (0.065)	2.892 (0.027)
	$n = 50$	1.286 (0.009)	1.572 (0.026)	2.188 (0.012)	0.980 (0.008)	1.511 (0.013)	1.878 (0.038)	2.491 (0.020)	1.079 (0.009)
	$n = 75$	0.894 (0.006)	1.119 (0.017)	2.068 (0.008)	0.724 (0.005)	0.986 (0.007)	1.264 (0.022)	2.244 (0.015)	0.768 (0.006)
	$n = 100$	0.706 (0.004)	0.974 (0.017)	2.009 (0.007)	0.597 (0.004)	0.761 (0.005)	1.081 (0.022)	2.155 (0.013)	0.628 (0.005)
例 5.5	例 5.6	$n = 30$	2.139 (0.016)	2.112 (0.040)	2.702 (0.016)	1.381 (0.011)	2.883 (0.031)	2.713 (0.071)	3.267 (0.032)
	$n = 50$	1.378 (0.009)	1.424 (0.020)	2.500 (0.013)	0.965 (0.007)	1.619 (0.013)	1.634 (0.027)	2.854 (0.024)	1.052 (0.009)
	$n = 75$	0.961 (0.006)	1.101 (0.016)	2.380 (0.009)	0.714 (0.005)	1.060 (0.008)	1.205 (0.019)	2.585 (0.018)	0.752 (0.006)
	$n = 100$	0.757 (0.005)	0.977 (0.016)	2.314 (0.008)	0.590 (0.004)	0.815 (0.006)	1.062 (0.019)	2.486 (0.016)	0.617 (0.005)

注: 在每种设置中, 表现最佳的结果以黑体标出.

型的响应变量 (例 5.1–5.3), 可以看到所提出的模型平均方法在几乎所有的设定下均比其他 3 种方法 (GFR, FRiSO 和 Equal) 表现更佳. 在例 5.2 中样本量 $n = 30$ 时, 等权重模型平均估计的 D_1 和 D_2 值比 MAGFR 略低, 但是等权重模型平均估计在其他设定下表现不佳. 尽管所有方法的 D_1 和 D_2 值随着样本量 n 增加都有所减小, 但 MAGFR 仍然具有几乎最小的 D_1 和 D_2 结果.

对于正定对称矩阵类型的响应变量 (例 5.4–5.6), MAGFR 在全部情形下均比其他 3 种方法 (GFR, FRiSO 和 Equal) 更具有优势. 在此类情形下, 等权重模型平均估计的表现最差, 并且随着样本量从 30 增加到 100, 其预测表现并未表现出明显提升. 相反, 本文所提出的方法的 D_1 和 D_2 值随着样本量的增加而显著减小. 通过比较各设定下的 D_1 和 D_2 值, 可以发现 D_2 的均值总是大于 D_1 的均值, 并且随着样本量的增加二者的均值越来越接近. 这种规律对例 5.1–5.3 也同样适用.

5.2 权重和变量重要性

为了研究 $\hat{w}$ 的过一致性, 本小节考虑包含不重要自变量的例 5.1 和 5.4. 图 1 展示了 $\hat{w}_C$ 在样本量 $n \in \{50, 100, 150, 200, 250, 300\}$ 取值下的平均值. 从图 1 中可以发现 $\hat{w}_C$ 关于样本量 n 单调增加且收敛到 1. 这验证了定理 3.5 的有效性.

接下来给出 4 个例子来评估所提出的变量重要性度量在衡量全局 Fréchet 回归自变量重要性时的表现.

例 5.7 除 $\rho \in \{0, 0.5, 0.9\}$ 外, 数据按照与例 5.1 相同的设定产生. 因此本例的重要变量为 X_1 , X_5 和 X_7 . 样本量设置为 $n \in \{50, 100, 200, 300\}$.

例 5.8 除 $\rho \in \{0, 0.5, 0.9\}$ 外, 数据按照与例 5.4 相同的设定产生. 因此本例的重要变量为 X_1 , X_5 和 X_7 . 样本量设置为 $n \in \{50, 100, 200, 300\}$.

例 5.7 和 5.8 基于 500 次重复实验的结果展示在表 3 中, 其对应的标准误差通常小于 0.01, 在此省略. 可以看到所提出的重要性度量方法可以显著地区分重要自变量和不重要自变量. 具体而言, 在所有情形下, 重要自变量所对应的重要性值都明显大于不重要的自变量所对应的重要性值. 当样本量

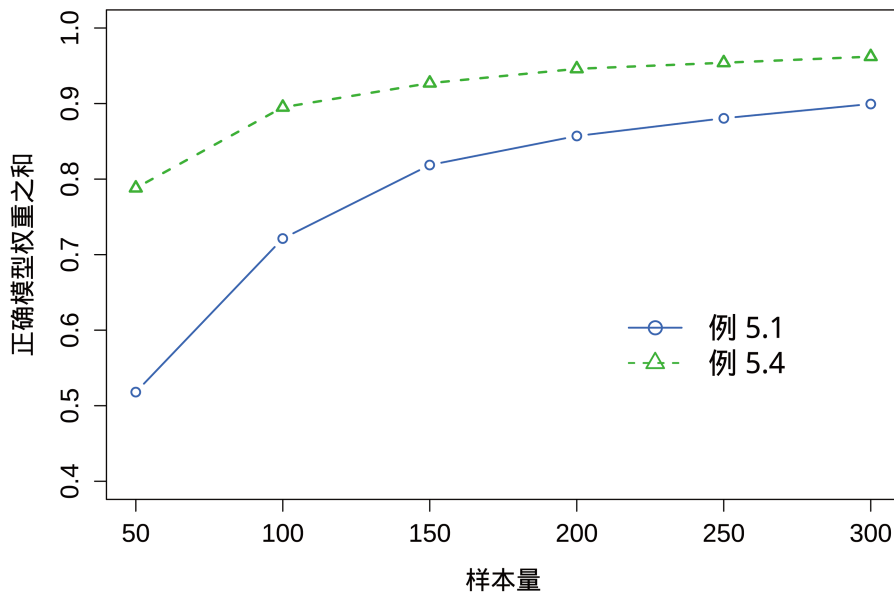


图 1 (网络版彩图) 正确模型的权重之和

表 3 例 5.7 和 5.8 的变量重要性结果

	ρ	n	X_1	X_2	X_3	X_4	X_5	X_6	X_7
例 5.7	0	50	0.920	0.155	0.185	0.172	0.894	0.172	0.618
		100	0.963	0.158	0.152	0.158	0.950	0.172	0.808
		200	0.983	0.152	0.161	0.157	0.977	0.164	0.917
		300	0.988	0.153	0.151	0.153	0.985	0.143	0.945
	0.5	50	0.889	0.169	0.178	0.174	0.841	0.196	0.578
		100	0.941	0.154	0.161	0.161	0.916	0.179	0.784
		200	0.967	0.147	0.142	0.162	0.955	0.156	0.895
		300	0.976	0.138	0.154	0.157	0.965	0.145	0.930
	0.9	50	0.718	0.233	0.217	0.209	0.509	0.233	0.369
		100	0.831	0.196	0.181	0.205	0.678	0.216	0.505
		200	0.896	0.145	0.148	0.183	0.816	0.214	0.716
		300	0.923	0.164	0.160	0.172	0.866	0.190	0.798
例 5.8	0	50	0.969	0.118	0.110	0.117	0.988	0.119	0.842
		100	0.986	0.104	0.123	0.118	0.994	0.113	0.923
		200	0.993	0.108	0.121	0.120	0.996	0.112	0.962
		300	0.995	0.114	0.104	0.110	0.997	0.116	0.975
	0.5	50	0.964	0.118	0.116	0.121	0.972	0.130	0.823
		100	0.982	0.103	0.131	0.112	0.984	0.128	0.912
		200	0.991	0.107	0.126	0.108	0.989	0.115	0.955
		300	0.993	0.113	0.113	0.117	0.991	0.107	0.970
	0.9	50	0.901	0.138	0.137	0.154	0.871	0.181	0.585
		100	0.947	0.120	0.127	0.127	0.919	0.154	0.768
		200	0.973	0.105	0.118	0.106	0.950	0.132	0.874
		300	0.979	0.122	0.123	0.120	0.963	0.109	0.913

从 50 逐渐增加到 300 时, 重要自变量的重要性值逐渐收敛到 1. 值得指出的是, 该重要性度量在自变量之间存在强相关时的表现仍然十分稳健.

例 5.9 产生 $(X_1, X_2, X_3) \sim N(0, \Sigma_\rho)$, 其中

$$\Sigma_\rho = \begin{pmatrix} 1 & 0 & \rho \\ 0 & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}.$$

此外独立产生 $(X_4, \dots, X_7) \sim N_4(0, \Sigma_r)$, 其中 $\Sigma_r = (r^{|i-j|})_{p \times p}$. Y 按照如下线性回归模型产生:

$$Y = -\rho X_1 - \rho X_2 + X_3 + \epsilon,$$

其中误差 ϵ 从 $N(0, 0.25)$ 中独立生成. 样本量设置为 $n \in \{50, 100, 200\}$, 相关系数 $\rho \in \{0.3, 0.5, 0.7\}$.

在本例中, $\text{Cov}(X_1, Y) = \text{Cov}(X_1, -\rho X_1 - \rho X_2 + X_3 + \epsilon) = -\rho - 0 + \rho = 0$ 且 $\text{Cov}(X_2, Y) = 0$, 即 X_1 和 X_2 与 Y 边际不相关. 同时, 由于 $E(Y | X_1, \dots, X_7) = E(Y | X_1, \dots, X_3)$, 故 X_1 和 X_2 是重要变量.

例 5.10 响应变量 $Y(\cdot)$ 按照如下方式产生: $Y^{-1}(\cdot) = (-\rho X_1 - \rho X_2 + X_3 + \epsilon)\Phi^{-1}(\cdot)$, 其余设置与例 5.9 相同.

表 4 展示了例 5.9 和 5.10 基于 500 次重复实验的结果. 为了节省空间, 这里只报告了 $r = 0.5$ 的结果, $r \in \{0, 0.9\}$ 的结果非常类似. 可以发现, 在所有设定下所提出的变量重要性方法可以显著区分

表 4 例 5.9 和 5.10 的变量重要性结果

	ρ	n	X_1	X_2	X_3	X_4	X_5	X_6	X_7
例 5.9	0.3	50	0.879	0.874	0.985	0.187	0.163	0.164	0.164
	0.3	100	0.950	0.951	0.994	0.168	0.162	0.163	0.173
	0.3	200	0.978	0.978	0.997	0.164	0.174	0.165	0.155
	0.5	50	0.936	0.934	0.972	0.185	0.161	0.164	0.164
	0.5	100	0.972	0.972	0.988	0.168	0.162	0.163	0.174
	0.5	200	0.987	0.987	0.994	0.164	0.174	0.165	0.155
	0.7	50	0.442	0.449	0.425	0.230	0.206	0.201	0.204
	0.7	100	0.615	0.619	0.620	0.196	0.190	0.190	0.205
	0.7	200	0.797	0.796	0.799	0.177	0.188	0.181	0.166
例 5.10	0.3	50	0.832	0.848	0.990	0.257	0.235	0.248	0.239
	0.3	100	0.907	0.907	0.998	0.229	0.216	0.220	0.226
	0.3	200	0.957	0.958	1.001	0.226	0.213	0.202	0.206
	0.5	50	0.917	0.922	0.981	0.252	0.243	0.225	0.232
	0.5	100	0.959	0.957	0.991	0.244	0.229	0.228	0.208
	0.5	200	0.983	0.985	0.997	0.228	0.229	0.232	0.210
	0.7	50	0.449	0.479	0.470	0.262	0.255	0.252	0.260
	0.7	100	0.654	0.629	0.670	0.244	0.244	0.235	0.219
	0.7	200	0.812	0.812	0.826	0.232	0.243	0.245	0.231

重要变量 $X_1, \dots, X_3$ 和不重要变量 $X_4, \dots, X_7$. 当样本量从 50 增加到 200 时, 尽管 X_1 和 X_2 与 Y 边际不相关, 但重要变量所对应的重要性值依旧迅速收敛到 1.

6 实际数据分析

本节展示所提出的模型平均方法在各国人类死亡数据上的表现. 在此实际数据分析中, 将死亡年龄分布作为响应变量, 并探索其与包括经济、社会和医疗保健指数在内的各种预测指标的关系. 与传统的汇总统计指标 (如粗死亡率) 不同, 将整个死亡年龄分布视为随机元可能会对人类寿命提供更全面的观点.

本分析使用 162 个国家在 2015 年的寿命表. 通过每个国家的寿命表计算死亡年龄直方图, 其条形宽度 (bin width) 等于 5 年, 然后使用 R 包 `frechet` 对其进行平滑处理, 以获得估计的死亡年龄概率密度函数. 经过此预处理步骤后得到的 162 个国家的死亡年龄概率密度展示在图 2 中. 此外, 还收集了 2015 年的 7 个自变量, 包括 Pop.density: 人口密度 (每平方公里上的人口数量)、Sex.ratio: 性别比 (人口中每 100 名女性对应的男性数量)、GDP: 人均国内生产总值, CPI: 消费者价格指数、Unemp: 失业率、Health.expen: 医疗支出占 GDP 的百分比、Arable.land: 耕地面积占总土地面积的百分比. 这些变量可能对死亡年龄分布有潜在的重要影响. 本数据集从联合国数据库 (<http://data.un.org/>) 和联合国 2019 年世界人口展望数据库 (<https://population.un.org/wpp/Download>) 获得 [35].

首先进行变量重要性分析, 表 5 报告了 7 个自变量的重要性结果. 所提的变量重要性方法识别出 3 个重要变量: Sex.ratio, GDP, 和 Unemp, 该结果与医学和人口统计学的领域知识相符. 实际上, Sex.ratio 的重要性反映了学术界已有充分研究的男女寿命性别差异现象: 有大量数据支持女性寿命长于男性 [3]. GDP 对人类死亡率的重要影响符合预期, 并与文献 [1, 7] 的结论一致. Unemp 的高重要性值验证了文献 [22] 的结论: 在纵向研究中, 失业男性的死亡率显著高于全男性群体的预期水平.

为了评估方法的预测表现, 本节随机选择 n_{train} 个样本作为训练集, 并将其余样本作为测试集. 进而采用均方预测误差 (MSPE) 来衡量每个方法的预测表现. MSPE 的定义为

$$\text{MSPE} = n_{\text{test}}^{-1} \sum_{i=1}^{n_{\text{test}}} d^2(Y_i, \hat{Y}_i),$$

其中 d 取定义于第 5 节的 2-Wasserstein 度量, Y_i 为第 i 个测试样本, $\hat{Y}_i$ 代表每个方法基于训练集给出的预测结果, n_{test} 表示测试集的样本量. 重复实验 500 次且训练集的样本量 $n_{\text{train}} \in \{100, 120, 140\}$. 本节将 FRiSO 生成的解路径上的所有模型作为候选模型集, 调节参数采用文献 [27] 的推荐设置. 类似的构造候选模型集的方法也被应用于研究 (参见文献 [32]).

表 6 总结了在 $m \in \{20, 50\}$ 的设定下各个方法的 MSPE 结果, 其中 m 表示用于计算 2-Wasserstein 距离对应的离散近似时所用到的节点个数 (具体细节参见第 5 节). 可以看到, 在所有情形下 MAGFR 的 MSPE 均小于 GFR, FRiSO 和 Equal, 这表明所提出的模型平均方法对人类死亡数据具有更好的预测表现, 而 GFR 的表现较差.

