arXiv:2304.08373v1 [math.ST] 17 Apr 2023

平均处理效应卡尺匹配估计量的渐近性质

Máté Kormos[1], Stéphanie van der Pas[2], 和 Aad van der Vaart[1]

[1]代尔夫特理工大学代尔夫特应用数学研究所，地址：荷兰代尔夫特 Mekelweg 4[2]流行病学与数据科学，阿姆斯特丹大学医学中心阿姆斯特丹自由大学分部，地址：荷兰阿姆斯特丹 De Boelelaan 1117；以及阿姆斯特丹公共卫生方法学，地址：荷兰阿姆斯特丹

摘要

卡尺匹配通过比较匹配后的处理组和对照组个体，利用观察性数据估计二元处理的因果效应。当个体的倾向得分——即给定处理前协变量时接受处理的条件概率——相距不超过某一特定距离（称为卡尺）时，就将其匹配。迄今为止，关于卡尺匹配的理论结果仍然缺乏，因此实践者只能凭经验选择卡尺和推断程序。为弥补这一空白，我们提出一种兼顾匹配质量与匹配数量的卡尺。我们证明，由此得到的平均处理效应和处理组平均处理效应估计量渐近无偏，并以参数率渐近正态。我们阐述了实现半参数效率所需的条件，并表明在估计参数化倾向得分时，这两个目标估计量的方差都会增大。最后，我们为这两个目标估计量构造渐近置信区间。

Máté Kormos 谢谢Hong Deng 给予她的帮助，也感谢Lisa Voois提供的建议（她们都来自廷伯根研究所和鹿特丹伊拉斯谟大学伊拉斯谟经济学院）。Stéphanie van der Pas 感谢荷兰研究理事会（NWO）通过 VI.Veni.192.087 号项目资助了本研究。此外，本研究还获得了荷兰研究理事会（NWO）斯宾诺莎基金的部分资助。

1. 引言

实证研究中常采用匹配方法，利用观察性数据估计二元处理的因果效应。估计值是匹配后的处理组和对照组个体在目标结果上的均值差异。匹配可以通过多种方式实现。本文考察基于倾向得分的匹配；倾向得分是给定观察到的处理前协变量时接受处理的条件概率（Rosenbaum and Rubin，1983年）。具体而言，我们考察卡尺匹配：当处理组个体与对照组个体的倾向得分相距不超过某一特定距离（称为卡尺）时，就将其匹配（Cochran and Rubin，1973年；Dehejia and Wahba，1998年）。卡尺匹配应用于劳动经济学（Dehejia and Wahba，2002年；Huber 等人，2015b年）、健康经济学（Erhardt，2017年；Salmasi and Pieroni，2015年；Keng and Sheu，2013年）、政策评估（Bannor 等人，2020年；Patel-Campillo and García，2022年）、商业与金融（Shen and Chang，2009年；Heese 等人，2017年）以及医疗保健（Capogrossi and You，2017年；Cho，2018年；Vecchio 等人，2018年；Izudi 等人，2019年；Wang 等人，2020年；Brenna，2021年；Krishnamoorthy and Rehman，2022年）等领域的实证研究中。尽管如此，关于卡尺选择和估计量极限分布，迄今仍没有严谨的理论结果。

我们的贡献包括提出一种由理论驱动的卡尺选择方法，推导基于倾向得分的卡尺匹配估计量的渐近分布，并构造渐近置信区间。我们研究平均处理效应（ATE）和处理组平均处理效应（ATT）的估计。我们证明，当卡尺的阶数随样本量n 以适当的速度递减时，即使参数化倾向得分是估计得到的，这两个目标估计量的估计量也渐近无偏，并以$\sqrt{n}$速率渐近服从正态分布。本节其余部分将阐明我们的贡献在相关文献中的定位。

匹配在文献中备受关注，其比较相似个体的思想至少可以追溯到Densen 等人（1952年）；参见Cochran （1953年）。Cochran 和 Rubin （1973年）回顾了当时适用于观察性研究的匹配方法。关于历史研究成果汇编，可参阅Rubin （2006年）；关于全面综述，可参阅Stuart （2010年）。Abadie and Imbens （2006年）提出了一项与本文密切相关的重要结果。他们研究了最近邻匹配，即根据协变量选取与给定个体最接近的个体进行匹配。他们证明，基于协变量的最近邻匹配具有渐近正态性，但只有在对标量变量（如倾向得分）进行匹配时才无偏。关于无偏性的识别结果以及倾向得分匹配的理论基础，则由 M Rosenbaum

和 Rubin （1983年）。Abadie and Imbens （2016）推导了基于估计的参数化倾向得分进行最近邻匹配的一些渐近性质。他们对倾向得分参数的最大似然估计量进行离散化，并证明所得匹配估计量会随着样本量增大而收敛到正态分布，且随着离散化逐渐变细，这一结论也成立。由于他们的方法改变了估计量，这一渐近结果并不等同于基于估计的参数化倾向得分进行最近邻匹配的渐近正态性。相比之下，我们既不改变估计量，也不诉诸离散化论证和双重极限。我们采用样本分割来估计倾向得分，并证明随着样本量增大，基于估计的参数化倾向得分进行的卡尺匹配具有渐近正态性。因此，我们能够构造 ATE 和 ATT 的置信区间，其中心为基于估计倾向得分的卡尺匹配估计量，并且随着样本量增大，这些区间会变得更加可靠。

卡尺匹配最早似乎见于Cochran andRubin （1973年）。文中针对几个特定模型对其进行了分析，并将其与最近邻等其他匹配方法进行了比较。卡尺的选择依据是处理组和对照组中结果变量的方差。 Rosenbaum and Rubin (1985)似乎是最早考虑同时涉及倾向得分和协变量的卡尺匹配的研究者。他们假设倾向得分服从逻辑斯蒂模型，并根据倾向得分的逻辑斯蒂变换进行匹配，即根据协变量的线性函数进行匹配。他们依据处理组和对照组中逻辑斯蒂变换的方差选择卡尺。Cochran and Rubin (1973年) 和 Rosenbaum and Rubin (1985) 的卡尺选择可能会产生足够多的匹配，从而降低卡尺匹配估计量的方差。然而，这些选择并不能使卡尺匹配估计量的偏差趋于零——除非将卡尺与最近邻匹配结合使用；参见下一段——。正如我们在本文中所展示的，要实现这一点，卡尺必须随样本量增大而缩小。

一些作者，包括 Rosenbaum and Rubin (1985)，使用“卡尺匹配”一词来指带有卡尺限制的最近邻匹配：匹配 M 最近的个体，但仅限于其距离在卡尺范围内的情况。另一些作者，例如 Dehejiaand Wahba (1998年)，则用该术语表示匹配卡尺范围内的所有个体，即使它们的权重可能不同。[1] 我们采用后一种方式，并使用统

[1]例如，在卡尺限制情形下，若将 M 取为 n，这两种解释便会一致。由于 M 通常设为一个独立于 n 的常数，因此区分这两种解释是合理的。

一权重，有时也称为半径匹配（Huber 等人，2015a），因为这种方法简单易行。卡尺匹配因此可视为一种采用矩形核、且带宽等于卡尺的核匹配方法。因此，Heckman 等人 (1998年）的开创性研究与我们的工作密切相关；该研究确立了核匹配估计量对处理组平均处理效应的渐近正态性，即使倾向得分是通过非参数方法估计的也是如此——带宽选择问题后来由 Frölich (2005) 进一步研究。然而，他们的结果并不适用于卡尺匹配，因为其要求核函数为利普希茨连续。矩形核不满足这一条件，因此无法得到核匹配估计量的渐近线性展开，而这正是他们论证的关键。 Lee (2018年）的研究思路与此相似。该研究将 Heckman 等人 (1998年）的研究扩展到平均效应以外、更丰富的一组目标估计量，并采用核匹配方法，但同样假设核函数光滑，因此排除了卡尺匹配所用的矩形核。

我们通过采用 Alexander (1987年) 和 Van der Vaart 和 Wellner (1996年中的经验过程理论，克服了矩形核的不光滑性。用经验测度表示匹配数量，使我们能够利用经验测度和经验过程的比率界和尾界来刻画卡尺匹配的渐近行为。此外，我们还可以确立卡尺匹配的效率性质。更高效的估计量具有更小的方差，因而能够得到更窄的置信区间。卡尺匹配的效率取决于待估参数、观测样本、结果变量关于协变量的回归，以及对倾向得分的了解。

首先，我们考虑倾向得分已知的情形。我们证明，如果在样本中只观测到倾向得分而未观测到协变量，或者结果变量关于协变量的回归仅通过倾向得分依赖于协变量，那么卡尺匹配估计量的极限方差满足以下结论：(i) 平均处理效应达到半参数下界；(ii) 处理组平均处理效应达到未知的倾向得分（Hahn，1998年）。后一个结果并非可能达到的最佳结果，因为与平均处理效应不同，当倾向得分已知时，处理组平均处理效应的下界更小（Hahn，1998年）。然而，我们表明，卡尺匹配比 Abadie and Imbens (2006，2016) 所研究的倾向得分最近邻匹配更有效，从而无论样本中是否观测到协变量，也无论结果回归依赖于协变量还是倾向得分，都能为平均处理效应和处理组平均处理效应提供更窄的置信区间。

其次，如果倾向得分未知，但我们假设并估计一个参数化设定，例如逻辑斯蒂模型或概率单位模型，那么卡尺

匹配估计量对两个目标估计量的极限方差通常比倾向得分已知时更大。因此，目前尚不清楚，基于估计倾向得分进行卡尺匹配还是最近邻匹配（Abadie and Imbens，2016）的效率更高。

我们的假设包括倾向得分满足通常的共同支持条件，以及结果变量的条件矩和倾向得分（的密度）满足光滑性条件。我们验证了这些假设适用于倾向得分采用逻辑斯蒂模型或 Probit 模型的情形，也适用于结果变量对协变量的回归为光滑、可能非线性且存在异方差，并且其密度在紧致支撑集上表现良好的情形。

本文其余部分安排如下。在第2节，我们介绍概念框架和卡尺匹配估计量。第3 节介绍我们的贡献、卡尺选择以及该估计量的渐近性质。第4 节作总结。

2. 预备知识

2.1. 框架

我们采用Neyman（1924年) 和 Rubin（1974年提出的潜在结果框架，并假设各个单位之间不存在干扰（稳定单元处理值假设，RosenbaumandRubin（1983年））。令 D 表示处理指示变量，其取值为 1 表示接受处理，取值为 0 表示处于对照组。实值变量 Y^1，Y^0 分别表示接受处理和处于对照组时的潜在结果。根据D，我们只能观测到 $Y = DY^1 + (1-D)Y^0$中的一个，即观测结果。我们关注的目标估计量——平均处理效应和处理组平均处理效应（ATT）——分别定义为

$$\tau := \mathbb{E}\left[Y^1 - Y^0\right], \quad \tau_{\mathrm{t}} := \mathbb{E}\left[Y^1 - Y^0 \,\middle|\, D = 1\right].$$

为了根据观察性数据识别平均处理效应和处理组平均处理效应，我们假设观测到的处理前协变量 X（取值属于 $\mathcal{X} \subset \mathbb{R}^K$）能够解释处理组和对照组单位之间所有系统性差异。形式上，假设给定协变量后，潜在结果与是否接受处理相互独立；这是因果推断中的一项标准假设（Rubin，1974年）。

假设1 (无混杂性)。 $Y^0 \perp\!\!\!\perp D \mid X$ 且 $Y^1 \perp\!\!\!\perp D \mid X$。

令 $\pi(x) := \mathbb{P}(D = 1 \mid X = x)$ 为倾向得分，其条件分布函数为 $F_d(p) := \mathbb{P}(\pi(X) \leq p \mid D = d)$。假设 F_d 满足假设 2。

假设2 （倾向得分分布）。 ***(i)*** F_0, F_1 分别具有密度函数 f_0, f_1。***(ii)*** f_0, f_1 具有相同的紧支集 $[\underline{p}, \bar{p}]$，$0 < \underline{p} < \bar{p} < 1$。***(iii)*** f_0, f_1 在其支撑集上严格为正。***(iv)*** f_0, f_1 在其支撑集上连续。

假设2对倾向得分分布提出了与Abadie and Imbens (2016)相同的要求，但还要求密度函数 f_0， f_1 严格为正。此要求可确保分位数函数F_d^{-1} 具有有界导数，我们利用这一点进行卡尺选择。它还在渐近正态性的证明中发挥作用，确保经验过程的比值界适用。[2]

假设2 意味着，若某个区域$[0,1]$, 内存在倾向得分处于该区域的个体，则以正概率能在其中找到来自另一处理组的个体。这确保了可以根据倾向得分对处理组和对照组个体进行比较。结合假设1，通过比较倾向得分相同的处理组和对照组个体，即可利用观测变量识别目标估计量（Rosenbaum andRubin，1983年）：

$$\tau = \mathbb{E}\left[\mathbb{E}\left[Y \mid D=1, \pi(X)\right] - \mathbb{E}\left[Y \mid D=0, \pi(X)\right]\right], \tag{1}$$

$$\tau_{\mathrm{t}} = \mathbb{E}\left[\mathbb{E}\left[Y \mid D=1, \pi(X)\right] - \mathbb{E}\left[Y \mid D=0, \pi(X)\right] \mid D=1\right]. \tag{2}$$

2.2. 卡尺匹配估计量

我们希望基于识别公式（1）和（2），利用来自分布$(Y,\ D,\ X)$的独立同分布（i.i.d.）样本构造估计量，记为 $((Y_i,\ D_i,\ X_i))^{i\in[n]}$，其中 $[n] :=\{1,2,\ldots,n\}$。这需要找到倾向得分取值相同的样本单位，但对于连续分布的倾向得分而言，这是不可行的。因此，匹配估计量转而寻找倾向得分相近的样本单位。卡尺匹配估计量通过卡尺 δ明确控制相似程度，其取值将在第 3 节后续讨论。

暂且假设倾向得分已知。给定 $\delta > 0$，卡尺匹配估计量构造匹配集 $\mathcal{J}(i) :=\{j \in [n] :\ D_j \neq D_i, |\pi(X_j) - \pi(X_i)| \leq \delta\}$ 个体 $i \in [n]$。接下来，它使用匹配集中个体的平均结果变量来估计该个体缺失的潜在结果。然后，对（估计的）潜在结果之差取平均，即可得到因果效应的估计值。令 $M_i := |\mathcal{J}(i)|$ 表示

[2] f_d 的严格正性意味着 $\inf_{p\in[\underline{p},\bar{p}]} \int_{p-\delta}^{p+\delta} f_d(\tilde{p})\mathrm{d}\tilde{p} \gtrsim \delta > 0$，因此经验测度与真实测度之比的分母有正下界，从而使这些比值保持有限。

个体 $i \in [n]$ 的匹配数量，并记 $N_0 := \sum_{i\in[n]}(1-D_i)$、$N_1 := \sum_{i\in[n]} D_i$ 分别表示对照组和接受处理者的数量。ATE 和 ATT 的估计量分别定义为

$$\hat{\tau}_\pi := \frac{1}{n}\sum_{i\in[n]}\left[D_i\left(Y_i - \frac{1}{M_i}\sum_{j\in\mathcal{J}(i)} Y_j\right) + (1-D_i)\left(\frac{1}{M_i}\sum_{j\in\mathcal{J}(i)} Y_j - Y_i\right)\right] \mathbb{I}^{M_i>0},$$

$$\hat{\tau}_{\mathrm{t},\pi} := \frac{1}{N_1}\sum_{i\in[n]} D_i\left(Y_i - \frac{1}{M_i}\sum_{j\in\mathcal{J}(i)} Y_j\right) \mathbb{I}^{M_i>0}.$$

指标 $\mathbb{I}^{M_i>0}$ 在单位 i 存在匹配时取 1，否则取 0，从而确保估计中仅包含存在匹配的单位。

在实践中，倾向得分通常是未知的。通常假设它遵循光滑的参数化模型，例如逻辑斯蒂模型或概率单位模型。遵循 Abadie and Imbens (2016) 的做法，我们也作出这一假设。

假设*3*（光滑参数化倾向得分）。*(i)* 对于参数模型 $\{\pi(\cdot,\theta)\colon\ \theta \in \Theta \subset \mathbb{R}^K\}$，倾向得分为 $\mathbb{P}(D=1\mid X) = \pi(X,\theta_0)$，其中 θ_0 位于 Θ 的内部。*(ii)* 对于所有 $x \in \mathcal{X}$，$\theta \mapsto \pi(x,\theta)$ 在 θ_0 的邻域内可微。*(iii)* 假设*3(ii)*中的导数在 θ_0 的邻域内关于 $x \in \mathcal{X}$ 一致有界。

卡尺匹配估计量随后通过代入规则定义。令 $\mathcal{J}_\theta(i) := \{j \in [n] : D_j \neq D_i, |\pi(X_j,\theta) - \pi(X_i,\theta)| \le \delta\}$ 为匹配集，$M_i(\theta) := |\mathcal{J}_\theta(i)|$ 为其基数 ˆ，其中某个 $\theta \in \Theta$。对于 θ 这一 θ_0 的估计量，平均处理效应和处理组平均处理效应的匹配估计量分别为，

$$\hat{\tau}_{\hat{\pi}} := \frac{1}{n}\sum_{i\in[n]}\left[D_i\left(Y_i - \frac{1}{M_i(\hat{\theta})}\sum_{j\in\mathcal{J}_{\hat{\theta}}(i)} Y_j\right) + (1-D_i)\left(\frac{1}{M_i(\hat{\theta})}\sum_{j\in\mathcal{J}_{\hat{\theta}}(i)} Y_j - Y_i\right)\right] \mathbb{I}^{M_i(\hat{\theta})>0},$$

$$\hat{\tau}_{\mathrm{t},\hat{\pi}} := \frac{1}{N_1}\sum_{i\in[n]} D_i\left(Y_i - \frac{1}{M_i(\hat{\theta})}\sum_{j\in\mathcal{J}_{\hat{\theta}}(i)} Y_j\right) \mathbb{I}^{M_i(\hat{\theta})>0}.$$

3. 渐近性质

本节陈述我们的主要结果：卡尺选择（第 3.1 节）、倾向得分已知（第 3.2 节）和估计得到（第 3.2 节）时，ATE 和 ATT 的卡尺匹配估计量的渐近正态性，以及方差估计（第 3.4 节）。

3.1. 卡尺选择

较小的卡尺意味着匹配的处理组和对照组个体的倾向得分更接近，因此匹配质量更好。与此同时，较小的卡尺会减少匹配数量。因此，卡尺直接控制匹配质量，并间接控制匹配数量，而这两者又决定了匹配估计量的性质。匹配质量决定偏差：比较差异较大的个体会威胁到（1）和（2）中目标估计量的识别。匹配数量通过排除没有匹配的个体来决定偏差，同时也决定估计量的方差：由于估计量涉及对匹配集取平均，匹配集较小会导致方差较大。

因此，恰当的卡尺选择必须兼顾匹配质量和匹配数量。随着样本量增加，我们预期在假设 2下，以不断增大的概率，可以在 $[\underline{p}, \bar{p}]$ 的每个区域中找到处理组和对照组个体。因此，在大样本极限下，力求为每个个体找到匹配对象是合理的。如果将卡尺设为 $\overline{\underline{\Delta}}_n := \max^{i\in[n]}\min_{j\in[n]:D_j\neq D_i}|\pi(X_i) - \pi(X_j)|$，即处理组和对照组个体之间最大的最近距离，那么每个个体至少都有一个匹配对象。借助 $\mathbb{E}\overline{\underline{\Delta}}_n$ 的结果，可以用样本量简洁地描述其阶数；相关结果依据 Shorack 和 Wellner (2009年年关于间距的研究（所有证明见附录 B 及补充材料）。

命题 1 (最大最近距离期望的阶)。 在假设 2 下，存在常数 $0 < n_0$， $c < \infty$ 使得 $\mathbb{E}\overline{\underline{\Delta}}_n \leq c\frac{\log n}{n}$ 对所有 $n \geq n_0$ 均成立。

这表明，对于 $n \geq 2$，卡尺选择

$$\delta := \delta_n := s\frac{\log n}{n} \quad \text{or} \quad \delta := \delta_n := \overline{\underline{\Delta}}_n \vee \frac{\log N_0}{N_0+1} \vee \frac{\log N_1}{N_1+1} \tag{3}$$

对于任意固定常数 $s > 0$，其渐近阶相同，且足够大，能够保证每个个体都有匹配；不过，数据依赖选择 $\delta_n = \overline{\underline{\Delta}}_n \vee \log\frac{N_0}{N_0+1} \vee \log\frac{N_1}{N_1+1}$ 能够更好地适应较小的样本，因此通常更受青睐。事实上，命题 2 表明，实际隐含的匹配数量阶为 $\log n$。

命题 2 (匹配数量)。 设卡尺满足 (3)。若假设 2 成立，则存在常数 $0 < c_l$， $c_u < \infty$ 使得

$$c_l(1 + o_P(1))\log n \leq \min_{i\in[n]} M_i \leq \max_{i\in[n]} M_i \leq c_u(1 + o_P(1))\log n$$

当 $n \to \infty$ 时。因此， $\mathbb{P}\left(\text{最小}_{i\in[n]} M_i \geq 1\right) \to 1$ 当 $n \to \infty$ 时。

3.2. 已知倾向得分

暂且假设倾向得分 $x \mapsto \pi(x)$ 已知。在这种情况下，我们推导卡尺匹配的渐近分布，并证明，卡尺选择（3）不仅会使匹配数量随样本量增加，还能使匹配估计量渐近无偏。

下面，我们提出一系列假设，这些假设共同保证卡尺匹配具有渐近正态性，并证明例如1 中的模型满足这些假设。常见模型，包括倾向得分的逻辑斯蒂模型和概率单位模型，以及平滑的异方差结果回归，均涵盖在1 中，只要协变量具有良好的密度。[3] 要求 $K \geq 2$ 协变量连续分布且倾向得分参数非零，这一条件并不严格；因为如果只有一个协变量，那么按倾向得分匹配与按协变量匹配几乎没有区别。[4] 关于例1的其他条件，$\nu_d \perp\!\!\!\perp D \mid X$ 蕴含假设1，而 $x \mapsto \mathbb{E}[\nu_d^2 \mid X = x]$ 的可微性则允许使用平滑的异方差模型。

例1 (可容许模型)。 设 $g\colon \mathbb{R} \to [0,1]$ 是一个严格递增函数，在 $\mathbb{R}$上二阶连续可微，其一阶导数 g' 满足 $\sup_{t\in\mathbb{R}} g'(t) < \infty$。这些$(K \geq 2)$维协变量的密度为 Ψ，该密度在紧支集 $\mathcal{X}$ 上严格为正且连续可微。倾向得分和潜在结果满足

$$\pi(x) = g(\theta_0^\intercal x)$$
$$Y^d = m_d(X) + \nu_d, \quad \mathbb{E}[\nu_d \mid X] = 0, \quad d \in \{0,1\},$$

其中， θ_0 位于 $\Theta \subset \mathbb{R}^K$的内部，且至少有两个非零坐标，$\Theta$ 有界，并且 m_d 连续可微。对于所有 $d \in \{0,1\}$、 $\nu_d \perp\!\!\!\perp D \mid X$， $x \mapsto \mathbb{E}[\nu_d^r \mid X = x]$、 $r \in \{2,4\}$在 $\mathcal{X}$上连续可微，且下确界$_{x\in\mathcal{X}} \mathbb{E}[\nu_d^2 \mid X = x] > 0$。

[3]为简化表述，我们假设全文中 X 不包含截距。通过重新定义关于 X 的分布假设，使其针对 X的非截距坐标，即可纳入截距。[4]在假设 2、4 和 5 中以标量协变量替代倾向得分，将得到命题 1 和 2 以及定理 1 和 2 的一个版本，其中倾向得分被协变量替代。

我们可以将$\hat{\tau}_{\pi}$、$\hat{\tau}_{t,\pi}$改写为结果变量Y的加权平均，如下所示：

$$\hat{\tau}_{\pi} = \frac{1}{n}\sum_{i\in[n]}(2D_i-1)(\mathbb{I}_{M_i>0}+w_i)Y_i, \quad \hat{\tau}_{t,\pi} = \frac{1}{N_1}\sum_{i\in[n]}(\mathbb{I}_{M_i>0}D_i-(1-D_i)w_i)Y_i,$$
$$w_i := \sum_{j\in\mathcal{J}(i)}\frac{1}{M_j},$$

其中$M_j = 0$仅当$\mathcal{J}(i)$为空时才成立，此时将w_i中的求和视为零。[5]

令$\mu^d(p) := \mathbb{E}[Y \mid D = d, \pi(X) = p]$为回归函数，$\varepsilon := Y - \mu^D(\pi(X))$为相应的扰动项，其条件方差为

$$\sigma_d^2(p) := \mathbb{V}[\varepsilon \mid D = d, \pi(X) = p] = \mathbb{V}[Y \mid D = d, \pi(X) = p], \quad d \in \{0,1\}.$$

当我们应用卡尺匹配来模拟（1）和（2）时，需要进行两项近似。首先，我们比较各单位的结果变量Y，而不是回归函数$\mu^D(\pi(X))$。由此产生的误差为ε。其次，我们比较倾向得分相近而非相同的单位。因此，必须对其大小ε以及μ^d的光滑性作出一些假设。ε的大小不能过大，但为避免极限退化，出于方便也不能过小。假设4和5与Abadie and Imbens (2006年中的假设4相同，只是将匹配对象从协变量$\pi(X)$改为倾向得分X。

假设4（扰动项）。 (*i*) σ_d^2满足下确界$_{d\in\{0,1\},\ p\in[\underline{p},\bar{p}]}\sigma_d^2(p) > 0$且上确界$_{d\in\{0,1\},\ p\in[\underline{p},\bar{p}]}\sigma_d^2(p) < \infty$. (*ii*) 上确界$_{d\in\{0,1\},\ p\in[\underline{p},\bar{p}]}\mathbb{E}[\varepsilon^4 \mid D = d, \pi(X) = p] < \infty$。

假设5（Lipschitz回归函数）。 μ^d是*Lipschitz*连续的：存在常数$0 < L_\mu < \infty$，使得$|\mu^d(p) - \mu^d(p')| \le L_\mu|p - p'|$对所有$p$，$p' \in [\underline{p}, \bar{p}]$以及所有$d \in \{0,1\}$均成立。

利普希茨连续性保证，当倾向得分$\pi(X_i)$和$\pi(X_j)$足够接近时（我们用δ_n控制这一点），$\mu^d(\pi(X_i))$与$\mu^d(\pi(X_j))$也会足够接近。这与识别公式(1)和(2)一致，从而带来渐近无偏性。与.....类似

