arXiv:2305.01350v1 [stat.ME] 2 May 2023

稳健自适应函数型逻辑回归

Ioannis Kalogridis

鲁汶大学数学系，比利时

2023年5月3日

摘要

我们提出并研究了一类用于函数型逻辑回归模型的稳健估计量，其稳健性能够根据数据自动调整，从而在干净数据上实现高效率，同时具有很强的抵御异常观测的能力。这些估计量基于密度之间的密度幂散度概念，并可根据需要结合任意形式的低秩近似与惩罚项。对于这些估计量，我们在相当一般的假设下，证明了它们相对于常用预测误差的一致收敛性和较快的收敛速率。模拟研究和一个包含异常观测的真实数据示例展示了该方法在实际应用中的出色表现。

关键词：函数型逻辑回归、稳健性、正则化、渐近理论 2020年数学主题分类：
62R10, 62G35, 62G20

1 引言

近年来，技术进步和存储能力的提升，使应用科学家能够观测并记录日益复杂的高维对象。这些对象可以被视为取值于某个性质良好的函数空间上的随机函数的实现。此类数据通常称为函数型数据。自 Ramsay 和 Dalzell (1991)、Ramsay 和 Silverman(2005) 以及 Ferraty 和 Vieu (2006) 等突破性研究问世以来，对函数型数据的研究备受关注。在许多情况下，人们希望将一个随机函数 X（其样本路径属于 $\mathcal{L}^2(\mathcal{I})$，其中 $\mathcal{I} \subset \mathbb{R}$ 是某个紧致 Y）与二元响应联系起来，具体通过以下模型

$$\mathbb{E}\{Y|X\} = H\left(\alpha_0 + \int_{\mathcal{I}} X(t)\beta_0(t)dt\right), \tag{1}$$

其中 $H\colon \mathbb{R} \to (0,\ 1)$ 是已知的双射“连接”函数，而 $(\alpha_0,\ \beta_0) \in \mathbb{R} \times \mathcal{L}^2(\mathcal{I})$ 是需要根据数据估计的未知量。该标量对函数型逻辑回归模型

最初由 James （2002）提出，时至今日仍广受欢迎，原因在于其具有广泛的应用范围，例如癌症基因检测和胎儿风险评估，分别参见Ratcliffe 等（2002）以及 Wei 等人 （2014）。

尽管业界人士持续关注，标量对函数型逻辑回归的理论研究却明显滞后，研究注意力反而大多集中在标量对函数线性回归模型上。对于这类模型，有意采用相关方法的研究者很容易找到大量理论基础扎实的估计技术。例如，基于主要函数主成分的回归（Cardot 等人，1999；Hall 等人，2007）、惩罚样条回归（Cardot 等人，2003、由 Crambes 等人（2009）和 Yuan 和 Cai （2010）提出并研究的全秩平滑样条回归，以及结合函数主成分与惩罚样条的混合方法（Reiss 和 Ogden，2007）。遗憾的是，前述研究中提出的理论结果都无法推广到函数型逻辑回归模型；据我们所知，该模型仅有的理论结果由 Cardot 和 Sarda （2005）以及 M¨uller 和 Stadtm¨uller （2005）提出。前者证明了基于惩罚似然的样条估计量具有一致性，后者则证明了由系数函数 β_0的截断函数主成分展开所导出的估计值具有一致性和渐近正态性。

Cardot 和 Sarda（2005）以及 M¨uller 和 Stadtm¨uller （2005）这两种方案在实际应用中的一个重要缺点，是无法抵抗异常观测。之所以缺乏这种抵抗力，是因为它们依赖每个观测的对数似然贡献（X_i， Y_i），因此，任何拟合不佳的观测对，例如具有较大杠杆值且被误分类的 X_i Y_i，都会显著扭曲对数似然函数及其导出的估计值。为解决这一缺乏抵抗力的问题，Denhere 和 Billor（2016）提出，使用 Hubert 等人提出的（2005）方法获得的前几个稳健主成分进行逻辑回归。该方法在实践中仅部分有效，因为它只处理离群预测函数，而不处理误分类响应；后者仍可能导致严重扭曲。最近，Mutis 等人（2022）在 Denhere 和 Billor（2016）的方案基础上作了扩展，将基于皮尔逊残差的权重纳入对数似然，从而获得了抵抗力更强的估计值。然而，这些方法的理论性质均尚未明了，因此无法确定最佳实践。

为填补文献中关于缺乏理论依据或无法抵御离群观测的估计量所留下的重要空白，本文提出并研究了一类灵活的惩罚低秩平滑器，兼具上述两种作用。通过采用由Basu等人(1998提出的密度幂散度目标函数实现稳健性。我们将其与系数函数的低秩表示及惩罚项灵活结合，从而获得易于计算且可避免过拟合的估计值。

我们以数据驱动的方式选择目标函数的调节参数，以便在干净数据中实现高效率，并显著抵御离群观测。在一般假设下（甚至允许调节参数为随机变量），我们证明了估计量的一致收敛，并将预测误差分解为方差以及由建模和惩罚化导致的两类偏差。这些理论结果是针对标量对函数逻辑回归模型的稳健估计量首次取得的成果；此外，我们的现代方法也可用于强化较早的研究结果，例如Cardot和Sarda(2005)关于惩罚最大似然估计的研究。

本文其余部分安排如下。第2 节详细阐述并说明我们的方案。第3 节研究所提出估计量类的渐近性质；我们分别讨论一致性以及关于一致误差和预测误差的收敛速率。第4节讨论近似子空间大小、惩罚参数以及目标函数调节常数的实际选择。第5节和6 节分别通过蒙特卡洛研究和一个涉及步态分析的真实数据示例，展示我们方案具有竞争力的有限样本性能。最后，第7节作简短的总结讨论，所有证明则汇集于附录中。

2 所提出的估计量族

为了详细说明所提出的估计量族，首先需要回顾密度幂散度的概念，该概念由Basu等人(1998提出。考虑关于计数测度的两个密度， h和 f。对于 $\kappa_0 \geq 0$密度幂散度$d_{\kappa_0}(h,\ g$定义如下:

$$d_{\kappa_0}(h,f) = \begin{cases} \sum_{y\in\mathbb{Z}} \left\{ f^{1+\kappa_0}(y) - \left(1+\frac{1}{\kappa_0}\right) h(y) f^{\kappa_0}(y) + \frac{1}{\kappa_0} h^{1+\kappa_0}(y) \right\} & \kappa_0 > 0 \\ \sum_{y\in\mathbb{Z}} h(y) \log\{h(y)/f(y)\} & \kappa_0 = 0. \end{cases} \tag{2}$$

可以证明， $d_{\kappa_0}(h,f) \geq 0$，并且当且仅当 $h = f$ 在整个 Z 上时， $d_{\kappa_0}(h,f) = 0$；参见附录中的引理1。假设 h 是真实密度，而 $\{f_\theta, \theta \in \Theta\}$ 是一族模型密度，则密度幂散度估计量渐近地使 $d_{\kappa_0}(h,f_\theta)$ 在 $\theta \in \Theta$ 上达到最小值。值得注意的是， $d_0(h,f_\theta)$ 其实就是 Kullback-Leibler散度，它由最大似然估计量渐近最小化。然而，众所周知，尽管最大似然估计量具有较高的效率，却无法抵御离群值和模型设定错误。基于 $\kappa_0 > 0$ 的密度幂散度估计量具有更强的抗干扰能力：其影响函数（Hampel等人，2011）有界，且对于较大的离群值会趋于零，因此更适合分析包含离群值的复杂数据集。在

事实上，正如对Kalogridis等人（2023的分析所展示的，密度幂散度估计量的稳健性会随着 κ_0取值增大而增强，但这会牺牲其效率，因此在实践中需要权衡效率与稳健性。

为了了解密度散度如何应用于函数型逻辑回归，我们假设拥有一个样本 $(X_1, Y_1) \ldots, (X_n, Y_n)$，由 $\mathcal{L}^2(\mathcal{I}) \times \{0, 1\}$ 中服从模型 (1) 的独立同分布元组组成。为简洁起见，且不失一般性，我们将 $\mathcal{I}$ 视为 $[0, 1]$区间。根据 (1)，我们可以假设，给定 X_i 后，每个二元响应 Y_i 都具有伯努利密度$f_{p_i(\alpha_0,\beta_0)}(y) = p_i(\alpha_0, \beta_0)^y(1 - p_i(\alpha_0, \beta_0))^{1-y}$其中，“成功”的概率 $p_i(\alpha_0, \beta_0)$ 为

$$p_i(\alpha_0, \beta_0) = H\left(\alpha_0 + \int_{[0,1]} X_i(t)\beta_0(t)dt\right), \quad (i = 1, \ldots, n).$$

我们在附录中证明，$M((\alpha, \beta) \kappa_0) := \mathbb{E}\{d_{\kappa_0}(f_{p_1(\alpha_0,\beta_0)} f_{p_1(\alpha,\beta)})\}$ 的下界为 0，上界为 $1+\kappa_0^{-1}$。当且仅当 $|p_1(\alpha_0, \beta_0)) - p_1(\alpha, \beta)| =$ 几乎处处等于 1 时，才能达到上界。此外，只要 H（即连接函数）也是连续函数，$M((\alpha, \beta) \kappa_0)$ 就是 (α, β) 的连续函数。这些性质使 $M((\alpha, \beta) \kappa_0)$ 成为稳健估计中颇具吸引力的目标函数。但遗憾的是，由于我们既不知道 X 的分布，也不知道总体参数 (α_0, β_0)，因此直接最小化 $M((\alpha, \beta) \kappa_0)$ 并不可行。因而，我们的方法是最小化这一准则的适当近似。

$M((\alpha, \beta) \kappa_0)$ 的有限样本近似可通过将 $\mathbb{E}\{\cdot\}$ 替换为其经验对应量获得，其中每个 X_i 赋予权重 n^{-1}；构造 M-估计量时通常采用这种做法（例如参见 van der Vaart，1998，第5章）。接下来，由（2）可知，我们可以删去不依赖于 (α, β) 的项，因为这些项不会影响极小值点。最后，对于每个 $\kappa_0 > 0$，$\sum_y f_{p_i}^{\kappa_0}(\alpha,\beta)(y) f_{p_i(\alpha_0,\beta_0)}(y)$ 的一个条件无偏估计量为 $f_{p_i}^{\kappa_0}(\alpha,\beta)(Y_i)$。综合以上各点，我们得到 $M((\alpha, \beta) \kappa_0)$ 的有限样本近似为

$$\frac{1}{n}\sum_{i=1}^{n} l_{\widehat{\kappa}_n}(Y_i, p_i(\alpha, \beta)) = \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{y\in\{0,1\}} f_{p_i(\alpha,\beta)}^{1+\widehat{\kappa}_n}(y) - \left(1 + \frac{1}{\widehat{\kappa}_n}\right) f_{p_i(\alpha,\beta)}^{\widehat{\kappa}_n}(Y_i)\right], \tag{3}$$

现在我们也允许调节参数 $\widehat{\kappa}_n$ 发生变化，甚至是随机的，并且这些参数可能依赖于样本本身。这些参数决定了稳健性与效率之间的权衡：当 $\widehat{\kappa}_n \to 0$ 时，我们得到的本质上是伯努利分布的对数似然；而当 $\widehat{\kappa}_n$ 较大时，我们得到的目标函数会导出稳健性更强、但效率较低的估计量。上述表述在某种意义上使我们能够兼顾两者，因为在干净样本中可以将 $\widehat{\kappa}_n$ 选得足够小，而在包含离群观测的数据集中则可以选得较大。一种可能的

我们发现，对于 $\widehat{\kappa}_n$ ，以下选择方案是有效的 p实践在第 4节中概述。

我们遵循Ramsay 和 Silverman(2005，第 15 章）的方案，从目标函数（3）中得到估计量。也就是说，我们不直接在整个$\mathbb{R} \times \mathcal{L}^2([0,1]$上最小化（3），以免产生过拟合的风险，而是将关注范围限制在小得多的空间$\mathbb{R} \times \Theta_{K_n}$，其中 Θ_{K_n} 是一个由已知实值函数 K_n维线性子空间 θ_1、……、θ_{K_n}张成。这些函数可以是 B样条或傅里叶基，它们都广泛应用于函数型数据分析。这里，近似子空间的维数 K_n取决于 n。为了使近似偏差在极限下消失，我们需要令 $K_n \to \infty$ 当 $n \to \infty$ 时成立，但其趋近速度不能太快，见下文定理1 。除了有限维子空间，我们还使用惩罚泛函 $\mathcal{J}$：$\Theta_{K_n} \to \mathbb{R}_+$ ，以进一步控制斜率估计值的粗糙度。通过这两项改进，我们的方法是对未知量进行如下估计：

$$(\widehat{\alpha}_n, \widehat{\beta}_n) = \underset{(\alpha,\beta)\in\mathbb{R}\times\Theta_{K_n}}{\operatorname{argmin}} \left[\frac{1}{n} \sum_{i=1}^{n} l_{\widehat{\kappa}_n}(Y_i, p_i(\alpha,\beta)) + \widehat{\lambda}_n \mathcal{J}(\beta) \right], \tag{4}$$

其中 λ_n 惩罚参数可能是随机的，它决定了对 β_n 粗糙度施加的惩罚程度，粗糙度由 $\mathcal{J}(\beta_n)$ 衡量。对于 $\lambda_n = 0$ ，估计量是 θ_j 的未惩罚线性组合；而对于 $\lambda_n \to \infty$ ，惩罚项主导目标函数，迫使斜率估计量落在其零空间中。$\mathcal{J}(\cdot)$ 的零空间通常维数较低，例如可能由不超过某一阶数的多项式构成，因此其中的函数往往形式简单。

值得注意的是，式（4）中的估计量在维数和复杂度方面都经过了正则化。根据我们的经验，为了确保实践中整体表现可靠，这两种正则化都不可或缺。具体而言，当前的双重正则化方案确保了通过取较大的 K_n 值，可以使估计量具有足够的灵活性；同时，借助 $\mathcal{J}(\cdot)$ 进行惩罚，又能保持其可解释性。此外，当前的设定非常灵活，应用研究者可以根据需要选用各种近似子空间 Θ_{K_n} 和惩罚项 $\mathcal{J}(\cdot)$ 。例如，常用的 B样条可以配合导数惩罚项或差分惩罚项；对于怀疑 l_1 不够平滑的情况，也可以采用小波并配合全变差惩罚项或 l_1 惩罚项。感兴趣的读者可参阅 Kalogridis 和 Van Aelst （2023），了解若干具有实际用途的替代方案。

乍看之下，所提出的估计量似乎计算成本较高，因为它取决于三个参数，即调节参数 $\widehat{\kappa}_n$、近似子空间的维数、K_n以及惩罚参数 λ_n。然而，低秩惩罚估计量的实际应用经验表明，K_n 并不像问题中的其他参数那样关键，因此可以采用半自动方式进行选择，也就是说，其选择几乎可以独立于样本（参见，例如 Ruppert 等人，2003；Wood，2017）。一种高效地选择 $\widehat{\kappa}_n$ 和 λ_n 的方法是

将在下文第 4节讨论。

3 理论性质

3.1 一致性

现在我们研究惩罚化密度幂散度估计量的渐近性质。在本节中，我们假设 $\alpha_0 = 0$，并且感兴趣的对象是系数函数β_0，这也是通常的情形。我们理论推导所需的假设如下。

(A1) 存在一个 $\kappa_0 \in (0, 1]$，使得 $\widehat{\kappa}_n \xrightarrow{\mathbb{P}} \kappa_0$，当 $n \to \infty$ 时。

(A2) 连接函数 $H : \mathbb{R} \to (0,1)$ 是双射、严格递增且可微的，其导数 H' 满足 $\sup_{x\in\mathbb{R}} H'(x) \leq c_0$，其中某个 $c_0 < \infty$。

(A3) 系数函数 β_0 属于一个巴拿赫函数空间 $\mathcal{B}([0,1])$，其范数为 $\|\cdot\|_{\mathcal{B}}$，该空间可嵌入 $\mathcal{C}([0,1])$，其范数为 $\|\cdot\|_\infty$，即 $\mathcal{B}([0,1]) \subset \mathcal{C}([0,1])$，并且存在一个有限的 $c_1 > 0$，使得

$$\|f\|_\infty \leq c_1 \|f\|_{\mathcal{B}}, \quad \forall f \in \mathcal{B}([0,1]).$$

~（A4） $\Theta_{K_n} \subset \mathcal{B}([0,1])$ 且对于某个 $\gamma \in (0, 1)$。令 $\beta_{K_n} = \text{argmin}_{\beta\in\Theta_{K_n}} \|\beta - \beta_0\|$，我们要求 $\|\widetilde{\beta}_{K_n} - \beta_0\| \to 0$ 和 $\widehat{\lambda}_n \mathcal{J}(\widetilde{\beta}_{K_n}) \xrightarrow{\mathbb{P}} 0$，因为 $n \to \infty$。

（A5）存在 $0 < C < \infty$ 使得 $\mathbb{P}(\|X\| \leq C) = 1$，并且存在 $0 \leq c_2 < 1$，使得对于每个 $\beta \in \mathcal{B}([0,1]$，满足 $\beta \neq 0$ 的情形下，均有 $\mathbb{P}(\langle X, \beta\rangle = 0) \leq c_2$。

假设 (A1) 要求随机调节参数 $\widehat{\kappa}_n$ 依概率收敛到某个正数 κ_0。这类假设常用于对涉及干扰（估计）参数的估计量进行理论分析，参见例如（van der Vaart，1998，第 5.4 节和第 5.8.1 节）。也可以确定性地选择调节参数，即 $\widehat{\kappa}_n = \kappa_0$ 对于某个$\kappa_0 > 0$ 对所有 $n \in \mathbb{N}$ 均成立；此时 (A1) 显然满足，且后续所有论证仍然有效。关于连接函数 H 的假设 (A2) 同样非常一般，涵盖所有常用的连接函数，即 logit、probit 和 Gumbel（互补对数-对数）连接函数。