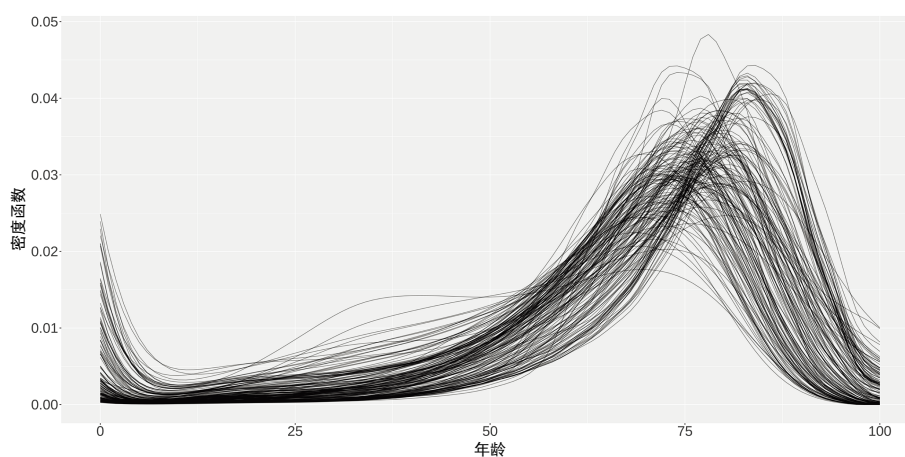


图 2 每个国家在年龄区间 $[0, 100]$ 内的估计死亡年龄概率密度

表 5 人类死亡数据的变量重要性结果

Pop.density	Sex.ratio	GDP	CPI	Unemp	Health.expen	Arable.land
0.059	0.123	0.941	0.019	0.639	0.040	3E-8

表 6 人类死亡数据的 MSPE

	n_{train}	GFR	FRiSO	Equal	MAGFR
$m = 20$	100	63.732	46.855	46.324	46.093
	120	61.512	46.569	45.684	45.611
	140	59.374	46.173	44.931	44.534
$m = 50$	100	64.685	47.434	46.957	46.528
	120	62.489	46.691	46.656	46.041
	140	60.256	45.944	45.549	44.911

7 结论

本文考虑了全局 Fréchet 回归的模型平均方法以解决模型选择不确定性问题. 在理论上, 本文建立了全局 Fréchet 回归估计量在各候选模型下的相合性并推导了收敛速度. 一方面, 在候选模型误设定下, 证明了基于留一交叉验证的模型平均估计可以实现最小的样本外预测损失. 另一方面, 如果候选模型集包含正确模型, 本文建立了权重的过一致性. 此外还得出权重的收敛速度. 本文发展了一种可靠的变量重要性度量来衡量全局 Fréchet 回归中自变量的重要性, 并从理论上证明了该重要性度量方法可以有效识别重要变量.

数值结果表明, 与常用的方法相比, 所提出的模型平均方法有更优秀的预测表现. 然而, 本文的理论结果是建立在自变量维数和候选模型数量固定的假设下. 将结论推广至高维 Fréchet 回归是未来研究的一个重要方向, 但在技术上面面临显著挑战. 近年来, Fréchet 回归被推广到半参数模型的框架中, 如文献 [4, 13] 提出的单指标 Fréchet 回归. 发展半参数 Fréchet 回归的模型平均方法是值得进一步研究的问题.

致谢 感谢编委和审稿人提出的宝贵意见.

参考文献

- 1 Alam M S, Shahbaz M, Paramati S R. The role of financial development and economic misery on life expectancy: Evidence from post financial reforms in India. *Soc Indic Res*, 2016, 128: 481–497
- 2 Ando T, Li K C. A model-averaging approach for high-dimensional regression. *J Amer Statist Assoc*, 2014, 109: 254–265
- 3 Austad S N, Fischer K E. Sex differences in lifespan. *Cell Metab*, 2016, 23: 1022–1033
- 4 Bhattacharjee S, Müller H G. Single index Fréchet regression. *Ann Statist*, 2023, 51: 1770–1798
- 5 Bigot J, Gouet R, Klein T, et al. Geodesic PCA in the Wasserstein space by convex PCA. *Ann Inst Henri Poincaré Probab Stat*, 2017, 53: 1–26
- 6 Chen Z, Liao J, Xu W, et al. Multifold cross-validation model averaging for generalized additive partial linear models. *J Comput Graph Statist*, 2023, 32: 1649–1659
- 7 Chetty R, Stepner M, Abraham S, et al. The association between income and life expectancy in the United States, 2001–2014. *J Anal Math*, 2016, 315: 1750–1766
- 8 Draper D. Assessment and propagation of model uncertainty. *J R Stat Soc Ser B Stat Methodol*, 1995, 57: 45–70
- 9 Dubey P, Müller H G. Fréchet analysis of variance for random objects. *Biometrika*, 2019, 106: 803–821
- 10 Fang K W. Symmetric Multivariate and Related Distributions. Boca Raton: CRC Press, 2018
- 11 Ferreira L K, Busatto G F. Resting-state functional connectivity in normal brain aging. *Neurosci Biobehav Rev*, 2013, 37: 384–400
- 12 Fréchet M. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann Inst Henri Poincaré*, 1948, 10: 215–310
- 13 Ghosal A, Meiring W, Petersen A. Fréchet single index models for object response regression. *Electron J Stat*, 2023, 17: 1074–1112
- 14 Han K, Müller H G, Park B U. Additive functional regression for densities as responses. *J Amer Statist Assoc*, 2020, 115: 997–1010
- 15 Hansen B E. Least squares model averaging. *Econometrica*, 2007, 75: 1175–1189
- 16 Hoeting J A, Madigan D, Raftery A E, et al. Bayesian model averaging: A tutorial. *Statist Sci*, 1999, 14: 382–401
- 17 Kurisu D, Otsu T. Model averaging for global Fréchet regression. *J Multivariate Anal*, 2025, 207: 105416
- 18 Li C, Li Q, Racine J S, et al. Optimal model averaging of varying coefficient models. *Stat Sinica*, 2018, 28: 2795–2809
- 19 Liao J, Zong X, Zhang X, et al. Model averaging based on leave-subject-out cross-validation for vector autoregressions. *J Econometrics*, 2019, 209: 35–60
- 20 Liao J, Zou G, Gao Y, et al. Model averaging prediction for time series models with a diverging number of parameters. *J Econometrics*, 2021, 223: 190–221
- 21 Lu X, Su L. Jackknife model averaging for quantile regressions. *J Econometrics*, 2015, 188: 40–58
- 22 Moser K A, Fox A J, Jones D R. Unemployment and mortality in the OPCS longitudinal study. *Lancet*, 1984, 324:

- 1324–1329
- 23 Mourtada J. Exact minimax risk for linear least squares, and the lower tail of sample covariance matrices. *Ann Statist*, 2022, 50: 2157–2178
 - 24 Peng J, Yang Y. On improvability of model selection by model averaging. *J Econometrics*, 2022, 229: 246–262
 - 25 Petersen A, Müller H G. Fréchet regression for random objects with Euclidean predictors. *Ann Statist*, 2019, 47: 691–719
 - 26 Petersen A, Zhang C, Kokoszka P. Fréchet regression for random objects with Euclidean predictors. *Econom Stat*, 2022, 21: 159–178
 - 27 Tucker D C, Wu Y, Müller H G. Variable selection for global Fréchet regression. *J Amer Statist Assoc*, 2023, 118: 1023–1037
 - 28 Van der Vaart A, Wellner J. *Weak Convergence and Empirical Processes*. New York: Springer, 1996
 - 29 Wang H, Marron J S. Object oriented data analysis: Sets of trees. *Ann Statist*, 2007, 35: 1849–1873
 - 30 Wang M, Zhang X, Wan A T K, et al. Jackknife model averaging for high-dimensional quantile regression. *Biometrics*, 2023, 79: 178–189
 - 31 Whittle P. Bounds for the moments of linear and quadratic forms in independent variables. *Theory Probab Appl*, 1960, 5: 302–305
 - 32 Yan X, Zhang X, Zhao P. Frequentist model averaging for global Fréchet regression. *IEEE Trans Inform Theory*, 2025, 71: 1994–2006
 - 33 Ye C, Yang Y, Yang Y. Sparsity oriented importance learning for high-dimensional linear regression. *J Amer Statist Assoc*, 2018, 113: 1797–1812
 - 34 Yuan Z, Yang Y. Combining linear regression models: When and how? *J Amer Statist Assoc*, 2005, 100: 1202–1214
 - 35 Zhang Q, Xue L, Li B. Dimension reduction for Fréchet regression. *J Amer Statist Assoc*, 2023, 119: 2733–2747
 - 36 Zhang X, Liu C A. Model averaging prediction by K-fold cross-validation. *J Econometrics*, 2023, 235: 280–301
 - 37 Zhang X, Wang W. Optimal model averaging estimation for partially linear models. *Stat Sinica*, 2019, 29: 693–718
 - 38 Zhang X, Yu D, Zou G, et al. Optimal model averaging estimation for generalized linear models and generalized linear mixed-effects models. *J Amer Statist Assoc*, 2016, 111: 1175–1790
 - 39 Zhang X, Zou G, Liang H, et al. Parsimonious model averaging with a diverging number of parameters. *J Amer Statist Assoc*, 2020, 115: 972–984
 - 40 Zhu R, Wan A T K, Zhang X, et al. A Mallows-type model averaging estimator for the varying-coefficient partially linear model. *J Amer Statist Assoc*, 2019, 114: 882–892

附录 A 定理的证明

本附录提供本文所有理论结果的证明.

记 $\ell^\infty(\mathbb{S})$ 为 $\mathbb{S}$ 上一致有界函数全体构成的空间. 令 $\|A\|_2$ 表示正定矩阵 A 的谱范数. 记 $u_{(k)}(X, x) = 1 + (X_{(k)} - \mu_{(k)})^T \Sigma_{(k)}^{-1} (x_{(k)} - \mu_{(k)})$ 与 $\hat{u}_{(k)i}(x) = 1 + (X_{(k)i} - \bar{X}_{(k)})^T \hat{\Sigma}_{(k)}^{-1} (x_{(k)} - \bar{X}_{(k)})$. 对任给的概率测度 P , 定义 $L_2(P) = \{f : \int f^2 dP < \infty\}$. 记 $\mathcal{F}$ 为 $L_2(P)$ 的子集, 其存在可测的包络函数 F (称函数 $x \mapsto F(x)$ 是函数类 $\tilde{\mathcal{F}}$ 的包络函数, 如果对任意的 $x \in \mathcal{X}$ 与 $f \in \tilde{\mathcal{F}}$, 有 $|f(x)| \leq F(x)$). 给定两个函数 l 和 u , 括号 $[l, u]$ 是由满足 $l \leq f \leq u$ 的 f 全体构成的函数集. 用 ε - 括号表示满足 $\|u - l\| < \varepsilon$ 的括号 $[l, u]$. $N_{[\cdot]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ 表示覆盖 $\mathcal{F}$ 所需要的 ε - 括号的最小数目. 定义 $L_2(P)$ - 括号积分为 $J_{[\cdot]}(\delta, \mathcal{F}, L_2(P)) = \int_0^\delta [1 + \log N_{[\cdot]}(\varepsilon \|F\|_{P,2}, \mathcal{F}, L_2(P))]^{1/2} d\varepsilon$, 其中 $\|F\|_{P,2} = (\int F^2 dP)^{1/2}$.

引理 A.1 令 M_n 表示以度量空间 Θ 为指标集的随机过程并用 $M : \Theta \mapsto R$ 表示某个确定函数. 假设 $\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \xrightarrow{P} 0$, 且存在 θ_0 使得对任意包含 θ_0 的开集 G , 均有

$$M(\theta_0) > \sup_{\theta \notin G} M(\theta),$$

则对任意满足 $M_n(\hat{\theta}_n) \geq \sup_{\theta} M_n(\theta) - o_p(1)$ 的序列 $\hat{\theta}_n$, 有 $\hat{\theta}_n \xrightarrow{P} \theta_0$.

证明 参见文献 [28, 推论 3.2.3]. □

附录 A.1 定理 3.1 的证明

首先证明对任意 $k \in \{1, \dots, K\}$ 和任意给定的 $x \in \mathcal{X}$, 均有

$$d(\widehat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) = o_p(1).$$

根据引理 A.1 和条件 3.1, 只需要证明 $\sup_{s \in \mathbb{S}} |W_{n(k)}(s, x) - W_{(k)}(s, x)| \xrightarrow{P} 0$ 对任意 $k \in \{1, \dots, K\}$ 成立. 为了得到 $\sup_{s \in \mathbb{S}} |W_{n(k)}(s, x) - W_{(k)}(s, x)| \xrightarrow{P} 0$, 需要证明, 对任意 $k \in \{1, \dots, K\}$, 在以上确界范数为度量的空间 $\ell^\infty(\mathbb{S})$ 中有 $W_{n(k)}(\cdot, x) \xrightarrow{d} W_{(k)}(\cdot, x)$. 事实上, 由于 $W_{(k)}(\cdot, x)$ 是非随机的, 等价地有 $W_{n(k)}(\cdot, x) - W_{(k)}(\cdot, x) \xrightarrow{d} 0$. 因为范数是连续的, 根据文献 [28, 定理 1.3.6] 可得 $\sup_{s \in \mathbb{S}} |W_{n(k)}(s, x) - W_{(k)}(s, x)| \xrightarrow{d} 0$, 从而有 $\sup_{s \in \mathbb{S}} |W_{n(k)}(s, x) - W_{(k)}(s, x)| \xrightarrow{P} 0$. 进而应用文献 [28, 定理 1.5.4 和 1.5.7] 以证明 $W_{n(k)}(\cdot, x) \xrightarrow{d} W_{(k)}(\cdot, x)$. 因此证明可分为如下两个部分:

(i) 证明对任意有限点集 $s_1, \dots, s_j \in \mathbb{S}$, 有

$$(W_{n(k)}(s_1, x), \dots, W_{n(k)}(s_j, x)) \xrightarrow{d} (W_{(k)}(s_1, x), \dots, W_{(k)}(s_j, x));$$

(ii) 证明 $W_{n(k)}(\cdot, x)$ 是依概率渐近等度连续, 即对任意的 $\varepsilon, \eta > 0$, 存在 $\delta > 0$ 使得

$$\limsup_{n \rightarrow \infty} P\left(\sup_{\substack{d(s_1, s_2) < \delta \\ s_1, s_2 \in \mathbb{S}}} |W_{n(k)}(s_1, x) - W_{n(k)}(s_2, x)| > \varepsilon\right) < \eta. \quad (\text{A.1})$$

为了证明 (i), 只需要证明 $W_{n(k)}(s, x) - W_{(k)}(s, x) = o_p(1)$ 对任意给定的 $s \in \mathbb{S}$ 成立. 事实上, 由此结果可立即得到

$$(W_{n(k)}(s_1, x), \dots, W_{n(k)}(s_j, x)) \xrightarrow{P} (W_{(k)}(s_1, x), \dots, W_{(k)}(s_j, x))$$

在 $\mathbb{R}^k$ 中成立. 因此可得 $(W_{n(k)}(s_1, x), \dots, W_{n(k)}(s_j, x)) \xrightarrow{d} (W_{(k)}(s_1, x), \dots, W_{(k)}(s_j, x))$. 记

$$W_{n(k)}^*(s, x) = n^{-1} \sum_{i=1}^n u_{(k)}(X_i, x) d^2(Y_i, s) \quad \text{与} \quad D = \sup_{s_1, s_2 \in \mathbb{S}} d(s_1, s_2).$$