[5]这源于卡尺匹配的对称性：当且仅当$i \in \mathcal{J}(j)$时，$j \in \mathcal{J}(i)$。

Abadie and Imbens (2006年)，将平均处理效应估计量写为

$$\hat{\tau}_\pi = \overline{\tau(\pi(X))} + E + B, \tag{4}$$

$$\overline{\tau(\pi(X))} := \frac{1}{n}\sum_{i\in[n]} \tau(\pi(X_i)), \quad \tau(\pi(X_i)) := \mu^1(\pi(X_i)) - \mu^0(\pi(X_i)), \tag{5}$$

$$E := \frac{1}{n}\sum_{i\in[n]} E_i, \quad E_i := (2D_i - 1)(\mathbb{I}_{M_i>0} + w_i)\varepsilon_i, \tag{6}$$

$$B := \frac{1}{n}\sum_{i\in[n]} B_i, \tag{7}$$

$$B_i := (2D_i - 1)\frac{\mathbb{I}_{M_i>0}}{M_i}\sum_{j\in\mathcal{J}(i)} (\mu^{1-D_i}(\pi(X_i)) - \mu^{1-D_i}(\pi(X_j))) + (2D_i - 1)(\mathbb{I}_{M_i>0} - 1)(\mu^{1-D_i}(\pi(X_i)) - \mu^{D_i}(\pi(X_i))). \tag{8}$$

第一项 $\overline{\tau(\pi(X))}$ 的均值为 τ，第二项 E 的均值为零。以 τ 为中心化点后，将证明前两项以 $\sqrt{n}$ 阶速率渐近联合正态且相互独立。第三项 B 存在两个偏差来源。(8) 式中的第一项偏差源于不完全匹配。如果匹配完全精确，该项将为零。根据假设5，该项的量级为 δ_n，因此即使乘以 $\sqrt{n}$，它也趋于零。(8) 式中的第二项是由于舍弃未匹配单位所致；采用卡尺选择 $\delta_n = s\frac{\log n}{n}$ 时可能发生这种情况，而采用数据依赖选择 $\delta_n = \overline{\Delta}_n \vee \log_{\frac{N_0}{N_0+1}} \vee \log_{\frac{N_1}{N_1+1}}$ 时则不会。这会导致偏差，因为我们根据 δ_n 引入了人为的样本选择。如果每个单位至少有一个匹配，就像数据依赖的卡尺选择那样，该项便为零。但是，正如命题 2 所示，这种情况会在大样本极限下发生，从而得到平均处理效应估计量 $\hat{\tau}_\pi$ 的渐近正态性。

定理1 （已知倾向得分下的渐近正态性（平均处理效应））。 假设$x \mapsto \pi(x)$ 已知，且卡尺 δ_n 满足 (3)。若假设 *1*、*2*、*4* 和 *5* 均成立，则

$$\sqrt{n}(\hat{\tau}_\pi - \tau) \rightsquigarrow \mathcal{N}(0, V) \quad as\ n \to \infty,$$

其中 $V := V_\tau + V_{\sigma,\pi}$ 满足 $V_\tau := \mathbb{E}[(\tau(\pi(X)) - \tau)^2]$ 且 $V_{\sigma,\pi} := \mathbb{E}\left[\frac{\sigma_0^2(\pi(X))}{1-\pi(X)} + \frac{\sigma_1^2(\pi(X))}{\pi(X)}\right]$。

Abadie and Imbens (2006年)证明，只有在标量协变量上进行匹配时，最近邻匹配才是渐近无偏的。卡尺匹配也非常相似。如果我们要在 K 维协变量上进行匹配，类似的论证表明，在正则条件下，$\sqrt{n}(\hat{\tau}_\pi - \tau)$ 的偏差阶为 $\sqrt{n}(\delta_n + \mathbb{I}_{\{\exists i\in[n]:\ M_i=0\}})$，而匹配数量的阶为 $n\delta_n^K$。这样一来，就不可能同时实现

对于 $K \geq 2$ 而言足够好的匹配质量和足够多的匹配，因此偏差 B 不会消失。因此，关键在于对标量倾向得分进行匹配。这样做时，ATT估计量 $\hat{\tau}_{\mathrm{t},\pi}$ 也渐近正态。

定理 2（已知倾向得分下的渐近正态性（处理组平均处理效应））。假设$x \mapsto \pi(x)$已知，且卡尺 δ_n 满足（3）。令 $p_1 := \mathbb{E}\pi(X)$ 若假设1、2、4 和5 均成立，则

$$\sqrt{n}(\hat{\tau}_{\mathrm{t},\pi} - \tau_{\mathrm{t}}) \rightsquigarrow \mathcal{N}(0, V_{\mathrm{t}}) \quad as\ n \to \infty,$$

其中 $V_{\mathrm{t}} := V_{\tau_{\mathrm{t}}} + V_{\mathrm{t},\sigma,\pi}$ 其中 $V_{\tau_{\mathrm{t}}} := \frac{1}{p_1^2}\mathbb{E}[\pi(X)(\tau(\pi(X)) - \tau_{\mathrm{t}})^2]$ 且

$$V_{\mathrm{t},\sigma,\pi} := \frac{1}{p_1^2}\mathbb{E}\left[\frac{\pi(X)^2\sigma_0^2(\pi(X))}{1-\pi(X)} + \pi(X)\sigma_1^2(\pi(X))\right].$$

为考察 $\hat{\tau}_{\pi}$ 和 $\hat{\tau}_{\mathrm{t},\pi}$的效率，令

$$\mu_{\mathcal{X}}^d(x) := \mathbb{E}[Y \mid D = d, X = x] \text{ and } \sigma_{\mathcal{X},d}^2(x) := \mathbb{V}[Y \mid D = d, X = x]$$

对于 $d \in \{0,1\}$。ATE 的半参数效率界为

$$V_{\mathrm{eff}} := \mathbb{E}\left[(\mu_{\mathcal{X}}^1(X) - \mu_{\mathcal{X}}^0(X) - \tau)^2 + \frac{\sigma_{\mathcal{X},0}^2(X)}{1-\pi(X)} + \frac{\sigma_{\mathcal{X},1}^2(X)}{\pi(X)}\right], \tag{9}$$

无论倾向得分是否已知（Hahn (1998年，定理1和定理2)）。处理组平均处理效应的半参数效率界为

$$V_{\mathrm{t,eff},\pi} := \frac{1}{p_1^2}\mathbb{E}\left[(\mu_{\mathcal{X}}^1(X) - \mu_{\mathcal{X}}^0(X) - \tau_{\mathrm{t}})^2\pi(X)^2 + \frac{\pi(X)^2\sigma_{\mathcal{X},0}^2(X)}{1-\pi(X)} + \pi(X)\sigma_{\mathcal{X},1}^2(X)\right]$$

如果倾向得分已知，则

$$V_{\mathrm{t,eff}} := \frac{1}{p_1^2}\mathbb{E}\left[(\mu_{\mathcal{X}}^1(X) - \mu_{\mathcal{X}}^0(X) - \tau_{\mathrm{t}})^2\pi(X) + \frac{\pi(X)^2\sigma_{\mathcal{X},0}^2(X)}{1-\pi(X)} + \pi(X)\sigma_{\mathcal{X},1}^2(X)\right]$$

当倾向得分未知时（Hahn（1998年，定理1和定理2））。V $\hat{\tau}_{\pi}$ 的极限方差类似于效率界 V_{eff}，但 V 涉及给定倾向得分 $\pi(X)$的结果变量条件矩，而不是像 V_{eff}中那样涉及给定协变量 X的条件矩。因此，如果我们的样本中只能观测到 $\pi(X)$，而非 X，那么$\hat{\tau}_{\pi}$ 将具有半参数效率，达到 V_{eff}。根据定理1和式（9）也可直接得出，如果对所有 $d \in \{0,1\}$，我们有 $\mu_{\mathcal{X}}^d(X) = \mu^d(\pi(X))$且$\sigma_{\mathcal{X},d}^2(X) = \sigma_d^2(\pi(X))$——也就是说，给定协变量的结果变量条件矩仅取决于倾向得分——那么平均处理效应估计量$\hat{\tau}_{\pi}$ 也将具有半参数效率。事实上，命题 3 给出了一个更精确的结果。

命题 3 (半参数效率)。 假设 *1* 成立。那么$V_{\text{eff}} \leq V$和$V_{\text{t, eff}} \leq V_{\text{t}}$，且当且仅当以下条件成立时，两种情形下等号均成立

$$\mu_{\mathcal{X}}^{D}(X) = \mu^{D}(\pi(X)) \quad almost\ surely. \tag{10}$$

假设（10）在命题 3 中成立。即便如此，与平均处理效应估计量$\hat{\tau}_{\pi}$相比，处理组平均处理效应估计量$\hat{\tau}_{\text{t, }\pi}$只能达到$V_{\text{t, eff}}$即未知倾向得分下的半参数效率界，该界大于$V_{\text{t, eff, }\pi}$已知倾向得分下的效率界。在（10）条件下，两者之差为

$$V_{\text{t,eff}} - V_{\text{t,eff,}\pi} = \frac{1}{p_1^2}\mathbb{E}\left[\pi(X)(1-\pi(X))(\tau(\pi(X)) - \tau_{\text{t}})^2\right] \geq 0. \tag{11}$$

由于 $\pi(X)(1-\pi(X)) \leq 1/2$，差值被$\frac{1}{2p_1^2}\mathbb{E}\left[(\tau(\pi(X)) - \tau_{\text{t}})^2\right]$所界定。因此，在不同 $\pi(X)$ 之间的处理效应越同质（等价地，在（10）条件下，在X之间越同质），且处理组 D 越同质，差值就越小。

效率损失（11）并非卡尺匹配所特有。事实上，定理 2 中 ATT 估计量的极限方差低于Abadie and Imbens (2016，命题 1）中最近邻匹配估计量的极限方差。两者之差为

$$\frac{1}{2Mp_1^2}\mathbb{E}\left[\sigma_0^2(\pi(X))\pi(X)\left(2 + \frac{\pi(X)}{1-\pi(X)}\right)\right] \geq 0, \tag{12}$$

其中 常数 M 表示要匹配的最近邻数量。这表明，对于较大的，卡尺匹配的效率增益（12）较小。然而，并没有证据表明令 趋于无穷大就能弥合差距，因为 Abadie and Imbens (2016 的结果以固定的 M 为前提。相比之下，采用定理 2 所述的卡尺选择时，根据命题 2，卡尺匹配的匹配数量趋于无穷大，从而降低方差。除非 $p \mapsto \sigma_0^2(p)$ 在 1 附近迅速下降（这被假设 4(i) 排除），否则当倾向得分趋近于 1 时，（12）会更大。在这种情况下，使用卡尺匹配而非最近邻匹配带来的收益更大，尽管此时$V_{\text{t、}\sigma\text{、}\pi}$，因而 V_{t} 也会增大。

最后，对于已知倾向得分的情形，我们通过验证示例 1 中模型的定理 1 和 2 所需的假设，完成论证。

命题 4 (可容许模型（已知倾向得分）)。 示例 *1* 中描述的模型族满足假设 *1*、*2*、*4* 和 *5*。

3.3. 估计倾向得分

假设倾向得分 $\pi(\cdot, \theta_0)$ （假设 3 ）是估计得到的。一个合理的 θ_0 估计量将收敛于 θ_0。因此，我们预期，如果假设 2，

4 和 5 在 θ_0的邻域内成立，那么基于估计倾向得分的卡尺匹配估计量也将渐近正态，前提是它们关于 θ足够光滑。

为此，我们要求条件分布 $F_{d,\,\theta}(p) := \mathbb{P}_{\theta_0}(\pi(X,\theta) \le p \mid D = d)$ 仅在局部上与真实倾向得分的分布相似。扩展假设 2，我们需要密度函数 $f_{0,\,\theta}, f_{1,\,\theta}$ 不仅连续且可微，而且关于 θ 平滑变化。对于某个任意固定常数 $\epsilon > 0$，令 $\mathrm{Nb}(\theta_0,\epsilon) := \{\theta \in \Theta : \|\theta - \theta_0\| < \epsilon\}$ 表示 θ_0 的一个邻域，并进一步令

$$\mathcal{S}_{\theta_0,\epsilon} := \left\{(\theta, p) : p \in [\underline{p}_\theta, \bar{p}_\theta], \theta \in \mathrm{Nb}(\theta_0,\epsilon)\right\}.$$

假设6（参数化倾向得分的分布）。 *(i)* $F_{0,\,\theta}$， $F_{1,\,\theta}$ 分别具有密度函数 $f_{0,\,\theta}, f_{1,\,\theta}$，对所有 $\theta \in \mathrm{Nb}(\theta_0,\epsilon)$均成立。
(ii) $f_{0,\theta}, f_{1,\theta}$ 具有相同的支撑集 $[\underline{p}_\theta, \bar{p}_\theta]$，其中 $0 < \underline{p}_\theta < \bar{p}_\theta < 1$ 对所有 $\theta \in Nb(\theta_0,\epsilon)$ 均成立。*(iii)* $f_{0,\theta}, f_{1,\theta}$ 均有正的下界：$inf_{\theta\in\mathrm{Nb}(\theta_0,\epsilon)} inf_{p\in[\underline{p}_\theta,\bar{p}_\theta]} f_{d,\theta}(p) > 0$ 对所有 $d \in \{0,1\}$均成立。*(iv)* $(\theta,p) \mapsto f_{d,\theta}(p)$ 在 $\mathcal{S}_{\theta_0,\epsilon}$ 上连续可微，对所有 $d \in \{0,1\}$均成立。

接下来，我们以一种取决于倾向得分参数 θ 的方式分解结果变量。我们不需要假设5所要求的连续性，而是要求回归函数 $\mu^d(\theta,p) := \mathbb{E}[Y \mid D = d, \pi(X,\theta) = p]$ 连续可微，且对 θ 也连续可微。结合假设3，假设7意味着 $\theta \mapsto \mu^d(\theta, \pi(x,\theta))$ 在 θ_0 的邻域内可以用误差阶为 $\|\theta - \theta_0\|$ 的近似表示。具体而言，它们意味着，对所有 $(\tilde{\theta}, x) \in \mathrm{Nb}(\theta_0, \epsilon) \times \mathcal{X}$，导数 $\theta \mapsto \mu^d(\theta, \pi(x,\theta))$ 均存在，且形式为 $\Lambda^d(\tilde{\theta}, x) := \frac{\partial \mu^d}{\partial \theta^\intercal}(\tilde{\theta}, \pi(x,\tilde{\theta})) + \frac{\partial \mu^d}{\partial p}(\tilde{\theta}, \pi(x,\tilde{\theta}))(\mathrm{D}_\theta \pi)(x,\tilde{\theta})$。

假设7 (回归函数的可微性)。 θ, p和 $\mapsto \mu^d(\theta,p)$ 在$\mathcal{S}_{\theta_0,\epsilon}$ 上连续可微，其偏导数 $\frac{\partial \mu^d}{\partial \theta}: \Theta \times [0,1] \to \mathbb{R}^K$ 和$\frac{\partial \mu^d}{\partial p}: \Theta \times [0,1] \to \mathbb{R}$ 在$\mathcal{S}_{\theta_0,\epsilon}$ 上对所有 $d \in \{0,1\}$均一致有界。

为确保光滑性并控制扰动项$\varepsilon_i(\theta) := Y_i - \mu^{D_i}(\theta, \pi(X_i,\theta))$ $i \in [n]$的大小，我们要求函数

$$\sigma_d^r(\theta,p) := \mathbb{E}\left[(Y - \mu^D(\theta,p))^r \,\middle|\, D = d, \pi(X,\theta) = p\right], \quad r \in \{2,4\}, d \in \{0,1\},$$

满足以下条件。

假设8（光滑参数化扰动项）。*(i)* 对所有 $p \in [\underline{p}_\theta, \bar{p}_\theta]$， σ_d^2 满足利普希茨条件 $|\sigma_d^2(\theta,p) - \sigma_d^2(\theta',p')| \le L_\sigma(\|\theta - \theta'\| + |p - p'|)$；并且对所有 θ， $p' \in [\underline{p}_{\theta'}, \bar{p}_{\theta'}]$ 满足 $Nb(\theta_0,\epsilon)$，其中常数 $0 < L_\sigma < \infty$，且下界 $inf_{p\in[\underline{p}_{\theta_0},\bar{p}_{\theta_0}]} \sigma_d^2(\theta_0,p) > 0$ 对所有 $d \in \{0,1\}$均成立。*(ii)* 对所有 $d \in \{0,1\}$，σ_d^4满足条件 $sup_{\theta\in\mathrm{Nb}(\theta_0,\epsilon)} sup_{p\in[\underline{p}_\theta,\bar{p}_\theta]} \sigma_d^4(\theta,p) < \infty$。

$\hat{}$ 最后，我们需要倾向得分参数的估计量 θ 以适当的方式收敛到 $\hat{}\theta_0$ 。例如，如果 θ 是最大似然估计量，那么在正则条件下，它会以适当的方式收敛。我们进一步假设， θ_0 是根据一个独立于 $((Y_i, D_i, X_i))^{i\in[n]}$ 的样本估计的。实践中，可以采用样本拆分来确保独立性：将一个规模为 $2n$ 的大样本分成两半，用第一半估计 θ_0，再将所得估计量 θ 代入第二半，以计算 $\hat{}\tau_{\hat{\pi}}$,$\hat{}\tau_{\mathrm{t},\ \hat{\pi}}$。

$\hat{}$**假设9** (倾向得分参数的估计量)。 ***(i)*** θ 渐近服从正态分布，且 $\sqrt{n}(\hat{}\theta - \theta_0) \rightsquigarrow \mathcal{N}(0, V_{\theta_0})$ 当 $n \to \infty$ 时，其中$\hat{}V_{\theta_0}$是有限可逆矩阵。***(ii)*** θ 与用于计算匹配估计量的数据集独立：$\hat{}\theta \perp\!\!\!\perp ((Y_i, X_i, D_i))^{i\in[n]}$。

为纳入倾向得分估计，我们引入

$$\underline{\widehat{\Delta}}_n := \max_{i\in[n]} \min_{j\in[n]:D_j\neq D_i} |\pi(X_i, \hat{\theta}) - \pi(X_j, \hat{\theta})|,$$

$\underline{\Delta}_n$的估计对应量，以及相应的卡尺选择

$$\delta_n := s\frac{\log n}{n} \quad \text{or} \quad \delta_n := \underline{\widehat{\Delta}}_n \vee \frac{\log N_0}{N_0+1} \vee \frac{\log N_1}{N_1+1} \tag{13}$$

对任意固定常数 $s > 0$。命题 5 表明，基于估计倾向得分和卡尺选择（13）的匹配数量也与命题 $\log n$ 2中的数量同阶。由此得到定理 3 和 4，证明了基于估计倾向得分的卡尺匹配具有渐近正态性。

命题5 (估计倾向得分的匹配数量)。 假设卡尺 δ_n 满足 (13)。若假设 *6* 和 *9* 成立，则存在常数 $0 < \bar{c}_l, \bar{c}_u < \infty$ 使得

$$\bar{c}_l(1 + o_P(1)) \log n \leq \min_{i\in[n]} M_i(\hat{\theta}) \leq \max_{i\in[n]} M_i(\hat{\theta}) \leq \bar{c}_u(1 + o_P(1)) \log n$$

当 $n \to \infty$ 时。因此，$\mathbb{P}\left(\text{最小值}_{i\in[n]}\, M_i(\hat{}\theta) \geq 1\right) \to 1$ 当 $n \to \infty$ 时。

定理 3 (估计倾向得分（平均处理效应）的渐近正态性)。 假设卡尺 δ_n 满足 (13)。若假设 *1*、*3* 和 *6* 至 *9* 均成立，则

$$\sqrt{n}(\hat{\tau}_{\hat{\pi}} - \tau) \rightsquigarrow \mathcal{N}(0, V_{\hat{\pi}}) \quad as\ n \to \infty,$$

其中 $V_{\hat{\pi}} := V_\tau + V_{\sigma,\ \pi} + (q_1 - q_0)^\intercal V_{\theta_0}(q_1 - q_0)$ 对于 V_τ， $V_{\sigma,\ \pi}$定理 *1*中的 V_{θ_0}假设 *9*，以及 $q_d \in \mathbb{R}^K$ 作为依概率极限得到的 $\frac{1}{n}\sum_{i\in[n]} \Lambda^d(\hat{}\theta,\ X_i)$ 当 $n \to \infty$ 对于$d \in \{0, 1\}$。

定理 4（估计倾向得分下处理组平均处理效应的渐近正态性）。 假设卡尺 δ_n 满足 (13)。如果假设 1、3 和 6 至 9 均成立，则

$$\sqrt{n}(\hat{\tau}_{\mathrm{t},\hat{\pi}} - \tau_{\mathrm{t}}) \rightsquigarrow \mathcal{N}(0, V_{\mathrm{t},\hat{\pi}}) \quad as\ n \to \infty,$$

其中 $V_{\mathrm{t},\hat{\pi}} := V_{\tau_{\mathrm{t}}} + V_{\mathrm{t},\sigma,\pi} + (1/p_1^2)(q_{\mathrm{t},1} - q_{\mathrm{t},0})^{\intercal} V_{\theta_0} (q_{\mathrm{t},1} - q_{\mathrm{t},0})$，对于 $V_{\tau_{\mathrm{t}}}$, $V_{\mathrm{t},\sigma,\pi}$ 以及定理 2 的 p_1、假设9 的 V_{θ_0}，还有 $q_{\mathrm{t},d} \in \mathbb{R}^K$，它们是作为 $\frac{1}{n}\sum_{i\in[n]} D_i \Lambda^d(\hat{\theta}, X_i) \xrightarrow{P} q^{\intercal}_{\mathrm{t}\,d}$ 在 $n \to \infty$ 下针对 $d \in \{0,1\}$ 的依概率极限而产生的。

与定理1和定理2相比，平均处理效应估计量 $\hat{\tau}_{\hat{\pi}}$ 的方差增加了 $(q_1 - q_0)^{\intercal} V_{\theta_0} (q_1 - q_0)$，而处理组平均处理效应估计量 $\hat{\tau}_{\mathrm{t},\hat{\pi}}$ 的方差增加了 $(1/p_1^2)(q_{\mathrm{t},1} - q_{\mathrm{t},0})^{\intercal} V_{\theta_0} (q_{\mathrm{t},1} - q_{\mathrm{t},0})$，这反映了倾向得分估计所带来的不确定性。倾向得分估计得越精确，V_{θ_0} 就越小，因此差异也越小。或者，如果 $q_1 \approx q_0$ 或 $q_{\mathrm{t},1} \approx q_{\mathrm{t},0}$，那么相应的增量也较小。如果导数 Λ^1 和 Λ^0 彼此接近，就会出现这种情况；然而，即使对于例 1 中的简单线性回归，也很难判断这种情况是否会发生以及何时发生。因此，当估计参数化倾向得分时，卡尺匹配或最近邻匹配（Abadie and Imbens，2016）哪种效率更高，仍不明确。

注 1（方差比较）。如果仅使用一半的大样本来计算 $\hat{\tau}_{\pi}$, $\hat{\tau}_{\mathrm{t},\pi}$ 和卡尺 (3)，那么 $\hat{\tau}_{\pi}$, $\hat{\tau}_{\mathrm{t},\pi}$ 与 $\hat{\tau}_{\hat{\pi}}$, $\hat{\tau}_{\mathrm{t},\hat{\pi}}$ 的渐近方差就会像前一段所述那样具有可比性；这是因为在计算 $\hat{\tau}_{\hat{\pi}}$, $\hat{\tau}_{\mathrm{t},\hat{\pi}}$ 和 (13) 时采用了样本分割。如果使用整个大样本来计算 $\hat{\tau}_{\pi}$, $\hat{\tau}_{\mathrm{t},\pi}$、(3)，并且仅使用 n 个观测值来评估 $\hat{\tau}_{\hat{\pi}}$, $\hat{\tau}_{\mathrm{t},\hat{\pi}}$、(13)——其余 n 个观测值则留作估计 θ_0——那么 $\hat{\tau}_{\pi}$ 的标准误为 $\sqrt{V_{\tau}+V_{\sigma},_{2n}}^{\pi}$，而 $\hat{\tau}_{\hat{\pi}}$ 的标准误为 $\sqrt{V_{\tau}+V_{\sigma},_{\pi}+(q_1-q_0)^{\intercal}V_{\theta_0}(q_1-q_0)}_{n}$，即使不计入 $(q_1 - q_0)^{\intercal} V_{\theta_0} (q_1 - q_0)$ 的贡献，也会增加 $\sqrt{2}$ 倍。$\hat{\tau}_{\mathrm{t},\pi}$ 和 $\hat{\tau}_{\mathrm{t},\hat{\pi}}$ 也~~同样如~~此。

Abadie and Imbens （2016）通过考虑（Y、 D、 X） $\sim \mathbb{P}_{\theta_0}$的移位分布，对倾向得分进行估计。他们假设，移位分布下的条件期望弱收敛于非移位分布下的条件期望。我们采用不同的方法。我们的假设6 至8 不涉及移位分布，而是要求条件期望关于 θ 具有光滑性，可视为假设2、4 和5 在 θ_0邻域内的局部版本。此外，我们验证了定理3 和 4 对例1中的模型成立。

命题6（可容许模型（估计倾向得分））。 考虑例 1中描述的模型族，其倾向得分模型为 $\{\pi(x,\theta) = g(\theta^{\intercal} x)\colon\ \theta \in$

$\Theta\}$ 在一个来自 n分布的独立同分布大样本上通过最大似然估计(D, X)。那么，假设*1*、*3* 和*6* 至*9* 均得到满足。

3.4. 方差估计

在本节中，我们为以下各项的组成部分提供相合估计量： $V_{\hat{\pi}}$ 和$V_{\mathrm{t},\ \hat{\pi}}$ 以便构造平均处理效应和处理组平均处理效应的渐近有效置信区间。为证明一致性，我们提出一些额外假设，这些假设均与示例1中的模型相符。

具体而言，我们需要某些估计量几乎处处有界；若结果变量几乎处处有界，则这一条件即可满足。此外， $\theta \mapsto \pi(\cdot, \theta)$通常可能有多种形式，这使得 Λ^d 难以处理。要求倾向得分遵循单指标模型，例如逻辑斯蒂模型或概率单位模型，并且协变量具有良好的密度，可以缓解这些困难，前提是结果回归足够光滑。假设协变量连续分布且满足某些光滑性条件，则 Λ^d 可表示为一种便于证明一致性的形式。

假设10 （结果变量与协变量分布）。 *(i)* 结果变量几乎必然有界：存在常数 $0 < \bar{y} < \infty$ 使得 $\mathbb{P}\left(|Y| > \bar{y}\right) = 0$。 *(ii)* 协变量向量 X 至少有 $K \geq 2$ 个坐标，且 X 在紧致集 Ψ 上有密度 $\mathcal{X}$； Ψ 如例 *1*所述。

假设11 （单指标倾向得分与光滑结果回归）。 *(i)* 假设 *3* 中的倾向得分模型为 $\pi(x, \theta) = g(\theta^{\intercal} x)$，其中 g 如例 *1*所述。 *(ii)* $m(x) := \mathbb{E}[Y \mid X = x]$ 有界，且存在两个协变量—— X_1 和 X_2 （不失一般性）——使得 $\frac{\partial m}{\partial x_1}$ 和 $\frac{\partial m}{\partial x_2}$ 对所有 $x \in \mathcal{X}$ 均定义良好且连续。