假设（A3）是关于系数函数的一项温和假设，β_0。要求β_0 属于定义在 $[0,1]$上的函数完备赋范空间，$\mathcal{B}([0,1])$，该空间是连续函数空间的子空间，$\mathcal{C}([0,1])$。此假设比 Kalogridis 和 Van Aelst (2023）中的相应假设宽泛得多，因为我们不要求 $\mathcal{B}([0,1])$ 嵌入 $\mathcal{C}([0,1])$是紧致的。因此，（A3）适用于更多种类的函数

空间。例如，在我们的框架中，可以取 $\mathcal{B}([0,1]) = \mathcal{C}([0,1])$，但这不符合 Kalogridis 和 Van Aelst (2023) 的假设，因为单位球 $\{f \in \mathcal{C}([0,1]): \|f\|_\infty \leq 1\}$ 在 $\mathcal{C}([0,1])$ 中不是紧致的。

假设（A4）要求近似子空间的维数 K_n 相对于样本量 n 的增长不能过快，以确保 $K_n/n \to 0$。这是基于近似子空间（筛法）的估计量常用的假设，参见例如 Eggermont 和 LaRiccia（2009，第 15 章）以及 Kalogridis 和 Van Aelst（2023）。此外，要求 β_0 能够在 $\mathcal{L}^2([0,1])$ 范数下由 Θ_{K_n} 中的一个元素良好逼近。β_0 在 Θ_{K_n}、$\widetilde{\beta}_{K_n}$ 中这一列确定性逼近的粗糙度（由 $\mathcal{J}(\cdot)$ 度量）应有限，从而使 $\widehat{\lambda}_n \mathcal{J}(\widetilde{\beta}_{K_n}) \xrightarrow{\mathbb{P}} 0$。因此，要求相对于 Θ_{K_n} 恰当选取 $\mathcal{J}(\cdot)$。许多常用的近似子空间和惩罚项都满足这两项要求，例如，采用基于导数的惩罚项的 B样条，以及采用导数惩罚或谐波加速惩罚的傅里叶基（Ramsay 和 Silverman，2005）。值得注意的是，与其他关于函数型线性回归模型的理论研究（如 Cardot 等人，2003；Yuan 和 Cai，2010；Shin 和 Lee，2016）不同，本文更现实地将 $\widehat{\lambda}_n$ 视为随机变量序列，而非常数序列。这一区别很重要，因为 λ_n 通常以数据驱动方式选取，因此是随机的，而非固定的。

最后，假设（A5）要求将协变量 X 视为 $\mathcal{L}^2([0,1])$ 中的随机元素时几乎处处有界。在许多关于函数型线性回归模型的论述中也存在类似假设，例如（Cardot 等人，2003；Kalogridis 和 Van Aelst，2023）。此外，还要求 X 不集中于 $\mathcal{B}([0,1])$ 的任何有限维子空间上。由于 $\mathcal{B}([0,1]) \subset \mathcal{L}^2([0,1])$，只要 X 的协方差算子的零空间仅由零元素 $\{0\}$ 构成，就能满足这一要求；当且仅当其可数无穷多个特征值均严格为正时，这一条件成立。该假设同样比 Kalogridis 和 Van Aelst（2023）的假设（A4）更为一般，因为它只涉及空间 $\mathcal{B}([0,1])$，而不是更大的 $\mathcal{L}^2([0,1])$。这是因为根据（A4），我们的估计量位于 $\mathcal{B}([0,1])$ 中，因此我们可以将注意力限制在这个较小的函数空间上。

如附录所示（见其中的引理 1 和引理 2），仅在假设(A1)、(A2)和(A5)成立时，我们的估计量便是 Fisher 一致的，也就是说，它在总体层面估计出正确的量。假设(A3)和（A4）的实际意义在于，它们使我们还能使我们的有限样本估计量以强一致性的形式 β_n 趋近总体函数 β_0。这一结果将在下文的定理 1 中给出。

定理 1。假设假设*(A1)–(A5)*均成立。那么，$\|\widehat{\beta}_n - \beta_0\|_{\mathcal{B}} \xrightarrow{\mathbb{P}} 0$，当 $n \to \infty$ 时。

定理1 证明了我们的估计量的收敛性 β_n 作为巴拿赫空间$\mathcal{B}([0,1])$中的一个元素。据我们所知，这是此前尚未针对任何稳健函数型回归估计量证明过的一种强收敛模式。特别是，Boente 等人 (2020)和Kalogridis 和 Van Aelst(2023)的结果仅能推出 $\|\beta_n - \beta_0\|_{\mathcal{B}} = O_{\mathbb{P}}(1)$，

当 $n \to \infty$时，这显然是一个较弱的结果。由于根据我们的假设， $\mathcal{B}([0,1]$ 可嵌入$\mathcal{C}([0,1])$中，定理 1 的结果至少意味着一致收敛 $\|\widehat{\beta}_n - \beta_0\|_\infty \xrightarrow{\mathbb{P}} 0$，因为$n \to \infty$；但如果 $\mathcal{B}([0,1])$具有更高的正则性，例如 $\mathcal{B}([0,1])$是一个赫尔德空间¨或索博列夫空间（使用 B 样条或傅里叶估计量时通常会作此假设），则它也可能意味着某些导数收敛。例如，如果 $\mathcal{B}([0,1])$是一个赫尔德空间，其导数直到阶¨均满足赫尔德条件，且指数为 m ，那么收敛是相对于该范数而言的

$$\|f\|_{\mathcal{B}} = \|f\|_\infty + \max_{0\le j\le m} \sup_{\substack{x,y\in[0,1]\\ x\neq y}} \left| \frac{f^{(j)}(x) - f^{(j)}(y)}{(x-y)^r} \right|, \quad f \in \mathcal{B}([0,1]).$$

除了其内在的理论意义外，这种收敛模式还为估计量的性能提供了重要的实际保证，即总体函数 β_0 及其导数在整个定义域内都能得到同样准确的估计。

3.2 收敛速率

为获得所提出的估计量族的收敛速率，我们采用由下式给出的距离准则：

$$\left|\pi(\widehat{\beta}_n, \beta_0)\right|^2 = \mathbb{E}\left\{ \left| \langle X, \widehat{\beta}_n - \beta_0 \rangle \right|^2 \right\},$$

该准则可改写为 $\pi(\widehat{\beta}_n, \beta_0) = \{\langle \Gamma(\widehat{\beta}_n - \beta_0)\ \widehat{\beta}_n - \beta_0 \rangle\}^{1/2}$，其中 Γ 表示 X的自伴希尔伯特–施密特协方差算子。这是一个自然的考量准则，在函数型线性模型的理论研究中已得到广泛应用，例如 Cardot 和 Sarda（2005）、Crambes 等（2009）、Boente 等人（2020）以及 Kalogridis 和 Van Aelst（2023）的研究。直观地说，$\pi(\beta_n, \beta_0)$对应于使用 $\langle X_{n+1}, \beta_n \rangle$ 预测$\langle X_{n+1}, \beta_0 \rangle$时产生的平均预测误差，其中 X_{n+1} 是一个新的随机函数，其分布与 X相同。

值得注意的是，在假设(A5)下， $\pi(\beta_n, \beta_0) \le C\|\beta_n - \beta_0\|_\infty$，因此，根据定理 1，$\pi(\widehat{\beta}_n, \beta_0) \xrightarrow{\mathbb{P}} 0$。下面的定理 2 旨在阐明子空间 Θ_{K_n}、惩罚泛函 $\mathcal{J}(\cdot)$ 以及惩罚参数 λ_n 在预测误差收敛至 0 过程中的作用。

定理 2。 假设假设*(A1)–(A5)*均成立。则，

$$\left|\pi(\widehat{\beta}_n, \beta_0)\right|^2 = O_{\mathbb{P}}\left(\frac{K_n \log n}{n}\right) + O_{\mathbb{P}}\left(\left\|\widetilde{\beta}_{K_n} - \beta_0\right\|^2\right) + O_{\mathbb{P}}\left(\widehat{\lambda}_n \mathcal{J}\left(\widetilde{\beta}_{K_n}\right)\right), \quad as\ n \to \infty.$$

定理 2 将预测误差表示为三个量的函数，分别代表方差、平方近似偏差和正则化偏差。这一分解与Kalogridis 和 Van Aelst(2023) 针对函数型线性回归估计量所得的分解相同。然而，正如前文所述，我们关于 β_0 和

X 的有限维分布显著弱于那些作者所作的假设。

详细考察预测误差分解可知，方差项 $K_n \log n/n$仅取决于近似子空间的大小， Θ_{K_n}，而不取决于其类型。这种情况在非参数回归中有一个众所周知的对应现象（Eggermont 和 LaRiccia, 2009，第 15 章）。$\log n$ 项在非参数情形中通常不会出现，而是在此处由于问题性质更加复杂而出现。特别是，它的出现是因为预测变量$\tilde{}$是无限维的 X。与方差不同，近似偏差 $\|\beta_{K_n} - \beta_0\|$ 不仅取决于近似子空间的类型，也取决于 β_0 其自身。为了使近似误差达到良好的衰减速率，我们需要选择一个能够很好地近似 β_0 所属函数类的子空间。例如，如果我们推测 β_0 也是$\hat{}$ $\tilde{}$二阶连续可微的，那么可以选择二阶连续可微函数构成的子空间。最后， $\lambda_n \mathcal{J}(\beta_{K_n})$ 衰减至 0 的方式在本质上取决于将 $\mathcal{J}(\cdot)$ 作为 Θ_{K_n}粗糙度度量的适当性。

作为一个具体例子，考虑这样的情形： $\mathcal{B}([0,1])$是由导数一致有界至 $r \geq 1$ 阶的函数构成的函数空间，且其 r阶导数满足阶数为 $v \in$ 的H¨older条件，其中 $v \in \in (0, 1)$。对于 Θ_{K_n}，我们选取由 K_n 个等距内节点生成的$K_n + p$维 $p > r$ 阶B样条子空间，这些节点位于 $[0,1]$ 中，且 $\mathcal{J}(\beta) = \|\beta^{(q)}\|^2$，其中 $q < p$。$\tilde{}$ $\tilde{}$于是， $\mathcal{J}(\beta_{K_n}) = O(1)$，并且如 de Boor（2001，第149页）所示， $\|\beta_{K_n} - \beta_0\| = O(K_n^{-r-v})$。因此，

$$\left|\pi(\widehat{\beta}_n, \beta_0)\right|^2 = O_{\mathbb{P}}\left(\frac{K_n \log n}{n}\right) + O_{\mathbb{P}}\left(K_n^{-2(r+v)}\right) + O_{\mathbb{P}}\left(\widehat{\lambda}_n\right).$$

对于 $K_n \asymp n^{1/(2(r+v)+1)}$ 和 $\widehat{\lambda}_n = O_{\mathbb{P}}(n^{-c})$，其中 $c \geq 2(r+v)/(2(r+v)+1)$ 可得最优速率$|\pi(\widehat{\beta}_n, \beta_0)|^2 = O_{\mathbb{P}}(n^{-2(r+v)/(2(r+v)+1)}$ 对数 n)。有趣的是，这一收敛速率显著快于 $O_{\mathbb{P}}(n^{-2(r+v)/(4(r+v)+1)})$所得到的速率；Cardot 和 Sarda (2005)为其惩罚样条似然估计量得到了该速率。差异源于我们在结果证明中采用了现代经验过程理论。

4 实践实现

4.1 计算算法

式（4）中定义的惩罚化密度幂散度估计量取决于近似子空间的选择、、其维数、调节参数和惩罚参数。本节将概述一种有效的选择策略，但首先讨论惩罚化密度幂散度估计值的计算。为构建我们的算法，不失一般性地，我们假设惩罚项()可以写成二次型，即可写为() P$\boldsymbol{\beta}$，其中$\mathbf{P}$为某个半正定矩阵， $\boldsymbol{\beta} \in \mathbb{R}^{K_n}$ 为 $\beta \in \Theta_{K_n}$的系数向量。尽管许多惩罚项（例如导数平方积分惩罚项）都满足这一假设，但 l_1型惩罚项并不满足。然而，对于此类惩罚项，我们可以采用计算上更为便利的

局部二次近似，该方法由Fan 和 Li （2004）提出，从而使我们算法的核心思想仍然有效。

对惩罚项进行二次简化后，我们使用修正的费舍尔评分程序求解（4），其更新步骤可按照 Green 和 Silverman （1994）所述的方式，简化为惩罚迭代重加权最小二乘更新。具体而言，本例中的权重为

$$w_i = (1+\widehat{\kappa}_n)\frac{\mathbb{E}_{p_i}\{f_{p_i}^{\widehat{\kappa}_n}(Y_i)|Y_i-p_i|^2|X_i\}}{p_i^2(1-p_i)^2}\left|\frac{\partial p_i}{\partial \eta_i}\right|^2, \quad (i=1,\ldots,n),$$

其中， $\mathbb{E}_{p_i}\{\cdot\}$ 表示相对于成功概率为 "success"$p_i = H(\eta_i)$ 的伯努利密度的期望，线性分量为 $\eta_i = \alpha + \langle X_i, \beta\rangle$， $i = 1$、.....、 n。同时，“工作”数据向量 $\mathbf{z} \in \mathbb{R}^n$ 给定为

$$z_i = \eta_i - (1+\widehat{\kappa}_n)\frac{\partial p_i}{\partial \eta_i}\frac{\mathbb{E}_{p_i}\{f_{p_i}^{\widehat{\kappa}_n}(Y_i)(Y_i-p_i)/(p_i(1-p_i))|X_i\} - f_{p_i}^{\widehat{\kappa}_n}(Y_i)(Y_i-p_i)/(p_i(1-p_i))}{w_i}.$$

当使用“典范”链接 $H(x) = e^x/(1+e^x)$ 时，表达式可进一步简化，此时 $\partial p_i/\partial \eta_i = p_i(1-p_i)$。我们以使用 $\widehat{\kappa}_n = 2$ 获得的密度幂散度估计作为费舍尔评分算法的初始值。对于 $\widehat{\kappa}_n = 0$, 不难看出，上述公式可化简为惩罚似然估计量的公式，如 Green 和 Silverman （1994，第5章）所给出，这再次表明密度幂散度估计是极大似然估计的扩展。

4.2 调节参数与平滑参数的选择

如前所述，在选择 Θ之后，估计量还需要确定 $\widehat{\kappa}_n$、 K_n和 λ_n，它们分别对应调节参数、近似子空间的维数和惩罚参数。在这些参数中， K_n 似乎对估计量能否成功最不关键，因为关于惩罚低秩估计量的大量经验表明，只要近似子空间足够丰富、但其维数仍小于样本量n，维数的选择影响就很小。根据我们的经验，采用诸如 $K = [\min\{30, n/4\}]$这样的选择是合适的，它可确保每个基函数至少对应 4 个观测值，并将基函数数量限制在 30 个以内。在复杂度很高的情况下，可以将基函数数量增加到 30 个以上，但实际中这类情况相当少见。因此，本小节的其余部分将专门讨论 $\widehat{\kappa}_n$ 和 λ_n的选择。

参数 $\widehat{\kappa}_n$ 尤为重要，因为它决定了我们的惩罚幂散度估计量在稳健性与效率之间的权衡，因此正确选择该参数至关重要。与Kalogridis等人(2023类似，我们选择 $\widehat{\kappa}_n$ 的策略是最小化均值积分平方误差（积分均方误差） $\mathbb{E}\{\|\widehat{\beta}_n - \beta_0\|^2\}$的近似值。为清楚地展示

$\widehat{\ }$ $\widehat{\ }$对 β_n 与 κ 之间依赖关系，本文用$\beta_{n,\ \kappa}$表示。标准计算表明，对于每个$\kappa \geq 0$，积分均方误差可分解为

$$\mathrm{MISE}(\kappa) = \left\| \mathbb{E}\{\widehat{\beta}_{n,\kappa}\} - \beta_0 \right\|^2 + \mathbb{E}\left\{ \left\| \widehat{\beta}_{n,\kappa} - \mathbb{E}\{\widehat{\beta}_{n,\kappa}\} \right\|^2 \right\}, \tag{5}$$

其中，右侧第一项表示积分平方偏差，第二项$\widehat{\ }$表示β_n的积分方差κ。由于$\widehat{\ }$ $\widehat{\ }$我们既不了解 β_0 也不了解 $\mathbb{E}\{\beta_{n,\kappa}\}$ （回想一下，β_n是κ 方程（4）的隐式解），因此这些项都无法显式计算。为了近似 MISE(κ中的偏差项，我们采用Warwick 和 Jones (2005)
$\widehat{\ }$ $\widehat{\ }$提出的策略，用其无偏估计量 β_n替代 $\mathbb{E}\{\beta_{n,\kappa}\}$ κ。此外，我们用一个“试点”估计量替代未知的 β_0。数值实验表明，调节参数足够大的任何密度幂散度估计量都能在实践中充当有效的“试点”估计量，$\widehat{\ }$因为它受离群观测的影响较小。因此，我们使用 $\beta_{n,\ 2}$ 作为 β_0的初始估计量。

为近似方差项，首先注意到 $\mathbb{E}\{\|\widehat{\beta}_{n,\kappa} - \mathbb{E}\{\widehat{\beta}_{n,\kappa}\}\|^2\} = \mathrm{Tr}\{\mathbf{P}_0 \mathrm{Cov}\{\widehat{\boldsymbol{\beta}}_{n,\kappa}\}\}$，其中$\mathbf{P}_0$ 由 $\langle \theta_i, \theta_j \rangle$，$i,\ j = \ldots,\ K_n$构成。接下来写作

$$l'_\kappa(Y_i, p_i(\alpha,\beta)) = \left.\frac{\partial l_\kappa(Y_i, x)}{\partial x}\right|_{x=p_i(\alpha,\beta)} \left.\frac{\partial H(x)}{\partial x}\right|_{x=\eta_i}, \quad (i = 1, \ldots, n),$$