$\mathcal{X}$ 的全有界性假设保证了 $D < \infty$. 此外, 因为 $\mathcal{X}$ 有界, 可以直接得到对任意给定的 $s \in \mathbb{S}$ 均有 $E(W_{n(k)}^*(s, x)) = W_{(k)}(s, x)$ 与

$$\text{Var}(W_{n(k)}^*(s, x)) \leq n^{-1} D^4 E[u_{(k)}(X, x)]^2 \leq 2n^{-1} D^4 (1 + (x_{(k)} - \mu_{(k)})^T \Sigma_{(k)}^{-1} (x_{(k)} - \mu_{(k)})) \rightarrow 0.$$

进而对任意给定的 $s \in \mathbb{S}$ 有 $W_{n(k)}^*(s, x) - W_{(k)}(s, x) = o_p(1)$. 直接计算可得

$$W_{n(k)}(s, x) - W_{n(k)}^*(s, x) = n^{-1} T_{1(k)} \sum_{i=1}^n d^2(Y_i, s) + n^{-1} T_{2(k)}^T \sum_{i=1}^n X_{(k)i} d^2(Y_i, s),$$

其中 $T_{1(k)} = \mu_{(k)}^T \Sigma_{(k)}^{-1} (x_{(k)} - \mu_{(k)}) - \bar{X}_{(k)} \hat{\Sigma}_{(k)}^{-1} (x_{(k)} - \bar{X}_{(k)})$, $T_{2(k)} = \hat{\Sigma}_{(k)}^{-1} (x_{(k)} - \bar{X}_{(k)}) - \Sigma_{(k)}^{-1} (x_{(k)} - \mu_{(k)})$. 因此需要证明该等式右边两项分别依概率收敛到 0. 一方面, $T_{1(k)}$ 有如下表示:

$$\begin{aligned} T_{1(k)} &= (\mu_{(k)} - \bar{X}_{(k)})^T \Sigma_{(k)}^{-1} (x_{(k)} - \mu_{(k)}) + (\bar{X}_{(k)}^T \Sigma_{(k)}^{-1} (x_{(k)} - \mu_{(k)}) \\ &\quad - \bar{X}_{(k)}^T \hat{\Sigma}_{(k)}^{-1} (x_{(k)} - \mu_{(k)})) - \bar{X}_{(k)}^T \hat{\Sigma}_{(k)}^{-1} (\mu_{(k)} - \bar{X}_{(k)}). \end{aligned}$$

令 $\|\cdot\|_F$ 表示矩阵的 Frobenius 范数. 根据 $\mathcal{X}$ 的有界性, 利用中心极限定理可得 $\|\bar{X}_{(k)} - \mu_{(k)}\| \leq \|\bar{X} - \mu\| = O_p(n^{-1/2})$, $\|\hat{\Sigma}_{(k)}^{-1} - \Sigma_{(k)}^{-1}\|_2 = \|\hat{\Sigma}_{(k)}^{-1}(\Sigma_{(k)} - \hat{\Sigma}_{(k)})\Sigma_{(k)}^{-1}\|_2 \leq O_p(\|\hat{\Sigma} - \Sigma\|_F) = O_p(n^{-1/2})$. 另一方面, $\|T_{2(k)}\| \leq \|(\hat{\Sigma}_{(k)}^{-1} - \Sigma_{(k)}^{-1})(x_{(k)} - \mu_{(k)})\| + \|\hat{\Sigma}_{(k)}^{-1}(\mu_{(k)} - \bar{X}_{(k)})\| = O_p(n^{-1/2})$. 同时, 大数定律和 Jensen 不等式保证了

$$n^{-1} \left\| \sum_{i=1}^n X_i d^2(Y_i, s) \right\| \leq D^2 n^{-1} \sum_{i=1}^n \|X_i\| = O_p(1).$$

事实上, $n^{-1} \sum_{i=1}^n \|X_i\| \xrightarrow{p} \mathbb{E}\|X\| \leq \sqrt{\mathbb{E}\|X\|^2} = \sqrt{\text{tr}(\mathbb{E}(XX^T))} < \infty$, 这里 $\text{tr}(A)$ 表示矩阵 A 的迹. 结合以上结果与 Cauchy-Schwarz 不等式, 可得

$$n^{-1} T_{2(k)}^T \sum_{i=1}^n X_{(k)i} d^2(Y_i, s) \leq \|T_{2(k)}\| \left\| n^{-1} \sum_{i=1}^n X_{(k)i} d^2(Y_i, s) \right\| = o_p(1).$$

因此, 由三角不等式可得 $W_{n(k)}(s, x) - W_{(k)}(s, x) = o_p(1)$ 对任意给定的 $s \in \mathbb{S}$ 成立.

现在证明 (A.1). 注意到对任意 $s_1, s_2 \in \mathbb{S}$, 均有

$$\begin{aligned} |W_{n(k)}(s_1, x) - W_{n(k)}(s_2, x)| &= \frac{1}{n} \sum_{i=1}^n |\hat{u}_{(k)i}(x)| |d^2(s_1, Y_i) - d^2(s_2, Y_i)| \\ &= \frac{1}{n} \sum_{i=1}^n |\hat{u}_{(k)i}(x)| |d(s_1, Y_i) - d(s_2, Y_i)| |d(s_1, Y_i) + d(s_2, Y_i)| \\ &\leq 2Dn^{-1} d(s_1, s_2) \sum_{i=1}^n |\hat{u}_{(k)i}(x)|, \end{aligned}$$

与 $n^{-1} \sum_{i=1}^n |\hat{u}_{(k)i}(x)| = n^{-1} \sum_{i=1}^n |u_{(k)}(X_i, x) + T_{1(k)} + T_{2(k)}^T X_{(k)i}| = O_p(1)$. 因此, $|W_{n(k)}(s_1, x) - W_{n(k)}(s_2, x)| = O_p(d(s_1, s_2))$ 对全体 $s_1, s_2 \in \mathbb{S}$ 一致成立. 该结果表明 $\sup_{d(s_1, s_2) \leq \delta} |W_{n(k)}(s_1, x) - W_{n(k)}(s_2, x)| = O_p(\delta)$. 结合以上结果可知对任意给定的 $s \in \mathbb{S}$ 有 $d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) = o_p(1)$.

为了证明 $\sup_{x \in \mathcal{X}} d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) = o_p(1)$, 定义随机过程 $\{\xi_{n(k)}(x) = d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) : x \in \mathcal{X}\}$. 该随机过程满足 $\xi_{n(k)}(x) = o_p(1)$ 对任意 $x \in \mathcal{X}$ 成立. 接下来证明

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{\substack{\|x_1 - x_2\| < \delta \\ x_1, x_2 \in \mathcal{X}}} |\xi_{n(k)}(x_1) - \xi_{n(k)}(x_2)| > \varepsilon \right) = 0.$$

由于

$$\begin{aligned} |\xi_{n(k)}(x_1) - \xi_{n(k)}(x_2)| &= |d(\hat{m}_{\oplus(k)}(x_1), m_{\oplus(k)}(x_1)) - d(m_{\oplus(k)}(x_1), \hat{m}_{\oplus(k)}(x_2))| \\ &\quad + |d(m_{\oplus(k)}(x_1), \hat{m}_{\oplus(k)}(x_2)) - d(\hat{m}_{\oplus(k)}(x_2), m_{\oplus(k)}(x_2))| \\ &\leq |d(\hat{m}_{\oplus(k)}(x_1), \hat{m}_{\oplus(k)}(x_2))| + |d(m_{\oplus(k)}(x_1), m_{\oplus(k)}(x_2))|, \end{aligned}$$

因此只需要证明

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{\substack{\|x_1 - x_2\| < \delta \\ x_1, x_2 \in \mathcal{X}}} d(\hat{m}_{\oplus(k)}(x_1), \hat{m}_{\oplus(k)}(x_2)) > \varepsilon/2 \right) = 0, \quad (\text{A.2})$$

且 $m_{\oplus(k)}(\cdot)$ 在 $\mathcal{X}$ 上一致连续.

根据 $W_{(k)}(\cdot, \cdot)$ 定义可得

$$\sup_{s \in \mathbb{S}} |W_{(k)}(s, x_1) - W_{(k)}(s, x_2)| \leq \sup_{s \in \mathbb{S}} |\mathbb{E}[(u_{(k)}(X, x_1) - u_{(k)}(X, x_2))d^2(Y, s)]|$$

$$\begin{aligned} &\leq D^2 [\mathbb{E}((X_{(k)} - \mu_{(k)})^T \Sigma_{(k)}^{-1} (x_{(k)1} - x_{(k)2}))^2]^{1/2} \\ &\leq D^2 \sqrt{\|\Sigma_{(k)}^{-1}\|_2} \|x_1 - x_2\|. \end{aligned}$$

这表明随着 $\delta \rightarrow 0$ 有 $\sup_{s \in \mathbb{S}} \sup_{x_1, x_2 \in \mathcal{X}} |W_{(k)}(s, x_1) - W_{(k)}(s, x_2)| \rightarrow 0$. 对某个给定的 $x \in \mathcal{X}$, 若 $m_{\oplus(k)}(\cdot)$ 在 x 处不连续, 则存在 $\varepsilon > 0$ 使得对任意 $\delta > 0$ 可以找到一个 $\tilde{x}$ 满足 $0 < \|x - \tilde{x}\| \leq \delta$ 与 $d(m_{\oplus(k)}(x), m_{\oplus(k)}(\tilde{x})) > \varepsilon > 0$, 从而有 $\lim_{\delta \rightarrow 0} \inf_{\{\tilde{x}: \|x - \tilde{x}\| \leq \delta\}} d(m_{\oplus(k)}(x), m_{\oplus(k)}(\tilde{x})) > \varepsilon > 0$. 根据条件 3.1, 对任意 $\delta > 0$, 存在 $\tilde{x}$ 使得 $W_{(k)}(m_{\oplus(k)}(\tilde{x}), x) - W_{(k)}(m_{\oplus(k)}(x), x) > 0$, $W_{(k)}(m_{\oplus(k)}(x), \tilde{x}) - W_{(k)}(m_{\oplus(k)}(\tilde{x}), \tilde{x}) > 0$ 且满足 $\|x - \tilde{x}\| \leq \delta$. 这意味着

$$W_{(k)}(m_{\oplus(k)}(x), \tilde{x}) - W_{(k)}(m_{\oplus(k)}(x), x) + W_{(k)}(m_{\oplus(k)}(\tilde{x}), x) - W_{(k)}(m_{\oplus(k)}(\tilde{x}), \tilde{x}) > 0,$$

与当 $\delta \rightarrow 0$ 时 $\sup_{s \in \mathbb{S}} \sup_{\|x_1 - x_2\| < \delta} |W_{(k)}(s, x_1) - W_{(k)}(s, x_2)| \rightarrow 0$ 相矛盾. 因此, $m_{\oplus(k)}(\cdot)$ 在紧空间 $\mathcal{X}$ 上连续从而一致连续.

最后来证明 (A.2). 假定 (A.2) 不成立, 则存在 $c > 0$ 使得对所有的 $\delta > 0$, 有

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{\substack{\|x_1 - x_2\| < \delta \\ x_1, x_2 \in \mathcal{X}}} d(\hat{m}_{\oplus(k)}(x_1), \hat{m}_{\oplus(k)}(x_2)) > \varepsilon/2 \right) > c.$$

不失一般性, 假定

$$\mathbb{P} \left(\sup_{\substack{\|x_1 - x_2\| < \delta \\ x_1, x_2 \in \mathcal{X}}} d(\hat{m}_{\oplus(k)}(x_1), \hat{m}_{\oplus(k)}(x_2)) > \varepsilon/2 \right) > c/2$$

对充分大的 n 均成立, 否则可抽取一个子列. 对任何满足 $\|x_1 - x_2\| < \delta$ 的 x_1 和 x_2 , $\sup_{s \in \mathbb{S}} |W_{n(k)}(s, x_1) - W_{n(k)}(s, x_2)|$ 存在一致上界

$$\begin{aligned} \sup_{s \in \mathbb{S}} |W_{n(k)}(s, x_1) - W_{n(k)}(s, x_2)| &= \frac{1}{n} \left| \sum_{i=1}^n (\hat{u}_{(k)i}(x_1) - \hat{u}_{(k)i}(x_2)) d^2(Y_i, s) \right| \\ &\leq D^2 n^{-1} \sum_{i=1}^n |(X_{(k)i} - \bar{X}_{(k)})^T \hat{\Sigma}_{(k)}^{-1} (x_{(k)1} - x_{(k)2})| \\ &\leq D^2 n^{-1} \sum_{i=1}^n \|\hat{\Sigma}_{(k)}^{-1} (X_{(k)i} - \bar{X}_{(k)})\| \|x_1 - x_2\| \\ &\leq \delta D^2 \|\hat{\Sigma}_{(k)}^{-1}\|_2 n^{-1} \sum_{i=1}^n \|X_{(k)i} - \bar{X}_{(k)}\| \\ &= O_p(\delta), \end{aligned}$$

即

$$\sup_{s \in \mathbb{S}} \sup_{\|x_1 - x_2\| \leq \delta} |W_{n(k)}(s, x_1) - W_{n(k)}(s, x_2)| = O_p(\delta). \quad (\text{A.3})$$

定义如下事件:

$$\begin{aligned} E_1 &= \left\{ \sup_{\substack{\|x_1 - x_2\| < \delta \\ x_1, x_2 \in \mathcal{X}}} d(\hat{m}_{\oplus(k)}(x_1), \hat{m}_{\oplus(k)}(x_2)) > \varepsilon/2 \right\}, \\ E_2 &= \left\{ \inf_{x \in \mathcal{X}} \inf_{d(s, \hat{m}_{\oplus(k)}(x)) > \varepsilon/2} W_{n(k)}(s, x) - W_{n(k)}(\hat{m}_{\oplus(k)}(x), x) \geq \tau \right\}, \end{aligned}$$

$$E_3 = \{W_{n(k)}(\widehat{m}_{\oplus(k)}(x_2), x_1) - W_{n(k)}(\widehat{m}_{\oplus(k)}(x_2), x_2) \\ + W_{n(k)}(\widehat{m}_{\oplus(k)}(x_1), x_2) - W_{n(k)}(\widehat{m}_{\oplus(k)}(x_1), x_1) \geq 2\tau\},$$

则根据条件 3.1 可得 $P(E_1) > c/2$ 和 $P(E_2) \rightarrow 1$ 成立. 注意到

$$P(E_3) \geq P(E_1 \cap E_2 \cap E_3) = P(E_1 \cap E_2) \geq P(E_1) - P(E_2^c) > c/3 > 0$$

对充分大的 n 成立. 然而 (A.3) 保证了

$$W_{n(k)}(\widehat{m}_{\oplus(k)}(x_2), x_1) - W_{n(k)}(\widehat{m}_{\oplus(k)}(x_2), x_2) + W_{n(k)}(\widehat{m}_{\oplus(k)}(x_1), x_2) - W_{n(k)}(\widehat{m}_{\oplus(k)}(x_1), x_1) \\ \leq 2 \sup_{\|x_1 - x_2\| \leq \delta} |W_{n(k)}(s, x_1) - W_{n(k)}(s, x_2)| = O_p(\delta),$$

这是一个矛盾. 因此论断 (A.2) 成立. 定理 3.1 证毕.