方差估计量为

$$\hat{V}_{\hat{\pi}} := \hat{V}_{\tau} + \hat{V}_{\sigma,\pi} + (\hat{q}_1 - \hat{q}_0)^{\intercal} \hat{V}_{\theta_0} (\hat{q}_1 - \hat{q}_0), \tag{14}$$

$$\hat{V}_{\mathrm{t},\hat{\pi}} := \hat{V}_{\tau_{\mathrm{t}}} + \hat{V}_{\mathrm{t},\sigma,\pi} + (1/\hat{p}_1^2)(\hat{q}_{\mathrm{t},1} - \hat{q}_{\mathrm{t},0})^{\intercal} \hat{V}_{\theta_0} (\hat{q}_{\mathrm{t},1} - \hat{q}_{\mathrm{t},0}), \tag{15}$$

各分量估计量如下。我们假设$\hat{}V_{\theta_0} \xrightarrow{P} V_{\theta_0}$ 是一个一致的$\hat{}$估计量，估计V_{θ_0}。在实践中，θ 通常是最大似然估计量，命题6支持这一点；此时，在假设11下，

$$\hat{V}_{\theta_0} := \left(\frac{1}{n} \sum_{i \in [n]} \frac{(g'(\hat{\theta}^{\intercal} X_i))^2}{g(\hat{\theta}^{\intercal} X_i)(1 - g(\hat{\theta}^{\intercal} X_i))} X_i X_i^{\intercal} \right)^{-1}$$

众所周知，它是 V_{θ_0}的一致估计量。 p_1 可通过$\hat{}p_1 := \frac{1}{n}\sum_{i\in[n]} D_i$ 利用大数定律进行一致估计。（14）和（15）中其余分量的非参数估计量见附录A。

命题*7*（相合方差估计量）。假设假设*1*、*3*和*6*至*11*均成立，卡尺 δ_n 满足（*13*），且 $\hat{}V_{\theta_0} \xrightarrow{P} V_{\theta_0}$ *as* $n \to \infty$。则 $\hat{V}_{\hat{\pi}} \xrightarrow{P} V_{\hat{\pi}}$ 和 $\hat{V}_{\mathrm{t},\hat{\pi}} \xrightarrow{P} V_{\mathrm{t},\hat{\pi}}$ *as* $n \to \infty$。特别地，（*14*）和（*15*）右侧的估计量均分别相合于其对应的目标估计量。

根据定理3 和4，一个直接推论是，我们可以为平均处理效应和处理组平均处理效应构造渐近置信区间。令 $z_{1-\alpha/2}$ 为标准正态分布 $\alpha \in (0,1)$ 的 $(1-\alpha/2)$ 分位数，并令 $[a \pm b]$ 表示区间 $[a-b, a+b]$，对应于 $a, b \in \mathbb{R}$, $b \geq 0$。那么，区间 $[\hat{}\tau_{\hat{\pi}} \pm z_{1-\alpha/2}(\hat{}V_{\hat{\pi}}/n)^{1/2}]$ 和 $[\hat{}\tau_{\mathrm{t},\hat{\pi}} \pm z_{1-\alpha/2}(\hat{}V_{\mathrm{t},\hat{\pi}}/n)^{1/2}]$分别是平均处理效应和处理组平均处理效应的渐近有效置信区间。

推论 1（渐近置信区间）。假设 *1*、*3* 以及 *6*至 *11* 均成立，卡尺 δ_n 满足 (13)，且当 $n \to \infty$ 时，$\hat{}V_{\theta_0} \xrightarrow{P} V_{\theta_0}$ 。那么$\mathbb{P}\left(\left[\hat{}\tau_{\hat{\pi}} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{}V_{\hat{\pi}}}{n}}\right] \ni \tau\right) \longrightarrow 1-\alpha$，且 $\mathbb{P}\left(\left[\hat{}\tau_{\mathrm{t},\hat{\pi}} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{}V_{\mathrm{t},\hat{\pi}}}{n}}\right] \ni \tau_{\mathrm{t}}\right) \to 1-\alpha$ 当$n \to \infty$时。

4. 结论

我们研究基于（估计的）倾向得分进行匹配时的卡尺匹配估计量。我们提出一种卡尺，并证明由此得到的平均处理效应（ATE）和处理组平均处理效应（ATT）估计量渐近无偏且服从正态分布。

当倾向得分已知时，在受限模型中，我们的平均处理效应估计量达到半参数下界；在该模型中，样本中只能观测到倾向得分而无法观测到协变量，或者协变量上的结果回归仅取决于倾向得分。在这一受限模型中，处理组平均处理效应的估计量只能达到更大的下界，该下界对应于倾向得分未知的情形。即使在非受限模型中，只要我们的假设成立，我们的两个估计量都比基于已知倾向得分的最近邻匹配估计量更有效率；因此，在大样本极限下，应优先采用我们的方法。当估计参数化倾向得分时，我们两个估计量的方差都会增大，因此目前尚不清楚卡尺匹配还是最近邻匹配的效率更高。

我们通过验证一类常用模型满足我们的假设，并构造平均处理效应的渐近置信区间，推动该估计量在实证研究中的应用。未来一个值得研究的方向是倾向得分的样本内估计，以及采用非参数倾向得分估计量。面临的主要挑战是弄清倾向得分估计所带来的不确定性如何传递到匹配估计量；对于非参数模型，这种影响更难量化。

A. 方差估计

在本节中，我们定义第3.4节中针对（14）和（15）中的各个组成部分的非参数方差估计量。令 $K(u) = (2\pi)^{-1/2}e^{-u^2/2}$、$u \in \mathbb{R}$ 表示高斯核，令 $K'(u)$ 表示其导数。令 $0 < \gamma_n \lesssim a_n$ 为两个任意序列 $\gamma_n := \kappa_0 n^{-\beta}$、$a_n := \kappa_1 n^{-\alpha}$，其中 $0 < \alpha < \beta < 1/4$ 和 κ_0 $\kappa_1 > 0$为固定有限常数。我们采用截断策略，以避免边界处的偏差。定义区间 $A_n := [\underline{p}_{\hat{\theta}} + a_n, \bar{p}_{\hat{\theta}} - a_n]$ 以及 $\hat{A}_n := [\min_{i\in[n]} g(\hat{\theta}^\intercal X_i) + a_n, \max_{i\in[n]} g(\hat{\theta}^\intercal X_i) - a_n]$；根据假设9，当 $\hat{\theta} \xrightarrow{P} \theta_0$ 时，这些区间以趋近于1的概率定义良好；参见命题7的证明。[6] 令 $\hat{N} := \sum_{i\in[n]} \mathbb{I}_{g(\hat{\theta}^\intercal X_i)\in \hat{A}_n}$。第一组分量的估计量为

$$\hat{V}_\tau := \left(\frac{1}{\hat{N}} \sum_{i\in[n]} [\hat{\mu}^1(\hat{\theta}, g(\hat{\theta}^\intercal X_i)) - \hat{\mu}^0(\hat{\theta}, g(\hat{\theta}^\intercal X_i))]^2 \mathbb{I}_{g(\hat{\theta}^\intercal X_i)\in \hat{A}_n} \right) - \hat{\tau}^2_{\hat{\pi}}, \tag{16}$$

$$\hat{V}_{\tau_{\mathrm{t}}} := \frac{1}{\hat{p}_1^2} \left(\frac{1}{\hat{N}} \sum_{i\in[n]} D_i [\hat{\mu}^1(\hat{\theta}, g(\hat{\theta}^\intercal X_i)) - \hat{\mu}^0(\hat{\theta}, g(\hat{\theta}^\intercal X_i))]^2 \mathbb{I}_{g(\hat{\theta}^\intercal X_i)\in \hat{A}_n} \right) - \frac{\hat{\tau}^2_{\mathrm{t},\hat{\pi}}}{\hat{p}_1},$$

其中

$$\hat{\mu}^d(\theta, p) := \frac{\hat{q}_{\mu,d}(\theta, p)}{\hat{h}_d(\theta, p)}, \quad \hat{q}_{\mu,d}(\theta, p) := \frac{1}{N_d \gamma_n} \sum_{j\in[n]} \mathbb{I}^{D_j=d} Y_j K\left(\frac{g(\theta^\intercal X_j) - p}{\gamma_n} \right),$$

$$\hat{h}_d(\theta, p) := \hat{f}_{\theta,d}(p) := \frac{1}{N_d \gamma_n} \sum_{j\in[n]} \mathbb{I}^{D_j=d} K\left(\frac{g(\theta^\intercal X_j) - p}{\gamma_n} \right), \quad d \in \{0, 1\};$$

而第二个分量的则为

$$\hat{V}_{\sigma,\pi} := \frac{1}{\hat{N}} \sum_{i\in[n]} \left(\frac{\hat{\sigma}_0^2(\hat{\theta}, g(\hat{\theta}^\intercal X_i))}{1 - g(\hat{\theta}^\intercal X_i)} + \frac{\hat{\sigma}_1^2(\hat{\theta}, g(\hat{\theta}^\intercal X_i))}{g(\hat{\theta}^\intercal X_i)} \right) \mathbb{I}_{g(\hat{\theta}^\intercal X_i)\in \hat{A}_n},$$

$$\hat{V}_{\mathrm{t},\sigma,\pi} := \frac{1}{\hat{p}_1^2 \hat{N}} \sum_{i\in[n]} \left(\frac{g(\hat{\theta}^\intercal X_i)^2 \hat{\sigma}_0^2(\hat{\theta}, g(\hat{\theta}^\intercal X_i))}{1 - g(\hat{\theta}^\intercal X_i)} + g(\hat{\theta}^\intercal X_i) \hat{\sigma}_1^2(\hat{\theta}, g(\hat{\theta}^\intercal X_i)) \right) \mathbb{I}_{g(\hat{\theta}^\intercal X_i)\in \hat{A}_n},$$

[6] 在实际应用中，尤其是样本量适中时，应谨慎选择 γ_n 和 a_n，以确保方差估计值为非负数。a_n 应选得足够小，以增大 A_n、$\hat{A}_n$；例如，可以将 κ_1 选为任意接近零的值，并令 $\alpha := 1/(4 + \varepsilon_\alpha)$，使 $\varepsilon_\alpha > 0$ 任意接近零。通常，γ_n 也应设得较小，以便依据标准的非参数理论，尽量减小方差估计值的偏差，从而避免出现负值；为适应 $\alpha < \beta$，可以令 $\beta = 1/(4 + \varepsilon_\beta)$，其中$0 < \varepsilon_\beta < \varepsilon_\alpha$，例如 $\varepsilon_\beta := \varepsilon_\alpha/2$。$\kappa_0$ 的选择应考虑在分别估计 μ^d 及其导数时，$(g(\hat{\theta}^\intercal X_i))_{i\in[n]}$ 与$(\hat{\theta}^\intercal X_i)_{i\in[n]}$ 所涉及的不同尺度。选择较小的 κ_0 是稳妥但较保守的做法。请注意，正如命题7所示，截断的影响在渐近意义下会消失$(\mathbb{E}\hat{N}/n \to 1)$。

其中

$$\hat{\sigma}_d^2(\theta,p) \coloneqq \hat{\mu}_2^d(\theta,p) - (\hat{\mu}^d(\theta,p))^2, \quad \hat{\mu}_2^d(\theta,p) \coloneqq \frac{\hat{q}_{\mu_2,d}(\theta,p)}{\hat{h}_d(\theta,p)},$$
$$\hat{q}_{\mu_2,d}(\theta,p) \coloneqq \frac{1}{N_d\gamma_n}\sum_{j\in[n]} \mathbb{I}^{D_j=d} Y_j^2 K\left(\frac{g(\theta^\intercal X_j)-p}{\gamma_n}\right), \quad d\in\{0,1\},$$

是 $\sigma_d^2(\theta,p) = \mu_2^d(\theta,p) - (\mu^d(\theta,p))^2$ 的估计量

$$\mu_2^d(\theta,p) \coloneqq \mathbb{E}\left[Y^2 \,\middle|\, D=d, \pi(X,\theta)=p\right], \quad d\in\{0,1\}.$$

最后，导数的概率极限由以下方法估计

$$\hat{q}_d^\intercal \coloneqq \frac{1}{\hat{N}}\sum_{i\in[n]} \hat{\Lambda}^d(\hat{\theta},X_i)_{\mathbb{I}^{g(\hat{\theta}^\intercal X_i)\in\hat{A}_n}}, \quad \hat{q}_{\mathrm{t},d}^\intercal \coloneqq \frac{1}{\hat{N}}\sum_{i\in[n]} D_i\hat{\Lambda}^d(\hat{\theta},X_i)_{\mathbb{I}^{g(\hat{\theta}^\intercal X_i)\in\hat{A}_n}},$$
$$\hat{\Lambda}^d(\theta,x) \coloneqq \widehat{\left(\frac{\partial\mu^d}{\partial\theta^\intercal}\right)}(\theta,g(\theta^\intercal x)) + \widehat{\left(\frac{\partial\mu^d}{\partial p}\right)}(\theta,g(\theta^\intercal x))g'(\theta^\intercal x)x^\intercal,$$

其中

$$\widehat{\left(\frac{\partial\mu^d}{\partial\theta_k}\right)}(\theta,p) \coloneqq \frac{\widehat{\left(\frac{\partial q_{\mu,d}}{\partial\theta_k}\right)}(\theta,p)\hat{h}_d(\theta,p) - \hat{q}_{\mu,d}(\theta,p)\widehat{\left(\frac{\partial h_d}{\partial\theta_k}\right)}(\theta,p)}{(\hat{h}_d(\theta,p))^2},$$
$$\widehat{\left(\frac{\partial q_{\mu,d}}{\partial\theta_k}\right)}(\theta,p) \coloneqq \frac{(g^{-1})'(p)}{N_d\gamma_n^2}\sum_{j\in[n]} \mathbb{I}^{D_j=d} Y_j X_{j,k} K'\left(\frac{\theta^\intercal X_j - g^{-1}(p)}{\gamma_n}\right),$$
$$\widehat{\left(\frac{\partial h_d}{\partial\theta_k}\right)}(\theta,p) \coloneqq \frac{(g^{-1})'(p)}{N_d\gamma_n^2}\sum_{j\in[n]} \mathbb{I}^{D_j=d} X_{j,k} K'\left(\frac{\theta^\intercal X_j - g^{-1}(p)}{\gamma_n}\right),$$

其中 θ_k ($X_{j,k}$) 是 θ (X_j) 的第 k 个坐标，其中 $k\in[K]$，并且

$$\widehat{\left(\frac{\partial\mu^d}{\partial p}\right)}(\theta,p) \coloneqq \frac{\left(\frac{\partial}{\partial p}\hat{q}_{\mu,d}(\theta,p)\right)\hat{h}_d(\theta,p) - \hat{q}_{\mu,d}(\theta,p)\frac{\partial}{\partial p}\hat{h}_d(\theta,p)}{(\hat{h}_d(\theta,p))^2},$$
$$\frac{\partial}{\partial p}\hat{q}_{\mu,d}(\theta,p) = -\frac{1}{N_d\gamma_n^2}\sum_{j\in[n]} \mathbb{I}^{D_j=d} Y_j K'\left(\frac{g(\theta^\intercal X_j)-p}{\gamma_n}\right),$$
$$\frac{\partial}{\partial p}\hat{h}_d(\theta,p) = -\frac{1}{N_d\gamma_n^2}\sum_{j\in[n]} \mathbb{I}^{D_j=d} K'\left(\frac{g(\theta^\intercal X_j)-p}{\gamma_n}\right)$$

对于 $d\in\{0,1\}$。通过定义 $X_{j,\,K+1} \coloneqq 1$，可以在倾向得分模型中纳入截距项；对于 $j\in[n]$，也考虑关于 θ_{K+1} 的导数。

B. 证明

本节证明主要结果，即命题 1、2 和4 至6 ，定理1至4 ，以及作为支撑的引理1 至4。命题3 和7 以及引理5 至7 的证明见补充材料。为简洁起见，我们给出卡尺选择 $\delta_n = s\frac{\log n}{n}$, $s := 1$ 的证明，并在必要时对选择 $\delta_n = \overline{\Delta}_n \vee \log\frac{N_0}{N_0+1} \vee \log\frac{N_1}{N_1+1}$和$\delta_n = \overline{\Delta}_n \vee \log\frac{N_0}{N_0+1} \vee \log\frac{N_1}{N_1+1}$ 作出说明。

我们采用以下记号。令 $D^{(n)} := (D_i)_{i\in[n]}$ 和 $PS^{(n)} := (\pi(X_i))_{i\in[n]}$。对于$a, b \in \mathbb{R}$, $a < b$，令

$$\begin{aligned} F_d[a,b] &:= \mathbb{P}\left(\pi(X) \in [a,b] \mid D = d\right), \\ \mathbb{E}_{N_d}[a,b] &:= \frac{1}{N_d}\sum_{i:D_i=d} \mathbb{I}^{\pi(X_i)\in[a,b]}, \end{aligned} \tag{17}$$

为区间的条件（经验）测度 $[a,b]$ 对于 $d \in \{0,1\}$。类似地，在假设 3下，定义条件（经验）测度

$$\begin{aligned} F_{d,\theta}[a,b] &:= \mathbb{P}\left(\pi(X,\theta) \in [a,b] \mid D = d\right), \\ \mathbb{F}_{N_d,\theta}[a,b] &:= \frac{1}{N_d}\sum_{i:D_i=d} \mathbb{I}^{\pi(X_i,\theta)\in[a,b]} \quad \text{for } \theta \in \Theta. \end{aligned} \tag{18}$$

令 $\mathbb{G}_{N_d} := \sqrt{N_d}(\mathbb{E}_{N_d} - F_d)$ 为 $((\pi(X_i))_{i:D_i=d} \mid D^{(n)}) \overset{\text{独立同分布}}{\sim} F_d$ 的经验过程，并 $[a \pm b]$ 表示区间 $[a-b, a+b]$。在证明中，常数的取值可能随公式而变化，恕不另行说明。

命题 1的证明。间距及其次序。设$U_{(1)} \le U_{(2)} \le \ldots \le U_{(N_1)}$ 是 $(U_1, \ldots, U_{N_1} \mid D^{(n)})$ 的次序统计量，其中各项独立同分布于 $\sim$均匀分布(0, 1)。设 $\tilde{U}_1 := U_{(1)}$, $\tilde{U}_i := U_{(i)} - U_{(i-1)}$ for $i = 2, \ldots, N_1$ and $U_{N_1+1} := 1 - U_{(N_1)}$ 是由 $(U_i)_{i\in[N_1]}$ 生成的间距。令 $\tilde{U}_{(1)} \le \tilde{U}_{(2)} \le \ldots \le \tilde{U}_{(N_1+1)}$ 为有序间距。Shorack 和 Wellner（2009年，第21章）证明了 $\mathbb{E}\left[\tilde{U}_{(N_1+1)} \,\middle|\, D^{(n)}\right] = \frac{1}{N_1+1}\sum_{i=1}^{N_1+1}\frac{1}{N_1+2-i}$，我们将如下应用这一结果。

利用间距进行界定。令 $\underline{\Delta}_i := \min_{j:D_j\neq D_i}|\pi(X_i) - \pi(X_j)|$，其中 $i \in [n]$，使得 $\overline{\Delta}_n \le \sum_{d\in\{0,1\}} \max_{i:D_i=d}\underline{\Delta}_i$。考虑 $\max_{i:D_i=0}\underline{\Delta}_i$。令 $\pi_{(1)} \le \pi_{(2)} \le \ldots \le \pi_{(N_1)}$ 为 $(\pi(X_i))_{i:D_i=1}$ 的次序统计量，并令 $\tilde{\pi}_1 := \pi_{(1)} - \underline{p}$, $\tilde{\pi}_i := \pi_{(i)} - \pi_{(i-1)}$，其中 $i \in \{2, \ldots, N_1\}$，且 $\tilde{\pi}_{N_1+1} := \bar{p} - \pi_{(N_1)}$ 为假设2中 $\underline{p}, \bar{p}$ 对应的间距。令$\tilde{\pi}_{(1)} \le \tilde{\pi}_{(2)} \le \ldots \le \tilde{\pi}_{(N_1+1)}$ 为这些间距的次序统计量。对照组单位的每个倾向得分 $\pi(X_j)$，要么位于 $\pi_{(1)}$ 的左侧，要么位于 $\pi_{(N_1)}$ 的右侧，要么位于某个 $i \in \{2, 3, \ldots, N_1\}$ 对应的两个倾向得分 $\pi_{(i)}$ 和 $\pi_{(i-1)}$ 之间。在这三种

情况下，与 $\pi(X_j)$ 最接近的接受处理者倾向得分处于 $\tilde{\pi}_{(N_1+1)}$ 距离以内。因此，$\max_{i:D_i=0} \underline{\Delta}_i \leq \tilde{\pi}_{(N_1+1)}$。

给定 $D^{(n)}$，限制在 i 上的 $\pi(X_i)$：$D_i = 1$是独立同分布的，且 $(\pi(X_i) \mid D_i = 1) \sim F_1$。根据假设2，$\inf_{p\in[\underline{p},\bar{p}]} f_1(p) > 0$，因此 F_1 在 $[\underline{p}, \bar{p}]$ 上严格递增，从而在 $[F_1(\underline{p}) \; F_1(\bar{p})]$上具有严格递增的逆函数 F_1^{-1}。由于 F_1^{-1}递增，根据分位数变换，$((\pi_{(i)})_{i\in[N_1]} \mid D^{(n)}) \sim ((F_1^{-1}(U_{(i)}))_{i\in[N_1]} \mid D^{(n)})$。我们区分三种情况。

- 情况1：对于某个 $i \in \{2, \ldots, N_1\}$，$\tilde{\pi}_{(N_1+1)} = \pi_{(i)} - \pi_{(i-1)}$。然后注意到 $(\pi_{(i)} - \pi_{(i-1)} \mid D_{(n)}) \sim (F_1^{-1}(U_{(i)}) - F_1^{-1}(U_{(i-1)}) \mid D_{(n)})$，由此可将 $\tilde{\pi}_{(N_1+1)}$ 进行界定；该式可由 $\|(F_1^{-1})'\|_\infty (U_{(i)} - U_{(i-1)}) = \|(F_1^{-1})'\|_\infty \tilde{U}_i$ 进行界定，因为 F_1^{-1}是常数为$\|(F_1^{-1})'\|_\infty$的利普希茨函数，且 $(F_1^{-1})'(u) = \frac{1}{f_1(F_1^{-1}(u))}$ 有限，这是因为根据假设2，$\inf_{p\in[\underline{p},\bar{p}]} f_1(p) > 0$。

- 情形 2：$\tilde{\pi}_{(N_1+1)} = \pi_{(1)} - \underline{p}$。接着写作 $\underline{p} = F_1^{-1}(F_1(\underline{p})) = F_1^{-1}(0)$，因此 $(\tilde{\pi}_{(N_1+1)} \mid D_{(n)})$ 服从一个由 $\|(F_1^{-1})'\|_\infty (U_{(1)} - 0) = \|(F_1^{-1})'\|_\infty \tilde{U}_1$ 界定的随机变量。

- 情形 3：$\tilde{\pi}_{(N_1+1)} = \bar{p} - \pi_{(N_1)}$。接着写作 $\bar{p} = F_1^{-1}(F_1(\bar{p})) = F_1^{-1}(1)$，因此 $(\tilde{\pi}_{(N_1+1)} \mid D_{(n)})$ 服从一个由 $\|(F_1^{-1})'\|_\infty (1 - U_{(N_1)}) = \|(F_1^{-1})'\|_\infty \tilde{U}_{N_1+1}$ 界定的随机变量。

由此可知，$(\tilde{\pi}_{(N_1+1)} \mid D^{(n)})$ 的分布与一个由$\|(F_1^{-1})'\|_\infty \tilde{U}_{(N_1+1)}$界定的随机变量相同。因此，

$$\begin{aligned}
\mathbb{E} \max_{i:D_i=0} \underline{\Delta}_i &\leq \mathbb{E}\mathbb{E}\left[\tilde{\pi}_{(N_1+1)} \,\middle|\, D^{(n)}\right] \lesssim \mathbb{E}\left[\frac{1}{N_1+1}\sum_{i=1}^{N_1+1}\frac{1}{N_1+2-i}\right] \\
&= \sum_{n_1=0}^{n}\left[\frac{1}{n_1+1}\sum_{i=1}^{n_1+1}\frac{1}{n_1+2-i}\right]\binom{n}{n_1}p_1^{n_1}(1-p_1)^{n-n_1} \\
&= (1-p_1)^n + \sum_{n_1=1}^{n}\left[\frac{1}{n_1+1}\sum_{i=1}^{n_1+1}\frac{1}{n_1+2-i}\right]\binom{n}{n_1}p_1^{n_1}(1-p_1)^{n-n_1} \qquad (19)
\end{aligned}$$

根据 Shorack 和 Wellner (2009年) 的结果，其中 $p_1 = \mathbb{P}(D = 1)$。(19) 中的第一项呈指数衰减。(19) 中的第二项里，方括号内的被积函数渐近于$\frac{\text{对数}\frac{n_1}{n_1+1}}{\text{对数}\frac{n}{n_1+1}} \leq$。也就是说，存在一些常数 $c, \bar{n}_1 > 0$ 使得 $\frac{1}{n_1+1}\sum_{i=1}^{n_1+1}\frac{1}{n_1+2-i} \leq c$ 若 $n_1 > \bar{n}_1$。由于方括号内的被积函数以 1 为上界，因此(19) 中的第二项有界于

$$\sum_{n_1=1}^{\bar{n}_1}\binom{n}{n_1}p_1^{n_1}(1-p_1)^{n-n_1} + c(\log n)\mathbb{E}\left[(1+N_1)^{-1}\right],$$

其中第一项为 $O\left(n^{\bar{n}_1}(1-p_1)^n\right) = O$(对数 n/n)，第二项为 O(对数 n/n)，根据 Cribari-Neto等人 (2000年)。通过对称性，可用类似论证处理 $\mathbb{E}$最大值$_{i:D_i=1}\underline{\Delta}_i$。■

命题 2 的证明。我们有引理 R_{di} （32）中的 1，

$$\min_{i\in[n]} M_i = \min_{i\in[n]} N_{1-D_i} F_{1-D_i}[\pi(X_i) \pm \delta_n](1 + R_{1-D_i,i}). \tag{20}$$

若 最小值$_{i\in[n]}$ $(1 + R_{1-D_i,\ i})$ ≥ 0，这以趋于 1 的概率发生，则（20）大于或等于

$$2\left(\min_{d\in\{0,1\}} \inf_{p\in[\underline{p},\bar{p}]} f_d(p)\right)(N_0 \wedge N_1)\delta_n \min_{i\in[n]}(1 + R_{1-D_i,i}).$$

根据假设2，$\inf_{p\in[\underline{p},\bar{p}]} f_d(p) > 0$。由强大数定律和连续映射定理，$(1/\log n)(N_0 \wedge N_1)$ $\delta_n = (N_0 \wedge N_1)/n \xrightarrow{a.s.} (1 - p_1) \wedge p_1 > 0$。由引理 1，$\max_{i\in[n]} |R_{1-D_i,i}| = o_P(1)$。于是 $\min_{i\in[n]} M_i \simeq (1 + o_P(1)) \log\ n$，由此可得命题 2 中的下界，进而得到 $\mathbb{P}\left(\min\right.$ $\left._{i\in[n]} M_i \geq 1\right) \to 1$。对于命题 2 中的上界，同样的推理也适用。