对于一阶导数，同理，二阶导数为 $l''_\kappa(Y_i, p_i(\alpha,\beta))$。定义 $n \times n$ 矩阵 $\mathsf{C}_\kappa = \mathrm{diag}\{|l'_\kappa(Y_i, p_i(\widehat{\alpha}_{n,\kappa}, \widehat{\beta}_{n,\kappa}))|^2\}$ 和 $\mathsf{D}_\kappa = \mathrm{diag}\{l''_\kappa(Y_i, p_i(\widehat{\alpha}_{n,\kappa}, \widehat{\beta}_{n,\kappa})\}$。(4) 的得分函数的一阶泰勒展开给出如下近似

$$\mathrm{Cov}\{(\widehat{\alpha}_{n,\kappa}, \widehat{\boldsymbol{\beta}}_{n,\kappa})\} \approx \left[\mathbf{B}^{\star\top}\mathbf{D}_\kappa\mathbf{B}^\star + 2\widehat{\lambda}_{n,\kappa}\mathbf{P}\right]^{-1} \mathbf{B}^{\star\top}\mathbf{C}_\kappa\mathbf{B}^\star \left[\mathbf{B}^{\star\top}\mathbf{D}_\kappa\mathbf{B}^\star + 2\widehat{\lambda}_{n,\kappa}\mathbf{P}^\star\right]^{-1},$$

其中 $\mathbf{B}^\star$有一列全为 1，且第（i,j）个元素为 $\langle X_i, \theta_j \rangle, i = 1, \ldots, n, j = 2 \ldots,$（$K_n + 1$），而$\mathbf{P}^\star$是增加了一行和一列零后的惩罚矩阵 $\mathbf{P}$ 。随后，可以从较大的协方差矩阵 $\mathsf{Cov}\{$（$\widehat{\alpha}_{n,\kappa}$,$\boldsymbol{\beta}_{n,\kappa}$）$\}$中选取相应元素，从而得到$\widehat{\ }\mathsf{Cov}\{\boldsymbol{\beta}_{n,\kappa}\}\}$ 的近似值。

利用上述偏差和方差近似值，我们得到近似均方误差（AMISE），其表达式为

$$\mathrm{AMISE}_2(\kappa) = \left(\widehat{\boldsymbol{\beta}}_{n,\kappa} - \widehat{\boldsymbol{\beta}}_{n,2}\right)^\top \mathbf{P}_0 \left(\widehat{\boldsymbol{\beta}}_{n,\kappa} - \widehat{\boldsymbol{\beta}}_{n,2}\right) + \mathrm{Tr}\left\{\mathbf{P}_0\, \mathrm{Cov}\{\widehat{\boldsymbol{\beta}}_{n,\kappa}\}\right\},$$

$\widehat{\ }$其中，下标 2 表示使用$\beta_{n,\ 2}$ 代替未知的 β_0。首先，我们通过在 $\widehat{\kappa}_n$ 上取 20 个等距值构成网格，并最小化渐近均方积分平方误差 $\mathrm{AMISE}_1(\kappa)$，以此选择 $[0, 2]$。该网格既包含惩罚最大似然估计量（$\kappa = 0$），也包含更稳健的估计量（$\kappa > 0$）。$\widehat{\ }$偏差近似以及随后对 κ 的选择都取决于先导估计量$\beta_{n,\ 2}$。为降低这种依赖性，我们进一步按照Basak 等人(2021年) 的建议，迭代进行选择

^过程。也就是说，在确定使渐近均方积分平方误差 AMISE$_2(\kappa)$ 最小的 κ 值后，我们将 $\beta_{n,\widehat{\kappa}_n}$ 用作新的先导估计量，并最小化 AMISE$_{\widehat{\kappa}_n}(\kappa)$。重复此过程，直至 $\widehat{\kappa}_n$ 收敛。^该过程只需针对不同的 κ 值计算若干个估计量（$\widehat{\alpha}_{n,\kappa}$，$\beta_{n,\kappa}$）。此外，得益于修正的费舍尔评分算法的快速收敛，相关计算负担极小。

对于给定的 κ，计算估计量需要适当的惩罚参数值 λ。为确定 λ，我们提出对Kalogridis等人(2023)所采用的AIC准则进行适当修改，即：

$$\mathrm{AIC}(\lambda) = 2\sum_{i=1}^{n} l_\kappa(Y_i, p_i(\widehat{\alpha}_{n,\kappa}, \widehat{\beta}_{n,k})) + 2\,\mathrm{Tr}\left\{\left[\mathbf{B}^{\star\top}\mathbf{D}_\kappa\mathbf{B}^\star + 2\lambda\mathbf{P}^\star\right]^{-1}\mathbf{B}^{\star\top}\mathbf{D}_\kappa\mathbf{B}^\star\right\}.$$

函数型密度幂散度的实现及示例说明可在R 编程语言（R Core Team，2023）中获得，网址为 https://github.com/ioanniskalogridis/Robust-and-Adaptive-Functional-Logistic-Regression。

5 数值示例

现在，我们通过数值研究考察所提出的惩罚化密度幂散度的实际表现。在数值实验中，我们主要关注评估系数函数 β_0 在各种情形下的表现，这些情形包括 β_0的局部行为、含噪声和离群预测变量 X_i 以及误分类响应 Y_i。为进行比较，我们考虑以下估计量：

- 上一节讨论的自适应密度幂散度估计量，记为DPD($\widehat{\kappa}_n$)。
- 与 $\kappa = 0$对应的自适应密度幂散度估计量，简称最大似然。
- 固定 $\kappa = 1$的密度幂散度估计量，简称 DPD(1)
- 固定 $\kappa = 2$的密度幂散度估计量，简称 DPD(2)。

请注意，自适应估计量 DPD($\widehat{\kappa}_n$) 使用 DPD(2) 作为 β_0 的初始猜测值，而 $\widehat{\kappa}_n$ 的候选值网格包含于 $[0, 2]$中。因此，我们有必要在实践中比较 DPD($\widehat{\kappa}_n$) 与估计量 ML、DPD(1) 和 DPD(2) 的表现。对于所有估计量，我们都使用了常用的三次B样条基，其中 $\mathcal{J}(f) = \|f''\|^2$。此外，在后续所有计算中，我们都使用了典型logit链接函数。

我们根据截断的 Karhunen-Lo`eve 分解生成了预测变量曲线—

其形式如下

$$X(t) = \sum_{j=1}^{50} j^{-1} Z_j \sqrt{2} \sin((j - 1/2)\pi t), \quad t \in [0, 1], \tag{6}$$

其中， Z_j 为相互独立的标准高斯随机变量。为便于实际处理这些曲线，我们在 $[0, 1]$区间内取了200个等距点 t_j 对其进行离散化，并在需要时使用黎曼和计算内积和范数。预测变量分别与以下三个系数函数相结合：

1. $\beta_1(t) = 3(t - 0.3)^2 + 12$. $\beta_2(t) = 3\sin(3.4t^2)$ 3. $\beta_3(t) = -\sin(5t/1.2)/0.5 - 1$.

随后，根据成功概率为 $H(\langle X, \beta \rangle)$的伯努利分布生成二元响应 Y_i ，其中 $H(x) = e^x/(1 + e^x)$。这些系数函数的复杂程度依次增加，因为 β_2 比 β_1 具有更多局部特征；同样， $\beta_3(t)$比 $\beta_2(t)$表现出更强的局部变化。

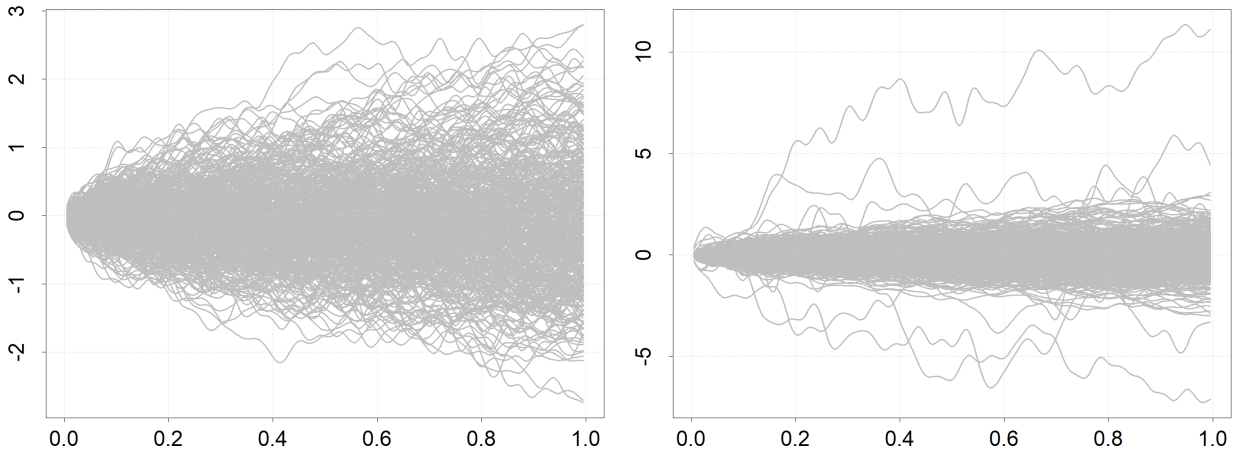

图 1：左侧和右侧分别展示了来自干净数据和受污染（$\epsilon = 0.02$）数据的两组代表性模拟曲线样本。

除了上述干净数据生成方式外，我们还以如下方式扰动了若干观测值（X_i， Y_i）。对于 $\epsilon \in \{0.01, 0.02, 0.05\}$，我们将 X_i中的 ϵ 比例替换为 $5X_i$，并将相应的 Y_i替换为 $1 - Y_i$，从而将 ϵ 比例的响应值从 0 改为 1，或从 1 改为 0。这种污染情景对应于

在数据中引入不良杠杆点；可以先验地预期，所有估计量的表现都会受到这些离群观测的影响。为了更好地理解乘以 5 对曲线形状和振幅的影响，图1 展示了两组具有代表性的曲线样本，其中 $\epsilon = 0$ （干净数据）和 $\epsilon = 0.02$ （轻度污染）。

对于每个系数函数 β 和污染程度 ϵ，我们根据以下均方误差评估各估计量的表现：

$$\text{MSE} = \frac{1}{200} \sum_{j=1}^{200} \left| \widehat{\beta}_n(t_j) - \beta(t_j) \right|^2,$$

它是 $\mathcal{L}^2([0,1])$-距离的一种近似。由于均方误差的下界为零，但上界不受限，因此其分布往往呈右偏且具有厚尾。基于此，我们发现中位数总体上比均值更能衡量集中趋势。因此，下方表1 报告了四种竞争估计量的均方误差中位数（中位数），以及基于 1000 个模拟数据集计算的自助法标准误（自助法标准误），其中 $n = 400$。

		DPD($\widehat{\kappa}_n$)		ML		DPD(1)		DPD(2)	
	ϵ	中位数	BSE	中位数	BSE	中位数	BSE	中位数	BSE
β_1	0	0.389	0.024	**0.385**	0.031	0.482	0.032	0.401	0.021
	0.01	**0.504**	0.027	2.398	0.337	0.726	0.056	0.506	0.025
	0.02	**0.714**	0.041	8.654	1.179	1.263	0.095	0.723	0.047
	0.05	**2.313**	0.139	6.670	0.888	5.327	0.284	2.322	0.142
β_2	0	1.180	0.021	**1.176**	0.025	1.307	0.021	1.195	0.019
	0.01	**1.347**	0.029	5.292	0.810	1.642	0.052	1.358	0.034
	0.02	**1.487**	0.040	20.19	3.736	1.932	0.098	1.505	0.040
	0.05	2.929	0.129	30.69	3.925	14.38	1.908	**2.927**	0.117
β_3	0	**1.038**	0.027	1.040	0.023	1.160	0.038	1.062	0.025
	0.01	**1.064**	0.029	2.520	0.266	1.196	0.045	1.096	0.024
	0.02	**1.278**	0.039	10.43	1.373	1.797	0.095	1.302	0.040
	0.05	**2.915**	0.180	11.15	1.617	8.473	2.057	2.946	0.195

表1：基于 10000 次自助法重复抽样得到的标准误，以及 1000 个大小为 $n = 400$ 的数据集上各竞争估计量均方误差的中位数。最佳中位数表现以粗体标出。

从表1中可以看出几个有趣的现象。显而易见，在没有污染的情况下，惩罚最大似然估计量能够有效估计真实的系数函数，但即使仅有

少量离群值，其表现也会大幅下降。尤其是，即使在污染程度最低时，惩罚最大似然估计量的表现也会下降至原来的二分之一甚至三分之一。基于最大似然的方法缺乏抗干扰能力，这与具有正调节参数的密度幂散度估计量的稳定性形成鲜明对比。在干净数据中，这些估计量的表现几乎与基于最大似然的方法相当；而在 $\epsilon \in \{0.01, 0.02\}$时，表现几乎没有下降。更高程度的污染确实会影响密度幂散度估计量的表现，但显然影响程度不及基于最大似然的方法。为直观展示基于最大似然的方法与密度幂散度估计量在实际表现上的差异，图2展示了干净数据和轻度污染数据下，ML 和 DPD($\beta_1(t)$ 方法所得的 1000 个估计值，分别列于第一行和第二行。$\widehat{\kappa}_n$)

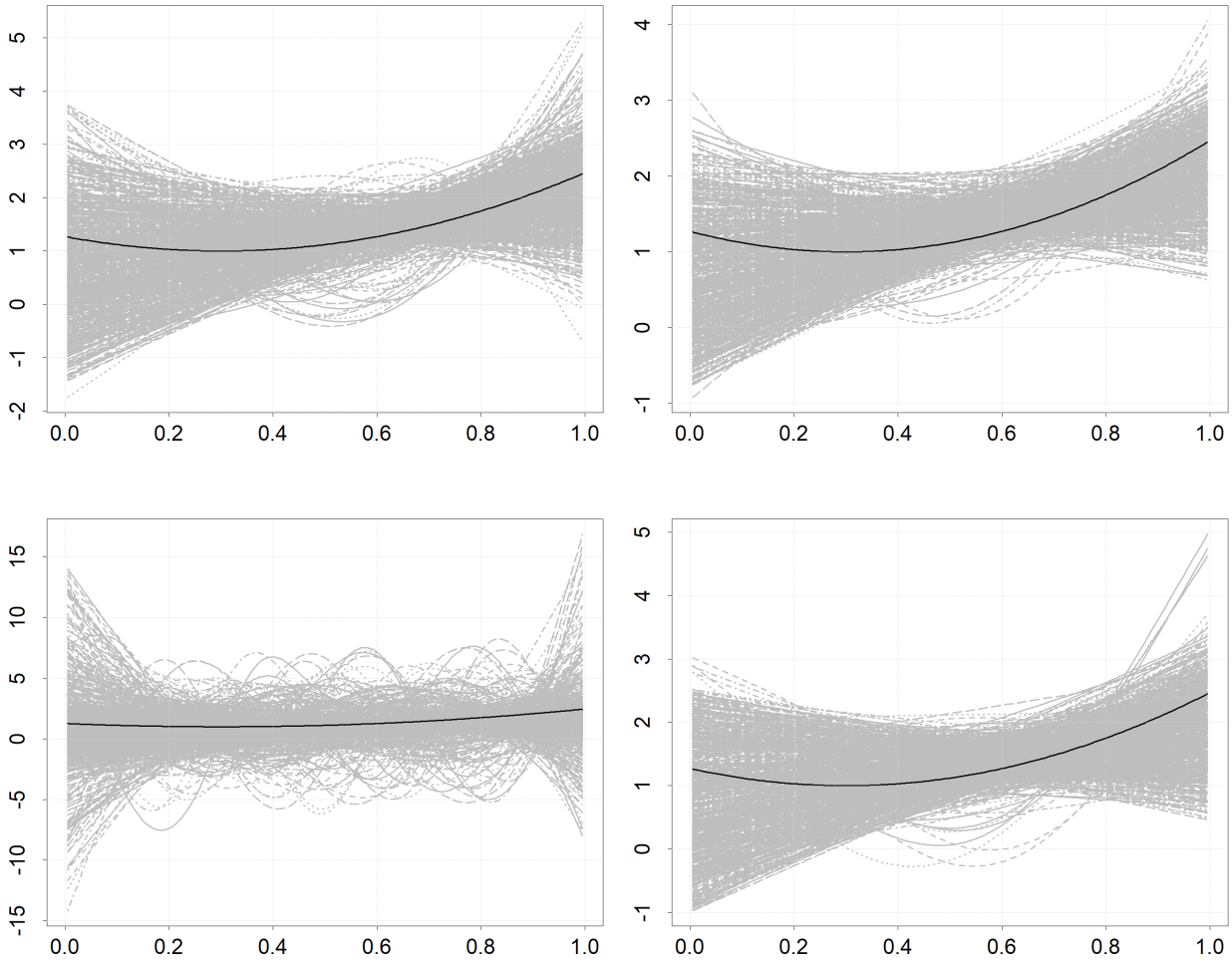

图2：上排：左图和右图分别显示 1000 个最大似然估计值和 DPD($\widehat{\kappa}_n$) 估计值，对应于 $\beta_1(t)$，其中 $\epsilon = 0$ 。下排：左图和右图分别显示 1000 个最大似然估计值和 DPD($\widehat{\kappa}_n$) 估计值，对应于 $\beta_1(t)$，其中 $\epsilon = 0.01$ 。黑色实线表示真实的系数函数 $\beta_1(t)$。