附录 A.2 定理 3.2 的证明

首先推导逐点收敛速度. 对于给定的 $x \in \mathcal{X}$, 记 $U_{n(k)}(s, x) = W_{n(k)}(s, x) - W_{(k)}(s, x)$. 令 $r_n = n^{1/2}$, 对充分大的 M , 考察概率 $P(d(\widehat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) > 2^M r_n^{-1})$. 对空间 $\mathbb{S}$ 作如下分解. 定义

$$S_{(k),j} = \{s \in \mathbb{S} : 2^{j-1} < r_n d(s, m_{\oplus(k)}(x)) \leq 2^j\},$$

则可得 $P(2^{j-1} r_n^{-1} < d(\widehat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) \leq 2^j r_n^{-1}) = P(\widehat{m}_{\oplus(k)}(x) \in S_{(k),j})$. 因为 $\widehat{m}_{\oplus(k)}(x)$ 是 $W_{n(k)}(s, x)$ 的极小值点, 所以可得

$$P(\widehat{m}_{\oplus(k)}(x) \in S_{(k),j}) \leq P\left(\inf_{s \in S_{(k),j}} (W_{n(k)}(s, x) - W_{n(k)}(m_{\oplus(k)}(x), x)) \leq 0\right). \quad (\text{A.4})$$

根据 $S_{(k),j}$ 的构造, 对于所有的 $s \in S_{(k),j}$, 可得 $d^2(s, m_{\oplus(k)}(x)) \geq 2^{2j-2} r_n^{-2}$. 因此, 由条件 3.2 可得, 当 $2^j \leq r_n \eta$ 时有

$$\inf_{s \in S_{(k),j}} (W_{(k)}(s, x) - W_{(k)}(m_{\oplus(k)}(x), x)) \geq C 2^{2j-2} r_n^{-2}. \quad (\text{A.5})$$

注意到

$$\sup_{s \in S_{(k),j}} |U_{n(k)}(s, x) - U_{n(k)}(m_{\oplus(k)}(x), x)| \\ \geq \sup_{s \in S_{(k),j}} (U_{n(k)}(m_{\oplus(k)}(x), x) - U_{n(k)}(s, x)) \\ \geq \sup_{s \in S_{(k),j}} (W_{n(k)}(m_{\oplus(k)}(x), x) - W_{n(k)}(s, x)) + \inf_{s \in S_{(k),j}} (W_{(k)}(s, x) - W_{(k)}(m_{\oplus(k)}(x), x)), \quad (\text{A.6})$$

因此, 本证明的关键在于控制随机过程 $U_{n(k)}(s, x) - U_{n(k)}(m_{\oplus(k)}(x), x)$. 记 $D_{(k)i}(s, x) = d^2(Y_i, s) - d^2(Y_i, m_{\oplus(k)}(x))$. 根据三角不等式, 可得

$$|U_{n(k)}(s, x) - U_{n(k)}(m_{\oplus(k)}(x), x)| \leq \left| \frac{1}{n} \sum_{i=1}^n (\widehat{u}_{(k)i}(x) - u_{(k)}(X_i, x)) D_{(k)i}(s, x) \right| \\ + \left| \frac{1}{n} \sum_{i=1}^n (u_{(k)}(X_i, x) D_{(k)i}(s, x) - E(u_{(k)}(X_i, x) D_{(k)i}(s, x))) \right|. \quad (\text{A.7})$$

因此, 对于使 $d(s, m_{\oplus(k)}(x))$ 较小的 s , 需要给出以上两项的一致上界. 对于第一项, 由先前的分析可知

$$\sup_{d(s, m_{\oplus(k)}(x)) < \delta} \left| \frac{1}{n} \sum_{i=1}^n (\hat{u}_{(k)i}(x) - u_{(k)}(X_i, x)) D_{(k)i}(s, x) \right| \leq \frac{2D\delta}{n} \sum_{i=1}^n |T_{1(k)} + T_{2(k)}^T X_{(k)i}| = O_p(\delta n^{-1/2}).$$

进而对给定的 $R > 0$, 定义事件

$$A(R) = \left\{ \sup_{d(s, m_{\oplus(k)}(x)) < \delta} \left| \frac{1}{n} \sum_{i=1}^n (\hat{u}_{(k)i}(x) - u_{(k)}(X_i, x)) D_{(k)i}(s, x) \right| \leq R\delta n^{-1/2} \right\}.$$

可以看到, 通过取充分大的 R 和 n , $P(A^c(R))$ 可以任意小.

接下来推导第二项在 $B(\delta; m_{\oplus(k)}(x))$ 的一致上界. 定义 $\mathcal{X} \times \mathbb{S} \mapsto R$ 的函数 $f_s(z, y) = u_{(k)}(z, x) d^2(y, s)$, 并定义与之相对应的函数族 $\mathcal{M}_\delta = \{f_s - f_{m_{\oplus(k)}} : d(s, m_{\oplus(k)}) < \delta\}$. 不难看出, $F_\delta(z, y) = 2D\delta|u_{(k)}(z, x)|$ 是 $\mathcal{M}_\delta$ 的包络函数且满足 $E(F_\delta(X, Y)^2) = O(\delta^2)$. 文献 [28, 定理 2.14.2] 给出了关于第二项的上界

$$\begin{aligned} E \left(\sup_{d(s, m_{\oplus(k)}(x)) < \delta} \left| \frac{1}{n} \sum_{i=1}^n (u_{(k)}(X_i, x) D_{(k)i}(s, x) - E(u_{(k)}(X_i, x) D_{(k)i}(s, x))) \right| \right) \\ \leq C J_{[\cdot]}(1, \mathcal{M}_\delta, L_2(P)) \sqrt{E(F_\delta(X, Y)^2)} / \sqrt{n}. \end{aligned}$$

因此, $J_{[\cdot]}(1, \mathcal{M}_\delta, L_2(P))$ 需要被控制. 因为 $|f_{s_1}(z, y) - f_{s_2}(z, y)| \leq d(s_1, s_2) 2D|u_{(k)}(z, x)|$, 文献 [28, 定理 2.7.11] 保证了 $N_{[\cdot]}(2M\epsilon, \mathcal{M}_\delta, L_2(P)) \leq N(\epsilon, B(\delta; m_{\oplus(k)}(x)), d)$, 其中 $M = 2D\|u_{(k)}(z, x)\|_{P,2}$. 注意到 $\|F_\delta\|_{P,2} = M\delta$, 从而有

$$\begin{aligned} J_{[\cdot]}(1, \mathcal{M}_\delta, L_2(P)) &= \int_0^1 [1 + \log N_{[\cdot]}(\epsilon M\delta, \mathcal{M}_\delta, L_2(P))]^{1/2} d\epsilon \\ &\leq \int_0^1 [1 + \log N(\delta\epsilon/2, B(\delta; m_{\oplus(k)}(x)), d)]^{1/2} d\epsilon \\ &= 2 \int_0^{\frac{1}{2}} [1 + \log N(\delta\epsilon, B(\delta; m_{\oplus(k)}(x)), d)]^{1/2} d\epsilon \\ &\leq 2 \int_0^1 [1 + \log N(\delta\epsilon, B(\delta; m_{\oplus(k)}(x)), d)]^{1/2} d\epsilon. \end{aligned}$$

结合条件 3.2 可得

$$E \left(\sup_{d(s, m_{\oplus(k)}(x)) < \delta} \left| \frac{1}{n} \sum_{i=1}^n (u_{(k)}(X_i, x) D_{(k)i}(s, x) - E(u_{(k)}(X_i, x) D_{(k)i}(s, x))) \right| \right) = O(\delta n^{-1/2}) \quad (\text{A.8})$$

对较小的 $\delta \leq \delta_0$ 成立. 由 (A.7) 和 $A(R)$ 的定义可知, 对 $\delta \leq \delta_0$ 有

$$E \left(I_{A(R)} \sup_{d(s, m_{\oplus(k)}(x)) < \delta} |U_{n(k)}(s, x) - U_{n(k)}(m_{\oplus(k)}(x), x)| \right) \leq C\delta n^{-1/2}. \quad (\text{A.9})$$

此外, 对任何正整数 L , $P(r_n d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) > 2^L)$ 有如下上界:

$$\begin{aligned} P(r_n d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) > 2^L) \\ \leq P(d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) \geq \eta) + P(A^c(R)) \end{aligned}$$

$$+ \sum_{\substack{j \geq L \\ 2^j \leq r_n \eta}} \mathbb{P} \left(\left(I_{A(R)} \sup_{s \in S_{(k),j}} |U_{n(k)}(s, x) - U_{n(k)}(m_{\oplus(k)}(x), x)| \right) \geq C \frac{2^{2(j-1)}}{r_n^2} \right). \quad (\text{A.10})$$

具体而言, (A.4)–(A.6) 蕴涵

$$\begin{aligned} & \mathbb{P}(\hat{m}_{\oplus(k)}(x) \in S_{(k),j}) \\ & \leq \mathbb{P} \left(\inf_{s \in S_{(k),j}} (W_{n(k)}(s, x) - W_{n(k)}(m_{\oplus(k)}(x), x)) \leq 0 \right) \\ & \leq \mathbb{P} \left(\inf_{s \in S_{(k),j}} (W_{n(k)}(s, x) - W_{n(k)}(m_{\oplus(k)}(x), x)) \leq \sup_{s \in S_{(k),j}} |U_{n(k)}(s, x) - U_{n(k)}(m_{\oplus(k)}(x), x)| \right) \\ & \leq \mathbb{P} \left(\sup_{s \in S_{(k),j}} |U_{n(k)}(s, x) - U_{n(k)}(m_{\oplus(k)}(x), x)| \geq C \frac{2^{2(j-1)}}{r_n^2} \right). \end{aligned}$$

定理 3.1 保证了 $\mathbb{P}(d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) \geq \eta)$ 收敛到 0. 同时, 选取充分大的 R 和 n 可以使 $\mathbb{P}(A^c(R))$ 任意小. 取 $\eta \leq \delta_0$ 且满足条件 3.2 的要求, 并取 $\delta = 2^j r_n^{-1}$. 根据条件 (A.9) 和 Markov 不等式可知, (A.10) 右边第三项的上界为

$$C \sum_{\substack{j \geq L \\ 2^j \leq r_n \eta}} \frac{r_n}{\sqrt{n}} 2^{-j} \leq C \sum_{j \geq L} 2^{-j}.$$

对于任意的 L , 该序列均收敛, 故选取充分大的 L 可以使其任意小. 综合以上结果可得逐点收敛速度

$$d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) = O_p(n^{-1/2}).$$

接下来给出 $\hat{m}_{\oplus(k)}(x)$ 在 $x \in \mathcal{X}$ 上的一致收敛速度. $\mathcal{X}$ 的有界性保证了 $T_{1(k)}$ 和 $\|T_{2(k)}\|$ 的阶数是 $O_p(n^{-1/2})$. 从而有

$$\sup_{x \in \mathcal{X}} \sup_{d(s, m_{\oplus(k)}(x)) < \delta} \left| \frac{1}{n} \sum_{i=1}^n (\hat{u}_{(k)i}(x) - u_{(k)}(X_i, x)) D_{(k)i}(s, x) \right| = O_p(\delta n^{-1/2}). \quad (\text{A.11})$$

类似地, 对 $R > 0$ 定义事件

$$B(R) = \left\{ \sup_{x \in \mathcal{X}} \sup_{d(s, m_{\oplus(k)}(x)) < \delta} \left| \frac{1}{n} \sum_{i=1}^n (\hat{u}_{(k)i}(x) - u_{(k)}(X_i, x)) D_{(k)i}(s, x) \right| \leq R \delta n^{-1/2} \right\}.$$

根据 Cauchy-Schwarz 和 l_2 范数不超过 l_1 这一事实, 我们得到 (A.7) 右边第二项的上界为

$$\begin{aligned} & \|\Sigma_{(k)}^{-1}(x_{(k)} - \mu_{(k)})\| \sum_{j=1}^{p_k} \left| \frac{1}{n} \sum_{i=1}^n (X_{(k)ij} - \mu_{(k)j}) D_{(k)i}(s, x) - \mathbb{E}((X_{(k)ij} - \mu_{(k)j}) D_{(k)i}(s, x)) \right| \\ & + \left| \frac{1}{n} \sum_{i=1}^n [D_{(k)i}(s, x) - \mathbb{E}(D_{(k)i}(s, x))] \right|. \end{aligned}$$

对 $\delta > 0$, 定义 $T_\delta = \{(x, s) : x \in \mathcal{X}, d(s, m_{\oplus(k)}(x)) < \delta\}$. 构造函数

$$f_{x,s}^j(z, y) = \begin{cases} d^2(m_{\oplus(k)}(x), y) - d^2(s, y), & j = 0, \\ (z_{(k)j} - \mu_{(k)j})[d^2(m_{\oplus(k)}(x), y) - d^2(s, y)], & j = 1, \dots, p_k, \end{cases}$$

并定义函数族 $\mathcal{F}_\delta^j = \{f_{x,s}^j(z, y) : (x, s) \in T_\delta\}$. 接下来证明条件 3.3 蕴涵

$$d(m_{\oplus(k)}(x_1), m_{\oplus(k)}(x_2)) \leq C \|x_1 - x_2\|^{1/2}.$$

事实上, 由条件 3.3 可知

$$\begin{aligned} W_{(k)}(m_{\oplus(k)}(x_2), x_1) - W_{(k)}(m_{\oplus(k)}(x_1), x_1) &\geq Cd^2(m_{\oplus(k)}(x_1), m_{\oplus(k)}(x_2)), \\ W_{(k)}(m_{\oplus(k)}(x_1), x_2) - W_{(k)}(m_{\oplus(k)}(x_2), x_2) &\geq Cd^2(m_{\oplus(k)}(x_1), m_{\oplus(k)}(x_2)). \end{aligned}$$

因此,

$$\begin{aligned} W_{(k)}(m_{\oplus(k)}(x_2), x_1) - W_{(k)}(m_{\oplus(k)}(x_2), x_2) + W_{(k)}(m_{\oplus(k)}(x_1), x_2) - W_{(k)}(m_{\oplus(k)}(x_1), x_1) \\ \geq 2Cd^2(m_{\oplus(k)}(x_1), m_{\oplus(k)}(x_2)). \end{aligned}$$

此外, 注意到

$$\begin{aligned} W_{(k)}(m_{\oplus(k)}(x_2), x_1) - W_{(k)}(m_{\oplus(k)}(x_2), x_2) &= E[(u_{(k)}(X, x_1) - u_{(k)}(X, x_2))d^2(Y, m_{\oplus(k)}(x_2))] \\ &\leq D^2[(x_{(k)1} - x_{(k)2})^T \Sigma_{(k)}^{-1} (x_{(k)1} - x_{(k)2})]^2]^{1/2} \\ &\leq D^2 \sqrt{\|\Sigma_{(k)}^{-1}\|_2} \|x_1 - x_2\|. \end{aligned}$$

以上结果表明 $d(m_{\oplus(k)}(x_1), m_{\oplus(k)}(x_2)) \leq C\|x_1 - x_2\|^{1/2}$.