当卡尺为 $\delta_n = \overline{\underline{\Delta}}_n \vee \log\frac{N_0}{N_0+1} \vee \log\frac{N_1}{N_1+1}$时，$\min_{i\in[n]} M_i \geq 1$。引理 1(iv)–(vi)结合大数定律和命题1，利用连续映射定理即可证明该结论。■

命题5的证明。证明思路与命题 2 和引理 1(iv)–(vi)的证明类似。当卡尺为 $\delta_n = \overline{\underline{\Delta}}_n \vee \log\frac{N_0}{N_0+1} \vee \log\frac{N_1}{N_1+1}$ 时，它有界于 $\overline{\underline{\Delta}}_n + \log\frac{N_0}{N_0+1} + \log\frac{N_1}{N_1+1}$，其中 $\mathbb{E}\log\frac{N_0}{N_0+1} = O\left(\frac{\log n}{n}\right)$由 Cribari-Neto 等人（2000）给出。对 $(\pi(X_i,\theta))_{i\in[n]}$应用间距论证，与命题 1 的证明类似，再结合假设 6 和 9，可得 $\overline{\underline{\Delta}}_n = O_P\left(\frac{\log n}{n}\right)$。然后，利用命题 2 和引理 1(iv)–(vi)中的论证即可证明命题5。■

定理1的证明。根据假设2，(5)中的 $\tau(\pi(X))$定义良好，且根据假设1，$\mathbb{E}\tau(\pi(X)) = \tau$。根据假设5，$\mu^d$ 在紧集$[\underline{p},\bar{p}]$上是 Lipschitz 连续的，因此有界，进而 $V_\tau < \infty$也有界。于是，由中心极限定理可得$\sqrt{n}(\tau(\pi(X)) - \tau)$ $\mathcal{N}(0, V_\tau)$。结合引理2，可得

$$\begin{bmatrix} V_\tau^{-1/2}\sqrt{n}(\overline{\tau(\pi(X))} - \tau) \\ V_E^{-1/2}\sqrt{n}E \end{bmatrix} \rightsquigarrow \mathcal{N}(0, \mathrm{I}_2),$$

沿子序列，其中 I_2 是 2×2 单位矩阵。根据引理 4，$V_E \xrightarrow{P} V_{\sigma,\ \pi}$，由假设 2 和 4可知其为有限值。于是，连续映射定理和 Slutsky 引理推出 $(V_\tau + V_{\sigma,\ \pi})^{-1/2}\sqrt{n}(\tau(\pi(X))$ $-\tau + E)$ $\mathcal{N}(0, 1)$。事件 $\{\min_{i\in[n]} M_i >$

0} 根据命题 2，该事件发生的概率趋于 1。在此事件下，$\sqrt{n}|B| \lesssim \sqrt{n}\delta_n = \text{对数}_{\sqrt{n}}^{n} = o(1)$ 根据假设 5。因此 $\sqrt{n}B = o_P(1)$。

当卡尺为 $\delta_n = \overline{\Delta}_n \vee \log\frac{N_0}{N_0+1} \vee \log\frac{N_1}{N_1+1}$ 时，引理2仍然适用。接下来，连续映射定理、大数定律和命题1 推出 $\sqrt{n}|B| = o_P(1)$。■

定理2的证明。类似于 (4)–(8) 的分解成立，因此

$$\begin{aligned}
\sqrt{n}(\hat{\tau}_{\mathrm{t},\pi} - \tau_{\mathrm{t}}) &= \sqrt{n}(\overline{\tau_{\mathrm{t}}(\pi(X))} - \tau_{\mathrm{t}}) + \sqrt{n}E_{\mathrm{t}} + \sqrt{n}B_{\mathrm{t}} \\
\overline{\tau_{\mathrm{t}}(\pi(X))} &:= \frac{1}{N_1}\sum_{i\in[n]} D_i\tau(\pi(X_i)) \\
E_{\mathrm{t}} &:= \frac{1}{N_1}\sum_{i\in[n]} E_{\mathrm{t},i}, \quad E_{\mathrm{t},i} := (\mathbb{I}_{M_i>0}D_i - (1-D_i)w_i)\varepsilon_i \\
B_{\mathrm{t}} &:= \frac{1}{N_1}\sum_{i\in[n]} B_{\mathrm{t},i}, \\
B_{\mathrm{t},i} &:= D_i(\mathbb{I}_{M_i>0} - 1)(\mu^{D_i}(\pi(X_i)) + \mu^{1-D_i}(\pi(X_i))) \\
&\quad + D_i\frac{\mathbb{I}_{M_i>0}}{M_i}\sum_{j\in\mathcal{J}(i)}(\mu^0(\pi(X_i)) - \mu^0(\pi(X_j))).
\end{aligned}$$

由于 $\mathbb{E}D(\tau(\pi(X)) - \tau_{\mathrm{t}}) = \mathbb{E}[\tau(\pi(X)) - \tau_{\mathrm{t}} \mid D=1]p_1 = 0$，根据假设 1 和 5，$\sqrt{n}(\tau_{\mathrm{t}}(\pi(X)) - \tau_{\mathrm{t}})$ 的均值为零且方差有限。由中心极限定理、连续映射定理和斯卢茨基引理，$\sqrt{n}(\tau_{\mathrm{t}}(\pi(X)) - \tau_{\mathrm{t}}) = (N_1/n)^{-1}n^{-1/2}\sum_{i\in[n]} D_i(\tau(\pi(X_i)) - \tau_{\mathrm{t}}) \xrightarrow{\mathcal{N}} 0\ V_{\tau_{\mathrm{t}}}$ $N_1/n \xrightarrow{a.s.} p_1$ $\sqrt{n}E_{\mathrm{t}}$) (,), 因为 ( ) 。该量的均值为零，且 Lindeberg-Feller 中心极限定理证明 $\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(V_{E_{\mathrm{t}}}^{-1/2}\sqrt{n}E_{\mathrm{t}} \le x \,\middle|\, D^{(n)}, PS^{(n)}\right) - \Phi(x)\right| \xrightarrow{P} 0$，其中 $V_{E_{\mathrm{t}}} := \frac{1}{n}\sum_{i\in[n]}(\mathbb{I}_{M_i>0}D_i - (1-D_i)w_i)^2\sigma^2_{D_i}(\pi(X_i))$，论证与引理2类似。通过与引理4类似的论证，$V_{E_{\mathrm{t}}} \xrightarrow{P} V_{\mathrm{t},\sigma,\pi}$。根据命题 2 以及定理1中的论证，$\sqrt{n}B_{\mathrm{t}} = o_P(1)$。■

定理3。令 $w_i(\hat{\theta}) := \sum_{j\in\mathcal{J}_{\hat{\theta}}(i)} \frac{1}{M_j(\hat{\theta})}\, i \in [n]$，并且，如（4）–（8）中所示，分解 se 证明

$$\sqrt{n}(\hat{\tau}_{\hat{\pi}} - \tau) = \frac{1}{\sqrt{n}}\sum_{i\in[n]}\left\{\mu^1(\hat{\theta},\pi(X_i,\hat{\theta})) - \mu^0(\hat{\theta},\pi(X_i,\hat{\theta})) - \tau\right\} \tag{21}$$

$$+\frac{1}{\sqrt{n}}\sum_{i\in[n]}(2D_i-1)(\mathbb{I}_{M_i(\hat{\theta})>0} + w_i(\hat{\theta}))\varepsilon_i(\hat{\theta}) \tag{22}$$

$$+\frac{1}{\sqrt{n}}\sum_{i\in[n]}(2D_i-1)(\mathbb{I}_{M_i(\hat{\theta})>0} - 1)(\mu^{1-D_i}(\hat{\theta},\pi(X_i,\hat{\theta})) - \mu^{D_i}(\hat{\theta},\pi(X_i,\hat{\theta}))) \tag{23}$$

$$+\frac{1}{\sqrt{n}}\sum_{i\in[n]}(2D_i-1)\frac{\mathbb{I}_{M_i(\hat{\theta})>0}}{M_i(\hat{\theta})}\sum_{j\in\mathcal{J}_{\hat{\theta}}(i)}\left[\mu^{1-D_i}(\hat{\theta},\pi(X_i,\hat{\theta})) - \mu^{1-D_i}(\hat{\theta},\pi(X_j,\hat{\theta}))\right]. \tag{24}$$

我们首先证明（21）和（22）渐近地服从联合正态分布且相互独立，然后证明（23）和（24）渐近可忽略。

条件（21）和（22）。我们将（21）和（22）分别对应于 V_n 和 W_n，并应用以下结果。设 V_n、W_n、$n=1$、2、.....是定义在某个概率空间上的两列随机变量。要证明 (V_n, W_n) (V, W) $\sim \mathcal{N}(0, \Sigma)$，对于对角矩阵 $\Sigma = \mathrm{diag}(\sigma_V^2, \sigma_W^2)$，只需证明对所有有界连续函数 $h_1, h_2 : \mathbb{R} \to \mathbb{R}$，均有 $\mathbb{E}h_1(V_n)h_2(W_n) \to (\mathbb{E}h_1(V))(\mathbb{E}h_2(W))$。设 $\mathcal{F}_{n0}$ 是一个子-σ-代数，使得对于所有 $n \geq 1$，V_n 均为 $\mathcal{F}_{n0}$-可测。由于 $\mathbb{E}h_1(V_n)h_2(W_n)$ $= \mathbb{E}\left[h_1(V_n)\mathbb{E}\left[h_2(W_n) \mid \mathcal{F}_{n0}\right]\right]$，根据 Portmanteau 引理，只需证明 $V_n \rightsquigarrow \mathcal{N}(0, \sigma_V^2)$，且对所有 $w \in \mathbb{R}$，均有 $\mathbb{P}(W_n \leq w \mid \mathcal{F}_{n0}) \xrightarrow{P} \Phi(w/\sigma_W)$。

收敛于(21)。 将 (21) 展开为

$$\frac{1}{\sqrt{n}} \sum_{i \in [n]} (\mu^1_{\theta_0}(\pi(X_i, \theta_0)) - \mu^0_{\theta_0}(\pi(X_i, \theta_0)) - \tau) \quad (25)$$

$$+ \frac{1}{\sqrt{n}} \sum_{i \in [n]} \left\{ \mu^1(\hat{\theta}, \pi(X_i, \hat{\theta})) - \mu^0(\hat{\theta}, \pi(X_i, \hat{\theta})) - \left[\mu^1_{\theta_0}(\pi(X_i, \theta_0)) - \mu^0_{\theta_0}(\pi(X_i, \theta_0))\right] \right\}. \quad (26)$$

根据假设 1 和 5，由标准中心极限定理，(25) 弱收敛于 $\mathcal{N}(0, V_\tau)$。根据假设3和假设7，对于所有 $(\tilde{\theta}, x) \in \mathrm{Nb}(\theta_0, \epsilon) \times \mathcal{X}$，$\Lambda^d(\tilde{\theta}, x)$ 都有定义。因此，根据中值定理，对于线段 $\hat{}\theta$ 与 θ_0 之间的某个 $\tilde{\theta}^d$，有 $\mu^d(\hat{}\theta, \pi(x, \hat{}\theta)) = \mu^d_{\theta_0}(\pi(x, \theta_0)) + \Lambda^d(\tilde{\theta}^d, x)(\hat{}\theta - \theta_0)$。将 (26) 重写为

$$\left(\frac{1}{n} \sum_{i \in [n]} \left[\Lambda^1(\tilde{\theta}^1, X_i) - \Lambda^0(\tilde{\theta}^0, X_i) \right] \right) \sqrt{n}(\hat{\theta} - \theta_0). \quad (27)$$

根据假设3和假设7，对于所有 $(\tilde{\theta}, x) \in \mathrm{Nb}(\theta_0, \epsilon) \times \mathcal{X}$，$\tilde{\theta} \mapsto \Lambda^d(\tilde{\theta}, x)$ 连续且一致有界，因此，对于某个有限的 $q_d \in \mathbb{R}^K$，$\frac{1}{n}\sum_{i \in [n]} \Lambda^d(\tilde{\theta}^d, X_i) \xrightarrow{P} q_d^\intercal$。于是，由 (27) 和斯卢茨基引理，根据假设9，(26) 弱收敛于 $\mathcal{N}(0, (q_1 - q_0)^\intercal V_{\theta_0}(q_1 - q_0))$ ^。另一方面，再次根据假设9，(25) 与 $\sqrt{n}(\theta - \theta_0)$ 独立，因此 (21) 是 (25) 和 (26) 之和，弱收敛于 $\mathcal{N}(0,$ $V_\tau + (q_1 - q_0)^\intercal V_{\theta_0}(q_1 - q_0))$。

^ ^ ^ (22) 的条件收敛。令 $\mathcal{F}_{n0} := \sigma\{D_1, \ldots, D_n, \pi(X_1, \theta) \ldots, \pi(X_n, \theta)\ \theta\}$，从而 (21) 是 $\mathcal{F}_{n0}$-可测的。我们证明，在给定 $\mathcal{F}_{n0}$ 的情况下，(22) 依概率弱收敛于一个正态随机变量。我们构造一个鞅阵列并应用引理7。令 ^ ^$\xi_{ni} := (2D_i - 1)(1 + w_i(\theta))\varepsilon_i(\theta)$ $/\sqrt{n}$ 并

$$\mathcal{F}_{ni} := \sigma\{D_1, \ldots, D_n, \pi(X_1, \hat{\theta}), \ldots, \pi(X_n, \hat{\theta}), \varepsilon_1(\hat{\theta}), \ldots, \varepsilon_i(\hat{\theta}), \hat{\theta}\}$$

对于 $i \in [n]$。暂且假设 $\min_{i \in [n]} M_i(\hat{}\theta) > 0$，因此（22）等于 $\sum_{i=1}^n \xi_{ni}$。可以验证，ξ_{n1}，ξ_{n2}，.....，ξ_{nn} 相对于滤过

$\mathcal{F}_{n1} \subset \mathcal{F}_{n2} \subset \ldots \subset \mathcal{F}_{nn}$，利用 $\mu^{D_i}(\hat{}\theta,\ \pi(X_i,\ \hat{}\theta)$ 是 $\mathcal{F}_{n可测的i-1}$，以及假设$\hat{}$ 9(ii)，这意味着给定 θ时，观测值是独立同分布的。

首先，我们验证引理7的方差条件（56）。考虑 $\sum_{i=1}^{n}\mathbb{E}\left[\xi_{ni}^2 \mid \mathcal{F}_{n,i-1}\right]$。我们有

$$\mathbb{E}\left[\xi_{ni}^2 \,\middle|\, \mathcal{F}_{n,i-1}\right] = (1+w_i(\hat{\theta}))^2\sigma_{D_i}^2(\hat{\theta},\pi(X_i,\hat{\theta}))/n,$$

$\hat{}$ 这里我们再次使用了假设9(ii)，以及 $w_i(\theta)$是$\mathcal{F}_{n,i-1}$-可测的。但此时，对于所有 $i \in [n]$ 以及所有 $n \geq 1$，$\mathbb{E}\left[\xi_{ni}^2 \mid \mathcal{F}_{n,i-1}\right]$ 都是 $\mathcal{F}_{n0}$-可测的，因此，对于条件（56），只需 $\sum_{i=1}^{n}\mathbb{E}\left[\xi_{ni}^2 \mid \mathcal{F}_{n,i-1}\right]$ 依概率 $\mathbb{P}$收敛到一个有限常数。假设3(iii)意味着，$\max_{i\in[n]}|\pi(X_i,\hat{}\theta) - \pi(X_i,\theta_0)| = O_P\left(\|\hat{}\theta-\theta_0\|\right)$。那么，在假设8下，我们可以写成 $\sigma_d^2(\hat{}\theta,\pi(X_i,\hat{}\theta)) = \sigma_d^2(\theta_0,\pi(X_i,\theta_0)) + S_i$，其中 $\max_{i\in[n]}|S_i| = O_P\left(\|\hat{}\theta-\theta_0\|\right)$。记

$$\sum_{i=1}^{n}\mathbb{E}\left[\xi_{ni}^2 \,\middle|\, \mathcal{F}_{n,i-1}\right] = \frac{1}{n}\sum_{i\in[n]}(1+w_i(\hat{\theta}))^2\sigma_{D_i}^2(\theta_0,\pi(X_i,\theta_0)) + \frac{1}{n}\sum_{i\in[n]}(1+w_i(\hat{\theta}))^2 S_i. \quad (28)$$

在假设 3、6 和 9的条件下，鉴于引理 4，其证明中的论证仍然适用。于是，由于引理 1(iv)–(vi)，对于任意随机变量序列 $(Q_i)^{i\in[n]}$ 以及任意固定常数 $r \in \mathbb{R}$，根据引理 1(iv)–(vi)，有：

$$\sum_{i\in[n]} w_i(\hat{\theta})^r Q_i = (1+o_P(1))\left\{\left(\frac{N_1}{N_0}\right)^r \sum_{i:D_i=0}\left(\frac{f_{1,\hat{\theta}}(\pi(X_i,\hat{\theta}))}{f_{0,\hat{\theta}}(\pi(X_i,\hat{\theta}))}\right)^r Q_i + \left(\frac{N_0}{N_1}\right)^r \sum_{i:D_i=1}\left(\frac{f_{0,\hat{\theta}}(\pi(X_i,\hat{\theta}))}{f_{1,\hat{\theta}}(\pi(X_i,\hat{\theta}))}\right)^r Q_i\right\}. \quad (29)$$

(28) 中的第二项有界于

$$O_P\left(\|\hat{\theta}-\theta_0\|\right)\left(1+\frac{1}{n}\sum_{i\in[n]} w_i(\hat{\theta})^2\right).$$

假设6通过在 $p \in [\underline{p}_\theta, \bar{p}_\theta]$ 和 $\theta \in \mathrm{Nb}(\theta_0,\epsilon)$中均匀地界定比率$f_{d,\theta}(p)/f_{1-d,\theta}(p)$，结合式（29）和 $(N_d/N_{1-d})^r \xrightarrow{a.s.} \left(\frac{p_d}{1-p_d}\right)^r$（其中 $p_0 := 1-p_1$），将$\frac{1}{n}\sum_{i\in[n]} w_i(\hat{}\theta)^2$ 界定为一个至多相差常数因子的（$1+o_P(1)$）项。因此，在假设9下，（28）中的第二项为 $O_P\left(\|\hat{}\theta-\theta_0\|\right) = o_P(1)$。令 $Q_i := \sigma_{D_i}^2(\theta_0,\pi_i)$，并在 θ 处围绕 θ_0 对（29）右侧作中值展开。在˜假设3和6下，这样做是可行的；这两个假设还将导数在 $(x,\theta) \in \mathcal{X}\times\mathrm{Nb}(\theta_0,\epsilon)$ 中均匀地界定。随后，假设8和9推出，在假设3下，（28）中的第一项为$V_{\sigma,\pi} + o_P(1)$，与引理4的证明相同，其中$V_{\sigma,\pi}$ 由假设6和8可知为有限值。由此可得（56）成立。

其次，我们验证引理7的林德伯格条件（57）。我们需要证明

$$\sum_{i=1}^{n} \mathbb{E}\left[\xi_{ni}^2 \mathbb{I}^{|\xi_{ni}|\geq\eta} \,\middle|\, \mathcal{F}_{n0}\right] \xrightarrow{P} 0 \quad \text{for each } \eta > 0.$$

由于 $\mathbb{I}^{|\xi_{ni}|\geq\eta}$ 被 ξ_{ni}^2/η^2 有界，

$$\sum_{i=1}^{n} \mathbb{E}\left[\xi_{ni}^2 \mathbb{I}^{|\xi_{ni}|\geq\eta} \,\middle|\, \mathcal{F}_{n0}\right] \leq \sum_{t=1}^{n} \frac{\mathbb{E}\left[\xi_{ni}^4 \mid \mathcal{F}_{n0}\right]}{\eta^2} = \frac{1}{n\eta^2} \sum_{i\in[n]} \frac{(1+\tilde{w}_i(\hat{\theta}))^4 \mathbb{E}\left[\varepsilon_i(\hat{\theta})^4 \,\middle|\, \mathcal{F}_{n0}\right]}{n},$$

其中 $\mathbb{E}\left[\varepsilon_i(\hat{\theta})^4 \,\middle|\, \mathcal{F}_{n0}\right] = \sigma_{D_i}^4(\hat{\theta}, \pi(X_i, \hat{\theta}))$ 根据假设 9(ii)，它在 $i \in [n]$ 上根据假设 8均匀有界。由 (29)，$\frac{1}{n^2}\sum_{i\in[n]}(1+\tilde{w}_i(\hat{\theta}))^4 = o_P(1)$，因此满足 (57)。

$\hat{}$ 得出结论：在暂时假设 $\min_{i\in[n]} M_i(\theta) > 0$ 下，引理7适用，因此 (22) 依概率弱收敛到 $\mathcal{N}(0, V_{\sigma,\pi})$。为去掉这一假设，定义 $\hat{}$ $\mathcal{F}_{n0}$-可测集 $A_n := \{\min_{i\in[n]} M_i(\theta) > 0\}$。在 A_n 上，(22) 等于 $\sum_{i\in[n]} \xi_{ni}$。由于$\mathbb{P}(A_n) \to 1$ 根据命题5，所需的收敛性即得证。

消失（23）。根据假设7，$p \mapsto \mu^d(\theta, p)$在紧集$[\underline{p}_\theta, \bar{p}_\theta]$ 上连续，对所有 $\theta \in \mathrm{Nb}(\theta_0, \epsilon)$ 均成立。根据假设9，$\mathbb{P}\left(\hat{\theta} \in \mathrm{Nb}(\theta_0, \epsilon)\right) \to 1$。根据命题5，$\mathbb{P}(A_n) \to 1$，因此（23）为 $o_P(1)$。

消失（24）。根据假设7，$|\mu^d(\theta, p') - \mu^d(\theta, p)| \leq L|p'-p|$ 对所有 p'，$p \in [\underline{p}_\theta, \bar{p}_\theta]$对所有 $\theta \in \mathrm{Nb}(\theta_0, \epsilon)$。然后

$$\max_{i\in[n]} \max_{j\in\mathcal{J}_{\hat{\theta}}(i)} |\mu^{1-D_i}(\hat{\theta}, \pi(X_i, \hat{\theta})) - \mu^{1-D_i}(\hat{\theta}, \pi(X_j, \hat{\theta}))| \lesssim \max_{i\in[n]} \max_{j\in\mathcal{J}_{\hat{\theta}}(i)} |\pi(X_i, \hat{\theta}) - \pi(X_j, \hat{\theta})|.$$

根据 $\mathcal{J}_{\hat{\theta}}(i)$的构造，右侧被 δ_n所界定，其中$\sqrt{n}\delta_n \to 0$。

当卡尺为 $\delta_n = \overline{\Delta}_n \vee \log\frac{N_0}{N_0+1} \vee \log\frac{N_1}{N_1+1}$ 时，以上论证仍然成立。$\hat{}$具体而言，在证明（22）的条件收敛时，$w_i(\hat{\theta})$ 仍然是$\mathcal{F}_{n,\,i-1}$-可测的；（23）恰好为零；而在（24）中，$\sqrt{n}\delta_n \leq \sqrt{n}\left(\widehat{\overline{\Delta}}_n + \log\frac{n}{N_0+1} + \log\frac{n}{N_1+1}\right) = o_P(1)$ 根据命题 5 的证明可知。■

定理 4 的证明。证明过程与定理 2 和定理 3相同。 ■

命题的证明 6。假设 1、3和 9(ii) 根据构造成立，因为 $\mathcal{X}$ 有界。例 1 的假设 9(i)由标准渐近理论推出（Van der Vaart，1998年）。下文假设协变量为 $K = 2$，因此 $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$；一般情形 $K \geq 2$ 可类似处理。首先，我们建立一些一般结果。令θ_k ($\theta_{0,k}$) 表示 θ (θ_0) 的第 k 个元素。对于 $t \in \mathcal{T} := \{\theta^\intercal x : \theta \in \Theta, x \in \mathcal{X}\}$，$\theta^\intercal X$ 有

密度 $f_{\theta^\intercal X}(t) = \int_{\mathcal{X}_1} \Psi\left(x_1, \frac{t-\theta_1 x_1}{\theta_2}\right) \mathrm{d}x_1$，根据关于 Ψ 的假设，该密度严格为正； X的密度。[7] 令 $h\colon \mathcal{X} \to \mathbb{R}^J$， $J \geq 1$为任意可积函数。我们有

$$\mathbb{E}\left[h(X_1, X_2) \mid \theta^\intercal X = t\right] = \frac{1}{f_{\theta^\intercal X}(t)} \int_{\mathcal{X}_1} h\left(x_1, \frac{t-\theta_1 x_1}{\theta_2}\right) \Psi\left(x_1, \frac{t-\theta_1 x_1}{\theta_2}\right) \mathrm{d}x_1. \tag{30}$$

结合期望的塔式性质，可得条件密度

$$f_{\theta^\intercal X|D}(t \mid d) = \begin{cases} \frac{1}{1-p_1} \int_{\mathcal{X}_1} (1 - g_{\theta_0}(x_1, \theta, t)) \Psi\left(x_1, \frac{t-\theta_1 x_1}{\theta_2}\right) \mathrm{d}x_1 & \text{if } d = 0, \\ \frac{1}{p_1} \int_{\mathcal{X}_1} g_{\theta_0}(x_1, \theta, t) \Psi\left(x_1, \frac{t-\theta_1 x_1}{\theta_2}\right) \mathrm{d}x_1 & \text{if } d = 1, \end{cases}$$

其中 $g_{\theta_0}(x_1, \theta, t) := g\left(\theta_{0,1} x_1 + \theta_{0,2} \frac{t-\theta_1 x_1}{\theta_2}\right) \in (0, 1)$，当 $0 < g(t') < 1$ 对有界 $\mathcal{T}$中的所有 t' 成立。因此， $f_{\theta^\intercal X|D}$ 严格为正。根据关于 g 和 Ψ的假设，它关于 t也是连续可微的。此外， $\theta \mapsto f_{\theta^\intercal X|D}(t \mid d)$ 在 $\theta \in \mathrm{Nb}(\theta_0, \epsilon)$ 处连续可微~，且对于一个 $\theta_{0,2} \neq 0$，其导数有界，因为 $\sup_{t \in \mathbb{R}} g'(t) < \infty$，且 Ψ 的导数有界。还可以证明， $\mathbb{E}\left[h(X_1, X_2) \mid D = d, \theta^\intercal X = t\right]$ 是

$$\begin{cases} \frac{1-p_1}{f_{\theta^\intercal X|D}(t \mid 0)} \int_{\mathcal{X}_1} h\left(x_1, \frac{t-\theta_1 x_1}{\theta_2}\right) (1 - g_{\theta_0}(x_1, \theta, t)) \Psi\left(x_1, \frac{t-\theta_1 x_1}{\theta_2}\right) \mathrm{d}x_1 & \text{如果 } d = 0, \\ \frac{p_1}{f_{\theta^\intercal X|D}(t \mid 1)} \int_{\mathcal{X}_1} h\left(x_1, \frac{t-\theta_1 x_1}{\theta_2}\right) g_{\theta_0}(x_1, \theta, t) \Psi\left(x_1, \frac{t-\theta_1 x_1}{\theta_2}\right) \mathrm{d}x_1 & \text{if } d = 1, \end{cases} \tag{31}$$