详细比较稳健估计量 DPD($\widehat{\kappa}_n$)、DPD(1) 和 DPD(2) 可以发现，DPD($\widehat{\kappa}_n$) 估计量具有自适应性，因此在干净数据中比其他方法效率更高；而当数据中存在离群观测时，也不会牺牲稳健性。事实上，在污染程度较高时，DPD($\widehat{\kappa}_n$) 的表现与 DPD(2) 相近，且显著优于稳健性较差的 DPD(1) 估计量。这种优异表现的原因可能在于 DPD($\widehat{\kappa}_n$) 估计量的构造方式：它能确保以一种使均值积分平方误差近似最小化的方式选择调节参数。在存在污染时，这一准则往往会得到较大的 κ 值，而在干净数据中则会得到较小的 κ 值，从而确保 DPD($\widehat{\kappa}_n$) 估计量整体表现良好。

6 应用：步态分析

人们通常会关注一个人是否正常行走，或其步行是否因某种潜在原因而受阻，例如腹部疼痛或脚趾受伤。本数据集包含166个观测，记录了个体进行数秒步态周期时的情况；在此期间，运动传感器在343个点记录了脚趾在垂直轴上的位置。这些轨迹绘于图3左侧面板中。每条轨迹还附有一个指示变量，其取值为1或0，分别表示步行是否正常。本研究旨在根据步态周期内脚趾的位置，确定正常步行的概率。该数据集以“ToeSegmentation2”为名，免费收录于UCR时间序列分类档案（Dau等人，2018）。

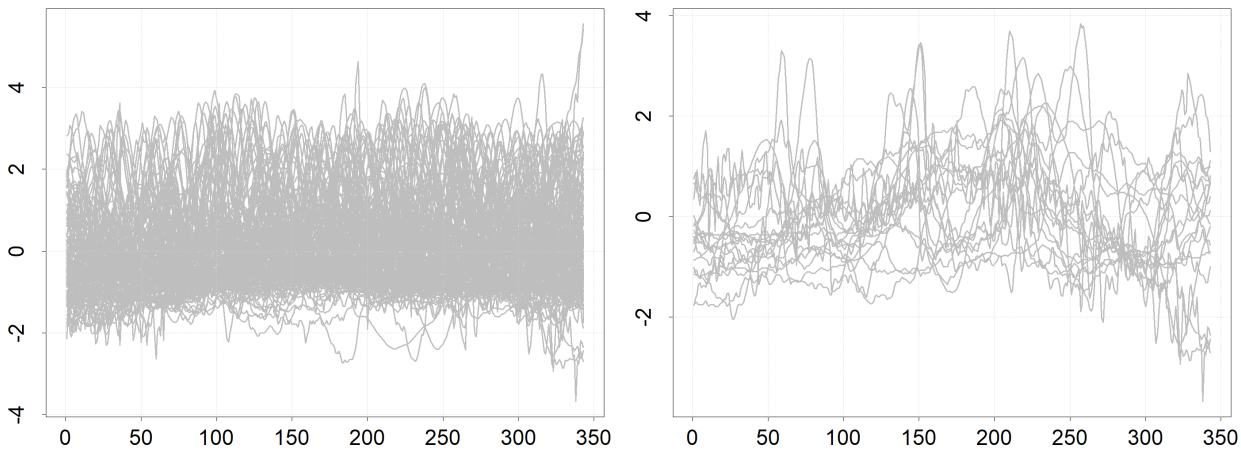

图 3：左：166 名受试者在一个步态周期内脚趾沿垂直轴的位置。右：根据 DPD($\widehat{\kappa}_n$) 估计值，Anscombe残差较大的观测值轨迹。

在大多数真实数据情境中，个体间的生理差异可能会导致出现若干离群观测，其表现形式为垂直离群值，也就是说，这些观测与模型并不完全吻合。在我们的案例中，这些观测可能来自步态模式正常、但感到疼痛或不适的个体，也可能来自步行异常、但除此之外身体健康的个体。为了展示此类观测对非稳健估计量的影响，我们计算了Cardot和Sarda（2005）提出的基于惩罚B样条、采用典范logit链接的最大似然估计量，以及与之类似的自适应密度幂散度估计量。

系数函数估计值分别以实线蓝色和虚线红色绘制于图4 的左侧面板中。该面板立即显示出这两种估计值之间的显著差异：DPD($\widehat{\kappa}_n$)估计值呈现出多个峰值，其中最明显的出现在周期内第100点和第300点附近；而最大似然估计值则基本形成一条直线，这表明存在严重的过度平滑。因此，最大似然估计量无法识别周期中那些对判定步态是否正常具有重要影响的点。鉴于前文讨论，我们可以推测，若干离群观测对最大似然估计值本身以及所选惩罚参数都产生了过度影响；在本例中，惩罚参数是以非稳健的方式选定的。

为了确定是否确有某些观测对ML估计产生了重要影响，我们可以查看稳健的DPD($\widehat{\kappa}_n$)估计值所产生的残差。第4 节所述算法选取了 $\widehat{\kappa}_n = 1.9$ ，这表明存在若干离群观测。为检测这些观测，我们可以查看Anscombe残差（McCullagh和Nelder，1983，第29页）；对于伯努利分布，其定义为

$$r_{A,i} = \frac{\mathrm{IB}(Y_i, 2/3, 2/3) - \mathrm{IB}(\widehat{\mu}_i, 2/3, 2/3)}{\widehat{\mu}_i^{1/6}(1-\widehat{\mu}_i)^{1/6}}, \quad (i = 1, \ldots, n),$$

其中 $\mathrm{IB}(x,\ a,\ b) = \int_0^x t^{a-1}(1-t)^{b-1}dt$ 且 $\widehat{\mu}_i$ 是 i第个拟合值，即 $\widehat{\mu}_i = H(\widehat{\alpha}_n + \int_1^{343} X_i(t)\,\widehat{\beta}_n(t)dt)$。与皮尔逊残差相比，Anscombe残差更接近标准高斯分布，因此我们可以采用传统的离群值检测规则。具体而言，如果$|r_{A,\ i}| \geq 2$，则可将某个观测判定为离群值。根据这一规则，DPD($\widehat{\kappa}_n$)估计值指向19个离群观测，其中13个的Anscombe残差大于10，因此可视为极端离群值。图4右侧面板绘出了检测到的19个离群观测的轨迹。有趣的是，其中18个观测被分类为正常步行，这表明在绝大多数情况下，被误分类为正常步行的是存在潜在问题的个体，而非相反情况。

作为敏感性检验，可以选择从数据中移除检测到的离群观测，并重新计算估计值。这样得到修订后的估计值，如

图 4右侧面板所示。该面板显示，尽管 DPD($\widehat{\kappa}_n$) 估计值几乎没有变化，ML 估计值却发生了显著变化，现在与 DPD($\widehat{\kappa}_n$) 估计值完全相同。因此，我们可以得出结论：离群值的存在对非稳健ML估计产生了负面影响，却没有影响稳健性高得多的幂密度散度估计；无论是否存在离群观测，这些估计都能提供可靠结果。此外，由于这些估计值不易被离群观测拉偏，也可用于离群值检测，从而识别出基于ML的估计可能会遗漏的有影响观测。

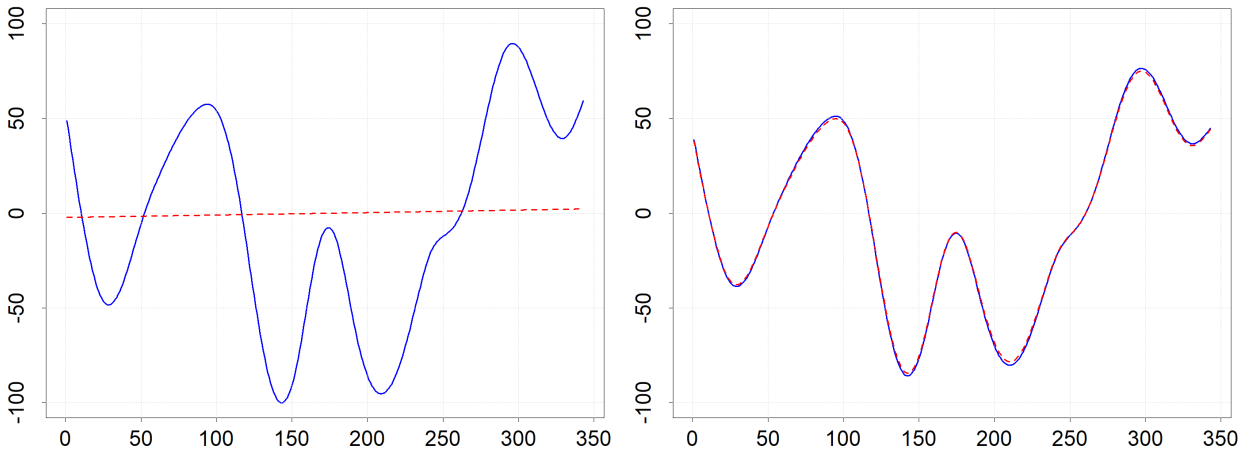

图4：左：基于完整数据集，分别使用 DPD($\widehat{\kappa}_n$) 和 ML 估计量得到的系数函数估计，以蓝色实线和红色虚线表示。右：根据 DPD($\widehat{\kappa}_n$) 估计值，移除 Anscombe残差较大的观测后得到的系数函数估计。

7 结论

本文为函数型逻辑回归模型提出了一类估计量，作为常用的基于最大似然的估计量的一种更加均衡的替代方案。后者缺乏稳健性，因而尤其容易受到离群观测的影响。通过适当调节，所提出的估计量族既能在干净数据上实现高效率，又能抵御离群观测的影响。此外，我们还提出了一种自动选择调节参数的算法，使应用研究者只需选择近似子空间和惩罚泛函。所提出的估计量具有良好的理论性质：它们一致收敛，即使调节参数是随机的，也能针对预测误差达到较高的收敛速率。这些性质使所提出的方法尤其适用于分析如今常见的许多复杂数据集，

例如第6节中的步态数据集。

未来的研究中，我们希望探讨若干有趣且具有实际应用价值的扩展，其中最值得关注的是将方法推广至其他广义线性模型（GLMs）。尽管基于惩罚化密度幂散度的方法不受响应变量所服从的指数族分布类型限制，可以直接应用，但本文建立的理论性质仅适用于伯努利分布，无法直接推广至其他广义线性模型。在适当条件下，这些理论性质很可能仍然成立；我们认为，探究此类条件是值得开展的工作。我们的研究还可扩展至更灵活的指数族分布，其中包含尺度参数，例如广义伽马分布和对数正态分布；这些分布常用于对保险业的理赔进行建模。鉴于此类数据集中经常存在离群观测，采用稳健方法是恰当的。我们确信，惩罚化密度幂散度同样能够得到计算上可行且具有良好理论性质的估计量。

致谢

作者衷心感谢 Research Foundation-Flanders 的资助（项目 1221122N）。

8 附录：理论结果的证明

引理1（密度幂散度的界）。 设 h 和 g 为伯努利密度，其“成功”概率分别为 p_h 和 p_g 。对于每个 $\kappa_0 \in (0,1]$ ，我们有

$$0 \leq d_{\kappa_0}(h,f) \leq 1 + \frac{1}{\kappa_0},$$

且当且仅当 $|p_f - p_h| = 1$ 时，上界取等号。

证明。 下界由算术—几何平均（AM-GM）不等式得出，因为密度非负，且对于每个 $y \in \{0,1\}$ ，我们有

$$h(y)f^{\kappa_0}(y) = \left\{h^{1+\kappa_0}(y)\right\}^{\frac{1}{1+\kappa_0}} \left\{f^{1+\kappa_0}(y)\right\}^{\frac{\kappa_0}{1+\kappa_0}} \leq \frac{1}{1+\kappa_0}h^{1+\kappa_0}(y) + \frac{\kappa_0}{1+\kappa_0}f^{1+\kappa_0}(y).$$

除非 $h = f$ 在 $\{0,1\}$上，否则不等式为严格不等式。两边乘以 $(\kappa_0+1)/\kappa_0 > 0$ ，即可得到所需结果。对于上界，写作

$$d_{\kappa_0}(h,f) = \sum_{y\in\{0,1\}} \{f(y) - h(y)\}f^{\kappa_0}(y) + \frac{1}{\kappa_0}\sum_{y\in\{0,1\}} \{h^{\kappa_0}(y) - f^{\kappa_0}(y)\}h(y),$$

并注意到

$$
\begin{aligned}
d_{\alpha_0}(h,f) &\le |p_h - p_f| \left| (1-p_f)^{\kappa_0} - p_f^{\kappa_0} \right| + \frac{1}{\kappa_0} \sum_{y\in\{0,1\}} |h^{\kappa_0}(y) - f^{\kappa_0}(y)|\, h(y) \\
&\le 1 + \frac{1}{\kappa_0},
\end{aligned}
$$

根据反向三角不等式以及 $|p_h - p_f| \le 1$ 和 $|h^{\kappa_0}(y) - f^{\kappa_0}(y)| \le 1$ 这些事实。为验证当 $|p_f - p_h| = 1$ 时上界可取到，注意此时必有 $p_f = 0$ 和 $p_h = 1$，或 $p_f = 1$ 和 $p_h = 1$。简单代数运算表明，在这两种情况下，上述不等式都取等号，反向蕴含也成立。

□

引理 2（费舍尔一致性）。对于任意 $\kappa_0 > 0$ 定义

$$
M(\beta, \kappa_0) = \mathbb{E}\{d_{\kappa_0}(f_{p_1(\beta_0)}, f_{p_1(\beta)})\}, \quad g \in \mathcal{B}([0,1]),
$$

其中 β_0 表示真实的系数函数。在假设 *(A2)* 和 *(A5)* 下，$M(\beta, \kappa_0) \ge 0$ 且 $M(\beta, \kappa_0) = 0$ 当且仅当 $\beta = \beta_0$。

证明。首先注意，根据引理 1，$M(\beta, \kappa_0) \ge 0$ 对于所有 $\beta \in \mathcal{B}([0,1])$，因为期望是单调运算。对于 $\beta \ne \beta_0$ 令 $A_\beta = \{\omega \in \Omega\colon \langle X_1(\omega)\ \beta - \beta_0\rangle = 0\}$。由于映射 $X_1\colon \Omega \to \mathcal{L}^2([0,1])$ 的可测性，我们有 $A_\beta \in \mathcal{A}$，并且根据 (A2)，$\mathbb{P}(A_\beta^c) > 0$。因此，

$$
\begin{aligned}
M(\beta,\kappa_0) &= \int_\Omega d_{\kappa_0}(f_{p_1(\beta_0)}, f_{p_1(\beta)}) d\mathbb{P} \\
&= \int_{A_\beta} d_{\kappa_0}(f_{p_1(\beta_0)}, f_{p_1(\beta)}) d\mathbb{P} + \int_{A_\beta^c} d_{\kappa_0}(f_{p_1(\beta_0)}, f_{\beta_1(g)}) d\mathbb{P} \\
&= \int_{A_\beta^c} d_{\kappa_0}(f_{p_1(\beta_0)}, f_{p_1(\beta)}) d\mathbb{P} \\
&> 0,
\end{aligned}
$$

因为根据（A2），$d_{\kappa_0}(f_{p_1(\beta_0)}\ f_{p_1(\beta)})$ 在 A_β 上消失，且 $d_{\kappa_0}(f_{p_1(\beta_0)}\ f_{p_1(\beta)}) > 0$ 在 A_β^c 上。然而，由于 $\mathbb{P}(A_\beta^c) > 0$，在 A_β^c 上积分会保持严格不等式。证明完毕。

□

对于每个固定的 $\kappa_0 > 0$ 和 $\beta \in \mathcal{L}^2([0,1])$，令 $M_n(\beta, \kappa_0)$ 表示 $M(\beta, \kappa_0)$ 的经验版本，即

$$
M_n(\beta,\kappa_0) = \frac{1}{n}\sum_{i=1}^{n} \left[\sum_{y\in\{0,1\}} f_{p_i(\beta)}^{1+\kappa_0}(y) - \left(1 + \frac{1}{\kappa_0}\right) f_{p_i(\beta)}^{\kappa_0}(Y_i) + \frac{1}{\kappa_0} \sum_{y\in\{0,1\}} f_{p_i(\beta_0)}^{1+\kappa_0}(y) \right] \tag{7}
$$

注意， $M_n(\beta, \kappa_0)$ 是有界独立同分布随机变量的平均值，因此，根据大数弱定律（弱大数定律），$M_n(\beta, \kappa_0) \xrightarrow{\mathbb{P}} M(\beta, \kappa_0)$。下面的引理 3 给出了一个一致定律，据此 $M_n(\beta, \kappa_0) - M(\beta, \kappa_0) \xrightarrow{\mathbb{P}} 0$ 在 $\beta \in \Theta_{K_n}$上均匀成立，当 $n \to \infty$。

引理 3. 假设（*A2*）和（*A4*）成立。那么，对每个 $0 < \kappa_0 < \infty$,

$$\sup_{\beta \in \Theta_{K_n}} |M_n(\beta, \kappa_0) - M(\beta, \kappa_0)| \xrightarrow{\mathbb{P}} 0.$$

证明。根据三角不等式和弱大数定律，只需证明由下式给出的函数类在 $\mathcal{L}^2([0,1]) \times \{0,1\}$ 上一致收敛:

$$\mathcal{F}_n = \left\{ f^{\kappa_0}_{H(\langle x,\beta \rangle)}(y), \beta \in \Theta_{K_n} \right\}$$
$$\mathcal{F}'_n = \left\{ \sum_{y \in \{0,1\}} f^{1+\kappa_0}_{H(\langle x,\beta \rangle)}(y), \beta \in \Theta_{K_n} \right\}.$$