定义 $d(T_\delta) = (\|x_1 - x_2\|^{1/2} + d(s_1, s_2))$. 对 $C > 1$ 和 $j = 0$, 令 $F_j = 2CD$; 对 $j = 1, \dots, p_k$, 令 $F_j = 2CD|z_{(k)j} - \mu_{(k)j}|$. 直接计算表明 $\mathcal{F}_\delta^j$ 关于指标集 $(x, s) \in T_\delta$ 是 Lipschitz 的, 更确切地, 有

$$|f_{x_1, s_1}^j(z, y) - f_{x_2, s_2}^j(z, y)| \leq F_j(z, y)d(T_\delta). \quad (\text{A.12})$$

进而, 对任意 $\epsilon > 0$, $N_{[\cdot]}(2M_j\epsilon, \mathcal{F}_\delta^j, L_2(P))$ 以 $N(\epsilon, T_\delta, d(T_\delta))$ 为上界, 其中 $M_j = \|F_j\|_{P,2}$. 事实上, 令 $\{(x_i, s_i), i = 1, \dots, N\}$ 表示 T_δ 在 $d(T_\delta)$ 下的 ϵ 覆盖. 对任意函数 $f_{x,s}^j \in \mathcal{F}_\delta^j$, 存在指标 $i \in \{1, \dots, N\}$ 使得 $d(T_\delta)((x, s), (x_i, s_i)) \leq \epsilon$. 根据 (A.12), 可得 $[f_{x_i, s_i}^j - \epsilon F_j, f_{x_i, s_i}^j + \epsilon F_j]$, $i = 1, \dots, N$ 覆盖 $\mathcal{F}_\delta^j$. 此外, 不难看到 $N(\epsilon, T_\delta, d(T_\delta)) \leq N(\epsilon/2, \mathcal{X}, \|\cdot\|^{1/2})(\sup_{x \in \mathcal{X}} N(\epsilon/2, B(\delta; m_{\oplus(k)}(x)), d))$ 且 $N(\epsilon^{1/2}, \mathcal{X}, \|\cdot\|^{1/2}) \leq N(\epsilon, \mathcal{X}, \|\cdot\|)$. 因为 $F_\delta^j(z, y)$ 满足 $F_\delta^j(z, y) = 2CD\delta$, $j = 0$ 和 $F_\delta^j = 2CD\delta|z_{(k)j} - \mu_{(k)j}|$, $j = 1, \dots, p_k$, 故其是 $\mathcal{F}_\delta^j$ 的包络函数, 所以 $\|F_\delta^j\|_{P,2} = \delta M_j$. 由此可得

$$\begin{aligned} J_{[\cdot]}(1, \mathcal{F}_\delta^j, L_2(P)) &= \int_0^1 [1 + \log N_{[\cdot]}(\epsilon M_j \delta, \mathcal{F}_\delta^j, L_2(P))]^{1/2} d\epsilon \\ &\leq 4 \int_0^1 [1 + \log N(\delta\epsilon, \mathcal{X}, \|\cdot\|^{1/2}) + \sup_{x \in \mathcal{X}} \log N(\delta\epsilon, B(\delta; m_{\oplus(k)}(x)), d)]^{1/2} d\epsilon \\ &\leq 4 \int_0^1 [\log N((\delta\epsilon)^2, \mathcal{X}, \|\cdot\|)]^{1/2} d\epsilon + 4 \int_0^1 \sup_{x \in \mathcal{X}} [1 + \log N(\delta\epsilon, B(\delta; m_{\oplus(k)}(x)), d)]^{1/2} d\epsilon. \end{aligned}$$

根据 $\mathbb{R}^p$ 中的任意有界集的 ϵ 覆盖数的阶数不超过 $(1/\epsilon)^p$ 这一事实可得

$$\int_0^1 [\log N((\delta\epsilon)^2, \mathcal{X}, \|\cdot\|)]^{1/2} d\epsilon \leq C + \sqrt{2p \log(1/\delta)} = O(\sqrt{\log(1/\delta)}).$$

结合以上结果和条件 3.3 可知, 随着 $\delta \rightarrow 0^+$, 有 $J_{[\cdot]}(1, \mathcal{F}_\delta^j, L_2(P)) = O(\sqrt{\log(1/\delta)})$. 因为 $\|F_\delta^j(X)\|_{P,2} = O(\delta)$, 文献 [28, 定理 2.14.2] 给出了上界

$$\begin{aligned} E \left(\sup_{x \in \mathcal{X}} \sup_{d(s, m_{\oplus(k)}(x)) < \delta} \left| \frac{1}{n} \sum_{i=1}^n (u_{(k)}(X_i, x) D_{(k)i}(s, x) - E(u_{(k)}(X_i, x) D_{(k)i}(s, x))) \right| \right) \\ = O \left(\frac{\sqrt{\log(1/\delta)} \delta}{\sqrt{n}} \right). \end{aligned} \quad (\text{A.13})$$

结合 (A.7), (A.11) 和 (A.13), 可得

$$\mathbb{E}\left(I_{B(R)} \sup_{x \in \mathcal{X}} \sup_{d(s, m_{\oplus(k)}(x)) < \delta} |U_{n(k)}(s, x) - U_{n(k)}(m_{\oplus(k)}(x), x)|\right) = O\left(\frac{\sqrt{\log(1/\delta)}\delta}{\sqrt{n}}\right).$$

取 $\tilde{r}_n = (n/\log n)^{1/2}$ 并定义 $S_{(k),j}(x) = \{s \in \mathbb{S} : 2^{j-1} < r_n d(s, m_{\oplus(k)}(x)) \leq 2^j\}$. 根据 $S_{(k),j}(x)$ 定义, 事件 $\{2^{j-1} < r_n \sup_{x \in \mathcal{X}} d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) \leq 2^j\}$ 包含于事件 $\{\inf_{x \in \mathcal{X}} \inf_{s \in S_{(k),j}(x)} (W_{n(k)}(s, x) - W_{n(k)}(m_{\oplus(k)}(x), x)) \leq 0\}$. 应用类似的证明方法可以得到相似的不等式

$$\begin{aligned} & \mathbb{P}(\tilde{r}_n \sup_{x \in \mathcal{X}} d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) > 2^L) \\ & \leq \mathbb{P}\left(\sup_{x \in \mathcal{X}} d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) \geq \tau\right) + \mathbb{P}(B^c(R)) \\ & \quad + \sum_{\substack{j \geq L \\ 2^j \leq \tilde{r}_n \tau}} \mathbb{P}\left(\left(I_{B(R)} \sup_{x \in \mathcal{X}} \sup_{s \in S_{(k),j}(x)} |U_{n(k)}(s, x) - U_{n(k)}(m_{\oplus(k)}(x), x)|\right) \geq C \frac{2^{2(j-1)}}{\tilde{r}_n^2}\right), \end{aligned}$$

其中 τ 已在条件 3.3 中定义. 上式右边的第三项的上界为

$$C \sum_{\substack{j \geq L \\ 2^j \leq \tilde{r}_n \tau}} \frac{\tilde{r}_n}{2^j} \frac{\sqrt{\log \tilde{r}_n}}{\sqrt{n}} \leq C \sum_{j \geq L} 2^{-j}.$$

因此, 结论 $\max_{k \in \{1, \dots, K\}} \sup_{x \in \mathcal{X}} d(\hat{m}_{\oplus(k)}(x), m_{\oplus(k)}(x)) = O_p((n/\log n)^{-1/2})$ 成立. 定理 3.2 证毕.

附录 A.3 定理 3.3 的证明

首先注意到条件 3.6 蕴涵了 $R_n > 0$ 对充分大的 n 成立. 为证明定理 3.3, 只需要证明以下结论成立:

$$\mathbb{P}\left(\inf_{w \in \mathcal{W}} L_n(w) > 0\right) \rightarrow 1, \quad (\text{A.14})$$

$$\sup_{w \in \mathcal{W}} |(\text{CV}(w) - L_n(w))/R_n(w)| \xrightarrow{P} 0, \quad (\text{A.15})$$

$$\sup_{w \in \mathcal{W}} |L_n(w)/R_n(w) - 1| \xrightarrow{P} 0. \quad (\text{A.16})$$

首先证明 (A.14) 成立. 回顾定义 $m_{\oplus}(x) = \arg \min_{s \in \mathbb{S}} \mathbb{E}(d^2(Y, s) \mid X = x)$. 令 $d_n = \{(x_1, y_1), \dots, (x_n, y_n)\}$ 表示 $\mathcal{D}_n$ 的实现值. 度量空间 $\mathbb{S}$ 的有界性保证了 $\mathbb{E}[d^2(Y, \hat{m}_{\oplus}(X; w))] \leq D^2 < \infty$. 记 $f_w((X, Y), D_n) = d^2(Y, \hat{m}_{\oplus}(X; w))$. 由 (X, Y) 和 $\mathcal{D}_n$ 相互独立可得

$$\mathbb{E}[d^2(Y, \hat{m}_{\oplus}(X; w)) \mid \mathcal{D}_n = d_n] = \mathbb{E}[f_w((X, Y), d_n)]$$

对任意 $w \in \mathcal{W}$ 和 d_n 成立, 这里, 期望 $\mathbb{E}[f_w((X, Y), d_n)]$ 是相对于 (X, Y) 的随机性而取的. 根据 $m_{\oplus}(x)$ 的定义, 可得

$$\begin{aligned} \mathbb{E}[d^2(Y, \hat{m}_{\oplus}(X; w)) \mid \mathcal{D}_n = d_n] &= \mathbb{E}\{\mathbb{E}[f_w((X, Y), d_n) \mid X]\} \\ &\geq \mathbb{E}\{\mathbb{E}[d^2(Y, m_{\oplus}(X)) \mid X]\} \\ &= \mathbb{E}[d^2(Y, m_{\oplus}(X))]. \end{aligned}$$

因此, 可得结论 $\inf_{w \in \mathcal{W}} (L_n(w)) \geq 0$ 与 $\inf_{w \in \mathcal{W}} (R_n(w)) \geq 0$. 结合条件 3.6 可知, 对充分大的 n 有 $\inf_{w \in \mathcal{W}} R_n(w) > 0$. 从而, 根据

$$\inf_{w \in \mathcal{W}} L_n(w) = \inf_{w \in \mathcal{W}} [R_n(w)(1 + (L_n(w) - R_n(w))/R_n(w))]$$

与 (A.16) (将于随后证明), 可以得到 $P(\inf_{w \in \mathcal{W}} L_n(w) > 0) \rightarrow 1$.

现在证明 (A.15) 成立. 对于任意给定的 $w \in \mathcal{W}$, 定义 $m_{\oplus}(x; w) = \sum_{k=1}^K w_k m_{\oplus(k)}(x)$. 注意到 $E[d^2(Y, m_{\oplus}(X))]$ 不依赖于 w , $n^{-1} \sum_{i=1}^n d^2(Y_i, \tilde{m}_{\oplus i}(w))$ 的极小值点也是 $n^{-1} \sum_{i=1}^n d^2(Y_i, \tilde{m}_{\oplus i}(w)) - E[d^2(Y, m_{\oplus}(X))]$ 在 $\mathcal{W}$ 上的极小值点. 因此, 为了证明方便, 记 $CV(w) = n^{-1} \sum_{i=1}^n d^2(Y_i, \tilde{m}_{\oplus i}(w)) - E[d^2(Y, m_{\oplus}(X))]$. 由三角不等式可得

$$\begin{aligned} |CV(w) - L_n(W)| &\leq \left| n^{-1} \sum_{i=1}^n d^2(Y_i, \tilde{m}_{\oplus}(X_i; w)) - n^{-1} \sum_{i=1}^n d^2(Y_i, m_{\oplus}(X_i; w)) \right| \\ &\quad + \left| n^{-1} \sum_{i=1}^n d^2(Y_i, m_{\oplus}(X_i; w)) - E[d^2(Y, \hat{m}_{\oplus}(X; w)) \mid \mathcal{D}_n] \right| \\ &= T_1(w) + T_2(w). \end{aligned}$$

因此只需要证明 $\sup_{w \in \mathcal{W}} T_1(w)/\xi_n \xrightarrow{P} 0$ 与 $\sup_{w \in \mathcal{W}} T_2(w)/\xi_n \xrightarrow{P} 0$. 首先推导 $T_1(w)$ 的上界. 直接计算可得

$$\begin{aligned} T_1(w) &\leq n^{-1} \sum_{i=1}^n |d^2(Y_i, \tilde{m}_{\oplus}(X_i; w)) - d^2(Y_i, m_{\oplus}(X_i; w))| \\ &= n^{-1} \sum_{i=1}^n |d(Y_i, \tilde{m}_{\oplus}(X_i; w)) - d(Y_i, m_{\oplus}(X_i; w))| |d(Y_i, \tilde{m}_{\oplus}(X_i; w)) + d(Y_i, m_{\oplus}(X_i; w))| \\ &\leq 2Dn^{-1} \sum_{i=1}^n d(\tilde{m}_{\oplus}(X_i; w), m_{\oplus}(X_i; w)), \end{aligned}$$

这里的最后一个不等式可由度量 d 的三角不等式和 $\mathbb{S}$ 的有界性推导出. 条件 3.4 保证了

$$\sup_{w \in \mathcal{W}} d(\tilde{m}_{\oplus}(X_i; w), m_{\oplus}(X_i; w)) \leq C \sum_{k=1}^K d(\tilde{m}_{\oplus(k)}^{-i}(X_i), m_{\oplus(k)}(X_i)).$$

故只需要证明 $\max_{k \in \{1, \dots, K\}} n^{-1} \sum_{i=1}^n d(\tilde{m}_{\oplus(k)}^{-i}(X_i), m_{\oplus(k)}(X_i))/\xi_n \xrightarrow{P} 0$. 考虑到 (X_i, Y_i) , $i = 1, \dots, n$ 独立同分布, 根据对称性, 有

$$E\left[n^{-1} \sum_{i=1}^n d(\tilde{m}_{\oplus(k)}^{-i}(X_i), m_{\oplus(k)}(X_i))\right] = E[d(\tilde{m}_{\oplus(k)}^{-1}(X_1), m_{\oplus(k)}(X_1))].$$

此外, 由定理 3.2 和条件 3.6 可得

$$\begin{aligned} \max_{k=1, \dots, K} d(\tilde{m}_{\oplus(k)}^{-1}(X_1), m_{\oplus(k)}(X_1)) &\leq \max_{k=1, \dots, K} \sup_{x \in \mathcal{X}} d(\tilde{m}_{\oplus(k)}^{-1}(x), m_{\oplus(k)}(x)) \\ &= O_p\left(\left(\frac{n-1}{\log(n-1)}\right)^{-1/2}\right) = O_p\left(\left(\frac{n}{\log n}\right)^{-1/2}\right) \\ &= o_p(\xi_n). \end{aligned}$$

进而, 条件 3.5 保证了

$$\mathbb{E}[n^{-1} \sum_{i=1}^n d(\tilde{m}_{\oplus(k)}^{-i}(X_i), m_{\oplus(k)}(X_i))/\xi_n] = o(1),$$

从而可得 $\sup_{w \in \mathcal{W}} T_1(w)/\xi_n \xrightarrow{P} 0$.