根据关于 g、Ψ 的假设以及上述推导出的 $f_{\theta^\intercal X|D}$ 的性质，只要 h 连续可微，它关于 t 和 θ 连续可微。现在我们可以验证其余假设了。

假设 6。这些分布是

$$F_{d,\theta}(p) = \begin{cases} \frac{1}{1-p_1} \int_{\mathcal{X}} \mathbb{I}^{g(\theta^\intercal x) \leq p} (1 - g(\theta_0^\intercal x)) \Psi(x) \mathrm{d}x & \text{if } d = 0, \\ \frac{1}{p_1} \int_{\mathcal{X}} \mathbb{I}^{g(\theta^\intercal x) \leq p} g(\theta_0^\intercal x) \Psi(x) \mathrm{d}x & \text{if } d = 1. \end{cases}$$

由于 g 递增，且 $\mathcal{T}_\theta := \{\theta^\intercal x\colon x \in \mathcal{X}\}$ 是紧致的，因此对所有 $\theta \in \Theta$，都存在 $0 < \underline{p}_\theta < \bar{p}_\theta < 1$ 使得 $F_{d,\theta}(\underline{p}_\theta) = 0$ 且 $F_{d,\theta}(\bar{p}_\theta) = 1$。具体而言， $\underline{p}_\theta = g(\inf \mathcal{T}_\theta)$ 和 $\bar{p}_\theta = g(\sup \mathcal{T}_\theta)$。在 $[\underline{p}_\theta, \bar{p}_\theta]$ 上， $F_{d,\theta}$ 具有密度函数

$$f_{d,\theta}(p) = f_{\theta^\intercal X|D}(g^{-1}(p) \mid d)(g^{-1})'(p) = \frac{f_{\theta^\intercal X|D}(g^{-1}(p) \mid d)}{g'(g^{-1}(p))},$$

~~[7]如果 X包含截距项，那么我们就有 $f_{\theta^\intercal X}(t) = \int_{\mathcal{X}_1} \Psi\left(x_1, \frac{t-\theta_3-\theta_1 x_1}{\theta_2}\right) \mathrm{d}x_1$，其中 θ_3 是截距项的系数。~~ 下面，(30) 右侧的 $f_{\theta^\intercal X|D}(t \mid D)$ 和 (31) 需要以类似方式调整，以纳入截距项。

其中 g^{-1} 定义良好，其导数也由反函数定理保证存在。然后，关于 g 的假设以及 $f_{\theta^\intercal X|D}$ 的性质意味着假设 6 成立。

假设 7。根据假设1，期望的塔式性质给出 s

$$\mu^d(\theta,p)=\mathbb{E}\left[Y\mid D=d,\theta^\intercal X=g^{-1}(p)\right]=\mathbb{E}\left[m_d(X)\mid D=d,\theta^\intercal X=g^{-1}(p)\right].$$

但这样一来，(31)的性质、 g 和 m_d 意味着假设7 成立。

假设8。(i)中的下界由例1中关于 ν_d 的假设得到。(i)和(ii)中的Lipschitz条件均可由中值定理推出，因为根据关于 m_d 和 ν_d的假设，σ_d^2和 σ_d^4是形式为 $\mathbb{E}\left[h(X_1,X_2)\mid D=d,\theta^\intercal X=g^{-1}(p)\right]$ 的多项式，其中 h 是连续可微函数，因此(31)适用。■

命题 4 的证明。证明过程与命题6相同。 ■

引理 1 (比率收敛)。 假设卡尺 δ_n 满足 (3)。对于 (17)中的测度，定义

$$R_{di}:=\frac{\mathbb{P}_{N_d}[\pi(X_i)\pm\delta_n]}{F_d[\pi(X_i)\pm\delta_n]}-1, \tag{32}$$

$$\tilde{R}_{di}:=\frac{F_d[\pi(X_i)\pm\delta_n]}{2\delta_n f_d(\pi(X_i))}-1, \tag{33}$$

$$\check{R}_{dji}:=\frac{f_d(\pi(X_j))}{f_d(\pi(X_i))}-1 \tag{34}$$

对于 $j\in\mathcal{J}(i)$， $i\in\{i\in[n]:\ D_i=1-d\}$ 且 $d\in\{0,1\}$。若假设 2 成立，则

(i) 最大值$_{i:D_i=1-d}|R_{di}|=o_P(1)$；且
(ii) 最大值$_{i:D_i=1-d}|R_{di}|=o_P(1)$；且
(iii) 最大值$_{i:D_i=1-d}$最大值$_{j\in\mathcal{J}(i)}|\check{R}_{dji}|=o_P(1)$

作为 $n\to\infty$ 对所有 $d\in\{0,1\}$。假设假设 3 成立且卡尺 δ_n 满足(13)。对于 (18)中的测度，定义

$$R_{di}^\theta:=\frac{\mathbb{P}_{N_d,\theta}[\pi(X_i,\theta)\pm\delta_n]}{F_{d,\theta}[\pi(X_i,\theta)\pm\delta_n]}-1, \tag{35}$$

$$\tilde{R}_{di}^\theta:=\frac{F_{d,\theta}[\pi(X_i,\theta)\pm\delta_n]}{2\delta_n f_{d,\theta}(\pi(X_i,\theta))}-1, \tag{36}$$

$$\check{R}_{dji}^\theta:=\frac{f_{d,\theta}(\pi(X_j,\theta))}{f_{d,\theta}(\pi(X_i,\theta))}-1 \tag{37}$$

对于 $j\in\mathcal{J}_\theta(i)$， $i\in\{i\in[n]:\ D_i=1-d\}$， $d\in\{0,1\}$ 以及 $\theta\in\mathrm{Nb}(\theta_0,\ \epsilon)$。若假设 6 成立，则对任意满足假设 θ 9，

(iv) 最大值$_{i:D_i=1-d}|R^{\theta}_{di}| = o_P(1)$；以及

(v) 最大值$_{i:D_i=1-d}|\tilde{R}^{\hat{\theta}}_{di}| = o_P(1)$；以及

(vi) $\max_{i:D_i=1-d}\max_{j\in\mathcal{J}_{\hat{\theta}}(i)}|\check{R}^{\hat{\theta}}_{dji}| = o_P(1)$

如 $n\to\infty$ 对所有 $d\in\{0,1\}$均成立。

证明。 断言 (i)。考虑

$$|R_{1i}| \le \sup_{p\in[\underline{p},\bar{p}]}\left|\frac{\mathbb{H}_{N_1}[p\pm\delta_n]}{F_1[p\pm\delta_n]}-1\right| = \sup_{p\in[\underline{p},\bar{p}]}\frac{|\mathbb{G}_{N_1}[p\pm\delta_n]|}{\sqrt{N_1}F_1[p\pm\delta_n]} =: W_1. \tag{38}$$

固定一个常数 $\zeta>0$。我们利用经验过程的比率界和尾界来界定 $\mathbb{P}\left(W_1>\zeta\mid D^{(n)}\right)$。为此，注意，对于任意有限的 $\delta_n>0$， $\mathcal{C}_{\delta_n}:=\{[p\pm\delta_n]:\ p\in[\underline{p},\bar{p}]\}$是一个 VC 类，其 VC 维数等于二（例如 Van der Vaart 和 Wellner (1996年，例 2.6.1）。令 $\gamma_n:=2\delta_n$ 下确界$_{p\in[\underline{p},\bar{p}]}f_1(p)$，于是 下确界$_{p\in[\underline{p},\bar{p}]}F_1[p\pm\delta_n]\ge\gamma_n$。事件 $\{W_1>\zeta\}$等于

$$\left\{\sup\left\{\frac{|\mathbb{G}_{N_1}[p\pm\delta_n]|}{F_1[p\pm\delta_n]}:p\in[\underline{p},\bar{p}],F_1[p\pm\delta_n]\ge\gamma_n\right\}>\sqrt{N_1}\zeta\right\}\subset$$

$$\left(\left\{\sup\left\{\frac{|\mathbb{G}_{N_1}[p\pm\delta_n]|}{F_1[p\pm\delta_n]}:p\in[\underline{p},\bar{p}],F_1[p\pm\delta_n]\ge\gamma_n,F_1[p\pm\delta_n]\le\frac12\right\}>\sqrt{N_1}\zeta\right\}\cup \tag{39}$$

$$\left\{\sup\left\{\frac{|\mathbb{G}_{N_1}[p\pm\delta_n]|}{F_1[p\pm\delta_n]}:p\in[\underline{p},\bar{p}],F_1[p\pm\delta_n]\ge\gamma_n,F_1[p\pm\delta_n]>\frac12\right\}>\sqrt{N_1}\zeta\right\}\right). \tag{40}$$

首先，我们对事件 (39) 的概率进行界定。取 $A:=\{p\in[\underline{p},\bar{p}]:\ F_1[p\pm\delta_n]\ge\gamma_n\}$, $B:=\{p\in[\underline{p},\bar{p}]:\ F_1[p\pm\delta_n]\le\frac12\}$。若 $\gamma_n>\frac12$, $A\cap B$为空，且根据 $\sup\emptyset=-\infty$ 的约定，集合 (39) 的测度为零。因此，不失一般性，假设 $\gamma_n\le F_1[p\pm\delta_n]\le\frac12$。于是，$\frac{\gamma_n}{2}\le\sigma_1^2[p\pm\delta_n]:=(F_1[p\pm\delta_n])(1-F_1[p\pm\delta_n])\le\frac14$，注意对于 $F_1[p\pm\delta_n]\ge\gamma_n>0$, $\sigma_1^2[p\pm\delta_n]<F_1[p\pm\delta_n]$。由于 $N_1\gamma_n\xrightarrow{a.s.}\infty$ 且$N_1^{-1}\log(e\vee\log(e\vee N_1))=o(\gamma_n)$ 几乎处处成立，Alexander（1987，定理 2.1）中的条件 (2.2)–(2.3) 几乎处处成立。因此，其中定理 5.1 的证明所给出的界适用。故如 Alexander（1987）中的 (7.66)–(7.67) 所示，

$$\begin{aligned}
&\mathbb{P}\left(\sup\left\{\frac{|\mathbb{G}_{N_1}[p\pm\delta_n]|}{F_1[p\pm\delta_n]}:p\in[\underline{p},\bar{p}],F_1[p\pm\delta_n]\ge\gamma_n,F_1[p\pm\delta_n]\le\frac12\right\}>\sqrt{N_1}\zeta\,\middle|\,D^{(n)}\right)\\
&\le\mathbb{P}\left(|\mathbb{G}_{N_1}[p\pm\delta_n]|>(\sigma_1^2[p\pm\delta_n])\sqrt{N_1}\zeta\text{ for some }p\in[\underline{p},\bar{p}]:\frac{\gamma_n}{2}\le\sigma_1^2[p\pm\delta_n]\le\frac14\,\middle|\,D^{(n)}\right)\\
&\le 36\int_{\gamma_n/2}^{1/4}t^{-1}e^{-\zeta^2N_1t/512}\mathrm{d}t+68e^{-\zeta N_1\gamma_n/256}\\
&\le\frac{36}{\zeta^2N_1\gamma_n}e^{-\zeta^2N_1\gamma_n/1024}+68e^{-\zeta N_1\gamma_n/256}.
\end{aligned}$$

其次，事件 (40) 的概率有如下界

$$\begin{aligned}
&\mathbb{P}\left(\sup\left\{\frac{|\mathbb{G}_{N_1}[p\pm\delta_n]|}{F_1[p\pm\delta_n]} : p\in[\underline{p},\bar{p}], F_1[p\pm\delta_n]>\frac{1}{2}\right\} > \sqrt{N_1}x \,\middle|\, D^{(n)}\right)\\
&\le \mathbb{P}\left(\sup\left\{|\mathbb{G}_{N_1}[p\pm\delta_n]| : p\in[\underline{p},\bar{p}], F_1[p\pm\delta_n]>\frac{1}{2}\right\} > \frac{\sqrt{N_1}}{2}x \,\middle|\, D^{(n)}\right)\\
&\le \mathbb{P}\left(\sup_{p\in[\underline{p},\bar{p}]}|\mathbb{G}_{N_1}[p\pm\delta_n]| > \frac{\sqrt{N_1}}{2}x \,\middle|\, D^{(n)}\right) \qquad (41)
\end{aligned}$$

因为在更大的集合上取上确界不会使其减小。Van der Vaart 和 Wellner (1996年, 定理 2.14.9) 对 (41) 给出的界为 $cN_1\zeta^2e^{-\frac{N_1\zeta^2}{2}}$。因此，

$$\mathbb{P}\left(W_1>\zeta \mid D^{(n)}\right) \le \frac{c_1}{\zeta^2N_1\gamma_n}e^{-\zeta^2N_1\gamma_n/1024} + c_2e^{-\zeta N_1\gamma_n/256} + c_3N_1\zeta^2e^{-N_1\zeta^2/2}, \qquad (42)$$

在 $N_0, N_1 \ge 1$ 成立的集合上，而这发生的概率趋于一。左侧以 1 为界，右侧依概率收敛于零；于是左侧的期望也收敛于零。对于 $W_0 := \sup_{p\in[\underline{p},\bar{p}]}\frac{|\mathbb{G}_{N_0}[|p\pm\delta_n|]|}{\sqrt{N_0}F_0[p\pm\delta_n]}$，也可用类似论证，以 $\max_{i:D_i=1}|R_{0i}|$ 为界。由此得出 $\max_{i:D_i=1-d}|R_{di}| = o_P(1)$。当卡尺为 $\delta_n = \overline{\Delta}_n \vee \log\frac{N_0}{N_0+1} \vee \log\frac{N_1}{N_1+1}$ 时，通过在界定 W_d, $d\in\{0,1\}$ 时令 $\gamma_n := \gamma_{N_d} := 2\inf_{p\in[\underline{p},\bar{p}]} f_d(p)\frac{\log N_d}{N_d+1}$，同样的论证可得该结论。

断言（ii）。根据假设2，f_d 在紧集 $[\underline{p},\bar{p}]$上连续，因此一致连续。由中值定理，f_d 的一致连续性意味着 F_d一致可微。由 F_0的一致可微性，对 $\delta_n = o_P(1)$，有 $\sup_{j:D_j=1}|F_0[\pi(X_j)\pm\delta_n] - 2\delta_nf_0(\pi(X_j))| = o_P(\delta_n)$。为此，固定一个常数 $\zeta>0$。根据 F_0的一致可微性，对于所有 ζ，存在一个常数 $\delta>0$，使得

$$\sup_{p\in[\underline{p},\bar{p}]}\frac{|F_0(p+\delta_n)-F_0(p)-\delta_nf_0(p)|}{\delta_n} \le \zeta$$

每当 $\delta_n\le\bar{\delta}$时。事件 $\{\sup_{p\in[\underline{p},\bar{p}]}\frac{|F_0(p+\delta_n)-F_0(p)-\delta_nf_0(p)|}{\delta_n} > \zeta\}$ 等于

$$\left\{\sup_{p\in[\underline{p},\bar{p}]}\frac{|F_0(p+\delta_n)-F_0(p)-\delta_nf_0(p)|}{\delta_n} > \zeta, \delta_n\le\bar{\delta}\right\}$$
$$\cup\left\{\sup_{p\in[\underline{p},\bar{p}]}\frac{|F_0(p+\delta_n)-F_0(p)-\delta_nf_0(p)|}{\delta_n} > \zeta, \delta_n>\bar{\delta}\right\}.$$

根据一致可微性，第一个事件的测度为零。第二个事件的概率由 $\mathbb{P}\left(\delta_n>\bar{\delta}\right)$ 控制，而 $\mathbb{P}\left(\delta_n>\bar{\delta}\right)$ 为 $o(1)$。然后，注意到$F_0[p\pm\delta_n] = F_0(p+\delta_n)-F_0(p)+F_0(p)-F_0(p-\delta_n)$，且 $\max^{i\in[n]}\max_{j\in\mathcal{J}(i)}|f_0(\pi(X_i))-f_0(\pi(X_j))| = o_P(1)$（见断言 (iii) 的证明），即可得出结论，因此 $2f_0(p) = f_0(p) + f_0(p-\delta_n) + f_0(p) -$

$f_0(p-\delta_n) = f_0(p) + f_0(p-\delta_n) + o_P(1)$。由于 $\inf_{p\in[\underline{p},\bar{p}]} f_0(p) > 0$ 根据假设 2，断言(ii)得证。当卡尺为 $\delta_n = \overline{\Delta}_n \vee \log\frac{N_0}{N_0+1} \vee \log\frac{N_1}{N_1+1}$ 时，命题 1、连续映射定理和大数定律推出 $\delta_n = o_P(1)$，由此结合上述论证即可得出结论。

断言 (iii)。由于 f_0 根据假设 2 是一致连续的，

$$\max_{i\in[n]} \max_{j\in\mathcal{J}(i)} |f_0(\pi(X_j)) - f_0(\pi(X_i))| = o_P(1).$$

为证明这一点，固定一个常数 $\zeta > 0$。由 f_0 的一致连续性，对于所有 ζ，存在一个 $\eta > 0$，使得当 $|p-p'| \le \eta$ 时，对于所有 p、$p' \in [\underline{p},\bar{p}]$，$|f_0(p) - f_0(p')| \le \zeta$。事件 $\{|f_0(\pi(X_i)) - f_0(\pi(X_j))| > \zeta\}$ 等于 $\{|f_0(\pi(X_i)) - f_0(\pi(X_j))| > \zeta, |\pi(X_i) - \pi(X_j)| \le \eta\} \cup \{|f_0(\pi(X_i)) - f_0(\pi(X_j))| > \zeta, |\pi(X_i) - \pi(X_j)| > \eta\}$。第一个事件的测度为零，这是由一致连续性得出的。由于 $j \in \mathcal{J}(i)$，第二个事件的概率不超过 $\mathbb{P}(|\pi(X_i) - \pi(X_j)| > \eta) \le \mathbb{P}(\delta_n > \eta)$，而其为 $o(1)$。因此，$\max_{i:D_i=0} \max_{j\in\mathcal{J}(i)} |\check{R}_{0ji}| = o_P(1)$，因为根据假设2，$\inf_{p\in[\underline{p},\bar{p}]} f_0(p) > 0$。当卡尺为 $\delta_n = \overline{\Delta}_n \vee \log\frac{N_0}{N_0+1} \vee \log\frac{N_1}{N_1+1}$ 时，证明断言 (ii) 的论证同样适用。

^断言(iv)、(v)、(vi) 沿用相同的论证，通过对 θ进行条件化，并利用假设6 和9，以及Alexander (1987年) 和Van der Vaart 和 Wellner(1996年) 的界限常数不依赖于底层分布。这⌢对于（13）中的任意卡尺选择均成立，因为 $\overline{\Delta}_n = o_P(1)$ 由命题5的证明可知。■

引理2(误差项)。假设*2* 和*4* 成立，且卡尺 δ_n 满足(3)。则

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(V_E^{-1/2}\sqrt{n}E^{CDO} \le x \,\middle|\, D^{(n)}, PS^{(n)} \right) - \Phi(x) \right| \xrightarrow{P} 0 \quad as\ n\to\infty,$$

对 $V_E :=$ $\frac{1}{n}\sum_{i\in[n]}(\mathbb{I}^{M_i>0} + w_i)^2\sigma^2_{D_i}(\pi(X_i))$ 和标准正态分布函数 Φ .

证明。 我们应用林德伯格–费勒中心极限定理，因为 E_i在给定 $(D^{(n)}, PS^{(n)})$的情况下，沿着 $i \in [n]$ (其中 M_i, w_i是给定 $(D^{(n)}, PS^{(n)})$的常数）相互独立但并非同分布，且均值为零。由假设 2可知， μ^d以及由此得到的 ε均有定义。根据 V_E, $\mathbb{V}\left[\sum_{i\in[n]} E_i/\sqrt{nV_E} \mid D^{(n)}\text{的定义}, \ PS^{(n)}\right] = 1$。因此，我们只需验证林德伯格–费勒条件:

$$\sum_{i\in[n]} \mathbb{E}\left[(E_i/\sqrt{nV_E})^2 \mathbb{I}_{|E_i/\sqrt{nV_E}|\ge\eta} \,\middle|\, D^{(n)}, PS^{(n)} \right] \xrightarrow{P} 0 \quad \text{for all constants } \eta > 0. \tag{43}$$

由于 $\mathbb{I}^{|E_i/\sqrt{nV_E}|\geq\eta}$ 由 $E_i^2/(\eta^2 nV_E)$ 在 $\{E_i^2/(\eta^2 nV_E)\geq 1\}$上界定，故（43）中的期望满足

$$\begin{aligned}\frac{1}{nV_E}\mathbb{E}\left[E_i^2\,\mathbb{I}^{|E_i|\geq\eta\sqrt{nV_E}}\,\middle|\,D^{(n)},PS^{(n)}\right]&\leq\frac{1}{nV_E}\mathbb{E}\left[E_i^4/(\eta^2 nV_E)\,\middle|\,D^{(n)},PS^{(n)}\right]\\&=\frac{\mathbb{E}\left[E_i^4\mid D^{(n)},PS^{(n)}\right]}{(\eta nV_E)^2},\end{aligned}\tag{44}$$

其中 $\mathbb{E}\left[E_i^4\mid D^{(n)},PS^{(n)}\right]=(\mathbb{I}^{M_i>0}+w_i)^4\mathbb{E}\left[\varepsilon_i^4\mid D_i,\pi(X_i)\right]$，满足 $\mathbb{E}\left[\varepsilon_i^4\mid D_i,\pi(X_i)\right]\leq \sup_{d\in\mathbb{B},p\in[0,1]}\mathbb{E}\left[\varepsilon_i^4\mid D_i=d,\pi(X_i)=p\right]<\infty$，根据假设4成立。由引理4，$V_E\xrightarrow{P}V_\sigma$，$_\pi\in(0,\infty)$，因此（44）至多为 $(1+o_P(1))n^{-1}\left(\frac{1}{n}\sum_{i\in[n]}(\mathbb{I}^{M_i>0}+w_i)^4\right)$，相差一个常数因子。固定一个常数 $C>0$。根据马尔可夫不等式，

$$\mathbb{P}\left(\frac{1}{n}\sum_{i\in[n]}(\mathbb{I}^{M_i>0}+w_i)^4>C\right)\leq C^{-1}\max_{i\in[n]}\mathbb{E}(\mathbb{I}^{M_i>0}+w_i)^4.$$

B下面，我们证明对于任意 $\varrho>0$，$n^{-\varrho}\max_{i\in[n]}\mathbb{E}(\mathbb{I}^{M_i>0}+w_i)^4=O(1)$，因此

$$\frac{1}{n}\sum_{i\in[n]}(\mathbb{I}^{M_i>0}+w_i)^4=O_P\left(n^{\varrho}\right).$$

那么，(44) 可由 $(1+o_P(1))O_P(n^{\varrho-1})$ 控制，而其为 $o_P(1)$ 对于 $\varrho<1$成立，因此满足 (43)。由于 $w_i\geq 0$ 且 $(1+x)^4\leq(2x)^4$ 对于 $x\geq 1$，

$$(\mathbb{I}^{M_i>0}+w_i)^4\leq 2^4+2^4 w_i^4\,\mathbb{I}^{w_i>1}.$$

我们有 $w_i\leq M_i\max_{j\in\mathcal{J}(i)}M_j^{-1}$，因此

$$\mathbb{E}w_i^4\,\mathbb{I}^{w_i>1}\leq\mathbb{E}\left[\mathbb{I}^{w_i>1}M_i^4\max_{j\in\mathcal{J}(i)}M_j^{-4}\right].$$

由引理 3 可得结论。

当卡尺为 $\delta_n=\underline{\Delta}_n\vee\log\frac{N_0}{N_0+1}\vee\log\frac{N_1}{N_1+1}$ 时，它是 $\sigma\left\{D^{(n)},PS^{(n)}\right\}$-可测的，因此 M_i，w_i 是由 $(D^{(n)},\ PS^{(n)})$ 给出的常数。引理 3 和 4 完成证明。■

引理3 (林德伯格–费勒界)。 假设卡尺 δ_n 满足 (3) 且假设 2 成立。那么，对于任意有限的固定常数整数 $r\geq 2$ 以及任意有限的固定常数 $\varrho>0$，

$$\max_{i\in[n]}\mathbb{E}_{\mathbb{I}^{w_i>1}}\max_{j\in\mathcal{J}(i)}\left(\frac{M_i}{M_j}\right)^r=o\left(n^{\varrho}\right).\tag{45}$$

证明。若 $w_i > 1$ 对某个 $i \in [n]$成立，则 $M_i \geq 1$，因此对所有 $j \in \mathcal{J}(i)$，$M_j \geq 1$ 也成立，因为 $j \in \mathcal{J}(i)$ 当且仅当 $i \in \mathcal{J}(j)$。这进而意味着 $N_0, N_1 \geq 1$。因此，对所有 $i \in [n]$，$\mathbb{I}^{w_i>1} \leq \mathbb{I}^{M_j \geq 1}\mathbb{I}^{N_0 \geq 1}\mathbb{I}^{N_1 \geq 1}$ 对所有 $j \in \mathcal{J}(i)$ 成立。根据定义，在（17）和（32）的记号下，（45）左侧有如下界

$$\sum_{d\in\{0,1\}} \max_{i:D_i=d} \mathbb{E}\left[\left(\frac{N_{1-d}\mathbb{I}^{N_d\geq 1}}{N_d}\right)^r \max_{j\in\mathcal{J}(i)}\left(\frac{F_{1-d}[\pi(X_i)\pm\delta_n]}{F_d[\pi(X_j)\pm\delta_n]}\frac{1+R_{1-d,i}}{1+R_{d,j}}\mathbb{I}^{N_{1-d}\geq 1}\mathbb{I}^{M_j\geq 1}\right)^r\right]$$
$$\lesssim \sum_{d\in\{0,1\}} \max_{i:D_i=d} \mathbb{E}\left[\left(\frac{N_{1-d}\mathbb{I}^{N_d\geq 1}}{N_d}\right)^r \max_{j\in\mathcal{J}(i)}\left(\frac{1+R_{1-d,i}}{1+R_{d,j}}\mathbb{I}^{N_{1-d}\geq 1}\mathbb{I}^{M_j\geq 1}\right)^r\right], \tag{46}$$

其中第二行由假设2推出：由于 f_0、f_1 具有相同的支撑集，且有界并远离零和无穷大，因此对于 $c \neq 0$, $0 < {}^{2c}\inf_{p\in[\underline{p},\bar{p}]} f_{1-d}(p)\ {}^{2c}\sup_{p\in[\underline{p},\bar{p}]} f_d(p) \leq F_{1-d}[a+c]/F_d[b+c] \leq {}^{2c}\sup_{p\in[\underline{p},\bar{p}]} f_{1-d}(p)\ {}^{2c}\inf_{p\in[\underline{p},\bar{p}]} f_d(p) < \infty$。我们讨论 (46) 中 $d = 0$ 的情形；$d = 1$的情形由对称性可得。对于 (46) 中的期望，应用两次柯西–施瓦茨不等式可得

$$\mathbb{E}\left[\left(\frac{N_1\mathbb{I}^{N_0\geq 1}}{N_0}\right)^r \max_{j\in\mathcal{J}(i)}\left(\frac{1+R_{1i}}{1+R_{0j}}\mathbb{I}^{N_1\geq 1}\mathbb{I}^{M_j\geq 1}\right)^r\right]$$
$$\leq \sqrt{\mathbb{E}(N_1\mathbb{I}^{N_0\geq 1}/N_0)^{2r}}\sqrt{\mathbb{E}(1+R_{1i})^{4r}\mathbb{I}^{N_1\geq 1}\mathbb{E}\left(\max_{j\in\mathcal{J}(i)}(\mathbb{I}^{M_j\geq 1}/(1+R_{0j}))^r\right)^4}. \tag{47}$$

这里 $\mathbb{E}(N_1\mathbb{I}^{N_0\geq 1}/N_0)^{2r} \leq n^{2r}\mathbb{E}(\mathbb{I}^{N_0\geq 1}/N_0)^{2r}$，其中

$$\begin{aligned}\mathbb{E}(\mathbb{I}^{N_0\geq 1}/N_0)^{2r} &= \sum_{n_0=1}^{n}\binom{n}{n_0}(1-p_1)^{n_0}p_1^{n-n_0}\left(\frac{1}{n_0}\right)^{2r}\\ &= \sum_{n_0=0}^{n-1}\frac{n}{n_0+1}\binom{n-1}{n_0}(1-p_1)^{n_0+1}p_1^{n-(n_0+1)}\left(\frac{1}{n_0+1}\right)^{2r}\\ &= n\frac{1-p_1}{p_1}\sum_{n_0=0}^{n-1}\binom{n-1}{n_0}(1-p_1)^{n_0}p_1^{n-n_0}\left(\frac{1}{n_0+1}\right)^{2r+1}.\end{aligned}$$

该直 线由 $O(n^{-2r})$ 提出，Cribari-Neto等人（2000年），因此 $\mathbb{E}(N_1\mathbb{I}^{N_0\geq 1}/N_0)^{2r} = O(1)$.