这里，与第 2 节 $f_{H(\langle x,\beta \rangle)}(y$ 中一样，表示成功概率为 $H(\langle x, \beta \rangle)$ 的伯努利密度。令 $\mathbb{P}_n$ 表示（X_i, Y_i），...，（X_n，Y_n）的经验测度。我们首先证明，$\mathcal{F}_n$ 是一个 Glivenko-Cantelli 函数类（van derVaart，1998，第 269 页），因此

$$\sup_{\beta \in \Theta_{K_n}} \left| \int f^{\kappa_0}_{H(\langle x,\beta \rangle)}(y) d\mathbb{P}_n - \int f^{\kappa_0}_{H(\langle x,\beta \rangle)}(y) d\mathbb{P} \right| \xrightarrow{\mathbb{P}} 0, \tag{8}$$

为证明（8），我们推导 ϵ-覆盖数的上界，该覆盖数对应于 $\mathcal{F}_n$ 在 $L_1(\mathbb{P}_n)$-范数下的覆盖数，$N_1(\epsilon, \mathcal{F}_n, \mathbb{P}_n)$。注意，$N_1(\epsilon, \mathcal{F}_n, \mathbb{P}_n)$ 是随机的，因为它取决于经验测度 $\mathbb{P}_n$。为界定该覆盖数，考虑由下式给出的 $\mathcal{L}^2([0,1])$ 上的实值函数集合:

$$\mathcal{G}_n = \{\langle x, \beta \rangle, \beta \in \Theta_{K_n}\}.$$

显然，$\mathcal{G}_n$ 至多是一个 K_n维函数空间 $\mathcal{L}^2([0,1])$，因此是一个 Vapnik-Chervonenkis（VC）函数类，其 VC 指数不超过 $K_n + 2$，参见例如（van der Vaart，1998，第275–276页）。函数类 $H(\mathcal{G}_n)$ 根据（A2）中假设的 H 单调性，也具有相同的 VC 指数，参见（Kosorok，2008）中的引理 9.9（viii）。此外，函数类 $H(\mathcal{G}_n)$ 的包络函数为1，因此，根据van der Vaart(1998）中的引理 19.15，存在一个通用常数 C 使得

$$\log(N_1(\epsilon, H(\mathcal{G}_n), \mathbb{P}_n)) \leq C(K_n + 1) \log\left(\frac{1}{\epsilon}\right),$$

对所有足够小的 $\epsilon > 0$。接下来注意，函数 $f_p(y) = p^y(1-p)^{1-y}$，$y \in \{0,1\}$ 关于 p 是利普希茨连续的，其利普希茨常数不超过 1，因此同样的界也适用于

定义在 $\mathcal{L}^2([0,1])$ 上的函数集合 $\times\{0,1\}$ 由 $\{f_{H(\langle x,\beta\rangle)}(y)\ \beta\in\Theta_{K_n}\}$给出。要据此推导 $\mathcal{F}_n$ 在 L_1 ($\mathbb{P}_n$-范数下的覆盖数界，只需注意函数 $x\mapsto x^{\kappa_0}$是递增的，并且对所有 $0\le x\le 1$均有界于 1。因此，再次由 van der Vaart 的引理19.15（1998），

$$\log(N_1(\epsilon,\mathcal{F}_n,\mathbb{P}_n))\le C(K_n+1)\log\left(\frac{1}{\epsilon}\right),$$

对于较小的 $\epsilon>0$。由（A4）可知，对于这样的 $\epsilon>0$，$\log(N_1(\epsilon,\mathcal{F})_n,\mathbb{P}_n))/n\xrightarrow{\mathbb{P}}0$。（8）中的一致收敛现在可由 van de Geer (2000，习题 3.6）得到，其中在该结果中取 $b_n=2$；这是因为 $\mathcal{F}$ 中的函数对所有 $\beta\in\Theta_{K_n}$均以 1 为上界。

为完成证明，我们还需要类似地证明函数集 $\mathcal{F}'$ 是 Glivenko-Cantelli 类。为此，我们可以分别考虑 $y=0$ 和 $y=1$ 这两种情况。证明可沿类似思路进行：首先给出 $H(\mathcal{G}_n)$ 的覆盖数上界，然后注意到函数$x\mapsto x^{k_0}$ 和 $x\mapsto(1-x)^{k_0}$ 对 $0\le x\le 1$单调且有界。因此，存在一个 $C>0$ 使得

$$\log(N_1(\epsilon,\mathcal{F}'_n,\mathbb{P}_n))\le C(K_n+1)\log\left(\frac{1}{\epsilon}\right),$$

对所有足够小的 $\epsilon>0$。根据（A4），再次通过引用 van de Geer (2000，习题 3.6)，即可得出该结果。 □

引理 4. 假设假设*(A1)–(A5)*成立。那么，对于每个 $\kappa_0>0$ ，

$$M(\widehat{\beta}_n,\kappa_0)\xrightarrow{\mathbb{P}}M(\beta_0,\kappa_0).$$

^证明。 根据引理 1, β_0 在所有 $\beta\in\mathcal{B}([0,1])$ 中唯一地使 $M(\beta,\kappa_0)$ 达到最小值。由于对于所有 $n\in\mathbb{N}$， $\beta_n\in\mathcal{B}([0,1])\subset\mathcal{L}^2([0,1])$，显然有

$$0<M(\widehat{\beta}_n,\kappa_0)-M(\beta_0,\kappa_0)=I+II+III, \tag{9}$$

其中

$$I=M(\widehat{\beta}_n,\kappa_0)-M_n(\widehat{\beta}_n,\kappa_0),\ II=M_n(\widehat{\beta}_n,\kappa_0)-M_n(\widehat{\beta}_n,\widehat{\kappa}_n),\ III=M_n(\widehat{\beta}_n,\widehat{\kappa}_n)-M(g_0,\kappa_0).$$

我们的证明只需说明右侧（9）中的每一项都依概率收敛到 0。^ ^从 I开始，根据 β_n 的定义，我们有 $\beta_n\in\Theta_{K_n}$，因此，由引理 2，

$$|I|\le\sup_{\beta\in\Theta_{K_n}}|M_n(\beta,\kappa_0)-M(\beta,\kappa_0)|\xrightarrow{\mathbb{P}}0.$$

为证明 II 收敛到 0，注意到，根据（A1），我们可以假设对所有充分大的 n，以高概率有 $\widehat{\kappa}_n \in (\kappa_0/2,\ 3\kappa_0/2)$。根据中值定理，对于每个 $x \in \mathcal{L}^2([0,1])$ 和 $y \in \{0,1\}$，存在一个 $\widetilde{\kappa}_n = \widetilde{\kappa}_n(x, y)$，使得 $|\widetilde{\kappa}_n - \kappa_0| \le |\widehat{\kappa}_n - \kappa_0| < \kappa_0/2$ 且

$$f^{\widehat{\kappa}_n}_{H(\langle x,\widehat{\beta}_n\rangle)}(y) - f^{\kappa_0}_{H(\langle x,\widehat{\beta}_n\rangle)}(y) = (\widehat{\kappa}_n - \kappa_0) f^{\widetilde{\kappa}_n}_{H(\langle x,\widehat{\beta}_n\rangle)}(y) \log\left(f_{H(\langle x,\widehat{\beta}_n\rangle)}(y)\right).$$

Now,

$$\begin{aligned}
\left|f^{\widetilde{\kappa}_n}_{H(\langle x,\widehat{\beta}_n\rangle)}(y) \log\left(f_{H(\langle x,\widehat{\beta}_n\rangle)}(y)\right)\right| &\le \sup_{|\kappa-\kappa_0|<\kappa_0/2,\ p\in[0,1]} \left|f^{\kappa}_{p}(y) \log\left(f_p(y)\right)\right| \\
&\le \sup_{|\kappa-\kappa_0|<\kappa_0/2,\ p\in[0,1]} (1-p)^{\kappa} \left|\log(1-p)\right| + \sup_{|\kappa-\kappa_0|<\kappa_0/2,\ p\in[0,1]} p^{\kappa} \left|\log(p)\right| \\
&\le \sup_{p\in[0,1]} (1-p)^{\kappa_0/2} \left|\log(1-p)\right| + \sup_{p\in[0,1]} p^{3\kappa_0/2} \left|\log(p)\right| \\
&\le M,
\end{aligned}$$

对某个有限的 $D = D(\kappa_0)$。结合以下事实应用这一点：对于每个 $p \in [0,1]$，$f_p(y) \le 1$ and $\sum_{y\in\{0,1\}} f_p(y) = 1$，可得

$$\begin{aligned}
|II| &\le \frac{1}{n}\sum_{i=1}^{n} \left[\sum_{y\in\{0,1\}} \left|f^{\widehat{\kappa}_n}_{H(\langle x,\widehat{\beta}_n\rangle)}(y) - f^{\kappa_0}_{H(\langle x,\widehat{\beta}_n\rangle)}(y)\right| f_{H(\langle x,\widehat{\beta}_n\rangle)}(y) + \frac{|\widehat{\kappa}_n - \kappa_0|}{\kappa_0 \widehat{\kappa}_n} f^{\widehat{\kappa}_n}_{H(\langle X_i,\widehat{\beta}_n\rangle)}(Y_i)\right] \\
&\quad + \left(1 + \frac{1}{\widehat{\kappa}_n}\right) \frac{1}{n}\sum_{i=1}^{n} \left|f^{\widehat{\kappa}_n}_{H(\langle X_i,\widehat{\beta}_n\rangle)}(Y_i) - f^{\kappa_0}_{H(\langle X_i,\widehat{\beta}_n\rangle)}(Y_i)\right| \\
&\le D\left|\widehat{\kappa}_n - \kappa_0\right| + \frac{|\widehat{\kappa}_n - \kappa_0|}{(\kappa_0 - \delta)\kappa_0} + \left(1 + \frac{1}{\kappa_0 - \delta}\right) D\left|\widehat{\kappa}_n - \kappa_0\right|,
\end{aligned}$$

以高概率，对所有充分大的 n 均成立。由此得出 $II \xrightarrow{\mathbb{P}} 0$。

为完成证明，我们现在证明 $III \xrightarrow{\mathbb{P}} 0$。为此，首先回顾一下，$\widehat{\beta}_n$ 在 $\beta \in \Theta_{K_n}$ 上最小化 $\sim M_n(\beta,\ \widehat{\kappa}_n) + \lambda_n \mathcal{J}(\beta)$，而 β_0 在 Θ_{K_n}，β_{K_n} 中的抽象近似满足 $\sim \beta_{K_n} \in \Theta_{K_n}$。因此，根据映射 $\beta \mapsto \lambda_n \mathcal{J}(\beta)$ 的非负性以及（A4），我们得到

$$\begin{aligned}
III &\le M_n(\widetilde{\beta}_{K_n}, \widehat{\kappa}_n) + \widehat{\lambda}_n \mathcal{J}(\widetilde{\beta}_{K_n}) - M(\beta_0, \kappa_0) \\
&= M_n(\widetilde{\beta}_{K_n}, \widehat{\kappa}_n) - M(\beta_0, \kappa_0) + o_{\mathbb{P}}(1) \\
&= \left\{M_n(\widetilde{\beta}_{K_n}, \widehat{\kappa}_n) - M_n(\beta_0, \kappa_0)\right\} + \left\{M_n(\beta_0, \kappa_0) - M(\beta_0, \kappa_0)\right\} + o_{\mathbb{P}}(1). \qquad (10)
\end{aligned}$$

由弱大数定律，$M_n(\beta_0,\ \kappa_0) \xrightarrow{\mathbb{P}} M(\beta_0,\ \kappa_0)$。为了证明右侧第一项的收敛，

(10) 至 0，注意到，根据三角不等式，

$$\left|M_n(\widetilde{\beta}_{K_n},\widehat{\kappa}_n)-M_n(\beta_0,\kappa_0)\right|\leq\left|M_n(\widetilde{\beta}_{K_n},\widehat{\kappa}_n)-M_n(\widetilde{\beta}_{K_n},\kappa_0)\right|+\left|M_n(\widetilde{\beta}_{K_n},\kappa_0)-M_n(\beta_0,\kappa_0)\right| \quad (11)$$

我们先前在处理 II 时采用的论证立即表明，式（11）右侧的第一项为 $o_{\mathbb{P}}(1)$，因为 $n\to\infty$。此外，对于每个固定的 $\kappa_0>0$，映射$[0,1]\to[0,1]:\ p\mapsto f_p^{\kappa_0}(y)\ y\in\{0,1\}$，是利普希茨连续的。因此，存在有限的 $D'=D'(\kappa_0)$，使得

$$\left|f^{\kappa_0}_{H(\langle X_i,\widetilde{\beta}_{K_n}\rangle)}(y)-f^{\kappa_0}_{H(\langle X_i,\beta_0\rangle)}(y)\right|\leq D'(\kappa_0)\left|H(\langle X_i,\widetilde{\beta}_{K_n}\rangle)-H(\langle X_i,\beta_0\rangle)\right|.$$

利用中值定理、（A2）和（A5），我们得到

$$\left|f^{\kappa_0}_{H(\langle X_i,\widetilde{\beta}_{K_n}\rangle)}(y)-f^{\kappa_0}_{H(\langle X_i,\beta_0\rangle)}(y)\right|\leq CD'(\kappa_0)\sup_{x\in\mathbb{R}}H'(x)\left\|\widetilde{\beta}_{K_n}-\beta_0\right\|,$$

由此可得

$$\left|M_n(\widetilde{\beta}_{K_n},\kappa_0)-M_n(\beta_0,\kappa_0)\right|\leq c_0\left\|\widetilde{\beta}_{K_n}-\beta_0\right\|,$$

对于某个 $c_0>0$。根据（A4），$\|\widetilde{\beta}_{K_n}-\beta_0\|\to 0$当 $K_n\to\infty$因此 $M_n(\widetilde{\beta}_{K_n},\ \kappa_0)-M_n(\beta_0,\ \kappa_0)\xrightarrow{\mathbb{P}}0$。因此，我们已证明（11）右侧依概率收敛于 0。考虑到 I 和 II 收敛至 0，我们看到

$$o_{\mathbb{P}}(1)<III\leq o_{\mathbb{P}}(1).$$

因此，$III=o_{\mathbb{P}}(1)$，证明完毕。 □

定理 1 的证明。证明依赖于引理2和引理4，以及 Kosorok（2008）中的引理14.3和定理2.12。首先注意，根据引理 2，$M(\beta,\kappa_0)$ 在 β_0 处取得唯一最小值；又由连接函数 H 的连续性、引理1、(A5) 和控制收敛定理可知，$M(\beta,\kappa_0)$ 在 $\mathcal{B}([0,1])$ 上连续。根据 Kosorok（2008）的引理14.3可知，对于任意序列 $\beta_k\in\mathcal{B}([0,1])$，若 $\lim_{k\to\infty}M(\beta_k,\kappa_0)=M(\beta_0,\kappa_0)$，则有 $\|\beta_k-\beta_0\|_{\mathcal{B}}\to 0$。根据 Kosorok（2008）的习题14.6.2，存在一个递增的右连续且有左极限函数 $f\colon[0,\infty]\to[0,\infty]$，满足 $f(0)=0$，在 0 处连续，并且对所有 $\beta\in\mathcal{B}([0,1])$，都有 $\|\beta-\beta_0\|_{\mathcal{B}}\leq f(|M(\beta,\kappa_0)-M(\beta_0,\kappa_0)|)$。注意到，对于所有$n\in\mathbb{N}$，我们的估计量 $\beta_n\in\mathcal{B}([0,1])$，可得

$$\left\|\widehat{\beta}_n-\beta_0\right\|_{\mathcal{B}}\leq f\left(\left|M(\widehat{\beta}_n,\kappa_0)-M(\beta_0,\kappa_0)\right|\right). \quad (12)$$

根据引理 4， $|M(\widehat{\beta}_n,\ \kappa_0) - M(\beta_0,\ \kappa_0)| \xrightarrow{\mathbb{P}} 0$。因此，由 f在 0 处的连续性以及 $f(0) = 0$这一事实，(12) 可推出 $\|\widehat{\beta}_n - \beta_0\|_{\mathcal{B}} \xrightarrow{\mathbb{P}} 0$，证毕。 □

现在引入一些有用的记号，以便证明定理 2。设 $\mathcal{G}$表示定义在 $\mathcal{L}^2([0,1]) \times \mathbb{R}$上的一类实值函数。对于 $g \in \mathcal{G}$ 我们定义

$$\|g\|_\infty = \sup_{x \in \mathcal{L}^2([0,1]), y \in \mathbb{R}} |g(x,y)|.$$

在该一致度量下的覆盖数， $N_\infty(\epsilon, \mathcal{G})$，定义为满足以下条件的序列 $\{g_j\}_{j=1}^N$ 存在时， $N \in \mathbb{N}$的最小值：

$$\sup_{g \in \mathcal{G}} \min_{j=1,\ldots,N} \|g - g_j\|_\infty \le \epsilon.$$

相应的熵， $H_\infty(\epsilon, \mathcal{G})$，定义为覆盖数的对数，即$H_\infty(\epsilon, \mathcal{G}) = \log(N_\infty(\epsilon, \mathcal{G}))$。

对于概率测度 $\mathbb{P}$，我们还定义 $\mathcal{L}^2(\mathbb{P})$-度量下的括号数， $N_B(\epsilon, \mathcal{G}, \mathbb{P})$，为满足以下条件的 $N \in \mathbb{N}$ 的最小值：存在 N 对函数对 $\{[g_j^L, g_j^U]\}$，使得$\|g_j^U - g_j^L\|_{\mathcal{L}^2(\mathbb{P})} \le \epsilon$ 对所有 $j = 1, \ldots, N$, 均成立；并且对于每个 $g \in \mathcal{G}$，都存在一个 $j = j(g) \in \{1, \ldots, N\}$，使得

$$g_j^L(x,y) \le g(x,y) \le g_j^U(x,y).$$

相应的括号熵 $H_B(\epsilon,\ \mathcal{G},\ \mathbb{P})$ 定义为括号数的对数，即 $H_B(\epsilon,\ \mathcal{G},\ \mathbb{P}) = \log(N_B(\epsilon,\ \mathcal{G},\ \mathbb{P}))$。