接下来证明 $\sup_{w \in \mathcal{W}} T_2(w)/\xi_n \xrightarrow{P} 0$. 由于

$$\begin{aligned} T_2(w) &\leq \left| n^{-1} \sum_{i=1}^n d^2(Y_i, m_{\oplus}(X_i; w)) - \mathbb{E} d^2(Y, m_{\oplus}(X; w)) \right| + |\mathbb{E} d^2(Y, m_{\oplus}(X; w)) - \mathbb{E}[d^2(Y, \hat{m}_{\oplus}(X; w)) \mid \mathcal{D}_n]| \\ &= T_{21}(w) + T_{22}(w), \end{aligned}$$

因此需要给出 $T_{21}(w)$ 和 $T_{22}(w)$ 在 $\mathcal{W}$ 上的一致上界. 后续将 $\mathcal{W}$ 看作度量空间 $\mathbb{R}^K$ 的子空间, 因此 $\mathcal{W}$ 是紧空间. 定义函数 $f_w(x, y) = d^2(y, m_{\oplus}(x; w))$ 并定义与其相关的函数族 $\mathcal{F}_{\mathcal{W}} = \{f_w : w \in \mathcal{W}\}$. 不难看出 $\mathcal{F}_{\mathcal{W}}$ 的包络函数可取为 $F = D^2$. 条件 3.4 保证了

$$\sup_{\substack{x \in \mathcal{X} \\ y \in \mathbb{S}}} |f_{w_1}(x, y) - f_{w_2}(x, y)| \leq 2Dd(m_{\oplus}(X_i; w_1), m_{\oplus}(X_i; w_2)) \leq 2CD\|w_1 - w_2\|^\alpha.$$

应用与定理 3.2 的证明中类似的手段可以得出 $N_{[\cdot]}(4C\epsilon^\alpha, \mathcal{F}_{\mathcal{W}}, L_2(P)) \leq N(\epsilon, \mathcal{W}, \|\cdot\|)$ 对某个 $C > 0$ 成立. 因此,

$$\begin{aligned} J_{[\cdot]}(1, \mathcal{F}_{\mathcal{W}}, L_2(P)) &= \int_0^1 [1 + \log N_{[\cdot]}(\epsilon D^2, \mathcal{F}_{\mathcal{W}}, L_2(P))]^{1/2} d\epsilon \\ &\leq \int_0^1 [1 + \log N((\epsilon D^2/(4C))^{1/\alpha}, \mathcal{W}, \|\cdot\|)]^{1/2} d\epsilon \\ &\leq \int_0^1 [1 + K\alpha^{-1} \log(4C/(D^2\epsilon))]^{1/2} d\epsilon \\ &= O(1). \end{aligned}$$

进而根据文献 [28, 定理 2.14.2], 有

$$\mathbb{E}\left(\sup_{w \in \mathcal{W}} T_{21}(w)\right) = \mathbb{E}\left(\sup_{f \in \mathcal{F}_{\mathcal{W}}} \left|\frac{1}{n} \sum_{i=1}^n (f(X_i, Y_i) - \mathbb{E}f(X, Y))\right|\right) = O(n^{-1/2}).$$

结合上式与条件 3.6, 可知 $\sup_{w \in \mathcal{W}} T_{21}(w) = o_p(\xi_n)$.

定义 $\mathcal{D}_\infty = \{(X_i, Y_i), i = 1, 2, \dots\}$. 因为对任意 $n \in N^+$, $\{(X_i, Y_i), i = n+1, \dots\}$ 和 $\{\mathcal{D}_n, (X, Y)\}$ 相互独立, 故可以得到

$$\mathbb{E}[d^2(Y, \hat{m}_{\oplus}(X; w)) \mid \mathcal{D}_n] = \mathbb{E}[d^2(Y, \hat{m}_{\oplus}(X; w)) \mid \mathcal{D}_\infty].$$

三角不等式和 $\mathbb{S}$ 的有界性保证了

$$|d^2(Y, \hat{m}_{\oplus}(X; w)) - d^2(Y, m_{\oplus}(X; w))| \leq 2Dd(\hat{m}_{\oplus}(X; w), m_{\oplus}(X; w)).$$

此外, 由条件 3.4, 3.6 和定理 3.2 可推得

$$\sup_{w \in \mathcal{W}} d(\hat{m}_{\oplus}(X; w), m_{\oplus}(X; w)) \leq CK \max_{1 \leq k \leq K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X)) = o_p(\xi_n).$$

至此已证明 $\sup_{w \in \mathcal{W}} |d^2(Y, \hat{m}_\oplus(X; w)) - d^2(Y, m_\oplus(X; w))|/\xi_n = o_p(1)$. 结合上式和条件 3.5 可得

$$\mathbb{E} \left[\sup_{w \in \mathcal{W}} |d^2(Y, \hat{m}_\oplus(X; w)) - d^2(Y, m_\oplus(X; w))|/\xi_n \right] \rightarrow 0. \quad (\text{A.17})$$

根据条件期望的控制收敛定理, 有

$$\begin{aligned} & \mathbb{E} \left[\sup_{w \in \mathcal{W}} |d^2(Y, \hat{m}_\oplus(X; w)) - d^2(Y, m_\oplus(X; w))|/\xi_n \mid \mathcal{D}_n \right] \\ &= \mathbb{E} \left[\sup_{w \in \mathcal{W}} |d^2(Y, \hat{m}_\oplus(X; w)) - d^2(Y, m_\oplus(X; w))|/\xi_n \mid \mathcal{D}_\infty \right] \xrightarrow{P} 0. \end{aligned}$$

此外, (X, Y) 和 $\mathcal{D}_\infty$ 的独立性保证了

$$\mathbb{E}[d^2(Y, m_\oplus(X; w)) \mid \mathcal{D}_\infty] = \mathbb{E}d^2(Y, m_\oplus(X; w)) = \mathbb{E}[d^2(Y, m_\oplus(X; w)) \mid \mathcal{D}_n].$$

同时, 不难得到

$$\begin{aligned} & \sup_{w \in \mathcal{W}} |\mathbb{E}d^2(Y, m_\oplus(X; w)) - \mathbb{E}[d^2(Y, \hat{m}_\oplus(X; w)) \mid \mathcal{D}_n]|/\xi_n \\ & \leq \mathbb{E} \left[\sup_{w \in \mathcal{W}} |d^2(Y, \hat{m}_\oplus(X; w)) - d^2(Y, m_\oplus(X; w))|/\xi_n \mid \mathcal{D}_n \right]. \end{aligned}$$

综合以上结果可知 $\sup_{w \in \mathcal{W}} T_{22}(w)/\xi_n \xrightarrow{P} 0$.

最后来证明 (A.16) 成立. 为此将证明 $\sup_{w \in \mathcal{W}} |L_n(w) - R_n(w)| = o_p(\xi_n)$. 根据三角不等式可得

$$\begin{aligned} |L_n(w) - R_n(w)| &= |\mathbb{E}[d^2(Y, \hat{m}_\oplus(X; w)) \mid \mathcal{D}_n] - \mathbb{E}[d^2(Y, \hat{m}_\oplus(X; w))]| \\ &\leq |\mathbb{E}[d^2(Y, \hat{m}_\oplus(X; w)) \mid \mathcal{D}_n] - \mathbb{E}[d^2(Y, m_\oplus(X; w))]| \\ &\quad + |\mathbb{E}[d^2(Y, m_\oplus(X; w))] - \mathbb{E}[d^2(Y, \hat{m}_\oplus(X; w))]|. \end{aligned}$$

对于上述不等式右边第一项, 已经证得

$$\sup_{w \in \mathcal{W}} T_{22}(w) = \sup_{w \in \mathcal{W}} |\mathbb{E}d^2(Y, m_\oplus(X; w)) - \mathbb{E}[d^2(Y, \hat{m}_\oplus(X; w)) \mid \mathcal{D}_n]| = o_p(\xi_n).$$

对于上述不等式右边第二项, 利用 (A.17), 可得

$$\begin{aligned} & \sup_{w \in \mathcal{W}} |\mathbb{E}d^2(Y, m_\oplus(X; w)) - \mathbb{E}[d^2(Y, \hat{m}_\oplus(X; w))]|/\xi_n \\ & \leq \mathbb{E} \left[\sup_{w \in \mathcal{W}} |d^2(Y, \hat{m}_\oplus(X; w)) - d^2(Y, m_\oplus(X; w))|/\xi_n \right] \\ & = o(1). \end{aligned}$$

定理 3.3 证毕.

附录 A.4 定理 3.4 的证明

取 $a_n = n^{-\frac{1-2\gamma}{2\beta}}$ 并定义 $\mathcal{W}(Ca_n; w^0) = \{w \in \mathcal{W} : \|w - w^0\| \leq Ca_n\}$ 作为可行域. 首先证明存在一个局部极小值点 $\hat{w}$, 即存在 $\hat{w} \in \mathcal{W}(Ca_n; w^0)$ 使得 $\text{CV}(\hat{w}) = \min_{w \in \mathcal{W}(Ca_n; w^0)} \text{CV}(w)$. 因为 $\mathcal{W}(Ca_n; w^0)$ 是紧集, 所以只需要证明 $\text{CV}(\cdot)$ 是连续函数. 应用条件 3.4 和三角不等式可得

$$\text{CV}(w_1) - \text{CV}(w_2) \leq 2Dn^{-1} \sum_{i=1}^n d^2 \left(\sum_{k=1}^K w_{1k} \tilde{m}_{\oplus(k)}^{-i}(X_i), \sum_{k=1}^K w_{2k} \tilde{m}_{\oplus(k)}^{-i}(X_i) \right)$$

$$\begin{aligned} &\leq 2D \sup_{\substack{s_k \in \mathbb{S} \\ k=1, \dots, K}} d\left(\sum_{k=1}^K w_{1k} s_k, \sum_{k=1}^K w_{2k} s_k\right) \\ &\leq C \|w_1 - w_2\|^\alpha. \end{aligned}$$

进而根据 $\lim_{\delta \rightarrow 0^+} \sup_{\|w_1 - w_2\| < \delta} |\text{CV}(w_1) - \text{CV}(w_2)| = 0$, 可得 $\text{CV}(\cdot)$ 是连续函数. 为了得到 $\hat{w}$ 的收敛速度, 只需要证明

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\inf_{\substack{w \in \mathcal{W}(Ca_n; w^0) \\ \|w - w^0\| = Ca_n}} \text{CV}(w) > \text{CV}(w^0) = 1\right).$$

事实上, 上式表明存在一个局部极小值点 $\hat{w}$ 满足 $\hat{w} \in \mathcal{W}$ 且 $\|\hat{w} - w^0\| < Ca_n$ 的概率趋于 1.

应用与 (A.15) 的证明中类似的方法, 根据条件 3.1, 3.3, 3.4 和 3.8, 可得

$$\sup_{w \in \mathcal{W}(Ca_n; w^0)} |\text{CV}(w) - R_n(w)| = o_p(n^{-1/2+\gamma}).$$

进而, 条件 3.7 保证对于充分大的 n ,

$$\begin{aligned} \text{CV}(w) &\geq R_n(w) + o_p(n^{-1/2+\gamma}) \geq R_n(w^0) + \|w - w^0\|^\beta + o_p(n^{-1/2+\gamma}) \\ &\geq \text{CV}(w^0) + o_p(n^{-1/2+\gamma}) + \|w - w^0\|^\beta \end{aligned} \quad (\text{A.18})$$

对 $w \in \mathcal{W}(Ca_n; w^0)$ 一致成立. 对任意的 $w \in \mathcal{W}(Ca_n; w^0)$ 且 $\|w - w^0\| = Ca_n$, 有 $\|w - w^0\|^\beta = (Ca_n)^\beta$. 这意味着 $o_p(n^{-1/2+\gamma}) + \|w - w^0\|^\beta = (Ca_n)^\beta(1 + o_p(1))$. 结合 (A.18) 可以推得

$$\inf_{\substack{w \in \mathcal{W}(Ca_n; w^0) \\ \|w - w^0\| = Ca_n}} \text{CV}(w) \geq \text{CV}(w^0) + n^{-1/2+\gamma}(1 + o_p(1)),$$

定理 3.4 证毕.

附录 A.5 定理 3.5 的证明

本证明采用反证法. 假定定理 3.5 的结论不真, 即存在 $\epsilon, \delta > 0$ 使得 $\limsup_{n \rightarrow \infty} \mathbb{P}(\hat{w}_{\mathcal{C}} \leq 1 - \epsilon) > \delta$, 下文将推导出一个矛盾.

采用与 (A.15) 和 (A.16) 类似的证明过程, 假定条件 3.1, 3.3, 3.4 和 3.10 成立, 可以证明

$$\sup_{w \in \mathcal{W}} |\text{CV}(w) - L_n(w)| = o_p(\tilde{\xi}_n(\epsilon)) \quad \text{且} \quad \sup_{w \in \mathcal{W}} |L_n(w) - R_n(w)| = o_p(\tilde{\xi}_n(\epsilon)).$$

由三角不等式可得

$$\sup_{w \in \mathcal{W}} |\text{CV}(w) - R_n(w)| = o_p(\tilde{\xi}_n(\epsilon)). \quad (\text{A.19})$$

定义 $\tilde{w} = (\tilde{w}_1, \dots, \tilde{w}_K)^T$, 其中 $\tilde{w}_k = 0$, 若 $k \notin \mathcal{C}$; $\tilde{w}_k = \hat{w}_k + (1 - \hat{w}_{\mathcal{C}})/|\mathcal{C}|$, 若 $k \in \mathcal{C}$, $|\mathcal{C}|$ 表示 $\mathcal{C}$ 的基数. 可以看出

$$\sum_{k=1}^K \tilde{w}_k = \sum_{k \in \mathcal{C}} \tilde{w}_k + \sum_{k \notin \mathcal{C}} \tilde{w}_k = \hat{w}_{\mathcal{C}} + |\mathcal{C}|(1 - \hat{w}_{\mathcal{C}})/|\mathcal{C}| = 1,$$

故 $\tilde{w} \in \mathcal{W}$. 条件 3.10 和 $L_n(w)$ 的非负性蕴涵 $\tilde{\xi}_n(\epsilon) > 0$ 对充分大的 n 成立. 当事件 $\{\hat{w}_{\mathcal{C}} \leq 1 - \epsilon\}$ 发生时, $R_n(\hat{w}) \geq \tilde{\xi}_n(\epsilon) > 0$. 进而根据 (A.19) 和 $\tilde{\xi}_n(\epsilon)$ 的定义可得

$$\text{CV}(\hat{w}) \geq R_n(\hat{w}) + o_p(\tilde{\xi}_n(\epsilon)) \geq \tilde{\xi}_n(\epsilon)(1 + o_p(1)). \quad (\text{A.20})$$

下面推导 $\text{CV}(\tilde{w})$ 的上界. 由文献 [27, 引理 1], 在条件 3.9 下可知 $m_{\oplus(k)}(x) = m_{\oplus}(x)$ 对一切 $k \in \mathcal{C}$ 和 $x \in \mathcal{X}$ 成立. 注意到

$$\sup_{w_{\mathcal{C}}=1} R_n(w) \leq 2D \sup_{w_{\mathcal{C}}=1} \mathbb{E}[d(\hat{m}_{\oplus}(X; w), m_{\oplus}(X; w))] = 2D \sup_{w_{\mathcal{C}}=1} \mathbb{E}\left[d\left(\sum_{k \in \mathcal{C}} w_k \hat{m}_{\oplus(k)}(X), \sum_{k \in \mathcal{C}} w_k m_{\oplus(k)}(X)\right)\right],$$

条件 3.4, 3.10 和定理 3.2 保证了

$$\sup_{w_{\mathcal{C}}=1} R_n(w) \leq 2CDKE \left[\max_{1 \leq k \leq K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X)) \right] = o(\tilde{\xi}_n(\epsilon)). \quad (\text{A.21})$$

结合上式和 (A.19), 可知

$$0 \leq \text{CV}(\tilde{w}) \leq R_n(\tilde{w}) + o_p(\tilde{\xi}_n(\epsilon)) \leq \sup_{w_{\mathcal{C}}=1} (R_n(w)) + o_p(\tilde{\xi}_n(\epsilon)) = o_p(\tilde{\xi}_n(\epsilon)).$$

然而, 因为 $\hat{w}$ 是极小值点, 故有 $\text{CV}(\hat{w}) \leq \text{CV}(\tilde{w}) = o_p(\tilde{\xi}_n(\epsilon))$, 这与 (A.20) 相矛盾, 证毕.