接下来，我们对（47）的第二个因子进行界定。我们有 $\mathbb{E}(1+R_{1i})^{4r}\mathbb{I}^{N_1\geq 1} \leq 2^{4r} + c\mathbb{E}|R_{1i}|^{4r}\mathbb{I}^{N_1\geq 1}$ 其中某个常数 $c > 0$满足 $\max_{i:D_i=0}|R_{1i}| \leq W_1$ 对于 W_1 属于（38）中的情形。由于 $W_1 \geq 0$,

$$\mathbb{E}\left[W_1^{4r}\,\middle|\, D^{(n)}\right] = \int_0^\infty \mathbb{P}\left(W_1^{4r} > w \mid D^{(n)}\right)\mathrm{d}w = \int_0^\infty \mathbb{P}\left(W_1 > w^{\frac{1}{4r}} \mid D^{(n)}\right)\mathrm{d}w.$$

根据引理 42）的（1,

$$\mathbb{E}\left[W_1^{4r} \mid D^{(n)}\right] \leq \frac{c_0}{N_1\gamma_n}\int_0^\infty \frac{1}{w^{\frac{2}{4r}}} e^{-w^{\frac{2}{4r}} N_1\gamma_n/1024}\mathrm{d}w + c_1\int_0^\infty e^{-w^{\frac{1}{4r}} N_1\gamma_n/256}\mathrm{d}w \tag{48}$$

$$+ cN_1\int_0^\infty w^{\frac{2}{4r}} e^{-\frac{N_1 w^{\frac{2}{4r}}}{2}}\mathrm{d}w. \tag{49}$$

(48) 式第一项中的积分为 $\frac{c_0}{N_1\gamma_n}\frac{1}{\lambda_n}$，$\int_0^\infty t^{-1}\lambda_{n,N_1} e^{-\lambda_{n,N_1} t} t^{2r-1}\mathrm{d}t$，其中 $\lambda_{n,N_1} := \frac{N_1\gamma_n}{1024}$ 对于 $N_1 \geq 1$ 严格为正。该积分是 Exponential(λ_{n,N_1}) 变量的第 $2r-2$ 阶矩，对于 $r \geq 2$ 有定义；对于有限整数 r，其绝对值在常数因子意义下被 $\left(\frac{1}{\lambda_{n,N_1}}\right)^{2r-2}$ 所界定。因此，(48) 式的第一项被 $c_0\left(\frac{1}{N_1\gamma_n}\right)^{2r} \simeq c_0\left(\frac{1}{N_1 \log n}\right)^{2r}$ 所界定。类似的论证表明，(48) 式和 (49) 式中的第二个积分分别被 $c_1\left(\frac{1}{N_1\gamma_n}\right)^{4r}$ 和 $c\left(\frac{1}{N_1}\right)^{2r}$ 所界定。由此可得

$$\mathbb{E}(1+R_{1i})^{4r}\mathbb{I}_{N_1\geq 1} \leq O(1) + c_0 n^{2r}\mathbb{E}\left(\frac{\mathbb{I}_{N_1\geq 1}}{N_1}\right)^{2r} + c_1 n^{4r}\mathbb{E}\left(\frac{\mathbb{I}_{N_1\geq 1}}{N_1}\right)^{4r}.$$

通过对 $\mathbb{E}(N_1\mathbb{I}_{N_0\geq 1}/N_0)^{2r}$ 的47式进行界定，右侧为 $O(1)$。

最后，47式的最后一个因子满足

$$\max_{j\in\mathcal{J}(i)}(\mathbb{I}_{M_j\geq 1}/(1+R_{0j}))^r = \max_{j\in\mathcal{J}(i)}\left(\frac{N_0 F_0[\pi(X_j)\pm\delta_n]}{M_j}\mathbb{I}_{M_j\geq 1}\right)^r \leq (2n\|f_0\|_\infty\delta_n)^r \tag{50}$$

$$\lesssim (\log n)^r.$$

由于任意$(\log n)^r = o(n^\varrho)$，因此 $\varrho > 0$，45式成立。

当卡尺为 $\delta_n = \underline{\Delta}_n \vee \log\frac{N_0}{N_0+1} \vee \log\frac{N_1}{N_1+1}$时，根据引理 1 的证明，(48) 式成立，其中 $\gamma_n := 2\inf_{p\in[\underline{p},\bar{p}]} f_1(p)\frac{\log N_1}{N_1+1}$，这表明 $\mathbb{E}(1+R_{1i})^{4r}\mathbb{I}_{N_1\geq 1} = O(1)$。考虑 (50) 式，其中 $\delta_n^r \leq 2^r\left(\underline{\Delta}_n^r + (2\log n)^r\left(\frac{1}{N_0+1}\right)^r + \left(\frac{1}{N_1+1}\right)^r\right)$。根据 Cribari-Neto 等人（2000），这里有 $\mathbb{E}\left(\frac{1}{N_d+1}\right)^r = O(n^{-r})$。命题1的证明和引理 5 中的论证表明$n^r\mathbb{E}\underline{\Delta}_n^r = O((\log n)^r)$。由此可得 $n^r\mathbb{E}\delta_n^r = O((\log n)^r)$，因此由 (50) 式可得 (45) 式。■

引理4 (半参数效率)。 若假设*2* 和*4* 成立，且卡尺 δ_n满足(3)，则 $V_E \xrightarrow{P} V_{\sigma,\pi}$ 当 $n\to\infty$。

证明。 根据命题 2， 最小值$_{i\in[n]} M_i > 0$ 以趋于 1 的概率成立，因此，以趋于 1 的概率，

$$V_E = \frac{1}{n}\sum_{i\in[n]}(1+w_i)^2\sigma_{D_i}^2(\pi(X_i)) = \frac{1}{n}\sum_{i\in[n]}(1+2w_i+w_i^2)\sigma_{D_i}^2(\pi(X_i))$$

$$= \mathbb{E}\sigma_D^2(\pi(X)) + o_P(1) + \frac{2}{n}\sum_{i\in[n]} w_i\sigma_{D_i}^2(\pi(X_i)) + \frac{1}{n}\sum_{i\in[n]} w_i^2\sigma_{D_i}^2(\pi(X_i)) \tag{51}$$

根据大数定律。在（17）的记号下，根据定义，我们有

$$\sum_{i\in[n]} w_i^r \sigma^2_{D_i}(\pi(X_i)) = \left(\frac{N_1}{N_0}\right)^r \sum_{i:D_i=0} \left(\frac{1}{N_1}\sum_{j\in\mathcal{J}(i)} \frac{1}{\mathbb{E}^{N_0}[\pi(X_j)\pm\delta_n]}\right)^r \sigma^2_{D_i}(\pi(X_i)) + \left(\frac{N_0}{N_1}\right)^r \sum_{i:D_i=1} \left(\frac{1}{N_0}\sum_{j\in\mathcal{J}(i)} \frac{1}{\mathbb{E}^{N_1}[\pi(X_j)\pm\delta_n]}\right)^r \sigma^2_{D_i}(\pi(X_i)). \quad (52)$$

˜令 $\mathbb{F}_{N_0}[\pi(X_j)\pm\delta_n] = (1+R_{0j})F_0[\pi(X_j)\pm\delta_n]$和 $F_0[\pi(X_j)\pm\delta_n] = 2\delta_n f_0(\pi(X_j))(1+R_{0j})$˜ 分别表示式 (32)、(33) 中的 R_{0j}, R_{0j}。由引理 1,

$$\max_{j:D_j=1} |R_{0j}| = o_P(1) \quad \text{and} \quad \max_{j:D_j=1} |\tilde{R}_{0j}| = o_P(1).$$

于是我们可以写成

$$\begin{aligned}\frac{1}{N_1}\sum_{j\in\mathcal{J}(i)} \frac{1}{\mathbb{E}^{N_0}[\pi(X_j)\pm\delta_n]} &= \frac{1}{N_1}\sum_{j\in\mathcal{J}(i)} \frac{1}{F_0[\pi(X_j)\pm\delta_n]}\frac{1}{1+R_{0j}} \\ &= \frac{1}{N_1}\sum_{j\in\mathcal{J}(i)} \frac{1}{2\delta_n f_0(\pi(X_j))}\frac{1}{1+\tilde{R}_{0j}}\frac{1}{1+R_{0j}} \\ &= \frac{1+o_P(1)}{N_1}\sum_{j\in\mathcal{J}(i)} \frac{1}{2\delta_n f_0(\pi(X_j))}\end{aligned}$$

其中 $o_P(1)$ 这些项关于 $i\in[n]$均匀成立。令 $f_0(\pi(X_j)) = f_0(\pi(X_i))(1+\check{}R_{0ji})$ 表示式 $\check{}R_{0ji}$（34）中的。由引理 1， $\max^{i:D_i=0}\max_{j\in\mathcal{J}(i)}|\check{}R_{0ji}| = o_P(1)$。于是，根据连续映射定理,

$$\begin{aligned}\frac{1}{N_1}\sum_{j\in\mathcal{J}(i)} \frac{1}{2\delta_n f_0(\pi(X_j))} &= \frac{1}{N_1}\sum_{j\in\mathcal{J}(i)} \frac{1}{2\delta_n f_0(\pi(X_i))}\frac{1}{1+\check{R}_{0ji}} \\ &= (1+o_P(1))\frac{f_1(\pi(X_i))}{f_0(\pi(X_i))}\frac{\mathbb{E}^{N_1}[\pi(X_i)\pm\delta_n]}{2\delta_n f_1(\pi(X_i))},\end{aligned}$$

因为 $\inf_{p\in[\underline{p},\bar{p}]} f_1(p) > 0$ 根据假设 2。令

$$\frac{\mathbb{E}^{N_1}[\pi(X_i)\pm\delta_n]}{2\delta_n f_1(\pi(X_i))} = \frac{\mathbb{E}^{N_1}[\pi(X_i)\pm\delta_n]}{F_1[\pi(X_i)\pm\delta_n]}\frac{F_1[\pi(X_i)\pm\delta_n]}{2\delta_n f_1(\pi(X_i))},$$

其中 最大值$_{i:D_i=0}\left|\frac{\mathbb{E}^{N_1}[\pi\ X_i\ \pm\delta_n]()()\pm\delta_n]}{F_1[\pi\ X_i}-1\right| = o_P(1)$ 且 最大值$_{i:D_i=0}\left|\frac{F_1[\pi(\quad X_i\ \pm\delta_n])((}{2\delta_n f_1\ \pi\ X_i}-1\right| = o_P(1)$根据引理 1。根据对称性，对（52）的第二项也适用类似的论证。于是（52），

除以 n，等于

$$\begin{aligned}
&\frac{1+o_P(1)}{n}\left[\left(\frac{N_1}{N_0}\right)^r \sum_{i:D_i=0}\left(\frac{f_1(\pi(X_i))}{f_0(\pi(X_i))}\right)^r \sigma^2_{D_i}(\pi(X_i))\right.\\
&\qquad\left.+\left(\frac{N_0}{N_1}\right)^r \sum_{i:D_i=1}\left(\frac{f_0(\pi(X_i))}{f_1(\pi(X_i))}\right)^r \sigma^2_{D_i}(\pi(X_i))\right]\\
&\xrightarrow{P} \mathbb{E}\left[(1-D)\left(\frac{p_1}{1-p_1}\frac{f_1(\pi(X))}{f_0(\pi(X))}\right)^r \sigma^2_D(\pi(X))\right] \qquad (53)\\
&\qquad +\mathbb{E}\left[D\left(\frac{1-p_1}{p_1}\frac{f_0(\pi(X))}{f_1(\pi(X))}\right)^r \sigma^2_D(\pi(X))\right], \qquad (54)
\end{aligned}$$

根据大数定律，当 $(N_d/N_{1-d})^r \xrightarrow{a.s.} \left(\frac{p_d}{1-p_d}\right)^r$，且 f_d/f_{1-d} 由假设 2 一致有界，且 σ^2_d 由假设 4 有界。对于 $r=1$，(53) 为

$$\begin{aligned}
\mathbb{E}\left[\frac{p_1}{1-p_1}\frac{f_1(\pi(X))}{f_0(\pi(X))}\sigma^2_0(\pi(X))\,\middle|\, D=0\right](1-p_1) &= \int_{\underline{p}}^{\bar{p}} p_1\frac{f_1(p)}{f_0(p)}\sigma^2_0(p)f_0(p)\mathrm{d}p\\
&= \mathbb{E}\pi(X)\sigma^2_0(\pi(X)),
\end{aligned}$$

其中我们利用了 $f_1(p)=\frac{p}{p_1}f_{\pi(X)}(p)$，其中 $f_{\pi(X)}$ 是 $\pi(X)$ 的密度。注意到 $f_0(p)=\frac{1-p}{1-p_1}f_{\pi(X)}(p)$，(54) 是 $\mathbb{E}(1-\pi(X))\sigma^2_1(\pi(X))$。对于 $r=2$，采用相同的论证可得，对于 (53)，有 $\mathbb{E}\frac{\pi(X)}{1-\pi(X)}\sigma^2_0(\pi(X))$；对于 (54)，有 $\mathbb{E}(1-\pi(X))\sigma^2_1(\pi(X))$。注意，$\mathbb{E}\sigma^2_D(\pi(X)) = \mathbb{E}(1-\pi(X))\sigma^2_0(\pi(X)) + \mathbb{E}\pi(X)\sigma^2_1(\pi(X))$。合并各项即可得到所述结论；根据引理 1，该结论对于 $\delta_n = \underline{\Delta}_n \vee \log\frac{N_0}{N_0+1} \vee \log\frac{N_1}{N_1+1}$ 同样成立。■

补充材料

命题 3 的证明。 半参数下界是有效影响函数的方差。当观测样本为 $((Y_i, D_i, X_i))^{i\in[n]}$ 时，ATE 的有效影响函数为

$$\chi_{\mathcal{X}}(Y, D, X) := \frac{D(Y - \mu^1_{\mathcal{X}}(X))}{\pi(X)} - \frac{(1-D)(Y - \mu^0_{\mathcal{X}}(X))}{1-\pi(X)} + \mu^1_{\mathcal{X}}(X) - \mu^0_{\mathcal{X}}(X) - \tau,$$

参见 Hahn (1998年)。那么 $V = V_{\text{eff}}$ 在命题 10) 下由式（3）可得。可以验证$V = \mathbb{V}[\chi(Y, D, \pi(X))]$, 其中

$$\begin{aligned}\chi(Y, D, \pi(X)) := &\frac{D(Y - \mu^1(\pi(X)))}{\pi(X)} - \frac{(1-D)(Y - \mu^0(\pi(X)))}{1-\pi(X)} \\ &+ \mu^1(\pi(X)) - \mu^0(\pi(X)) - \tau.\end{aligned}$$

假设 1 和 $\mathbb{E}[D \mid X] = \pi(X)$ 意味着

$$\mathbb{E}\chi_{\mathcal{X}}(Y, D, X)(\chi(Y, D, \pi(X)) - \chi_{\mathcal{X}}(Y, D, X)) = 0,$$

thus

$$\begin{aligned}V = \mathbb{V}[\chi(Y, D, \pi(X))] &= \mathbb{V}[\chi(Y, D, \pi(X)) - \chi_{\mathcal{X}}(Y, D, X)] + \mathbb{V}[\chi_{\mathcal{X}}(Y, D, X)] \\ &\geq \mathbb{V}[\chi_{\mathcal{X}}(Y, D, X)] = V_{\text{eff}}.\end{aligned}$$

对于处理组平均处理效应，$V_{\text{t, eff}} \leq V_{\text{t}}$ 可类似地得到，因为未知倾向得分下处理组平均处理效应的有效影响函数（Hahn，1998年）为

$$\begin{aligned}\chi_{\text{t},\mathcal{X}}(Y, D, X) := &\frac{D}{p_1}(Y - \mu^1_{\mathcal{X}}(X)) - \frac{1-D}{p_1}\frac{\pi(X)}{1-\pi(X)}(Y - \mu^0_{\mathcal{X}}(X)) \\ &+ \frac{D}{p_1}(\mu^1_{\mathcal{X}}(X) - \mu^0_{\mathcal{X}}(X) - \tau_{\text{t}}).\end{aligned}$$

在假设1下，可以验证 $V_{\text{t}} = \mathbb{V}[\chi_{\text{t}}(Y, D, \pi(X))]$, 其中

$$\begin{aligned}\chi_{\text{t}}(Y, D, \pi(X)) := &\frac{D}{p_1}(Y - \mu^1(\pi(X))) - \frac{1-D}{p_1}\frac{\pi(X)}{1-\pi(X)}(Y - \mu^0(\pi(X))) \\ &+ \frac{D}{p_1}(\mu^1(\pi(X)) - \mu^0(\pi(X)) - \tau_{\text{t}}),\end{aligned}$$

并且

$$\mathbb{E}\chi_{\text{t},\mathcal{X}}(Y, D, X)(\chi_{\text{t}}(Y, D, \pi(X)) - \chi_{\text{t},\mathcal{X}}(Y, D, X)) = 0,$$

从而证明了该断言。

■

引理 5（有序均匀间距的矩界）。 设$(U_1, U_2, \ldots, U_n)$为独立同分布$\sim$均匀分布$(0,1)$的次序统计量，记为 $U_{(1)} \le U_{(2)} \le \ldots \le U_{(n)}$。令 $\tilde{U}_1 := U_{(1)}$，$\tilde{U}_i := U_{(i)} - U_{(i-1)}$（$i = 2, \ldots, n$）以及 $U_{n+1} := 1 - U_{(n)}$ 为$(U_i)^{i\in[n]}$生成的间距。令 $\tilde{U}_{(1)} \le \tilde{U}_{(2)} \le \ldots \le \tilde{U}_{(n+1)}$ 为有序间距。那么，对于任意有限的固定整数$1 \le a < \frac{n+1}{2}$和 $n \ge 2$,

$$\mathbb{E}\tilde{U}^a_{(r)} = \begin{cases} O\left(\left(\frac{1}{n}\right)^{2a}\right) & for\ r = 1 \\ O\left(\left(\frac{\log r}{n}\right)^{a}\right) & for\ all\ r = 2, 3, \ldots, n+1. \end{cases}$$

特别地，$\mathbb{E}\tilde{U}^a_{(r)} = o(1)$对所有 $r \in [n+1]$ 以及有限的固定整数 $1 \le a < \frac{n+1}{2}$均成立。

证明。根据Shorack 和 Wellner（2009年，第21章），$\tilde{U}_{(r)}$服从$\frac{Z_{r:n+1}}{\sum_{i\in[n+1]} Z_i}$，其中$(Z_1, Z_2, \ldots, Z_{n+1})$独立同分布$\sim$指数分布(1)，次序统计量为 $Z_{1:n+1} \le Z_{2:n+1} \le \ldots \le Z_{n+1:n+1}$。由柯西–施瓦茨不等式可得

$$\mathbb{E}\tilde{U}^a_{(r)} \le \sqrt{\mathbb{E}Z^{2a}_{r:n+1}\mathbb{E}\left(\sum_{i\in[n+1]} Z_i\right)^{-2a}}.$$

这里，对于 $r \ge 2$，以及由引理6可知的有限固定整数 $a \ge 1$，$\mathbb{E}Z^{2a}_{1:n+1} = O(n^{-2a})$和 $\mathbb{E}Z^{2a}_{r:n+1} = O\left((\log r)^{2a}\right)$。$n+1$ 个独立同分布的指数分布 (1) 随机变量之和服从 Gamma$(n+1, 1)$ 分布。因此，$(\sum_{i\in[n+1]} Z_i)^{-1}$服从 Inverse-Gamma$(n+1, 1)$ 分布，其 $2a$ 阶矩等于 $(n-2a)!\,/\,n! \lesssim n^{-2a}$，其中 $2a < n+1$，最后一个不等式由 $\sqrt{2\pi}n^{n+1/2}e^{-n} \le n! \le en^{n+1/2}e^{-n}$（Robbins，1955年）得出，因为对于正整数 a，$n - 2a + 1/2 \ge 0$。■

引理6 (指数分布(1)次序统计量)。 设 $(Z_1, Z_2, \ldots, Z_n)$ 是独立同分布 $\sim$指数分布*(1)*，其次序统计量为 $Z_{1:n} \le Z_{2:n+1} \le \ldots \le Z_{n:n}$。那么，对于有限的固定整数$k \ge 1$，对所有 $n \ge 2$,

$$\mathbb{E}Z^k_{r:n} = k! \sum_{t_1=1}^{r} \frac{1}{n+1-t_1} \sum_{t_2=1}^{t_1} \frac{1}{n+1-t_2} \cdots \sum_{t_{k-1}=1}^{t_{k-2}} \frac{1}{n+1-t_{k-1}} \sum_{t_k=1}^{t_{k-1}} \frac{1}{n+1-t_k} \tag{55}$$

对所有 $r = 1, 2, \ldots, n$。式 (55) 右侧为 $O\left((\log r)^k\right)$ 当 $r \ge 2$时，且 $\mathbb{E}Z^k_{1:n} = O\left(n^{-k}\right)$。

证明。（55）右侧和 $\mathbb{E}Z^k_{1:n} = O\left(n^{-k}\right)$ 是通过求解 Balakrishnan 和 Gupta（1998年，定理1和定理2）中的递推关系得到的。（55）中的最内层求和满足

$\sum_{t_k=1}^{t_{k-1}} \frac{1}{n+1-t_k} \leq \sum_{j=1}^{r} \frac{1}{n+1-j}$，因为 $t_k \leq r$。由于（55）中的每个分数均为正，因此我们可以用以下表达式作为（55）右侧的上界

$$k! \left(\sum_{j=1}^{r} \frac{1}{n+1-j} \right) \left(\sum_{t_1=1}^{r} \frac{1}{n+1-t_1} \sum_{t_2=1}^{t_1} \frac{1}{n+1-t_2} \cdots \sum_{t_{k-1}=1}^{t_{k-2}} \frac{1}{n+1-t_{k-1}} \right).$$

对其余的 $k-1$ $1 \leq t_1 \leq t_1 \leq \ldots \leq t_{k-1} \leq$ 个和应用相同的界，并注意到 r，即可得到（）上的界 k！$\left(\sum_{j=1}^{r} \frac{1}{n+1-j} \right)^k$。55 证明完成，因为对于 $r \in [n]$，$\sum_{j=1}^{r} \frac{1}{n+1-j}$ 为 $O(\log r)$，当 $n \to \infty$ 时。■

引理7（条件鞅中心极限定理）。 设 $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ 为一列概率空间。设 $\xi_{n1}, \xi_{n2}, \ldots, \xi_{nn}$：$\Omega_n \to \mathbb{R}$ 是相对于子σ-代数 $\mathcal{F}_{n1} \subset \mathcal{F}_{n2} \subset \ldots \subset \mathcal{F}_{nn} \subset \mathcal{F}_n$的鞅差。设 $\mathcal{F}_{n0} \subset \mathcal{F}_{n1}$ 为一个子σ-代数。对于 $k = 1, 2, \ldots, n$，设 $\sigma_{nk}^2 := \mathbb{E}_n \left[\xi_{nk}^2 \mid \mathcal{F}_{n,k-1} \right]$。若存在有限常数 $\sigma > 0$ 使得

$$\mathbb{P}_n \left(\left| \sum_{k=1}^{n} \sigma_{nk}^2 - \sigma^2 \right| > \epsilon \,\middle|\, \mathcal{F}_{n0} \right) \xrightarrow{\mathbb{P}_n} 0 \quad \textit{for all constants } \epsilon > 0 \textit{ and} \tag{56}$$

$$\sum_{k=1}^{n} \mathbb{E}_n \left[\xi_{nk}^2 \mathbb{I}_{|\xi_{nk}| \geq \eta} \,\middle|\, \mathcal{F}_{n0} \right] \xrightarrow{\mathbb{P}_n} 0 \quad \textit{for all constants } \eta > 0, \tag{57}$$

则 $\mathbb{P}_n \left(\sigma^{-1} \sum_{k=1}^{n} \xi_{nk} \leq x \mid \mathcal{F}_{n0} \right) \xrightarrow{\mathbb{P}_n} \Phi(x)$ 对于所有 $x \in \mathbb{R}$ 均成立，其中 Φ 是标准正态分布函数。

证明。 通过对 Billingsley (1995年，定理 35.12）推导得出，整个过程中以 $\mathcal{F}_{n0}$ 为条件。■

ˆ ˆ命题 7 的证明。、A_n 的存在性。我们证明 A_n、A_n 以趋于 1 的概率有良好定义。对于 A_n，当且仅当以下条件成立时，这种情况才会发生

$$\mathbb{P} \left(\underline{p}_{\hat{\theta}} + a_n < \bar{p}_{\hat{\theta}} - a_n \right) = \mathbb{P} \left(2a_n < \bar{p}_{\hat{\theta}} - \underline{p}_{\hat{\theta}} \right) \to 1.$$