为方便起见，在余下的证明中，只要某些正数常量的具体取值对我们的论证并不重要，我们就用 c_0 表示这些常量。因此， c_0 在不同位置出现时可能取值不同。

引理 5 （括号熵）。对于 $\beta \in \Theta_{K_n}$ 和 $\kappa \in (\kappa_0 - \delta_2,\ \kappa_0 + \delta_2)$，对于满足 $\kappa_0 - \delta_2 > 0$ 的 $\delta_2 > 0$，定义实值函数 z： $\mathcal{L}^2([0,1]) \times \{0,1\} \to \mathbb{R}$,

$$z_{\beta,\kappa}(x,y) = \sum_{y \in \{0,1\}} \left\{ f^{1+\alpha}_{H(\langle x, \widetilde{\beta}_{K_n}\rangle)}(y) - f^{1+\kappa}_{H(\langle x,\beta\rangle)}(y) \right\} - \left(1 + \frac{1}{\kappa}\right) \left\{ f^{\kappa}_{H(\langle x, \widetilde{\beta}_{K_n}\rangle)}(y) - f^{\kappa}_{H(\langle x,\beta\rangle)}(y) \right\},$$

以及函数类

$$\mathcal{Z}_{n,\delta_1,\delta_2} = \left\{ z_{\beta,\kappa}(x,y), \beta \in \Theta_{K_n}, \|\beta - \widetilde{\beta}_{K_n}\| \le \delta_1, |\kappa - \kappa_0| < \delta_2 \right\}.$$

令 $\mathbb{P}_{X,Y}$ 表示由随机变量 $(X,Y) \in \mathcal{L}^2([0,1]) \times \{0,1\}$ 所诱导的概率测度（分布）。那么，对于每个足够小的 $\delta_1 > 0$ 和 $\delta_2 > 0$，都存在一个正的

常数 $A = A(\delta_1, \delta_2)$，使得对所有 $n \geq n_0$ 都有

$$H_B\left(\epsilon, \mathcal{Z}_{n,\delta_1,\delta_2}, \mathbb{P}_{X,Y}\right) \leq 3K_n \log\left(1 + \frac{A}{\epsilon}\right), \quad \epsilon > 0.$$

证明。根据 van de Geer (2000) 中的引理 2.1，我们有

$$H_B\left(\epsilon, \mathcal{Z}_{n,\delta_1,\delta_2}, \mathbb{P}_{X,Y}\right) \leq H_\infty\left(\epsilon/2, \mathcal{Z}_{n,\delta_1,\delta_2}\right), \quad \epsilon > 0.$$

因此，只需对一致熵进行界定。由 (A5)，不失一般性，我们可以假设 $\|x\| \leq C$。对于 $\mathcal{Z}_{n,\delta_1,\delta_2}$, $z_{\beta_1,\kappa_1}(x,y)$ 和 $z_{\beta_2,\kappa_2}(x,y)$ 中的任意两个函数，我们有

$$\begin{aligned}
|z_{\beta_1,\kappa_1}(x,y) - z_{\beta_2,\kappa_2}(x,y)| &\leq |z_{\beta_1,\kappa_1}(x,y) - z_{\beta_1,\kappa_2}(x,y)| + |z_{\beta_1,\kappa_2}(x,y) - z_{\beta_2,\kappa_2}(x,y)| \\
&\leq \sum_{y\in\{0,1\}} \left\{ \left| f^{1+\kappa_1}_{H(\langle x,\beta_1\rangle)}(y) - f^{1+\kappa_2}_{H(\langle x,\beta_1\rangle)}(y) \right| + \left| f^{1+\kappa_1}_{H(\langle x,\widetilde{\beta}_{K_n}\rangle)}(y) - f^{1+\kappa_2}_{H(\langle x,\widetilde{\beta}_{K_n}\rangle)}(y) \right| \right\} \\
&\quad + \frac{|\kappa_1 - \kappa_2|}{|\kappa_0 - \delta|^2} + \left| f^{\kappa_1}_{H(\langle x,\beta_1\rangle)}(y) - f^{\kappa_2}_{H(\langle x,\beta_1\rangle)}(y) \right| + \left| f^{\kappa_1}_{H(\langle x,\widetilde{\beta}_{K_n}\rangle)}(y) - f^{\kappa_2}_{H(\langle x,\widetilde{\beta}_{K_n}\rangle)}(y) \right| \\
&\quad + \sum_{y\in\{0,1\}} \left| f^{1+\kappa_2}_{H(\langle x,\beta_1\rangle)}(y) - f^{1+\kappa_2}_{H(\langle x,\beta_2\rangle)}(y) \right| + \frac{\kappa_2 + 1}{\kappa_0 - \delta} \left| f^{\kappa_2}_{H(\langle x,\beta_1\rangle)}(y) - f^{\kappa_2}_{H(\langle x,\beta_2\rangle)}(y) \right|
\end{aligned}$$

~对于所有充分小的 $\delta_1 > 0$, $\|\beta - \beta_{K_n}\| < \delta_1$ 意味着，对于所有充分大的 n, $H(\langle x,\beta\rangle)$ 与 0 保持正距离。为此，首先注意到，由 (A2), $H(\langle x,\beta_0\rangle)$ 与 0 和 1 均保持正距离，因为 $|\langle x,\beta_0\rangle| \leq C\|\beta_0\|$ 和 H 是双射。此外，

$$\begin{aligned}
|H(\langle x,\beta\rangle) - H(\langle x,\beta_0\rangle)| &\leq \left| H(\langle x,\beta\rangle) - H(\langle x,\widetilde{\beta}_{K_n}\rangle) \right| + \left| H(\langle x,\widetilde{\beta}_{K_n}\rangle) - H(\langle x,\beta_0\rangle) \right| \\
&\leq \sup_{x\in\mathbb{R}} H'(x) C\delta_1 + o(1),
\end{aligned}$$

当 $n \to \infty$ 时。现在，对于远离 0 和 1 的 p，例如 $0 < p_{\min} \leq p \leq p_{\max} < 1$，容易看出，对于所有 $\kappa \in (\kappa_0 - \delta_2, \kappa_0 + \delta_2)$，函数 $p \mapsto f^\kappa_p(y)$ 都是利普希茨连续的，其利普希茨常数等于 $(k_0 + \delta_2)(1 + p_{\max})/(p_{\min}(1 - p_{\max}))$。因此，我们进一步得到

$$\begin{aligned}
|z_{\beta_1,\kappa_1}(x,y) - z_{\beta_2,\kappa_2}(x,y)| &\leq 4 \sup_{y\in\{0,1\},p\in[0,1],|\kappa-\kappa_0|<\delta_2} \left| f^{1+\kappa}_p(y)\log(f_p(y)) \right| |\kappa_1 - \kappa_2| + \frac{|\kappa_1 - \kappa_2|}{|\kappa_0 - \delta_2|^2} \\
&\quad + 2 \sup_{y\in\{0,1\},p\in[0,1],|\kappa-\kappa_0|<\delta_2} \left| f^{\kappa}_p(y)\log(f_p(y)) \right| |\kappa_1 - \kappa_2| \\
&\quad + \sup_{x\in\mathbb{R}} H'(x) C \frac{(1 + k_0 + \delta_2)(1 + p_{\max})}{p_{\min}(1 - p_{\max})} \|\beta_1 - \beta_2\| \\
&\leq c_0 |\kappa_1 - \kappa_2| + c_0 \left\| \beta_1 - \beta_{\widetilde{K}_n} \right\| + c_0 \left\| \beta_2 - \beta_{\widetilde{K}_n} \right\|
\end{aligned}$$

存在某个全局的 $c_0 > 0$，它仅取决于 δ_2。由此可得

$$H_\infty\left(\epsilon, \mathcal{Z}_{n,\delta_1,\delta_2}\right) \leq H\left(\epsilon/(2c_0), \{|\kappa - \kappa_0| \leq \delta_2\}\right) + 2H\left(\epsilon/(4c_0), \{\beta \in \Theta_{K_n}, \|\beta - \widetilde{\beta}_{K_n}\| \leq \delta_1\}\right).$$

根据 vande Geer （2000）的引理 2.5 和推论 2.6，这两个熵数分别可以得到界，从而有

$$H_\infty\left(\epsilon, \mathcal{Z}_{n,\delta_1,\delta_2}\right) \leq \log\left(\frac{8c_0\delta_2}{\epsilon} + 1\right) + 2K_n \log\left(\frac{16c_0\delta_1}{\epsilon} + 1\right) \leq 3K_n \log\left(\frac{A}{\epsilon} + 1\right),$$

对于 $A = 8c_0\delta_2 \vee 16c_0\delta_1$。证明完毕。 □

$\widehat{\ }$ $\widehat{\ }$定理 2 的证明。注意，由于 β_n 使 $M_n(\beta, \widehat{\kappa}_n) + \lambda_n \mathcal{J}(\beta)$ 在 $\beta \in \Theta_{K_n}$ 和 $\widetilde{g}_{K_n} \in \Theta_{K_n}$ 上取得最小值，因此有

$$M\left(\widehat{\beta}_n, \widehat{\kappa}_n\right) - M\left(\widetilde{\beta}_{K_n}, \widehat{\kappa}_n\right) \leq U_n\left(\widetilde{\beta}_{K_n}, \widehat{\beta}_n, \widehat{\kappa}_n\right) + \widehat{\lambda}_n \mathcal{J}(\widetilde{\beta}_{K_n}), \tag{13}$$

$\widehat{\ }$其中，我们舍去了非负项 $\lambda_n \mathcal{J}(\widehat{g}_n)$ 并重新整理了各项。这里，$U_n(f, g, \alpha)$: $\Theta_{K_n} \times \Theta_{K_n} \times \mathbb{R}$ 表示由下式给出的中心化过程

$$U_n(f, g, \alpha) = \{M_n(f, \alpha) - M(f, \alpha\} - \{M_n(g, \alpha) - M(g, \alpha)\}.$$

我们的证明分为两步。第一步，我们建立 LHS$\widehat{\ }$ $\widetilde{\ }$ (13) 关于 $\pi(\beta_n, \beta_{K_n})$ 的二次下界。具体而言，我们将证明存在常数 $\eta, L > 0$ 使得

$$M\left(\widehat{\beta}_n, \widehat{\kappa}_n\right) - M\left(\widetilde{\beta}_{K_n}, \widehat{\kappa}_n\right) \geq \eta\left|\pi(\widehat{\beta}_n, \widetilde{\beta}_{K_n})\right|^2 - L\left\|\widetilde{\beta}_{K_n} - \beta_0\right\| \pi(\widehat{\beta}_n, \widetilde{\beta}_{K_n}), \tag{14}$$

对所有充分大的 n，以高概率成立。第二步，我们推导$U_n\left(\widetilde{\beta}_{K_n}, \widehat{\beta}_n, \widehat{\kappa}_n\right)$ 关于 $\pi(\widehat{\beta}_n, \widetilde{\beta}_{K_n})$ 的连续性模，并证明

$$U_n\left(\widetilde{\beta}_{K_n}, \widehat{\beta}_n, \widehat{\kappa}_n\right) = O_{\mathbb{P}}(1)\left\{\gamma_n \pi(\widehat{\beta}_n, \widetilde{\beta}_{K_n}) \vee \gamma_n^2\right\}, \tag{15}$$

其中 $\gamma_n^2 = K_n$ 对数 n/n。综合起来，(14) 和 (15) 得出定理的结论。为此，将 (14) 和 (15) 代入 (13)，并重新整理可得

$$|\pi(\widehat{g}_n, \widetilde{g}_{K_n})|^2 \leq L'\left(\gamma_n \pi(\widehat{\beta}_n, \widetilde{\beta}_{K_n}) \vee \gamma_n^2\right) + L'\left\|\widetilde{\beta}_{K_n} - \beta_0\right\| \pi(\widehat{\beta}_n, \widetilde{\beta}_{K_n}) + L'\widehat{\lambda}_n \mathcal{J}\left(\widetilde{\beta}_{K_n}\right), \tag{16}$$

对于某个充分大的 L' 以及所有足够大的 n 以高概率成立。于是请注意，我们只需证明定理对 $\gamma_n \pi(\widehat{g}_n, \widetilde{g}_{K_n}) \geq \gamma_n^2$成立，或者等价地，对 $\pi(\widehat{g}_n, \widetilde{g}_{K_n}) \geq \gamma_n$成立；否则结论显然成立。在此简化下，（16）是不等式 $x_0^2 \leq bx_0 + c$ 对于

$x_0 = \pi(\widehat{\beta}_n,\ \widetilde{\beta}_{K_n}) \geq 0,\quad b = L'(\left\|\widetilde{\beta}_{K_n} - \beta_0\right\| + \gamma_n)$ 以及 $c = L'\mathcal{J}\left(\widetilde{\beta}_{K_n}\right)$。这意味着 x_0必须小于或等于 $x^2 - bx - c = 0$ 的正根，即

$$0 \leq x_0 \leq \frac{b + \sqrt{b^2 + 4c}}{2} \leq b + \sqrt{c},$$

并在代入 x_0、 b 和 c后，

$$\pi(\widehat{\beta}_n, \widetilde{\beta}_{K_n}) \leq L'\left(\left\|\widetilde{\beta}_{K_n} - \beta_0\right\| + \gamma_n\right) + \sqrt{L'\widehat{\lambda}_n\mathcal{J}\left(\widetilde{\beta}_{K_n}\right)}.$$

将两边平方，并两次使用不等式 $(x+y)^2 \leq 2x^2 + 2y^2$，得到

$$\left|\pi(\widehat{\beta}_n, \widetilde{\beta}_{K_n})\right|^2 \leq O_{\mathbb{P}}(1)\gamma_n^2 + O_{\mathbb{P}}(1)\left\|\widetilde{\beta}_{K_n} - \beta_0\right\|^2 + O_{\mathbb{P}}(1)\widehat{\lambda}_n\mathcal{J}\left(\widetilde{\beta}_{K_n}\right).$$

$\widehat{\ }\ \widetilde{\ }$ $\widehat{\ }$要从 $\pi(\beta_n,\ \beta_{K_n})$ 推出 $\pi(\beta_n,\ g_0)$，只需注意到

$$\begin{aligned}\left|\pi(\widehat{\beta}_n, \beta_0)\right|^2 &\leq 2\left|\pi(\widehat{\beta}_n, \widetilde{\beta}_{K_n})\right|^2 + 2C^2\left\|\widetilde{\beta}_{K_n} - \beta_0\right\|^2 \\ &= O_{\mathbb{P}}(1)\gamma_n^2 + O_{\mathbb{P}}(1)\left\|\widetilde{\beta}_{K_n} - \beta_0\right\|^2 + O_{\mathbb{P}}(1)\widehat{\lambda}_n\mathcal{J}\left(\widetilde{\beta}_{K_n}\right),\end{aligned}$$

这正是我们要证明的。

为证明(14)和(15)，首先注意到，根据定理1 和(A1)，我们可以将讨论限制在集合 $F_\delta \subset \mathcal{B}([0,1]) \times \mathbb{R}$ 内，该集合定义如下：

$$F_\delta = \left\{\|\beta - \beta_0\|_\infty < \delta, |\kappa - \kappa_0| < \delta\right\}, \tag{17}$$

对于某个充分小的 $\delta > 0$。因此，要证明(14)，只需证明

$$M(\beta,\kappa) - M\left(\widetilde{\beta}_{K_n}, \kappa\right) \geq \eta\left|\pi(\beta, \widetilde{\beta}_{K_n})\right|^2 - L\left\|\beta - \widetilde{\beta}_{K_n}\right\|\pi(\beta, \widetilde{\beta}_{K_n}) \tag{18}$$

对于某个 η， $L > 0$，在 $(\beta,\ \kappa)$ 上一致地 $\in F_\delta$。为证明 (18)，首先注意到

$$\begin{aligned}M(\beta,\kappa) - M\left(\widetilde{\beta}_{K_n}, \kappa\right) = &\sum_{y\in\{0,1\}}\mathbb{E}\left\{f^{1+\kappa}_{p_1(\beta)}(y) - f^{1+\kappa}_{p_1(\widetilde{\beta}_{K_n})}(y)\right\} \\ &- \left(1 + \frac{1}{\kappa}\right)\sum_{y\in\{0,1\}}\mathbb{E}\left\{f_{p_1(\beta_0)}(y)\left[f^{\kappa}_{p_1(\beta)}(y) - f^{\kappa}_{p_1(\widetilde{\beta}_{K_n})}(y)\right]\right\}.\end{aligned} \tag{19}$$

接下来考虑一个 fi rst 阶泰勒展开及拉格朗日余项，涉及 $f^{1+\kappa}_{p_1\ \beta\,()}(y)$ 和 $f^{\kappa}_{p_1(\beta)}(y$)

分别关于 $f^{1+\kappa}_{p_1\ (\widetilde{\beta}_{K_n})}(y)$ 和 $f^{\kappa}_{p_1(\widetilde{\beta}_{K_n})}(y)$。利用以下公式

$$\frac{\partial f_p^\kappa(y)}{\partial p} = \kappa f_p^\kappa(y)\left\{\frac{y-p}{p(1-p)}\right\}$$
$$\frac{\partial^2 f_p^\kappa(y)}{\partial p^2} = \kappa\frac{f_p^\kappa(y)}{p^2(1-p)^2}\left\{(\kappa-1)p^2+2(1-\kappa)yp+\kappa y^2-y\right\},$$