附录 A.6 推论 3.1 的证明

为证明推论 3.1, 需要推导 $L_n(\hat{w})$ 的上界和 $\inf_{w \in \mathcal{W}_\epsilon} L_n(w)$ 的下界. 对任意给定的 $0 < \epsilon \leq 1$, 由 (A.16) 可得

$$\inf_{w \in \mathcal{W}_\epsilon} L_n(w) \geq \inf_{w \in \mathcal{W}_\epsilon} R_n(w) - o_p(\tilde{\xi}_n(\epsilon)) \geq \tilde{\xi}_n(\epsilon)(1 + o_p(1)).$$

条件 3.10 保证了事件 $\{\inf_{w \in \mathcal{W}_\epsilon} L_n(w) > \tilde{\xi}_n(\epsilon)/2 > 0\}$ 发生的概率收敛到 1. 接下来推导 $L_n(\hat{w})$ 的上界. 由 $\text{CV}(\hat{w})$ 的极小性和 (A.21) 可得

$$\text{CV}(\hat{w}) \leq \sup_{w_{\mathcal{C}}=1} \text{CV}(w) \leq \sup_{w_{\mathcal{C}}=1} R_n(w) + o_p(\tilde{\xi}_n(\epsilon)) = o_p(\tilde{\xi}_n(\epsilon)).$$

因此, $L_n(\hat{w}) \leq \text{CV}(\hat{w}) + |L_n(\hat{w}) - \text{CV}(\hat{w})| = o_p(\tilde{\xi}_n(\epsilon))$. 此外, 已经证得 $\inf_{w \in \mathcal{W}} L_n(w) \geq 0$, 从而有 $L_n(\hat{w}) \geq 0$. 综合以上结果可知, 事件

$$0 \leq \frac{L_n(\hat{w})}{\inf_{w \in \mathcal{W}} L_n(w)} \leq \frac{L_n(\hat{w})}{\tilde{\xi}_n(\epsilon)/2} = o_p(1)$$

发生的概率收敛到 1. 证毕.

附录 A.7 定理 4.1 的证明

根据正确模型的定义, 可知每一个正确模型都包含全部的重要变量. 因此, 对于每个 $j \in \mathcal{I}$, 有

$$S_j = \sum_{k=1}^K \hat{w}_k I(j \in M_k) = \sum_{k \in \mathcal{C}} \hat{w}_k I(j \in M_k) + \sum_{k \notin \mathcal{C}} \hat{w}_k I(j \in M_k) \geq \hat{w}_{\mathcal{C}}.$$

应用定理 3.5 得到 $\hat{w}_{\mathcal{C}} \xrightarrow{P} 1$. 这表明 $\min_{j \in \mathcal{I}} S_j \xrightarrow{P} 1$.

将类似的方法应用于 $j \in \mathcal{I}_C^-$, 可以证明

$$\max_{j \in \mathcal{I}_C^-} S_j = \max_{j \in \mathcal{I}_C^-} \left[\sum_{k \in \mathcal{C}} \hat{w}_k I(j \in M_k) + \sum_{k \notin \mathcal{C}} \hat{w}_k I(j \in M_k) \right].$$

根据 $\mathcal{I}_C^-$, 可知 $\mathcal{I}_C^- \cap (\bigcup_{k \in \mathcal{C}} M_k) = \emptyset$. 这意味着对于任意 $j \in \mathcal{I}_C^-$ 和 $k \in \mathcal{C}$, 必然有 $I(j \in M_k) = 0$. 从而得到 $\max_{j \in \mathcal{I}_C^-} \sum_{k \in \mathcal{C}} \hat{w}_k I(j \in M_k) = 0$. 再注意到

$$\max_{j \in \mathcal{I}_C^-} \sum_{k \notin \mathcal{C}} \hat{w}_k I(j \in M_k) \leq \sum_{k \notin \mathcal{C}} \hat{w}_k = 1 - \hat{w}_C \xrightarrow{P} 0,$$

可得 $\max_{j \in \mathcal{I}_C^-} S_j \xrightarrow{P} 0$. 定理 4.1 证毕.

附录 B 假设条件验证与合理性说明

本附录给出本文中假设条件的合理性说明, 并提供相关示例.

附录 B.1 满足条件 3.1–3.3 的具体示例

首先证明 2-Wasserstein 空间和任意有限维 Hilbert 空间满足条件 3.1–3.3. 鉴于 2-Wasserstein 空间分析起来比较困难, 在此引入分位函数空间 $(\mathbb{Q}, \|\cdot\|_2)$, 该空间由 $L^2([0, 1])$ 空间中非降左连续函数构成并配备 L^2 范数. 根据三角不等式和等式

$$\int_0^1 (F^{-1}(t))^2 dt = \int_{\mathbb{R}} x^2 dF(x) = E(X^2)$$

可知 $\mathbb{S}$ 有界. 分析的关键在于度量空间 $(\mathbb{S}, W_2)$ 和 $(\mathbb{Q}, \|\cdot\|_2)$ 一一对应且等距同构. 因此, $(\mathbb{S}, W_2)$ 可视为 $L^2([0, 1])$ 空间的一个有界子空间. 对任意的分布 $s \in \mathbb{S}$, 令 $Q(s)$ 表示相应的分位函数. 对于任意 $x \in \mathcal{X}$, 由 $\mathcal{X}$ 的有界性和 Cauchy-Schwarz 不等式可知 $E(u(X, x)\langle Q(Y), \cdot \rangle)$ 是有界线性泛函, 其中 $\langle \cdot, \cdot \rangle$ 为 L^2 范数导出的内积, $u(X, x) = 1 + (X - \mu)^T \Sigma^{-1}(x - \mu)$. 从而, Riesz 表示定理保证了存在 $f_x \in L^2([0, 1])$ 使得

$$E(u(X, x)\langle Q(Y), q \rangle) = \langle f_x, q \rangle$$

对所有的 $q \in \mathbb{Q}$ 成立. 由内积的线性性质直接计算可得

$$\begin{aligned} & \|Q(Y) - f_x - (Q(s) - f_x)\|_2^2 \\ &= \|Q(Y) - f_x\|_2^2 + \|Q(s) - f_x\|_2^2 + 2(\langle Q(Y), Q(s) \rangle - \langle f_x, Q(s) \rangle - \langle Q(Y), f_x \rangle + \|f_x\|_2^2). \end{aligned}$$

进而有 $W(s, x) = E(u(X, x)\|Q(Y) - f_x - (Q(s) - f_x)\|_2^2) = E(u(X, x)\|Q(Y) - f_x\|_2^2) + \|Q(s) - f_x\|_2^2$. 这表明 $m_{\oplus}(x) = \operatorname{argmin}_{s \in \mathbb{S}} W(s, x) = Q^{-1}(\operatorname{argmin}_{q \in \mathbb{Q}} \|q - f_x\|_2^2)$.

不难看出空间 $\mathbb{Q}$ 在 L^2 度量下是闭的且 L^2 收敛保持单调性和左连续性. 此外, 直接验算可知 $\mathbb{Q}$ 是凸的且目标函数 $\|q - f_x\|_2^2$ 为严格凸. 因此, 对所有的 $x \in \mathcal{X}$, $m_{\oplus}(x)$ 存在且唯一. 注意到对所有的 $x \in \mathcal{X}$ 与 $k \in \{1, \dots, K\}$ 有 $\|x_{(k)}\| \leq \|x\|$, 因此 $m_{\oplus(k)}(x)$ 同样存在且唯一. 定义 $\hat{f}_{(k)x} = n^{-1} \sum_{i=1}^n \hat{u}_{(k)i}(x) Q(Y_i)$. 根据内积的线性性质可得

$$\frac{1}{n} \sum_{i=1}^n \hat{u}_{(k)i}(x) \langle Q(Y_i), q \rangle = \left\langle \frac{1}{n} \sum_{i=1}^n \hat{u}_{(k)i}(x) Q(Y_i), q \right\rangle,$$

其中 $\hat{u}_{(k)i}(x) = 1 + (X_{(k)i} - \bar{X}_{(k)})^T \hat{\Sigma}_{(k)}^{-1}(x_{(k)} - \bar{X}_{(k)})$. 这表明

$$W_{n(k)}(s, x) = \frac{1}{n} \sum_{i=1}^n \hat{u}_{(k)i}(x) \|Q(Y_i) - Q(s)\|_2^2 = \frac{1}{n} \sum_{i=1}^n \hat{u}_{(k)i}(x) \|Q(Y_i) - \hat{f}_{(k)x}\|_2^2 + \|Q(s) - \hat{f}_{(k)x}\|_2^2.$$

通过类似的方法可以证得 $\hat{m}_{\oplus(k)}(x)$ 和 $\tilde{m}_{\oplus(k)}^{-i}(x)$ 的存在性和唯一性.

进一步地, 根据 Hilbert 空间投影定理, $Q(m_{\oplus(k)}(x))$ 对全部 $q \in \mathbb{Q}$ 满足如下最优性条件:

$$\langle f_x - Q(m_{\oplus(k)}(x)), q - Q(m_{\oplus(k)}(x)) \rangle \leq 0.$$

由此立即可得

$$W_{(k)}(s, x) - W_{(k)}(m_{\oplus(k)}(x), x) = \|Q(s) - f_{(k)x}\|_2^2 - \|Q(m_{\oplus(k)}(x)) - f_{(k)x}\|_2^2 \geq d^2(m_{\oplus(k)}(x), s)$$

对所有的 $x \in \mathcal{X}$ 和 $k \in \{1, \dots, K\}$ 成立, 其中 $f_{(k)x}$ 定义方式与 f_x 相似. 类似地, 对于任意 $\varepsilon \geq 0$,

$$\inf_{x \in \mathcal{X}} \inf_{d(s, \hat{m}_{\oplus(k)}(x)) > \varepsilon} (W_{n(k)}(s, x) - W_{n(k)}(\hat{m}_{\oplus(k)}(x), x)) \geq d^2(\hat{m}_{\oplus(k)}(x), s) > \varepsilon^2$$

成立.

至此, 条件 3.1 以及条件 3.2 和 3.3 的前半部分已得证. 为了证明条件 3.2 和 3.3 的后半部分成立, 只需要证明

$$\max_{k=1, \dots, K} \int_0^1 \sup_{x \in \mathcal{X}} [1 + \log N(\delta\epsilon, B(\delta; m_{\oplus(k)}(x)), d)]^{1/2} d\epsilon = O(1).$$

为此需要一个额外的假设, 即 $\mathbb{S}$ 由具有 R 上的有界支撑 $\mathcal{K}$ 的一维概率分布构成. 令 $\mathcal{M}([0, 1])$ 为 $[0, 1]$ 的任一单调且一致有界的函数空间. 根据文献 [28, 定理 2.7.5], 可得

$$N(\epsilon, \mathbb{S}, d) = N(\epsilon, \mathbb{Q}, \|\cdot\|_2) \leq N(\epsilon/2, \mathcal{M}([0, 1]), \|\cdot\|_2) \leq e^{K/\epsilon},$$

其中 K 仅取决于 $\mathcal{M}([0, 1])$. 定义 $\mathcal{M}_-([0, 1]) = \{f - g : f, g \in \mathcal{M}([0, 1])\}$. 利用三角不等式即得

$$N(\epsilon, \mathcal{M}_-([0, 1]), \|\cdot\|_2) \leq N^2(\epsilon, \mathcal{M}([0, 1]), \|\cdot\|_2) \leq e^{2K/\epsilon}.$$

由此可知, 对于任意 $f \in \mathcal{M}_-([0, 1])$, 存在 $\{f_1, \dots, f_N\}$ 的 ϵ 覆盖, 这里 $N = N(\epsilon, \mathcal{M}_-([0, 1]), \|\cdot\|_2)$. 因此, 对于任意 $g \in B(\delta; q, \|\cdot\|_2)$, 有 $g = q + \delta(g - q)/\delta$ 且满足 $\|(g - q)/\delta\|_2 \leq 1$. 由于 $(g - q)/\delta \in \mathcal{M}_-([0, 1])$ 成立, 故而 $\{B(\delta\epsilon; q + \delta f_i, \|\cdot\|_2), i = 1, \dots, N\}$ 覆盖了 $B(\delta, q, \|\cdot\|_2)$. 综合以上可得

$$\begin{aligned} \max_{k=1, \dots, K} \sup_{x \in \mathcal{X}} \log N(\delta\epsilon, B(\delta; m_{\oplus(k)}(x)), d) &\leq \sup_{q \in \mathbb{Q}} \log N(\delta\epsilon, B(\delta; q, \|\cdot\|_2) \cap \mathbb{Q}, \|\cdot\|_2) \\ &\leq 2 \log N(\epsilon/2, \mathcal{M}_-([0, 1]), \|\cdot\|_2) \\ &\leq 4K/\epsilon. \end{aligned}$$

进而有

$$\max_{k=1, \dots, K} \int_0^1 \sup_{x \in \mathcal{X}} [1 + \log N(\delta\epsilon, B(\delta; m_{\oplus(k)}(x)), d)]^{1/2} d\epsilon \leq \int_0^1 \sqrt{1 + 4(K/\epsilon)} d\epsilon \leq 1 + 4\sqrt{K}.$$

因此, 已证得 2-Wasserstein 空间满足条件 3.1–3.3.

接下来将证明条件 3.1–3.3 对任意有界且有限维的 Hilbert 空间成立. 令 $\mathbb{S}$ 表示配备了范数 $\|\cdot\|$ 的 p 维 Hilbert 空间, 并用 $\langle \cdot, \cdot \rangle$ 表示相应的内积. 利用 Riesz 表示定理和内积的线性性质即得如下两个等式:

$$W(s, x) = E(u(X, x)\|Y - f_x\|^2) + \|s - f_x\|^2, \quad W_{n(k)}(s, x) = \frac{1}{n} \sum_{i=1}^n \hat{u}_{(k)i}(x) \|Y_i - \hat{f}_{(k)x}\|^2 + \|s - \hat{f}_{(k)x}\|^2.$$

进一步地, 由类似的方法可以证得 $m_{\oplus(k)}(x)$, $\hat{m}_{\oplus(k)}(x)$ 与 $\tilde{m}_{\oplus(k)}^{-i}(x)$ 的存在性和唯一性. 另外, 不难得到 $W_{(k)}(s, x) - W_{(k)}(m_{\oplus(k)}(x), x) \geq d^2(m_{\oplus(k)}(x), s)$ 和

$$\min_{k=1, \dots, K} \inf_{x \in \mathcal{X}} \inf_{d(s, \hat{m}_{\oplus(k)}(x)) > \varepsilon} (W_{n(k)}(s, x) - W_{n(k)}(\hat{m}_{\oplus(k)}(x), x)) \geq \varepsilon^2$$

对任意 $\varepsilon \geq 0$ 成立.

还需要证明

$$\max_{k=1, \dots, K} \int_0^1 \sup_{x \in \mathcal{X}} [1 + \log N(\delta\epsilon, B(\delta; m_{\oplus(k)}(x)), \|\cdot\|)]^{1/2} d\epsilon = O(1).$$

注意到 $N(\delta\epsilon, B(\delta; m_{\oplus(k)}(x)), \|\cdot\|) = N(\epsilon, B(1; m_{\oplus(k)}(x)), \|\cdot\|)$. 利用体积比方法可得

$$\max_{k=1, \dots, K} \sup_{x \in \mathcal{X}} N(\epsilon, B(1; m_{\oplus(k)}(x)), \|\cdot\|) \leq p \log(1 + (2/\epsilon)) \leq 2p/\epsilon.$$

由此可得

$$\max_{k=1, \dots, K} \int_0^1 \sup_{x \in \mathcal{X}} [1 + \log N(\delta\epsilon, B(\delta; m_{\oplus(k)}(x)), \|\cdot\|)]^{1/2} d\epsilon \leq 1 + 2\sqrt{2p},$$

即条件 3.1–3.3 对任意有界且有限维的 Hilbert 空间成立.