若 $\underline{p}_{\hat{\theta}} = \underline{p} + o_P(1)$ 且 $\bar{p}_{\hat{\theta}} = \bar{p} + o_P(1)$，则概率为 $\mathbb{P} \left(2a_n < \bar{p} - \underline{p} + o_P(1) \right) = \mathbb{P} \left(2 < a_n^{-1} (\bar{p} - \underline{p}) + o_P(a_n^{-1}) \right)$，当 $\bar{p} > \underline{p}$ 和 $a_n \downarrow 0$时，该概率趋于 1。为证明这些条件成立，定义 $T_n(x) := \theta^\intercal x$ 和 $T(x) := \theta_0^\intercal x$，根据定义有 $\bar{p} = \sup_{x \in \mathcal{X}} g(T(x)) = g(\sup_{x \in \mathcal{X}} T(x))$，且 $\bar{p}_{\hat{\theta}} = \sup_{x \in \mathcal{X}} g(T_n(x)) = g(\sup_{x \in \mathcal{X}} T_n(x))$，因为根据假设11， g 是递增的。根据假设11， g 是连续的，因此根据连续映射定理，只需证明 $\sup_{x \in \mathcal{X}} T_n(x) \xrightarrow{P} \sup_{x \in \mathcal{X}} T(x)$。由于根据假设9， $\mathcal{X}$ 有界且 $\hat{\theta} \xrightarrow{P} \theta_0$ˆ，因此 $\sup_{x \in \mathcal{X}} |T_n(x) - T(x)| \lesssim \|\theta - \theta_0\| = o_P(1)$，从而 $\bar{p}_{\hat{\theta}} = \bar{p} + o_P(1)$。类似论证可得 $\underline{p}_{\hat{\theta}} = \underline{p} + o_P(1)$。

ˆ ˆ 对于 A_n，所需结果可由上文关于 A_n 的结果推出，并且 $\min^{i\in[n]} g(\theta^\intercal X_i) = \underline{p}_{\hat\theta} + o_P(1)$，$\max^{i\in[n]} g(\hat\theta^\intercal X_i) = \bar p_{\hat\theta} + o_P(1)$。由于 $F^{-1}\hat{}_{\hat\theta}$，$F_{\hat\theta}(p)$ =ˆ ˆ$p_1 F_1$ 的逆函数 $\hat\theta(p) + (1-p_1)F_0$，即在假设 9 下 $(\pi(X,\theta) \mid \theta)$ 的分布，依据假设 6 是严格递增的，因此 $(\min^{i\in[n]} \pi(X_i, \hat\theta) \mid \hat\theta)$ 的分布与 $F^{-1}\hat{}_{\hat\theta}(U_{(1)})$ 相同，其中 $U_{(1)}$ 是 $((U_1, U_2, \ldots, U_n) \mid \hat\theta)$ 中的样本最小值，这些变量是独立同分布的，服从均匀分布$(0, 1)$。于是，$(\min^{i\in[n]} \pi(X_i, \hat\theta) - \underline{p}_{\hat\theta} \mid \hat\theta)$ 的分布与 $F^{-1}\hat{}_{\hat\theta}(U_{(1)}) - F^{-1}\hat{}_{\hat\theta}(F_{\hat\theta}(\underline{p}_{\hat\theta})) = F^{-1}\hat{}_{\hat\theta}(U_{(1)}) - F^{-1}\hat{}_{\hat\theta}(0)$ 在给定 $\hat\theta$. 时的分布相同。根据假设6，$F^{-1}\hat{}_{\hat\theta}$ 是利普希茨函数，其常数为 $\left\|(F^{-1}\hat{}_{\hat\theta})'\right\|_\infty$，其中 $(F^{-1}\hat{}_{\hat\theta})'(u) = {}^{1}_{f_{\hat\theta}(F^{-1}\hat{}_{\hat\theta}(u))}$ 在 $\inf^{p\in\underline{p}_{\hat\theta},\bar p_{\hat\theta}} f_{\hat\theta}(p) > 0$ 的条件下有限；对于 $\hat\theta \in \mathrm{Nb}(\theta_0, \epsilon)$，此结论由假设6得出。因此，$F^{-1}\hat{}_{\hat\theta}(U_{(1)}) - F^{-1}\hat{}_{\hat\theta}(0) \lesssim U_{(1)}$。这里，$\mathbb{E}U_{(1)}$ 根据命题 1 的证明趋于零。因此，根据假设11，同理可得 $\min^{i\in[n]} g(\theta^\intercal X_i) = \underline{p}_{\hat\theta} + o_P(1)$ 且 $\max^{i\in[n]} g(\theta^\intercal X_i) = \bar p_{\hat\theta} + o_P(1)$。下文证明 $V_{\hat\pi}$ 的方差分量估计量的一致性；类似论证可得 $V_{\mathrm{t},\hat\pi}$ 的结果。我们将在下文证明 $\hat N/n \xrightarrow{P} 1$ (因为 (62) 是 $o_P(1)$)。因此，下文证明估计量 V_τ，$V_{\tau\mathrm{t}}, V_{\sigma,\pi}, V_{\mathrm{t},\sigma,\pi}$，$\hat q_d$ˆ 以及 $\hat q_{\mathrm{t}}$ 的一致性，这些估计量以 n 而非 N 进行标准化。

的一致性 ˆ V_τ。根据定理 3 以及连续映射定理，$(\hat\tau_{\hat\pi})^2 \xrightarrow{P} \tau^2$。简言之，令 $\pi_i := g(\theta_0^\intercal X_i)$，$\hat\pi_i := g(\hat\theta^\intercal X_i)$，以及 $\hat{\mathbb{I}}^i := {}_{\mathbb{I}g(\hat\theta^\intercal X_i)\in\hat A_n}$。式（16）中的第一项为

$$\frac{1}{n}\sum_{i\in[n]}[\mu^1(\hat\theta,\hat\pi_i) - \mu^0(\hat\theta,\hat\pi_i)]^2\hat{}_{\mathbb{I}^i} \tag{58}$$

$$+\frac{1}{n}\sum_{i\in[n]}[\hat\mu^1(\hat\theta,\hat\pi_i) - \hat\mu^0(\hat\theta,\hat\pi_i) - (\mu^1(\hat\theta,\hat\pi_i) - \mu^0(\hat\theta,\hat\pi_i))]^2\hat{}_{\mathbb{I}^i} \tag{59}$$

$$+\frac{2}{n}\sum_{i\in[n]}[\hat\mu^1(\hat\theta,\hat\pi_i) - \hat\mu^0(\hat\theta,\hat\pi_i) - (\mu^1(\hat\theta,\hat\pi_i) - \mu^0(\hat\theta,\hat\pi_i))][\mu^1(\hat\theta,\hat\pi_i) - \mu^0(\hat\theta,\hat\pi_i)]\hat{}_{\mathbb{I}^i}. \tag{60}$$

这里，（58）收敛于 $\mathbb{E}(\mu^1(\theta_0,\pi_i) - \mu^0(\theta_0,\pi_i))^2$。为说明这一点，首先注意到在假设 9 下，

$$\left|\frac{1}{n}\sum_{i\in[n]}[\mu^1(\hat\theta,\hat\pi_i) - \mu^0(\hat\theta,\hat\pi_i)]^2\hat{}_{\mathbb{I}^i} - \frac{1}{n}\sum_{i\in[n]}[\mu^1(\theta_0,\pi_i) - \mu^0(\theta_0,\pi_i)]^2\hat{}_{\mathbb{I}^i}\right| \xrightarrow{P} 0,$$

这由 $[\mu^1(\hat\theta,\hat\pi_i) - \mu^0(\hat\theta,\hat\pi_i)]^2$ 关于 $\hat\theta$ 的中值展开得出，这与定理 3 证明中对式 (26) 的处理类似。其次，由于 $\mu^d(\theta,\cdot)$、$\theta \in \mathrm{Nb}(\theta_0,\epsilon)$ 受假设5限制，

$$\left|\frac{1}{n}\sum_{i\in[n]}[\mu^1(\theta_0,\pi_i) - \mu^0(\theta_0,\pi_i)]^2\hat{}_{\mathbb{I}^i} - \frac{1}{n}\sum_{i\in[n]}[\mu^1(\theta_0,\pi_i) - \mu^0(\theta_0,\pi_i)]^2\right| \tag{61}$$

其阶为

$$\frac{1}{n}\sum_{i\in[n]}(1-\hat{\mathbb{I}}^{i}) = \frac{1}{n}\sum_{i\in[n]}\left(\mathbb{I}^{g(\hat{\theta}^{\intercal}X_i)\notin\hat{A}_n} - \mathbb{E}\left[\mathbb{I}^{g(\hat{\theta}^{\intercal}X_i)\notin\hat{A}_n}\,\middle|\,\hat{\theta}\right]\right) + \frac{1}{n}\sum_{i\in[n]}\mathbb{E}\left[\mathbb{I}^{g(\hat{\theta}^{\intercal}X_i)\notin\hat{A}_n}\,\middle|\,\hat{\theta}\right]. \tag{62}$$

以趋近于1的概率。这里第一项的均值为零，方差有界于 $1/n$，因此它依均值收敛于零，进而依概率收敛于零。根据假设9，(62)的第二项为

$$\mathbb{P}\left(g(\hat{\theta}^{\intercal}X_i)\notin\hat{A}_n\,\middle|\,\hat{\theta}\right) = \mathbb{P}\left(g(\hat{\theta}^{\intercal}X_i) < \min_{i\in[n]} g(\hat{\theta}^{\intercal}X_i) + a_n\,\middle|\,\hat{\theta}\right) \tag{63}$$

$$+\,\mathbb{P}\left(\max_{i\in[n]} g(\hat{\theta}^{\intercal}X_i)) - a_n < g(\hat{\theta}^{\intercal}X_i)\,\middle|\,\hat{\theta}\right) \tag{64}$$

对于某个 $i\in[n]$。令 $\underline{G}_{\hat{\theta}} :=$ 最小值$_{i\in[n]}\, g(\theta^{\intercal}X_i)$。为界定(63)，由 Shanmugam和 Arnold (1988年)，

$$\mathbb{P}\left(g(\hat{\theta}^{\intercal}X_i)\le \underline{G}_{\hat{\theta}} + a_n\,\middle|\,\hat{\theta},\underline{G}_{\hat{\theta}}\right) = \frac{1}{n} + \frac{n-1}{n}\frac{F_{\hat{\theta}}(\underline{G}_{\hat{\theta}}+a_n) - F_{\hat{\theta}}(\underline{G}_{\hat{\theta}})}{1-F_{\hat{\theta}}(\underline{G}_{\hat{\theta}})} \tag{65}$$

在假设9和假设11下，其中 $F_{\hat{\theta}} := p_1F_{\hat{\theta},1} + (1-p_1)F_{\hat{\theta},0}$是给定 θ时 $g(\theta^{\intercal}X)$的分布函数。根据假设6， $F_{\hat{\theta}}$ 是连续的，并根据上述关于 A_n 的论证$\underline{G}_{\hat{\theta}} = \underline{p}_{\hat{\theta}} + o_P(1)$。于是 $a_n\downarrow 0$ 意味着(65)是 $o_P(1)$，进而意味着(63)也是 $o_P(1)$，因为它有界于1。通过类似的论证可知，项(64)为 $o_P(1)$；注意到 $\left\{\max_{i\in[n]} g(\hat{\theta}^{\intercal}X_i)) - a_n < g(\hat{\theta}^{\intercal}X_i)\right\} = \left\{\min_{i\in[n]} -g(\hat{\theta}^{\intercal}X_i)) + a_n > -g(\hat{\theta}^{\intercal}X_i)\right\}$。因此，(62)为 $o_P(1)$。由此可得，(58)依概率收敛于 $V_\tau+\tau^2$。将(59)写作

$$\frac{1}{n}\sum_{i\in[n]}[\hat{\mu}^1(\hat{\theta},\hat{\pi}_i)-\mu^1(\hat{\theta},\hat{\pi}_i)]^2\hat{\mathbb{I}}^{i} + \frac{2}{n}\sum_{i\in[n]}(\hat{\mu}^1(\hat{\theta},\hat{\pi}_i)-\mu^1(\hat{\theta},\hat{\pi}_i))(\hat{\mu}^0(\hat{\theta},\hat{\pi}_i)-\mu^0(\hat{\theta},\hat{\pi}_i))\hat{\mathbb{I}}^{i} + \frac{1}{n}\sum_{i\in[n]}[\hat{\mu}^0(\hat{\theta},\hat{\pi}_i)-\mu^0(\hat{\theta},\hat{\pi}_i)]^2\hat{\mathbb{I}}^{i}.$$

由于假设 10(i) 同时限制了 $\hat{\mu}^d$和 μ^d，上确界$_{p\in A_n}|\hat{\mu}^d(\hat{\theta},\ p) - \mu^d(\hat{\theta},\ p)| \xrightarrow{P} 0$意味着最后一个展示式中的三项均为 $o_P(1)$（注意， $A_n\subset A_n$）。这种收敛性可参照下文详述的其他估计量的证明思路来确立。类似的论证表明，（60）为 $o_P(1)$。由此得出 $\hat{V}_\tau \xrightarrow{P} V_\tau$。

的一致性：$V_{\sigma,\ \pi}$。首先， $0 < g(\theta^{\intercal}x) < 1$ 对所有 x 属于紧致集 $\mathcal{X}$ 且对所有 $\theta\in\mathrm{Nb}(\theta_0,\ \epsilon)$成立。然后，在假设11下，由中值展开和$\hat{\theta}\xrightarrow{P}\theta_0$可得

ˆ由假设9可得上确界$_{x\in\mathcal{X}}|1/g(\theta^\intercal x) - 1/g(\theta_0^\intercal x)| = o_P(1)$ 且类似地，对ˆ$1/(1-g($ $\theta^\intercal x))$。其次，假设8(i) 中的利普希茨条件意味着

$$\left|\frac{1}{n}\sum_{i\in[n]}\sigma_d^2(\hat{\theta},\hat{\pi}_i)\hat{\mathbb{I}}^i - \frac{1}{n}\sum_{i\in[n]}\sigma_d^2(\theta_0,\pi_i)\hat{\mathbb{I}}^i\right| \xrightarrow{P} 0,$$

并且，与上文（61）中的情况一样，我们还有

$$\left|\frac{1}{n}\sum_{i\in[n]}\sigma_d^2(\theta_0,\pi_i)\hat{\mathbb{I}}^i - \frac{1}{n}\sum_{i\in[n]}\sigma_d^2(\theta_0,\pi_i)\right| \xrightarrow{P} 0.$$

因此，若 $\sup_{p\in A_n}|\hat{}\sigma_d^2(\hat{}\theta,p) - \sigma_d^2(\hat{}\theta,p)| \xrightarrow{P} 0$，则由大数定律可得 $\hat{}V_{\sigma,\pi} \xrightarrow{P} V_{\sigma,\pi}$。只要 $\sup_{p\in A_n}|\hat{}\mu^d(\hat{}\theta,p) - \mu^d(\hat{}\theta,p)| \xrightarrow{P} 0$ 和 $\sup_{p\in A_n}|\hat{}\mu_2^d(\hat{}\theta,p) - \mu_2^d(\hat{}\theta,p)| \xrightarrow{P} 0$ 均成立，这一结论便成立，因为前者在假设10(i)下意味着 $\sup_{p\in A_n}|(\hat{}\mu^d(\hat{}\theta,p))^2 - (\mu^d(\hat{}\theta,p))^2| \xrightarrow{P} 0$。关于这类收敛性的证明（例如证明（103）），见下文。

一致性ˆq_d。根据 q_d 的定义，条件

$$\|\frac{1}{n}\sum_{i\in[n]}\Lambda^d(\hat{\theta},X_i) - \frac{1}{n}\sum_{i\in[n]}\Lambda^d(\hat{\theta},X_i)\hat{\mathbb{I}}^i\| \xrightarrow{P} 0 \tag{66}$$

$$\sup_{p\in A_n}\left|\widehat{\left(\frac{\partial\mu^d}{\partial p}\right)}(\hat{\theta},p) - \frac{\partial\mu^d}{\partial p}(\hat{\theta},p)\right| \xrightarrow{P} 0 \tag{67}$$

$$\sup_{p\in A_n}\left|\widehat{\left(\frac{\partial\mu^d}{\partial\theta_k}\right)}(\hat{\theta},p) - \frac{\partial\mu^d}{\partial\theta_k}(\hat{\theta},p)\right| \xrightarrow{P} 0 \quad \text{for all } k=1,2,\ldots,K, \tag{68}$$

结合假设 10 和11，这意味着 $\mathcal{X}$ 和 g'有界，从而保证ˆ$q_d \xrightarrow{P} q_d$。假设 7、11(i)以及假设10 意味着 Λ^d 是一致有界的。因此，满足式（66），因为在上文讨论ˆ $\mathbb{I}^i$ 的情形下ˆ V_τ 时所用的论证（例如式（61））同样适用。

接下来，我们证明（67）和（68）成立，同时详细证明上述假设的非参数估计量的一致收敛。为简化表述，我们只给出 $\mu(\theta,p) := \mathbb{E}[Y \mid g(\theta^\intercal X) = p] = \mathbb{E}[Y \mid \theta^\intercal X = g^{-1}(p)]$ 和 $\mu_2(\theta,p) := \mathbb{E}[Y^2 \mid g(\theta^\intercal X) = p]$ 的（导数的）证明。若同时对 $D=d$ 进行条件化，证明思路相同。为此，令 $h(\theta,p) := p_1h_1(\theta,p) + (1-p_1)h_0(\theta,p)$ 为 $\pi(X,\theta)$ 的密度，令 $q_\mu(\theta,p) := \mu(\theta,p)h(\theta,p)$、$q_{\mu_2}(\theta,p) := \mu_2(\theta,p)h(\theta,p)$ 及其相应估计量，分别通过在ˆ h_d、ˆq_μ、$_d$、ˆq_{μ_2}、$_d$ 的公式中对所有 $j\in[n]$ 令 $\mathbb{I}^{D_j=d} := 1$ 得到。简言之，我们还令 $h'(\theta,p) := (\partial/\partial p)h(\theta,p)$，$\widehat{\quad}\hat{}h'(\theta,p) := (\partial/\partial p)\hat{}h(\theta,p)$，$q'_\mu$ 和 ˆq'_μ 也同理。进一步，$\widehat{\left(\frac{\partial h}{\partial\theta_k}\right)}$ 和 $\widehat{\left(\frac{\partial q_\mu}{\partial\theta_k}\right)}$ 分别通过在 $\widehat{\left(\frac{\partial h_d}{\partial\theta_k}\right)}$ 和 $\widehat{\left(\frac{\partial q_{\mu,d}}{\partial\theta_k}\right)}$ 的公式中令 $\mathbb{I}^{D_j=d} := 1$ 得到。

简言之，我们记 $\dot{h}_k(\theta,p) := (\partial/\partial\theta_k)h(\theta,p)$、$\hat{\dot{h}}_k(\theta,p) := \left(\frac{\partial \hat{h}}{\partial \theta_k}\right)(\theta,p)$，对 $\hat{\dot{q}}_{\mu,k}(\theta,p)$ $\dot{q}_{\mu,k}(\theta,p)$ 也作类似记法。对于函数 $r\colon \Theta\times[0,1]\to\mathbb{R}$，我们令 $\|r\|_{A_n} := \sup_{p\in A_n}|r(\theta,p)|$。

条件 (67)。首先，我们证明分子的

$$\widehat{\left(\frac{\partial\mu}{\partial p}\right)}(\hat{\theta},p) := \frac{\hat{q}'_\mu(\hat{\theta},p)\hat{h}(\hat{\theta},p) - \hat{q}_\mu(\hat{\theta},p)\hat{h}'(\hat{\theta},p)}{(\hat{h}(\hat{\theta},p))^2} \tag{69}$$

收敛到.....的分子

$$\frac{\partial\mu}{\partial p}(\hat{\theta},p) = \frac{q'_\mu(\hat{\theta},p)h(\hat{\theta},p) - q_\mu(\hat{\theta},p)h'(\hat{\theta},p)}{h(\hat{\theta},p)^2}. \tag{70}$$

考虑

$$\begin{aligned}
\|\hat{q}'_\mu\hat{h} - \hat{q}_\mu\hat{h}' - (q'_\mu h - q_\mu h')\|_{A_n} &\le \|\hat{q}'_\mu\hat{h} - q'_\mu h\|_{A_n} + \|\hat{q}_\mu\hat{h}' - q_\mu h'\|_{A_n}\\
\|\hat{q}'_\mu\hat{h} - q'_\mu h\|_{A_n} &= \|(\hat{q}'_\mu - q'_\mu + q'_\mu)(\hat{h} - h + h) - q'_\mu h\|_{A_n}\\
&\le \|\hat{q}'_\mu - q'_\mu\|_{A_n}\|\hat{h} - h\|_{A_n} + \|\hat{q}'_\mu - q'_\mu\|_{A_n}\|h\|_{A_n}\\
&\quad + \|q'_\mu\|_{A_n}\|\hat{h} - h\|_{A_n}\\
\|\hat{q}_\mu\hat{h}' - q_\mu h'\|_{A_n} &\le \|\hat{h}' - h'\|_{A_n}\|\hat{q}_\mu - q_\mu\|_{A_n} + \|\hat{h}' - h'\|_{A_n}\|q_\mu\|_{A_n}\\
&\quad + \|h'\|_{A_n}\|\hat{q}_\mu - q_\mu\|_{A_n}.
\end{aligned}$$

根据假设 6(iv)，$\|h\|_{A_n}$ 以趋于一的概率有界，当 $\hat{\theta}\xrightarrow{P}\theta_0$ 并结合假设 7 和 11(ii)，同样也适用于 $\|h'\|_{A_n}$，$\|q_\mu\|_{A_n} = \|\mu h\|_{A_n}$和 $\|q'_\mu\|_{A_n} = \|\mu' h + \mu h'\|_{A_n}$。由此可知，如果所有

$$\|\hat{q}'_\mu - q'_\mu\|_{A_n}\xrightarrow{P}0, \tag{71}$$

$$\|\hat{h}' - h'\|_{A_n}\xrightarrow{P}0, \tag{72}$$

$$\|\hat{q}_\mu - q_\mu\|_{A_n}\xrightarrow{P}0, \tag{73}$$

$$\|\hat{h} - h\|_{A_n}\xrightarrow{P}0, \tag{74}$$

那么，(69) 的分子一致收敛于 (70) 的分子，关于 $p\in A_n$。(69) 的分母满足

$$\|(\hat{h})^2 - h^2\|_{A_n} = \|(\hat{h} - h)(\hat{h} + h + h - h)\|_{A_n} \le \|\hat{h} - h\|_{A_n}\left\{\|\hat{h} - h\|_{A_n} + 2\|h\|_{A_n}\right\}.$$

根据假设 6(iv)，$\|h\|_{A_n} < \infty$ 以概率趋于 1，当 $\hat{\theta}\xrightarrow{P}\theta_0$时，因此（74）意味着（69）的分母均匀收敛于（70）的分母，其中 $p\in A_n$。因此，（69）的分子和分母分别收敛于（70）的分子和分母。因为

ˆ ˆ$\inf_{p\in[\underline{p}_{\hat\theta},\bar p_{\hat\theta}]} h(\theta,\ p) > 0$ 对所有 $\theta \in \mathrm{Nb}(\theta_0,\ \epsilon)$，由此可得，（69）在 70）的意义下均匀收敛，其中 $p \in A_n$。下文将证明（71）–（74）成立。

条件 (67)，第 (71)。证明的关键在于证明

$$\mathbb{E}\mathbb{E}\left[\sup_{p\in A_n}\left|\hat q'_\mu(\hat\theta,p) - \mathbb{E}\left[\hat q'_\mu(\hat\theta,p)\,\middle|\,\hat\theta\right]\right|\,\middle|\,\hat\theta\right] \to 0 \quad \text{and} \tag{75}$$

$$\sup_{p\in A_n}\left|\mathbb{E}\left[\hat q'_\mu(\hat\theta,p)\,\middle|\,\hat\theta\right] - q'_\mu(\hat\theta,p)\right| \xrightarrow{P} 0. \tag{76}$$

我们沿用 Bierens（1994）的方法证明（75）。对于虚数单位 i，令 $\psi(t) := \int e^{itx}K(x)\mathrm{d}x$,$t \in \mathbb{R}$ 为 K 的特征函数；对于高斯核 K，它为 $\psi(t) = e^{-t^2/2}$，因此根据特征函数的反演公式，$K(x) = (2\pi)^{-1}\int e^{-itx}\psi(t)\mathrm{d}t$。于是 $K'(x) = (2\pi)^{-1}\int(-it)e^{-itx}\psi(t)\mathrm{d}t$，因此

$$\begin{aligned}\hat q'_\mu(\hat\theta,p) &= -\frac{1}{n\gamma_n^2}\sum_{j\in[n]} Y_j(2\pi)^{-1}\int(-it)e^{-it(g(\hat\theta^\intercal X_j)-p)/\gamma_n}\psi(t)\mathrm{d}t\\ &= (2\pi)^{-1}\int\left(\frac{1}{n}\sum_{j\in[n]} Y_j e^{-itg(\hat\theta^\intercal X_j)}\right)e^{itp}it\psi(\gamma_n t)\mathrm{d}t,\end{aligned}$$

这里我们使用了变量替换和 Fubini定理（根据假设 $|Y_j| < \bar y$，该积分几乎必然有界 10(i)，$|it| \le 1$，上确界$_{t\in\mathbb{R}}|\psi(t)| < \infty$，指数分布也有界）。由于 $|e^{itp}| \le 1$ 且 $|it| \le |t|$，我们有

$$\begin{aligned}&\mathbb{E}\left[\sup_{p\in A_n}\left|\hat q'_\mu(\hat\theta,p) - \mathbb{E}\left[\hat q'_\mu(\hat\theta,p)\,\middle|\,\hat\theta\right]\right|\,\middle|\,\hat\theta\right]\\ &\le (2\pi)^{-1}\int\mathbb{E}\left[\left|\left(\frac{1}{n}\sum_{j\in[n]} Y_j e^{-itg(\hat\theta^\intercal X_j)}\right) - \mathbb{E}\left[Y_j e^{-itg(\hat\theta^\intercal X_j)}\,\middle|\,\hat\theta\right]\right|\,\middle|\,\hat\theta\right]|t||\psi(\gamma_n t)|\mathrm{d}t.\end{aligned}$$

由于当 $a \in \mathbb{R}$ 时 $e^{-ia} = \cos(a) - i\sin(a)$，且对于任意平方可积随机变量 W，$\mathbb{E}|W| \le \sqrt{\mathbb{E}W^2}$，上一公式右侧的期望有界于

$$\mathbb{V}\left[\frac{1}{n}\sum_{j\in[n]} Y_j\cos(g(\hat\theta^\intercal X_j))\,\middle|\,\hat\theta\right]^{1/2} + \mathbb{V}\left[\frac{1}{n}\sum_{j\in[n]} Y_j\sin(g(\hat\theta^\intercal X_j))\,\middle|\,\hat\theta\right]^{1/2} \le 2\sqrt{\frac{\mathbb{E}Y_j^2}{n}},$$