对每个 $\kappa>0$ 均成立，因此我们可以推知存在随机变量 $p_\star = p_\star(X_1,\ y,\ \kappa,\ \beta)$~ ~ 和 $p_{\star\star} = p_{\star\star}(X_1,\ Y_1,\ \kappa,\ \beta)$，使得 $\max\{|p_\star - p_1(\beta_{K_n})|,\ |p_{\star\star}-p_1(\beta_{K_n})|\} \le |p_1(\beta) - p_1(\beta_{K_n})|$，并且

$$\begin{aligned}
f^{1+\kappa}_{p_1(\beta)}(y) - f^{1+\kappa}_{p_1(\widetilde{\beta}_{K_n})}(y) &= (1+\kappa)f^{1+\kappa}_{p_1(\widetilde{g}_{K_n})}(y)\left\{\frac{y-p_1(\widetilde{\beta}_{K_n})}{p_1(\widetilde{\beta}_{K_n})(1-p_1(\widetilde{\beta}_{K_n}))}\right\}(p_1(\beta)-p_1(\widetilde{\beta}_{K_n}))\\
&\quad+(1+\kappa)\frac{f^{1+\kappa}_{p_\star}(y)}{2p_\star^2(1-p_\star)^2}\left\{\alpha p_\star^2-2\alpha y p_\star+(\kappa+1)y^2-y\right\}\left|p_1(\beta)-p_1(\widetilde{\beta}_{K_n})\right|^2\\
&= (1+\kappa)f^{1+\kappa}_{p_1(\beta_0)}(y)\left\{\frac{y-p_1(\beta_0)}{p_1(\beta_0)(1-p_1(\beta_0))}\right\}(p_1(\beta)-p_1(\widetilde{\beta}_{K_n}))\\
&\quad+(1+\kappa)\frac{f^{1+\kappa}_{p_\star}(y)}{2p_\star^2(1-p_\star)^2}\left\{\kappa p_\star^2-2\kappa y p_\star+(\kappa+1)y^2-y\right\}\left|p_1(\beta)-p_1(\widetilde{\beta}_{K_n})\right|^2\\
&\quad+(1+\kappa)\left[f^{1+\kappa}_{p_1(\widetilde{\beta}_{K_n})}(y)\left\{\frac{y-p_1(\widetilde{\beta}_{K_n})}{p_1(\widetilde{\beta}_{K_n})(1-p_1(\widetilde{\beta}_{K_n}))}\right\}-f^{1+\kappa}_{p_1(\beta_0)}(y)\left\{\frac{y-p_1(\beta_0)}{p_1(\beta_0)(1-p_1(\beta_0))}\right\}\right]\\
&\quad\times(p_1(\beta)-p_1(\widetilde{\beta}_{K_n}))
\end{aligned}$$

以及

$$\begin{aligned}
f^{\kappa}_{p_1(\beta)}(y) - f^{\kappa}_{p_1(\widetilde{\beta}_{K_n})}(y) &= \kappa f^{\kappa}_{p_1(\widetilde{\beta}_{K_n})}(y)\left\{\frac{y-p_1(\widetilde{\beta}_{K_n})}{p_1(\beta_{K_n})(1-p_1(\beta_{K_n}))}\right\}\Big(p_1(\beta)-p_1(\widetilde{\beta}_{K_n}))\Big)\\
&\quad+\kappa\frac{f^{\kappa}_{p_{\star\star}}(y)}{2p_{\star\star}^2(1-p_{\star\star})^2}\left\{(\kappa-1)p_{\star\star}^2+2(1-\kappa)yp_{\star\star}+\kappa y^2-y\right\}\left|p_1(\beta)-p_1(\widetilde{\beta}_{K_n})\right|^2\\
&= \kappa f^{\kappa}_{p_1(\beta_0)}(y)\left\{\frac{y-p_1(\beta_0)}{p_1(\beta_0)(1-p_1(\beta_0))}\right\}(p_1(\beta)-p_1(\widetilde{\beta}_{K_n}))\\
&\quad+\kappa\frac{f^{\kappa}_{p_{\star\star}}(y)}{2p_{\star\star}^2(1-p_{\star\star})^2}\left\{(\kappa-1)p_{\star\star}^2+2(1-\kappa)yp_{\star\star}+\kappa y^2-y\right\}\left|p_1(\beta)-p_1(\widetilde{\beta}_{K_n})\right|^2\\
&\quad+\kappa\left[f^{\kappa}_{p_1(\widetilde{\beta}_{K_n})}(y)\left\{\frac{Y_1-p_1(\widetilde{\beta}_{K_n})}{p_1(\widetilde{\beta}_{K_n})(1-p_1(\widetilde{\beta}_{K_n}))}\right\}-f^{\kappa}_{p_1(\beta_0)}(y)\left\{\frac{y-p_1(\beta_0)}{p_1(\beta_0)(1-p_1(\beta_0))}\right\}\right]\\
&\quad\times\left\{p_1(\beta)-p_1(\widetilde{\beta}_{K_n})\right\}.
\end{aligned}$$

将这些展开式代入（19）并化简可得

$$M(\beta,\kappa)-M\left(\widetilde{\beta}_{K_n},\kappa\right)=(1+\kappa)\sum_{y\in\{0,1\}}\mathbb{E}\left\{\left(u_n(X_1,y,\widetilde{\beta}_{K_n},\kappa)-v_n(X_1,y,\widetilde{\beta}_{K_n},\kappa)\right)\left(p_1(\beta)-p_1(\widetilde{\beta}_{K_n})\right)\right\}$$
$$+(1+\kappa)\sum_{y\in\{0,1\}}\mathbb{E}\left\{\left(u_n'(X_1,y,\beta,\kappa)-v_n'(X_1,y,\beta,\kappa)\right)\left|p_1(\beta)-p_1(\widetilde{\beta}_{K_n})\right|^2\right\},$$

其中，为方便起见，我们定义了

$$u_n(X_1,y,\widetilde{\beta}_{K_n},\alpha)=f^{1+\kappa}_{p_1(\widetilde{\beta}_{K_n})}(y)\left\{\frac{y-p_1(\widetilde{\beta}_{K_n})}{p_1(\widetilde{\beta}_{K_n})(1-p_1(\widetilde{\beta}_{K_n}))}\right\}-f^{1+\kappa}_{p_1(\beta_0)}(y)\left\{\frac{y-p_1(\beta_0)}{p_1(\beta_0)(1-p_1(\beta_0))}\right\}$$
$$v_n(X_1,y,\widetilde{\beta}_{K_n},\kappa)=f^{\kappa}_{p_1(\widetilde{\beta}_{K_n})}(y)f_{p_1(\beta_0)}(y)\left\{\frac{y-p_1(\widetilde{\beta}_{K_n})}{p_1(\widetilde{\beta}_{K_n})(1-p_1(\widetilde{\beta}_{K_n}))}\right\}-f^{1+\kappa}_{p_1(\beta_0)}(y)\left\{\frac{y-p_1(\beta_0)}{p_1(\beta_0)(1-p_1(\beta_0))}\right\}$$
$$u_n'(X_1,y,\beta,\kappa)=\frac{f^{1+\kappa}_{p_\star}(y)}{2p_\star^2(1-p_\star)^2}\left\{\kappa p_\star^2-2\kappa y p_\star+(\kappa+1)y^2-y\right\}$$
$$v_n'(X_1,y,\beta,\kappa)=\frac{f^{\kappa}_{p_{\star\star}}(y)f_{p_1(\beta_0)}(y)}{2p_{\star\star}^2(1-p_{\star\star})^2}\left\{(\kappa-1)p_{\star\star}^2+2(1-\kappa)yp_{\star\star}+\kappa y^2-y\right\},$$

其中 $u_n'(X_1,\ y,\ \beta,\ \kappa)$ 和 $v_n'(X_1,\ y,\ \beta,\ \kappa)$ 分别通过 $p_\star$ 和 $p_{\star\star}$ 依赖于 β。

我们首先证明，在 F_δ 的定义中，当 δ 足够小时，存在一个 $\eta>0$

$$(1+\kappa)\sum_{y\in\{0,1\}}\mathbb{E}\left\{\left(u_n'(X_1,y,\beta,\kappa)-v_n'(X_1,y,\beta,\kappa)\right)\left|p_1(\beta)-p_1(\widetilde{\beta}_{K_n})\right|^2\right\}\geq\eta\left|\pi(\beta,\widetilde{\beta}_{K_n})\right|^2,\quad(20)$$

对于所有足够大的 n，在 $(\beta,\ \kappa)\in F$ 中一致成立。为此，首先注意到，根据 $p_\star$ 的定义和我们的假设，我们得到

$$\begin{aligned}|p_\star-p_1(\beta_0)|&\leq\left|p_\star-p_1(\widetilde{\beta}_{K_n})\right|+\left|p_1(\widetilde{\beta}_{K_n})-p_1(\beta_0)\right|\\&\leq c_0\left\|\beta-\beta_0\right\|_\infty+c_0\left\|\widetilde{\beta}_{K_n}-\beta_0\right\|\\&\leq c_0\delta+o(1)\end{aligned}$$

类似的论证还表明 $|p_{\star\star}-p_1(g_0)|<c_0\delta+o(1)$。接下来回顾一下，$p_1(\beta_0)$ 几乎必然远离 0 和 1（参见引理 5 的证明）。因此，$p_\star$ 和 $p_{\star\star}$ 对所有 $\delta\leq\delta_0$ 和 $n\geq n_0$ 也几乎必然远离 0 和 1。由连续性可知，存在

一个 $\delta_0' \le \delta_0$，使得对于所有 $\delta \le \delta_0'$，我们都有

$$\begin{aligned} u_n'(X_1, y, \beta, \kappa) - v_n'(X_1, y, \beta, \kappa) &\ge \frac{f_{p_1(\beta_0)}^{1+\kappa}(y)}{8p_1^2(\beta_0)(1-p_1(\beta_0))^2} \left\{p_1^2(\beta_0) - 2yp_1(\beta_0) + y^2\right\} \\ &= \frac{f_{p_1(\beta_0)}^{1+\kappa}(y)}{8p_1^2(\beta_0)(1-p_1(\beta_0))^2} \left|y - p_1(\beta_0)\right|^2, \end{aligned}$$

对所有 $n \ge n_0$，几乎必然成立。对 $y \in \{0,1\}$ 求和可知，存在一个 $c_0 > 0$，使得

$$\sum_{y\in\{0,1\}} \left\{u_n'(X_1, y, \beta, \kappa) - v_n'(X_1, y, \beta, \kappa)\right\} \ge c_0$$

对所有 $n \ge n_0$，几乎必然成立。因此，

$$\sum_{y\in\{0,1\}} \mathbb{E}\left\{\left(u_n'(X_1, y, \beta, \kappa) - v_n'(X_1, y, \beta, \kappa)\right)\left|p_1(\beta) - p_1(\beta_{K_n})\right|^2\right\} \ge c_0 \mathbb{E}\left\{\left|p_1(\beta) - p_1(\widetilde{\beta}_{K_n})\right|^2\right\}.$$

~现在，根据中值定理，存在一个（随机的） $x^\star \in \mathbb{R}$，使得 $|x^\star - \langle X_1, \beta_{K_n}\rangle| \le_\sim |\langle X_1, \beta\rangle - \langle X_1, \beta_{K_n}\rangle|$，并且

$$\left|p_1(\beta) - p_1(\widetilde{\beta}_{K_n})\right|^2 = \left|H\left(\langle X_1, \beta\rangle\right) - H\left(\langle X_1, \widetilde{\beta}_{K_n}\rangle\right)\right|^2 = \left|H'(x^\star)\right|^2 \left|\langle X_1, \beta - \widetilde{\beta}_{K_n}\rangle\right|^2.$$

对于 $\beta \in \{f \in \mathcal{B}([0,1]) \colon \|\beta - \beta_0\|_\infty < \delta\}$，由 (A5) 和 （A4），可得

$$|x^\star - \langle X_1, \beta_0\rangle| \le c_0\delta + o(1),$$

对于某个 $c_0 > 0$，该结论几乎必然成立。由于 $|\langle X_1, \beta_0\rangle|$ 几乎必然有界且远离零和一，并且 H 严格递增，因此存在一个 δ_0''，使得 $H'(x^\star) \ge c_0 > 0$ 对所有 $\delta \le \delta_0'' \le \delta_0'$ 和 $n \ge n_0'$几乎必然成立，由此可得

$$\mathbb{E}\left\{\left|p_1(\beta) - p_1(\widetilde{\beta}_{K_n})\right|^2\right\} \ge c_0^2 \mathbb{E}\left\{\left|\langle X_1, \beta - \widetilde{\beta}_{K_n}\rangle\right|^2\right\} = c_0^2 |\pi(\beta, \widetilde{\beta}_{K_n})|^2,$$

对所有足够大的 n，$\delta \le \delta_0''$ 和 $n \ge n_0'$ 几乎必然同时成立。结合上述下界，我们得到

$$(1+\kappa) \sum_{y\in\{0,1\}} \mathbb{E}\left\{\left(u_n'(X_1, y, \beta, \kappa) - v_n'(X_1, y, \beta, \kappa)\right)\left|p_1(\beta) - p_1(\beta_{K_n})\right|^2\right\} \ge \eta \left|\pi(\beta, \widetilde{\beta}_{K_n})\right|^2,$$

对于所有 $n \ge n_0 \vee n_0'$ 和 $\delta \le \delta_0''$。这里，$\eta = c_0(1 + \kappa_0 - \delta) > 0$ 和 $(1 + \kappa_0 - \delta)$ 对所有 $\delta \le (1 + \kappa_0)/2$ 都严格为正，因为 $c_0 > 0$。选择 $\delta \le \delta_0'' \wedge (\kappa_0 + 1)/2$ 可确保 $\eta > 0$ 的正性，

完成（20）的证明。

接下来证明

$$\sum_{y\in\{0,1\}}\left|\mathbb{E}\left\{\left(u_n(X_1,y,\widetilde{\beta}_{K_n},\kappa)-v_n(X_1,y,\widetilde{\beta}_{K_n},\kappa)\right)\left(p_1(\beta)-p_1(\widetilde{\beta}_{K_n})\right)\right\}\right|$$
$$\leq L\left\|\widetilde{\beta}_{K_n}-\beta_0\right\|\pi(\beta,\widetilde{\beta}_{K_n}), \tag{21}$$

对某个有限的 $L>0$，一致地对 $(\beta,\ \kappa)\in F_\delta$ 成立。为证明这一点，首先注意到，由于 $p_1(\beta_0)$ 与 0 和 1 的距离有界，且 $|p_1(\beta_{K_n})-p_1(\beta_0)|\to 0$ 随着 $n\to\infty$ 以概率 $\widetilde{}$ 1 成立，因此也可以假设，对所有充分大的 n，$p_1(\beta_{K_n})$ 与 0 和 1 的距离也有界。当 p 与 0 和 1 的距离有界时，对于所有 $\kappa\in(\kappa_0-\delta,\ \kappa_0+\delta)$（其中 $\delta>0$ 足够小）以及所有 $y\in\{0,1\}$，函数 $p\mapsto f_p^\kappa(y)$ 和 $p\mapsto(y-p)/p(1-p)$ 都有界且满足利普希茨条件。由有界性可知，它们的乘积也满足利普希茨条件。由此可得，对所有 $n\geq n_0''$，存在一个仅依赖于 δ 的 $c_0>0$，使得

$$\left|u_n(X_1,y,\widetilde{\beta}_{K_n},\kappa)-v_n(X_1,y,\widetilde{\beta}_{K_n},\kappa)\right|\leq c_0\left|p_1(\widetilde{\beta}_{K_n})-p_1(\beta_0)\right|\leq c_0\left\|\widetilde{\beta}_{K_n}-\beta_0\right\|,$$

几乎必然。此外，由（A2）和 Schwarz 不等式可得

$$\mathbb{E}\left\{\left|p_1(\widetilde{\beta}_{K_n})-p_1(\beta)\right|\right\}\leq c_0\mathbb{E}\left\{\left|\langle X_1,\widetilde{\beta}_{K_n}-\beta\rangle\right|\right\}\leq c_0\pi(\widetilde{\beta}_{K_n},\beta).$$

这两个界共同推出（18），从而完成证明的第一部分。

为完成证明，我们现在证明（15），为此，我们采用 Kalogridis 和 Van $\widetilde{}$ Aelst （2023）中的论证。首先注意，在 F_δ 上，我们有 $\|\beta-\beta_{K_n}\|\leq\delta+o$（1）$\leq 2\delta$ 对所有充分大的 n均成立，这由（A4）以及一致范数支配 $\mathcal{L}^2([0,1])$-范数这一事实可得。因此，

$$\left|\frac{U_n(\widetilde{\beta}_{K_n},\widehat{\beta}_n,\widehat{\kappa}_n)}{\gamma_n\pi\left(\widetilde{\beta}_{K_n},\widehat{\beta}_n\right)\vee\gamma_n^2}\right|\leq\sup_{\substack{\beta\in\Theta_{K_n}:\|\beta-\widetilde{\beta}_{K_n}\|\leq 2\delta\\ |\kappa-\kappa_0|\leq\delta}}\left|\frac{U_n(\widetilde{\beta}_{K_n},\beta,\kappa)}{\gamma_n\pi\left(\widetilde{\beta}_{K_n},\beta\right)\vee\gamma_n^2}\right|, \tag{22}$$

对所有足够大的 n，以高概率成立。我们将证明（22）右侧的随机变量依概率有界，这等价于证明

$$\lim_{T\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(\sup_{\substack{\beta\in\Theta_{K_n,\delta},\kappa\in V_\delta\\ \pi(\widetilde{\beta}_{K_n},\beta)\leq\gamma_n}}\left|U_n(\widetilde{\beta}_{K_n},\beta,\alpha)\right|\geq T\gamma_n^2\right)=0, \tag{23}$$