附录 B.2 满足条件 3.4–3.6 的具体示例

令 $\mathbb{S}$ 表示配备有范数 $\|\cdot\|$ 的有界赋范线性空间. 接下来证明 $(\mathbb{S}, \|\cdot\|)$ 满足条件 3.4. 事实上, 对任意的 $s_k, \tilde{s}_k \in \mathbb{S}$ ($k = 1, \dots, K$), 成立

$$d\left(\sum_{k=1}^K w_k s_k, \sum_{k=1}^K w_k \tilde{s}_k\right) = \left\| \sum_{k=1}^K w_k (s_k - \tilde{s}_k) \right\| \leq \sum_{k=1}^K \|s_k - \tilde{s}_k\| = \sum_{k=1}^K d(s_k, \tilde{s}_k).$$

此外, 对所有的 $s_k \in \mathbb{S}$, 可以得到

$$\begin{aligned} \sup_{\substack{s_k \in \mathbb{S} \\ k=1, \dots, K}} d\left(\sum_{k=1}^K w_{1k} s_k, \sum_{k=1}^K w_{2k} s_k\right) &\leq \sup_{\substack{s_k \in \mathbb{S} \\ k=1, \dots, K}} \sum_{k=1}^K |w_{1k} - w_{2k}| \|s_k\| \\ &\leq \sqrt{K} \|w_1 - w_2\| \sup_{\substack{s_k \in \mathbb{S} \\ k=1, \dots, K}} \|s_k\| \\ &= C \|w_1 - w_2\|, \end{aligned}$$

其中最后一个不等式成立是因为 K 是固定的且 $\mathbb{S}$ 有界. 因此, 条件 3.4 对 $(\mathbb{S}, \|\cdot\|)$ 成立. 作为特例, 当 $\mathbb{S}$ 取为 Euclid 空间 $\mathbb{R}$ 的有界子空间或为 Frobenius 范数下固定维数的相关矩阵空间时, 条件 3.4 成立. 特别地, 当 $\mathbb{S}$ 为 2-Wasserstein 空间时, $(\mathbb{S}, W_2)$ 可以被看作有界赋范线性空间 $L^2([0, 1])$ 的子空间, 故满足条件 3.4.

下面验证条件 3.5. 考虑经典线性回归模型

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon, \quad (\text{B.1})$$

其中 $\mathbf{Y} = (y_1, \dots, y_n)^T$ 为 n 维响应向量, $\mathbf{X}$ 为设计矩阵, $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T \sim N(0, \sigma^2 I)$ 为独立于 $\mathbf{X}$ 的随机误差向量. 令 (X, Y) 为 $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ 的独立同分布样本. 下面证明: 存在正整数 n_0 使得

$$\lim_{M \rightarrow \infty} \sup_{n \geq n_0} \mathbb{E}[n^{1/2} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X)) I_{\{n^{(1/2)} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X)) > M\}}] = 0, \quad (\text{B.2})$$

即 $\{n^{1/2} \max_{k=1, \dots, K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))\}_{n \geq n_0}$ 一致可积. 事实上, 在此情形下, 可得

$$\mathbb{E}[d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))^2] = \mathbb{E}(|X_{(k)}^T(\hat{\beta}_{(k)} - \beta_{(k)}^*)|^2) \leq \mathbb{E}(\|X_{(k)}\|^2 \|\hat{\beta}_{(k)} - \beta_{(k)}^*\|^2),$$

其中

$$\hat{\beta}_{(k)} = (\mathbf{X}_{(k)}^T \mathbf{X}_{(k)})^{-1} \mathbf{X}_{(k)}^T \mathbf{Y},$$

$\beta_{(k)}^* = \mathbb{E}(X_{(k)} X_{(k)}^T)^{-1} \mathbb{E}(XY)$, $\mathbf{X}_{(k)}$ 表示第 k 个候选模型所对应的设计矩阵. 由于 $X \in \mathcal{X}$ 是 $\mathbb{R}^p$ 的有界子集且 $X_{(k)}$ 是 X 的子向量, 所以可以推得 $\|X_{(k)}\| \leq C$. 进而, 有

$$\mathbb{E}[d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))^2] \leq C \mathbb{E}[\epsilon^T \mathbf{X}_{(k)} (\mathbf{X}_{(k)}^T \mathbf{X}_{(k)})^{-2} \mathbf{X}_{(k)}^T \epsilon].$$

由于 ϵ 和 $\mathbf{X}$ 相互独立, 根据重期望公式可得

$$\mathbb{E}[\epsilon^T \mathbf{X}_{(k)} (\mathbf{X}_{(k)}^T \mathbf{X}_{(k)})^{-2} \mathbf{X}_{(k)}^T \epsilon] = \mathbb{E}\{\mathbb{E}[\text{tr}(\epsilon \epsilon^T \mathbf{X}_{(k)} (\mathbf{X}_{(k)}^T \mathbf{X}_{(k)})^{-2} \mathbf{X}_{(k)}^T) \mid \mathbf{X}_{(k)}]\} = \sigma^2 \mathbb{E}[\text{tr}((\mathbf{X}_{(k)}^T \mathbf{X}_{(k)})^{-1})],$$

这里 $\text{tr}(A)$ 表示矩阵 A 的迹. 令 $\hat{\Sigma}_{(k)} = n^{-1} \mathbf{X}_{(k)}^T \mathbf{X}_{(k)}$. 注意到

$$\mathbb{E}(\|n^{1/2}(\hat{\beta}_{(k)} - \beta_{(k)}^*)\|^2) = \sigma^2 \mathbb{E}[\text{tr}((n^{-1} \mathbf{X}_{(k)}^T \mathbf{X}_{(k)})^{-1})] \leq C \mathbb{E}[\|\hat{\Sigma}_{(k)}^{-1}\|_2],$$

其中 $\|A\|_2$ 表示矩阵 A 的算子范数. 直接计算表明

$$\mathbb{E}[\|\hat{\Sigma}_{(k)}^{-1}\|_2] = \int_0^\infty \mathbb{P}(\|\hat{\Sigma}_{(k)}^{-1}\|_2 \geq t) dt = \int_0^\infty \mathbb{P}(\lambda_{\min}(\hat{\Sigma}_{(k)}) \leq t^{-1}) dt = \int_0^\infty \mathbb{P}(\lambda_{\min}(\hat{\Sigma}_{(k)}) \leq u) u^{-2} du,$$

其中最后一个等式利用了变量替换 $u = t^{-1}$. 假设 X 满足文献 [23] 中较弱的小球条件 (small-ball condition). 根据文献 [23, 定理 4] 可得

$$\mathbb{P}(\lambda_{\min}(\hat{\Sigma}_{(k)}) \leq u) \leq (C_1 u)^{C_2 n}.$$

注意到

$$\begin{aligned} \int_0^\infty \mathbb{P}(\lambda_{\min}(\hat{\Sigma}_{(k)}) \leq u) u^{-2} du &\leq \int_0^{C_1^{-1}} (C_1 u)^{C_2 n} u^{-2} du + \int_{C_1^{-1}}^\infty u^{-2} du \\ &\leq C_1 \int_0^1 v^{C_2 n - 2} dv + C_1 \leq \infty, \end{aligned}$$

从而有 $\sup_{n \geq n_0} \mathbb{E}[\|\hat{\Sigma}_{(k)}^{-1}\|_2] \leq C$ 对某个正整数 n_0 成立. 进而由 Cauchy-Schwarz 不等式和 Markov 不等式保证了

$$\begin{aligned} &\sup_{n \geq n_0} \mathbb{E}[n^{1/2} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X)) I_{\{n^{1/2} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X)) > M\}}] \\ &\leq \sup_{n \geq n_0} \{\mathbb{E}[n^{1/2} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))]^2 \mathbb{P}(n^{1/2} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X)) > M)\}^{1/2} \\ &\leq CM^{-1} \end{aligned}$$

随着 $M \rightarrow \infty$ 收敛到 0. 综上所述, 在条件 3.6 成立时 $\{\xi_n^{-1} \max_{k=1, \dots, K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))\}_{n \geq n_0}$ 与 $\{\xi_{n+1}^{-1} \max_{k=1, \dots, K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))\}_{n \geq n_0}$ 一致可积.

为验证条件 3.6, 考虑经典线性回归模型 (B.1). 不失一般性, 假设响应变量和自变量均为中心化的. 假设所有的候选模型均遗漏了某些真实变量, 即存在一个非空集 $\mathcal{A} \subseteq \{1, \dots, p\}$ 满足 $\mathcal{A} \cap (\bigcup_{k=1}^K M_k) = \emptyset$

且 $\sum_{i \in \mathcal{A}} \beta_i^2 > 0$. 对候选模型 M_k , 用 $\hat{\beta}_{(k)}$ 表示 β 相应于 M_k 的最小二乘估计. 则 β 对应于权重 w 的模型平均估计量可以表示为

$$\hat{\beta}(w) = \sum_{k=1}^K w_k \begin{pmatrix} \hat{\beta}_{(k)} \\ 0 \end{pmatrix},$$

进而 $m_{\oplus}(x) = E(Y | X = x) = x^T \beta$ 的模型平均估计为 $x^T \hat{\beta}(w)$.

在经典线性回归框架下, 有

$$\begin{aligned} R_n(w) &= E[(Y - X^T \hat{\beta}(w))^2 - (Y - X^T \beta)^2] = E\{[X^T(\beta - \hat{\beta}(w))][2\epsilon + X^T(\beta - \hat{\beta}(w))]\} \\ &= E[X^T(\beta - \hat{\beta}(w))]^2, \end{aligned}$$

其中最后一个等式成立是根据 $E(X\epsilon) = 0$. 利用 $\mathcal{D}_n$ 和 (X, Y) 独立性可以得到

$$E[X^T(\beta - \hat{\beta}(w))]^2 = E_{\mathcal{D}_n}[(\beta - \hat{\beta}(w))^T E(XX^T)(\beta - \hat{\beta}(w))] \geq \lambda_{\min}(\Sigma) E\|\beta - \hat{\beta}(w)\|^2,$$

这里 $\lambda_{\min}(\cdot)$ 表示最小特征值. 注意到 $\hat{\beta}_{\mathcal{A}}(w) = 0$ 与 $\beta_{\mathcal{A}} \neq 0$, 有

$$\|\beta - \hat{\beta}(w)\|^2 \geq \|\beta_{\mathcal{A}} - \hat{\beta}_{\mathcal{A}}(w)\|^2 = \|\beta_{\mathcal{A}}\|^2 > 0.$$

结合以上结果可知

$$\xi_n = \inf_{w \in \mathcal{W}} R_n(w) > 0,$$

从而条件 3.6 成立.

附录 B.3 条件 3.8 和 3.10 的合理性说明

对于线性回归模型 (B.1), 已证得 (B.2), 即

$$\left\{ n^{1/2} \max_{k=1, \dots, K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X)) \right\}_{n \geq n_0}$$

一致可积, 因此条件 3.10 成立.

若 $\mathcal{C}$ 为空集 (或 $\epsilon = 1$), 则条件 3.10 的第一部分等价于条件 3.6, 即所有的候选模型都被错误设定. 此外, 在线性回归模型 (B.1) 下, 对 $w \in \mathcal{W}_\epsilon$, 有

$$R_n(w) \geq CE \|\beta - \hat{\beta}(w)\|^2 = CE \left\| \sum_{k=1}^K w_k (\beta - \hat{\beta}_{(k)}) \right\|^2.$$

令集合 $\mathcal{A}$ 表示被全部候选模型遗漏的真变量并用 $\hat{\beta}_{(k)j}$ 表示 $\hat{\beta}_{(k)}$ 的第 j 个分量. 进而可以给出 $R_n(w)$ 的下界:

$$\begin{aligned} R_n(w) &\geq CE \left[\sum_{j \in \mathcal{A}} \left\{ \sum_{k \in \mathcal{C}^c} w_k \beta_j + \sum_{k \in \mathcal{C}} w_k (\beta_j - \hat{\beta}_{(k)j}) \right\}^2 \right] \\ &= CE \left[\sum_{j \in \mathcal{A}} \left\{ \left(\sum_{k \in \mathcal{C}^c} w_k \beta_j \right)^2 + 2 \sum_{k \in \mathcal{C}^c} w_k \beta_j \sum_{k \in \mathcal{C}} w_k (\beta_j - \hat{\beta}_{(k)j}) + \left(\sum_{k \in \mathcal{C}} w_k (\beta_j - \hat{\beta}_{(k)j}) \right)^2 \right\} \right]. \end{aligned}$$

当 $k \in \mathcal{C}$ 时有 $E(\hat{\beta}_{(k)j}) = \beta_j$, 故对 $w \in \mathcal{W}_\epsilon$ 有 $R_n(w) \geq C \sum_{j \in \mathcal{A}} \epsilon^2 \beta_j^2$ 成立. 对于给定的 $\epsilon > 0$, 当 $(n/\log n)^{-1/2}/(\sum_{j \in \mathcal{A}} \epsilon^2 \beta_j^2) = o(1)$ 时条件 3.10 的第一部分成立. 此外, 条件 3.10 的第一部分和 $\{n^{1/2} \max_{k=1, \dots, K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))\}_{n \geq n_0}$ 一致可积保证了 $\{\tilde{\xi}_n^{-1}(\epsilon) \max_{k=1, \dots, K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))\}_{n \geq n_0}$ 与 $\{\tilde{\xi}_{n+1}^{-1}(\epsilon) \max_{k=1, \dots, K} d(\hat{m}_{\oplus(k)}(X), m_{\oplus(k)}(X))\}_{n \geq n_0}$ 一致可积. 因此, 在线性回归模型 (B.1) 下, 条件 3.10 的第二部分成立.

Optimal model averaging for global Fréchet regression

Jianqiang Zhang, Ze Chen, Jun Liao & Wangli Xu

Abstract Global Fréchet regression provides a powerful tool for studying complex non-Euclidean data, which are frequently encountered in many areas of applications. However, it is often difficult and challenging for global Fréchet regression to determine the covariates in practice. To alleviate the issue of model selection uncertainty, we propose a feasible model averaging approach coupled with a data-driven cross-validation weight choice criterion. Theoretically, we establish the consistency of the global Fréchet regression estimators and derive the optimal pointwise convergence rate and a near-optimal uniform convergence rate even when the candidate models are misspecified. Furthermore, we establish the asymptotic optimality under the model misspecification and the rate of convergence for the weight estimator obtained by cross-validation. In addition, if the candidate model set includes at least one correct model, the weights chosen by cross-validation are shown to be asymptotically concentrated on all correct models. Moreover, based on the proposed model averaging approach, we develop a reliable variable importance procedure to measure the importance of each covariate in global Fréchet regression. This procedure is proven to asymptotically identify all the variables in the true model. Simulation experiments and an empirical application illustrate the superiority of the proposed model averaging method over some common model selection and averaging methods.

Keywords asymptotic optimality, Fréchet regression, metric space-valued data, model averaging, variable importance

MSC(2020) 62J99

doi: 10.1360/SCM-2024-0265