以及

$$\lim_{T\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(\sup_{\substack{\beta\in\Theta_{K_n,\delta},\kappa\in V_\delta\\ \pi\left(\widetilde{\beta}_{K_n},\beta\right)>\gamma_n}}\left|\frac{U_n(\widetilde{\beta}_{K_n},\beta,\kappa)}{\pi\left(\widetilde{\beta}_{K_n},\beta\right)}\right|\geq T\gamma_n\right)=0, \tag{24}$$

为简化记号，我们定义这些球

$$\Theta_{K_n,\delta}=\left\{\beta\in\Theta_{K_n}:\|\beta-\widetilde{\beta}_{K_n}\|\leq 2\delta\right\}$$
$$V_\delta=\{\kappa\in\mathbb{R}:|\kappa-\kappa_0|\leq\delta\}.$$

为证明 (23) 和 (24，我们将利用以下事实：对所有 $\epsilon>0$ 充分小的$\epsilon\leq\epsilon_0$，例如，存在一个 $B>0$ 使得

$$\int_0^\epsilon\log^{1/2}\left(1+\frac{1}{u}\right)du\leq B\epsilon\log^{1/2}\left(\frac{1}{\epsilon}\right).$$

(例如，使用洛必达法则)。为证明 (23)，我们将对van de Geer (2000) 中的定理 5.11 应用于 ~过程 $U_n(\beta_{K_n},\beta,\kappa)$，其中 $\beta\in\Theta_{K_n,\delta}$, $\kappa\in V_\delta$和 $\pi(\beta,\beta_{K_n})\leq\gamma_n$。为准备这一应用，~将 $U_n(\beta_{K_n},\beta,\kappa)$ 写为

$$U_n(\widetilde{\beta}_{K_n},\beta,\kappa)=\int z_{\beta,\kappa}d\left(\mathbb{P}_n-\mathbb{P}_{X,Y}\right)=n^{-1/2}v_n(z_{\beta,\kappa}),$$

其中 $z_{\beta,\kappa}$ 和 $\mathbb{P}_{X,Y}$ 与引理 5 中相同， $v_n(\cdot)$ 表示经验过程。根据我们的定义，$z_{\beta,\kappa}\in\mathcal{Z}(n,2\delta,\delta)$ 以及每个 $z_{\beta,\kappa}$ 都一致有界于 $2+1/(\kappa_0-\delta)$。令 $\mathcal{H}_{B,K}(\epsilon,\mathcal{Z}(n,2\delta,\delta)\,\mathbb{P}_{X,Y})$ 表示函数类 $\mathcal{Z}(n,2\delta,\delta)$ 的括号广义熵（关于 Bernstein“范数” ρ_K），其定义见 van de Geer (2000) 中的定义 5.1。由 $\mathcal{Z}(n,2\delta,\delta)$ 中元素的一致有界性，van de Geer (2000) 的引理 5.10 得到

$$\mathcal{H}_{B,8+4/(\alpha_0-\delta)}\left(\epsilon,\mathcal{Z}(n,2\delta,\delta)\right)\leq H_B(\epsilon/\sqrt{2},\mathcal{Z}(n,2\delta,\delta),\mathbb{P}_{X,Y}),\quad\epsilon>0. \tag{25}$$

~接下来注意到，根据我们的假设，对所有 $\|\beta-\beta_{K_n}\|\leq 2\delta$，我们有 $|H(\langle X,\beta\rangle)-H(\langle X,\beta_0\rangle)|\leq c_0\delta+o(1)$，因为 $n\to\infty$。如前所述， $H(\langle X,\beta_0\rangle)$ 几乎必然与 0 和 1 保持正距离。因此，当 δ 足够小且 n 几乎必然足够大时，对所有 $\beta\in\Theta_{K_n,\delta}$， $H(\langle X,\beta\rangle)$ 也与 0 和 1 保持正距离。由此可推得，对所有 $\kappa\in(\kappa_0-\delta,\kappa_0+\delta)$ 和 $y\in\{0,1\}$， $f^\kappa_{H(\langle X,\beta\rangle)}(y)$ 以及 $f^{1+\kappa}_{H\,(\langle X,g\rangle)}(y)$ 关于 ~ 概率 $H(\langle X,\beta\rangle)$ 是利普希茨连续的，其中 $\|\beta-\beta_{K_n}\|\leq 2\delta$。经过简单计算即可证明存在

一个全局的 $c_0 > 0$，使得

$$|z_{\beta,\kappa}(X,Y)| \leq c_0 \left|\langle X, \beta - \widetilde{\beta}_{K_n}\rangle\right|.$$

取期望可得

$$\mathbb{E}\left\{|z_{\beta,\kappa}(X,Y)|^2\right\} \leq c_0^2 \left|\pi\left(\beta, \widetilde{\beta}_{K_n}\right)\right|^2 \leq c_0^2 \gamma_n^2. \tag{26}$$

由 van de Geer (2000 的引理 5.8 可知，我们可以在 $R_n = c_0\gamma_n$ 中取 $R_n = c_0\gamma_n$，该结果见 van de Geer (2000) 的定理 5.11。假设（A5）意味着 $R_n \to 0$ 当 $n \to \infty$ 时成立。因此，对于所有足够大的 n，有 $R_n \leq \epsilon_0$。于是，取平方根，对（25）从 0 到 R_n 积分，并应用引理 5，得到

$$\int_0^{R_n} \mathcal{H}^{1/2}_{B,8+4/(\alpha_0-\delta)}(u, \mathcal{Z}(n, 2\delta, \delta)du \leq c_0 K_n^{1/2}\gamma_n \log^{1/2}(n) = c_0 \frac{K_n \log(n)}{n^{1/2}}.$$

若取$C_0 = T/c_0$和 $C_1 = (8 + 4/(\alpha_0 - \delta))T/c_0^2$，则定理 5.11 中的条件（5.31）、（5.33）和（5.34）vandeGeer(2000) 均得到满足。应用该定理可得

$$\begin{aligned}\mathbb{P}\left(\sup_{\substack{\beta\in\Theta_{K_n,\delta},\kappa\in V_\delta \\ \pi(\widetilde{\beta}_{K_n},\beta)\leq\gamma_n}} \left|U_n(\widetilde{\beta}_{K_n}, \beta, \kappa)\right| \geq T\gamma_n^2\right) &= \mathbb{P}\left(\sup_{\substack{\beta\in\Theta_{K_n,\delta},\kappa\in V_\delta \\ \pi(\widetilde{\beta}_{K_n},\beta)\leq\gamma_n}} |v_n(z_{\beta,\kappa})| \geq Tn^{1/2}\gamma_n^2\right) \\ &\leq C\exp\left[-\frac{T^2 K_n \log(n)}{C^2(C_1+1)}\right].\end{aligned}$$

对于某个全局 $C > 0$。根据（A4），指数项趋于 0，因为 $K_n \log n \to \infty$。此外，由于这对所有足够大的 T 都成立，因此（23）成立，从而完成了证明的第一步。

最后，我们证明（24）。首先注意到，由（A5）可知，$\pi(\beta, \beta_{K_n}) \leq C\|\beta - \beta_{K_n}\| \leq 2C\delta \leq \epsilon_0/c_0$ 当$\beta \in \Theta_{K_n,\delta}$ 其中 $\delta \leq \epsilon_0/(c_0 2C)$。这里，$c_0$ 是（26）中出现的常数。因此，只需证明

$$\lim_{T\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\sup_{\substack{\beta\in\Theta_{K_n,\delta},\kappa\in V_\delta \\ \gamma_n < \pi(\widetilde{\beta}_{K_n},\beta)\leq\epsilon_0/c_0}} \left|\frac{U_n(\widetilde{\beta}_{K_n}, \beta, \kappa)}{\pi\left(\widetilde{\beta}_{K_n}, \beta\right)}\right| \geq T\gamma_n\right) = 0. \tag{27}$$

令 $S = \min\{s > 1 : 2^{-s}\epsilon_0/c_0 < \gamma_n\}$。根据（A4），$K_n \asymp n^\gamma$ 对某个 $\gamma \in (0,1)$ 成立。因此，我们可以得到

$S \leq \lceil c_0 \log_2(n) + 1 \rceil$ 对某个 $c_0 > 0$成立。Boole 不等式给出

$$\mathbb{P}\left(\sup_{\substack{\beta \in \Theta_{K_n,\delta}, \kappa \in V_\delta \\ \gamma_n < \pi\left(\widetilde{\beta}_{K_n}, g\right) \leq \epsilon_0/c_0}} \left| \frac{U_n(\widetilde{\beta}_{K_n}, \beta, \kappa)}{\pi\left(\widetilde{\beta}_{K_n}, \beta\right)} \right| \geq T\gamma_n \right) \leq \sum_{s=1}^{S} \mathbb{P}\left(\sup_{\substack{\beta \in \Theta_{K_n,\delta}, \kappa \in V_\delta \\ 2^{-s}\epsilon_0/c_0 < \pi\left(\widetilde{\beta}_{K_n}, \beta\right) \leq 2^{-s+1}\epsilon_0/c_0}} \left| \frac{U_n(\widetilde{\beta}_{K_n}, \beta, \kappa)}{\pi\left(\widetilde{\beta}_{K_n}, \beta\right)} \right| \geq T\gamma_n \right)$$

$$\leq \sum_{s=1}^{S} \mathbb{P}\left(\sup_{\substack{\beta \in \Theta_{K_n,\delta}, \kappa \in V_\delta \\ \pi\left(\widetilde{\beta}_{K_n}, \beta\right) \leq 2^{-s+1}\epsilon_0/c_0}} \left| U_n(\widetilde{\beta}_{K_n}, \beta, \alpha) \right| \geq T 2^{-s} \gamma_n \epsilon_0 / c_0 \right)$$

显然，对于每个 $s \geq 1$，我们有 $2^{-s+1}\epsilon_0 \leq \epsilon_0$。因此，由（26）和引理 5,

$$\int_0^{2^{-s+1}\epsilon_0} \mathcal{H}^{1/2}_{B,8+4/(\alpha_0-\delta)}\left(u, \mathcal{Z}(n, 2\delta, \delta)\right) du \leq 2^{-s+1} B^\star K_n^{1/2} \log^{1/2}(n),$$

对于某个全局 $B^\star > 0$，这里我们用到了这样的事实：对于所有充分大的 n， $2^s \leq 2^S \leq n^{2c}$。我们将对每个加项分别应用 van de Geer（2000）的定理 5.11，另见（23）的证明。对于 $C_0 = T\epsilon_0/(2B^\star c_0)$ 和 $C_1 = T(8 + 4/(\alpha_0 - \delta))/c_0$，当 T 足够大时，该定理的条件均得到满足；根据 S 的定义，对于所有 $s = 1, \ldots, S$，我们有 $\gamma_n \leq 2^{-s+1}\epsilon_0/c_0$。因此，由 van de Geer（2000）的定理 5.11，存在一个全局常数 $C' > 0$，使得

$$\mathbb{P}\left(\sup_{\substack{\beta \in \Theta_{K_n,\delta}, \kappa \in V_\delta \\ \pi\left(\widetilde{\beta}_{K_n}, \beta\right) \leq 2^{-s+1}\epsilon_0/c_0}} \left| U_n(\widetilde{\beta}_{K_n}, \beta, \kappa) \right| \geq T 2^{-s} \gamma_n \epsilon_0 / c_0 \right) \leq C' \exp\left[-\frac{T^2 K_n \log(n)}{4|C'|^2 (C_1 + 1) c_0^2} \right].$$

该指数项不依赖于 $s = 1, \ldots, S$，因此对其求和并回想 $S \leq \lceil c_0 \log_2(n) + 1 \rceil$ 对于某个 $c_0 > 0$ 成立，可得

$$\mathbb{P}\left(\sup_{\substack{\beta \in \Theta_{K_n,\delta}, \kappa \in V_\delta \\ \gamma_n < \pi\left(\widetilde{\beta}_{K_n}, \beta\right) \leq \epsilon_0/c_0}} \left| \frac{U_n(\widetilde{\beta}_{K_n}, \beta, \kappa)}{\pi\left(\widetilde{\beta}_{K_n}, \beta\right)} \right| \geq T\gamma_n \right) \leq C' \lceil c_0 \log_2(n) + 1 \rceil \exp\left[-\frac{T^2 K_n \log(n)}{4|C'|^2 (C_1 + 1) c_0^2} \right],$$

对于所有充分大的 n。对于所有 T 足够大的，譬如 $T \geq T_0$，右端趋于 0，因为 $n \to \infty$，由 $K_n \asymp n^\gamma$可知。至此我们证明了 (24)，从而完成证明。 □

参考文献

Basak, S., Basu, A., and Jones M.C. (2021年) 关于“最优”密度幂散度调节参数，J. Appl. Stat. 48536–556。

Basu, A., Harris, I.R., Hjort, N.L, and Jones, M.C. (1998) 通过最小化密度幂散度实现稳健高效估计，Biometrika 85 549–559。Boente, G., Salibi´an-Barrera, M., and Vena, P. (2020) 半函数型线性回归模型的稳健估计，Comp. Statist. Data Anal. 152。Cardot, H., Ferraty, F., and Sarda, P. (1999) 函数型线性模型，Statist. Probab. Lett. 45 11–22。Cardot, H., Ferraty, F., and Sarda, P. (2003) 函数型线性模型的样条估计量，Statist. Sinica 13 571–591。Cardot, H., and Sarda, P. (2005) 通过惩罚似然估计函数型数据的广义线性模型，J. Multivariate Anal. 92 24–41。Crambes, C., Kneip, A., and Sarda, P. (2009) 函数型线性回归的平滑样条估计量，Ann. Statist. 37 35–72。Dau, H.A., Keogh, E., Kamgar, K., Yeh, C.-C. M., Zhu, Y, Gharghabi, S., Ratanamahatana, C. A., Yanping, D., Hu, B., Begum, N., Bagnall, A., Mueen, A., Batista, G., and Hexagon-ML (2018) UCR时间序列分类档案，https://www.cs.ucr.edu/~eamonn/time_series_data_2018/. de Boor, C. (2001) 样条实用指南，修订版，施普林格，纽约。Denhere, M., and Billor, N. (2016) 稳健主成分函数型逻辑回归，Comm. Statist. Simulation Comput. 45 264–281。Eggermont, P.P.B., and LaRiccia, V.N. (2009) 最大惩罚似然估计，第二卷：回归，施普林格，纽约。Fan, J., and Li, R. (2001) 通过非凹惩罚似然进行变量选择及其预言机性质，J. Amer. Statist. Assoc. 96 1348–1360。Ferraty, F., and Vieu, P. (2006) 非参数函数型数据分析：理论与实践，施普林格，纽约。Green, P.J., and Silverman, B.W. (1994) 非参数回归与广义线性模型，Chapman & Hall，伦敦。Hall, P., and Horowitz, J.L. (2007) 函数型线性回归的方法论与收敛速率，Ann. Statist. 35 70–91。Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (2011) 稳健统计：基于影响函数的方法，Wiley，纽约。

Hubert, M., Rousseeuw, P.J. 和 Vanden Branden, K. (2005) ROBPCA：一种稳健主成分分析的新方法，Technometrics 47 64–79。James, G.M. (2002) 带函数型预测变量的广义线性模型，J. R. Stat. Soc. Ser. B. Stat. Methodol. 64 411–432。Kalogridis, I., Claeskens, G. 和 Van Aelst, S. (2022) 非参数广义线性模型的稳健高效估计，TEST，已接收。Kalogridis, I. 和 Van Aelst, S. (2023) 函数型线性回归的稳健惩罚估计量，J. Multivariate Anal. 194。Kosorok, M.R. (2008) 经验过程与半参数推断导论，施普林格，纽约。McCullagh, P. 和 Nelder, J.A. (1983) 广义线性模型，Chapman & Hall，伦敦。M¨uller, H.-G. 和 Stadtm¨uller, U. (2005) 广义函数型线性模型，The Annals of Statistics 33 774–805。Mutis, M., Beyaztas, U., Simsek, G.G. 和 Shang, H.L. (2022) 一种用于分类的稳健标量对函数型逻辑回归，Comm. Statist. Theory Methods，已在线发表。R Core Team (2023). R：一种用于统计计算的语言和环境。R 统计计算基金会，维也纳，奥地利。Ramsay, J.O. 和 Dalzell, C.J. (1991) 函数型数据分析的一些工具，J. R. Stat. Soc. Ser. B. Stat. Methodol 53 539–561。Ramsay, J.O. 和 Silverman, B.W. (2005) 函数型数据分析，Wiley，纽约。Ratcliffe, S.J., Heller, G.Z. 和 Leader, L.R. (2002) 函数型数据分析及其在周期性刺激胎儿心率数据中的应用。II：函数型逻辑回归，Stat. Med. 21 1115–1127。Reiss, P.T. 和 Ogden, R.T. (2007) 函数型主成分回归与函数型偏最小二乘，Journal of the American Statistical Association 102 984–996。Ruppert, D., Wand, M.P. 和 Carroll, R.J. (2003) 半参数回归，剑桥，纽约。Shin, H. 和 Lee, S. (2016) 一种用于稳健函数型线性回归的再生核希尔伯特空间方法，Statistica Sinica 26 255–272。

van de Geer, S. (2000) M估计中的经验过程，剑桥大学出版社，纽约州纽约市。van der Vaart, A.W. (1998) 渐近统计学，剑桥大学出版社，纽约州纽约市。Warwick, J. 和 Jones, M.C. (2005) 稳健性调节参数的选择，J. Stat. Comput. Simul. 75 581–588。Wei, P., Tang, H. 和 Li, D. (2014) 在病例对照研究中检测基因与纵向环境暴露交互作用的函数型逻辑回归方法，Gen. Epidem. 38 638–651。Wood, S. (2017) 广义加性模型，第2版，CRC出版社，Boca Raton, FL。Yuan, M. 和 Cai, T.T. (2010) 一种用于函数型线性回归的再生核希尔伯特空间方法，The Annals of Statistics 38 3412–3444。