ˆ这里我们利用了假设 9(ii)：给定 θ，求和中的各项独立同分布，因此协方差为零；并且 $\mathbb{V}\left[Y_j\cos(g(\hat\theta^\intercal X_j))\,\middle|\,\hat\theta\right] \le \mathbb{E}\left[Y_j^2\,\middle|\,\hat\theta\right] = \mathbb{E}Y_j^2$。因此，

$$\begin{aligned}\mathbb{E}\left[\sup_{p\in A_n}\left|\hat q'_\mu(\hat\theta,p) - \mathbb{E}\left[\hat q'_\mu(\hat\theta,p)\,\middle|\,\hat\theta\right]\right|\,\middle|\,\hat\theta\right] &\le \sqrt{\frac{\mathbb{E}Y_j^2}{\pi^2 n}}\int|t||\psi(\gamma_n t)|\mathrm{d}t\\ &\le \sqrt{\frac{\mathbb{E}Y_j^2}{\pi^2 n\gamma_n^4}}\int|t||\psi(t)|\mathrm{d}t.\end{aligned}$$

由于 $\mathbb{E}Y_j^2 < \infty$ 根据假设 10(i) 以及 $\int |t||\psi(t)|\mathrm{d}t = \int |t|e^{-t^2/2} < \infty$ 对于高斯K，右侧的阶为 $1/(\gamma_n^2\sqrt{n}) = (\kappa_0)^{-2}n^{2\beta-1/2} = o(1)$ 当 $\beta < 1/4$。

ˆ接下来，我们证明（76）。由于在给定 θ 的情况下，各加项具有相同的分布，根据假设9(ii)，由期望的塔式性质可得

$$\begin{aligned}
\mathbb{E}\left[\hat{q}'_\mu(\hat{\theta},p)\,\middle|\,\hat{\theta}\right] &= -\frac{1}{\gamma_n^2}\mathbb{E}\left[\mathbb{E}\left[Y \mid g(\hat{\theta}^\intercal X),\hat{\theta}\right]K'((g(\hat{\theta}X)-p)/\gamma_n)\,\middle|\,\hat{\theta}\right] \\
&= -\frac{1}{\gamma_n^2}\mathbb{E}\left[\mu(\hat{\theta},g(\hat{\theta}^\intercal X))K'((g(\hat{\theta}X)-p)/\gamma_n)\,\middle|\,\hat{\theta}\right] \\
&= -\frac{1}{\gamma_n^2}\int_{\underline{p}_{\hat{\theta}}}^{\bar{p}_{\hat{\theta}}}\mu(\hat{\theta},\tilde{p})K'((\tilde{p}-p)/\gamma_n)h(\hat{\theta},\tilde{p})\mathrm{d}\tilde{p} \\
&= -\frac{1}{\gamma_n^2}\int_{\underline{p}_{\hat{\theta}}}^{\bar{p}_{\hat{\theta}}}q_\mu(\hat{\theta},\tilde{p})K'((\tilde{p}-p)/\gamma_n)\mathrm{d}\tilde{p} \\
&= -\frac{1}{\gamma_n}\int_{(\underline{p}_{\hat{\theta}}-p)/\gamma_n}^{(\bar{p}_{\hat{\theta}}-p)/\gamma_n}q_\mu(\hat{\theta},\gamma_n v+p)K'(v)\mathrm{d}v
\end{aligned}$$

ˆ ˆˆ根据 $h(\theta,\cdot)$ 作为$(g(\theta^\intercal X) \mid \theta)$ 在假设9(ii) 和11(i)下的密度的定义，以及 $q_\mu = \mu h$。分部积分可得

$$\mathbb{E}\left[\hat{q}'_\mu(\hat{\theta},p)\,\middle|\,\hat{\theta}\right] = -\frac{1}{\gamma_n}\left\{q(\hat{\theta},\bar{p}_{\hat{\theta}})K((\bar{p}_{\hat{\theta}}-p)/\gamma_n) - q(\hat{\theta},\underline{p}_{\hat{\theta}})K((\underline{p}_{\hat{\theta}}-p)/\gamma_n)\right\} \tag{77}$$

$$+\int_{(\underline{p}_{\hat{\theta}}-p)/\gamma_n}^{(\bar{p}_{\hat{\theta}}-p)/\gamma_n}q'_\mu(\hat{\theta},\gamma_n v+p)K(v)\mathrm{d}v. \tag{78}$$

ˆ ˆ由于 $q_\mu(\theta,\cdot)$ 有界，且 $\theta \in \mathrm{Nb}(\theta_0,\epsilon)$以趋于一的概率成立（77）的阶为

$$\gamma_n^{-1}\left[\sup_{p\in A_n}K((\bar{p}_{\hat{\theta}}-p)/\gamma_n)+\sup_{p\in A_n}K((\underline{p}_{\hat{\theta}}-p)/\gamma_n)\right] = 2\gamma_n^{-1}K(a_n/\gamma_n).$$

由于 $\gamma_n^{-1}K(a_n/\gamma_n) \to 0$，(77) 是 $o_P(1)$。现在我们通过调整Schuster and Yakowitz (1979年，引理 1）的证明，来证明(78)均匀收敛到ˆ$q'_\mu(\theta,p)$ 。核 K是积分为 1 的密度，因此

$$q'_\mu(\hat{\theta},p) = q'_\mu(\hat{\theta},p)\gamma_n^{-1}\left\{\int_{-\infty}^{p-\bar{p}_{\hat{\theta}}}K(u/\gamma_n)\mathrm{d}u+\int_{p-\bar{p}_{\hat{\theta}}}^{p-\underline{p}_{\hat{\theta}}}K(u/\gamma_n)\mathrm{d}u+\int_{p-\underline{p}_{\hat{\theta}}}^{\infty}K(u/\gamma_n)\mathrm{d}u\right\}.$$

将此结果与对(78)作变量代换（令 $u := -\gamma_n v$ 注意到 K 关于零点对称），可得

$$\sup_{p\in A_n}\left|\int_{(\underline{p}_{\hat\theta}-p)/\gamma_n}^{(\bar p_{\hat\theta}-p)/\gamma_n} q'_\mu(\hat\theta,\gamma_n v+p)K(v)\mathrm{d}v - q'_\mu(\hat\theta,p)\right|$$

$$\le \sup_{p\in A_n}\left|\int_{-\infty}^{p-\bar p_{\hat\theta}} q'_\mu(\hat\theta,p)\gamma_n^{-1}K(u/\gamma_n)\mathrm{d}u\right| + \sup_{p\in A_n}\left|\int_{p-\underline{p}_{\hat\theta}}^{\infty} q'_\mu(\hat\theta,p)\gamma_n^{-1}K(u/\gamma_n)\mathrm{d}u\right| \tag{79}$$

$$+\sup_{p\in A_n}\left|\int_{p-\bar p_{\hat\theta}}^{p-\underline{p}_{\hat\theta}}[q'_\mu(\hat\theta,p-u)-q'_\mu(\hat\theta,p)]\gamma_n^{-1}K(u/\gamma_n)\mathrm{d}u\right|. \tag{80}$$

式（79）中的两项分别为 $o_P(1)$。第一项有如下界：

$$\sup_{p\in A_n}|q'_\mu(\hat\theta,p)|\sup_{p\in A_n}\int_{-\infty}^{(p-\bar p_{\hat\theta})/\gamma_n}K(v)\mathrm{d}v = \sup_{p\in A_n}|q'_\mu(\hat\theta,p)|\sup_{p\in A_n}\int_{-\infty}^{-a_n/\gamma_n}K(v)\mathrm{d}v,$$

$\hat{}$它趋于零：一方面，$\sup_{p\in A_n}|q'_\mu(\theta,p)|<\infty$ 以概率趋于一，当$\hat{}\theta \xrightarrow{P} \theta_0$ 时，这由假设 6(iv) 和 7得出；另一方面，$a_n/\gamma_n\to\infty$ 意味着高斯 K 的积分趋于零。式（79）中的第二项有如下界：

$$\sup_{p\in A_n}|q'_\mu(\hat\theta,p)|\sup_{p\in A_n}\int_{(p-\underline{p}_{\hat\theta})/\gamma_n}^{\infty}K(v)\mathrm{d}v = \sup_{p\in A_n}|q'_\mu(\hat\theta,p)|\sup_{p\in A_n}\int_{a_n/\gamma_n}^{\infty}K(v)\mathrm{d}v,$$

因此，采用相同论证，它也为 $o_P(1)$。为证明式（80）也趋于零，令 $\rho_n>0$ 为满足 $\rho_n/\gamma_n\to\infty$ 和 $\rho_n<a_n$的一个序列，即 $\gamma_n\ll\rho_n<a_n$ （当 $\gamma_n=\kappa_0 n^{-\beta}$ 且 $a_n=\kappa_1 n^{-\alpha}$时，$\beta>\alpha$，我们可取 $\rho_n=\kappa_2 n^{-(\beta+\alpha)/2}$，$0<\kappa_2<\kappa_1$）。那么，式（80）有如下界：

$$\sup_{p\in A_n}\left|\int_{[p-\bar p_{\hat\theta},p-\underline{p}_{\hat\theta}]\cap\{|u|\le\rho_n\}}[q'_\mu(\hat\theta,p-u)-q'_\mu(\hat\theta,p)]\gamma_n^{-1}K(u/\gamma_n)\mathrm{d}u\right| \tag{81}$$

$$+\sup_{p\in A_n}\left|\int_{[p-\bar p_{\hat\theta},p-\underline{p}_{\hat\theta}]\cap\{|u|>\rho_n\}}[q'_\mu(\hat\theta,p-u)-q'_\mu(\hat\theta,p)]u^{-1}u\gamma_n^{-1}K(u/\gamma_n)\mathrm{d}u\right|. \tag{82}$$

这里，(81) 有界于

$$\sup_{p\in A_n}\sup_{|u|\le\rho_n}|q'_\mu(\hat\theta,p-u)-q'_\mu(\hat\theta,p)|\sup_{p\in A_n}\int_{-\infty}^{\infty}\gamma_n^{-1}K(u/\gamma_n)\mathrm{d}u$$

$$\le \sup_{p\in A_n}\sup_{|u|\le\rho_n}|q'_\mu(\hat\theta,p-u)-q'_\mu(\hat\theta,p)|.$$

$\hat{}$ 根据假设 6(iv) 和 7，$q'_\mu(\theta,\cdot)$ 在紧集 $[\underline{p}_{\hat\theta},\bar p_{\hat\theta}]\supset A_n$ 上连续，因此一致连续。因此，对于足够小的 $\rho_n<a_n$，$|u|\le\rho_n$ 意味着 $\hat{}\hat{}p-u,p\in[\underline{p}_{\hat\theta},\bar p_{\hat\theta}]$ 对所有 $p\in A_n$ 都成立。因此，$\sup_{p\in A_n}\sup_{|u|\le\rho_n}|q'_\mu(\theta,p-u)-q'_\mu(\theta,p)|=o_P(1)$。

在 (82) 的积分中，$u \in [p-\bar{p}_{\hat{\theta}}, p-\underline{p}_{\hat{\theta}}]$，因此 $p-u \in [\underline{p}_{\hat{\theta}}, \bar{p}_{\hat{\theta}}]$，所以 (82) 有界ˆ，其上界为 $2\sup_{p\in[\underline{p}_{\hat{\theta}},\bar{p}_{\hat{\theta}}]}|q'_\mu(\theta,p)|$ 倍

$$\sup_{p\in A_n}\left|\int_{[p-\bar{p}_{\hat{\theta}},p-\underline{p}_{\hat{\theta}}]\cap\{|u|>\rho_n\}}\frac{K(u/\gamma_n)}{\gamma_n}\mathrm{d}u\right| \le \sup_{p\in A_n}\int_{\{|u|>\rho_n\}}\frac{K(u/\gamma_n)}{\gamma_n}\mathrm{d}u$$
$$\le \int_{\{|v|>\rho_n/\gamma_n\}} K(v)\mathrm{d}v,$$

这里我们用到了 $K>0$。由于 $\gamma_n \ll \rho_n$，且 $K(v)\downarrow 0$ 由于 $|v|\to\infty$，右侧积分趋于ˆ零。由假设 6(iv) 和 7 控制 $\sup_{p\in[\underline{p}_{\hat{\theta}},\bar{p}_{\hat{\theta}}]}|q'_\mu(\theta,\ p)|$，(82) 是 $o_P(1)$。因此，(76) 成立。由此可得 (71) 成立。

条件 (67)，第 (72)部分。按照 (71) 的思路，令 $Y_j := 1$ 对所有 $j\in[n]$，代入 (71) 的公式即可直接得到。

条件 (67)，第 (73)部分。与上文的 $\hat{}q'_\mu$ 类似，我们可以写出

$$\hat{q}_\mu(\hat{\theta},p) = \frac{1}{n\gamma_n}\sum_{j\in[n]} Y_j(2\pi)^{-1}\int e^{-it(g(\hat{\theta}^\intercal X_j)-p)/\gamma_n}\psi(t)\mathrm{d}t$$
$$= (2\pi)^{-1}\int\left(\frac{1}{n}\sum_{j\in[n]} Y_j e^{-itg(\hat{\theta}^\intercal X_j)}\right)e^{itp}\psi(\gamma_n t)\mathrm{d}t,$$

然后

$$\mathbb{E}\left[\sup_{p\in A_n}\left|\hat{q}_\mu(\hat{\theta},p)-\mathbb{E}\left[\hat{q}_\mu(\hat{\theta},p)\,\middle|\,\hat{\theta}\right]\right|\,\middle|\,\hat{\theta}\right] \le \sqrt{\frac{\mathbb{E}Y_j^2}{\pi^2 n}}\int|\psi(\gamma_n t)|\mathrm{d}t \le \sqrt{\frac{\mathbb{E}Y_j^2}{\pi^2 n\gamma_n^2}}\int|\psi(t)|\mathrm{d}t.$$

假设10(i) 和 $\int|\psi(t)|\mathrm{d}t = \int e^{-t^2/2} = 2\pi$ 对于高斯核 K 意味着右侧的阶为 $1/(\gamma_n\sqrt{n}) = (\kappa_0)^{-1}n^{\beta-1/2} = o(1)$ 对于 $\beta<1/4$。接下来，对于ˆq'_μ,

$$\mathbb{E}\left[\hat{q}_\mu(\hat{\theta},p)\,\middle|\,\hat{\theta}\right] = \frac{1}{\gamma_n}\int_{\underline{p}_{\hat{\theta}}}^{\bar{p}_{\hat{\theta}}}\mu(\hat{\theta},\tilde{p})K((\tilde{p}-p)/\gamma_n)h(\hat{\theta},\tilde{p})\mathrm{d}\tilde{p}$$
$$= \frac{1}{\gamma_n}\int_{\underline{p}_{\hat{\theta}}}^{\bar{p}_{\hat{\theta}}} q_\mu(\hat{\theta},\tilde{p})K((\tilde{p}-p)/\gamma_n)\mathrm{d}\tilde{p} = \int_{(\underline{p}_{\hat{\theta}}-p)/\gamma_n}^{(\bar{p}_{\hat{\theta}}-p)/\gamma_n} q_\mu(\hat{\theta},\gamma_n v+p)K(v)\mathrm{d}v.$$

ˆ根据与（ $q_\mu(\theta,p)$ 在 $p\in A_n$ 中的论证相同，78）收敛到ˆ $\hat{}q'_\mu(\theta,p)$，给定假设 6 和 7 确保 $q_\mu(\theta,\cdot)$ 以趋于 $\hat{}\theta \xrightarrow{P} \theta_0$的概率有界且连续。

条件(67)，第(74)部分。由（73）令 $Y_j := 1$ 对所有 $j\in[n]$。

条件 (68)。为证明

$$\widehat{\left(\frac{\partial\mu}{\partial\theta_k}\right)}(\hat{\theta},p) := \frac{\hat{\dot{q}}_{\mu,k}(\hat{\theta},p)\hat{h}(\hat{\theta},p)-\hat{q}_\mu(\hat{\theta},p)\hat{\dot{h}}_k(\hat{\theta},p)}{(\hat{h}(\hat{\theta},p))^2} \tag{83}$$

to

$$\frac{\partial \mu}{\partial \theta_k}(\hat{\theta}, p) := \frac{\dot{q}_{\mu,k}(\hat{\theta}, p)h(\hat{\theta}, p) - \hat{q}_\mu(\hat{\theta}, p)\dot{h}_k(\hat{\theta}, p)}{(\hat{h}(\hat{\theta}, p))^2} \tag{84}$$

在 $p \in A_n$中的一致收敛，我们可以沿用导出条件 (71)–(74) 的相同步骤，因为假设 6(iv) 和7 确保以趋近于一的概率， $\|h\|_{A_n}$， $\|h_k\|_{A_n}$， $\|q_\mu\|_{A_n}$， $\|\dot{q}_{\mu,\ k}\|_{A_n}$ 均有界。由于上文已证明 (73) 和 (74)，因此只需证明以下两者

$$\|\hat{\dot{q}}_{\mu,k} - \dot{q}_{\mu,k}\|_{A_n} \xrightarrow{P} 0 \quad \text{and} \tag{85}$$

$$\|\hat{\dot{h}}_k - \dot{h}_k\|_{A_n} \xrightarrow{P} 0. \tag{86}$$

条件(68)，第 (85) 部分。我们接下来证明

$$\mathbb{E}\mathbb{E}\left[\sup_{p\in A_n}\left|\hat{\dot{q}}_{\mu,k}(\hat{\theta}, p) - \mathbb{E}\left[\hat{\dot{q}}_{\mu,k}(\hat{\theta}, p) \,\middle|\, \hat{\theta}\right]\right| \,\middle|\, \hat{\theta}\right] \to 0 \quad \text{and} \tag{87}$$

$$\sup_{p\in A_n}\left|\mathbb{E}\left[\hat{\dot{q}}_{\mu,k}(\hat{\theta}, p) \,\middle|\, \hat{\theta}\right] - \dot{q}_{\mu,k}(\hat{\theta}, p)\right| \xrightarrow{P} 0. \tag{88}$$

与上面的 $\hat{}q'_\mu$ 一样，

$$\hat{\dot{q}}_{\mu,k}(\hat{\theta}, p) = \frac{(g^{-1})'(p)}{n\gamma_n^2}\sum_{j\in[n]} Y_j X_{j,k}(2\pi)^{-1}\int (-it)e^{-it(\hat{\theta}^\intercal X_j - g^{-1}(p))/\gamma_n}\psi(t)\mathrm{d}t \tag{89}$$

$$= (g^{-1})'(p)(2\pi)^{-1}\int\left(\frac{1}{n}\sum_{j\in[n]} Y_j X_{j,k}e^{-it\hat{\theta}^\intercal X_j}\right)e^{itg^{-1}(p)}(-it)\psi(\gamma_n t)\mathrm{d}t. \tag{90}$$

由于 $|-it| \le |t|$ 和 $\mathcal{X}$ 由假设 10有界，欧拉公式和假设9(ii) 意味着，75的界在此处也成立（相差一个常数因子），从而界定$\mathcal{X}$，乘以 $\sup_{p\in A_n}|(g^{-1})'(p)|$。由于 $\sup_{p\in A_n}|(g^{-1})'(p)| \le 1/\|g'\|_\infty < \infty$ 根据假设 11，式（87）成立。

^ ^条件(68)，第 (85)、(88)。 我们首先推导 $\dot{q}_{\mu,\ k}(\theta, p) = ((\partial/\partial\theta_k)(\mu h))(\theta, p)$。为简单起见，假设我们只有两个协变量，且二者均为连续变量，并且我们关注 θ ($k := 1$) 关于第一个坐标的导数。将下文的论证推广到一般情形是直接的。根据期望的塔式性质，

$$\mu(\theta, p) = \mathbb{E}\left[Y \mid \theta^\intercal X = g^{-1}(p)\right] = \mathbb{E}\left[m(X) \mid \theta^\intercal X = g^{-1}(p)\right].$$

鉴于命题 30 的6，我们可以写成

$$\mathbb{E}\left[m(X) \mid \theta^\intercal X = t\right] = \frac{1}{v(\theta, t)}\int_{\mathcal{X}_1} m\left(x_1, \frac{t-\theta_1 x_1}{\theta_2}\right)\Psi\left(x_1, \frac{t-\theta_1 x_1}{\theta_2}\right)\mathrm{d}x_1 \tag{91}$$

$$v(\theta, t) := \int_{\mathcal{X}_1}\Psi\left(x_1, \frac{t-\theta_1 x_1}{\theta_2}\right)\mathrm{d}x_1, \tag{92}$$

其中 $v(\theta,\ \cdot)$ 是 $\theta^\intercal X$ 的密度，满足 $v(\theta,\ \cdot)\ >0$（根据假设10）。因此，$h(\theta,\ \cdot)$ 作为 $g(\theta^\intercal X)$ 的密度，等于

$$h(\theta,p)=(g^{-1})'(p)\int_{\mathcal{X}_1}\Psi\left(x_1,\frac{g^{-1}(p)-\theta_1x_1}{\theta_2}\right)\mathrm{d}x_1. \tag{93}$$

^根据假设 9(ii), (91) 和 (93) 在将 θ 替换为 θ后仍然成立。因此，对于连续可微的 Ψ 和 m (假设 10 和 11)，当 $k=1$ 时，

$$\begin{aligned}
q_\mu(\theta,p)&=\mu(\theta,p)h(\theta,p)\\
&=(g^{-1})'(p)\int_{\mathcal{X}_1}m\left(x_1,\frac{g^{-1}(p)-\theta_1x_1}{\theta_2}\right)\Psi\left(x_1,\frac{g^{-1}(p)-\theta_1x_1}{\theta_2}\right)\mathrm{d}x_1\\
\dot{q}_{\mu,k}(\hat{\theta},p)&=(g^{-1})'(p)\int_{\mathcal{X}_1}\frac{\mathrm{d}}{\mathrm{d}\theta_1}\left[m\left(x_1,\frac{g^{-1}(p)-\hat{\theta}_1x_1}{\hat{\theta}_2}\right)\Psi\left(x_1,\frac{g^{-1}(p)-\hat{\theta}_1x_1}{\hat{\theta}_2}\right)\right]\mathrm{d}x_1\\
&=\frac{\mathrm{d}}{\mathrm{d}p}\int_{\mathcal{X}_1}-x_1m\left(x_1,\frac{g^{-1}(p)-\hat{\theta}_1x_1}{\hat{\theta}_2}\right)\Psi\left(x_1,\frac{g^{-1}(p)-\hat{\theta}_1x_1}{\hat{\theta}_2}\right)\mathrm{d}x_1,
\end{aligned}\tag{94}$$

其中，最后一步用到了 $(g^{-1})'\ >\ 0$。接下来，我们证明所需的收敛性。令 $[\underline{t}_{\hat{\theta}},\bar{t}_{\hat{\theta}}]$: $=[g^{-1}(\underline{p}_{\hat{\theta}})\ g^{-1}(\bar{p}_{\hat{\theta}})]$。根据塔式性质，对于 $k=1$，我们有

$$\begin{aligned}
\mathbb{E}\left[\hat{\dot{q}}_{\mu,k}(\hat{\theta},p)\,\middle|\,\hat{\theta}\right]&=\frac{g^{-1}(p)}{\gamma_n^2}\mathbb{E}\left[Y_jX_{j,k}K'((\hat{\theta}^\intercal X_j-g^{-1}(p)/\gamma_n)\,\middle|\,\hat{\theta}\right]\\
&=\frac{(g^{-1})'(p)}{\gamma_n^2}\mathbb{E}\left[\mathbb{E}\left[m(X_j)X_{j,1}\mid\hat{\theta}^\intercal X_j,\hat{\theta}\right]K'((\hat{\theta}^\intercal X_j-g^{-1}(p))/\gamma_n)\,\middle|\,\hat{\theta}\right]\\
&=\frac{(g^{-1})'(p)}{\gamma_n^2}\int_{\underline{t}_{\hat{\theta}}}^{\bar{t}_{\hat{\theta}}}K'\left(\frac{t-g^{-1}(p)}{\gamma_n}\right)\int_{\mathcal{X}_1}x_1m\left(x_1,\frac{t-\hat{\theta}_1x_1}{\hat{\theta}_2}\right)\Psi\left(x_1,\frac{t-\hat{\theta}_1x_1}{\hat{\theta}_2}\right)\mathrm{d}x_1\,\mathrm{d}t\\
&=(g^{-1})'(p)\\
&\times\int_{\mathcal{X}_1}x_1\left[\frac{1}{\gamma_n^2}\int_{\underline{t}_{\hat{\theta}}}^{\bar{t}_{\hat{\theta}}}K'\left(\frac{t-g^{-1}(p)}{\gamma_n}\right)m\left(x_1,\frac{t-\hat{\theta}_1x_1}{\hat{\theta}_2}\right)\Psi\left(x_1,\frac{t-\hat{\theta}_1x_1}{\hat{\theta}_2}\right)\mathrm{d}t\right]\mathrm{d}x_1,
\end{aligned}$$

其中我们利用了命题 6：(91) 对 m 成立，对 X 的任意一般函数也成立。令 $\lambda_n(x_1,t)$ $:=m\left(x_1,{}^{t-\hat{}}\frac{\theta_1x_1}{\theta_2}\right)\Psi\left(x_1,{}^{t-\hat{}\hat{}}\frac{\theta_1x_1}{\theta_2}\right)\hat{}$。我~~们证~~明，方括~~号中~~的项收敛于

$$-\frac{\partial}{\partial t}\lambda_n(x_1,g^{-1}(p))=-\left(\frac{\mathrm{d}}{\mathrm{d}p}\lambda_n(x_1,g^{-1}(p))\right)/[(g^{-1})'(p)]$$

对 $x_1 \in \mathcal{X}_1$ 和 $p \in A_n$一致成立，结合式 (94)，这便完成了式 (88) 的证明。令 $\lambda_n'(x_1,t) := \frac{\partial}{\partial t}\lambda_n(x_1,t)$。分部积分得

$$\begin{aligned}\frac{1}{\gamma_n^2}\int_{\underline{t}_{\hat{\theta}}}^{\bar{t}_{\hat{\theta}}} K'\left(\frac{t-g^{-1}(p)}{\gamma_n}\right)\lambda_n(x_1,t)\mathrm{d}t &= \frac{1}{\gamma_n}\int_{(\underline{t}_{\hat{\theta}}-g^{-1}(p))/\gamma_n}^{(\bar{t}_{\hat{\theta}}-g^{-1}(p))/\gamma_n} K'(v)\lambda_n(x_1,\gamma_n v+g^{-1}(p))\mathrm{d}t \\ &= \frac{1}{\gamma_n}\left\{\lambda_n(x_1,\bar{t}_{\hat{\theta}})K((\bar{t}_{\hat{\theta}}-g^{-1}(p))/\gamma_n) - \lambda_n(x_1,\underline{t}_{\hat{\theta}})K((\underline{t}_{\hat{\theta}}-g^{-1}(p))/\gamma_n)\right\} \quad (95)\\ &\quad - \int_{(\underline{t}_{\hat{\theta}}-g^{-1}(p))/\gamma_n}^{(\bar{t}_{\hat{\theta}}-g^{-1}(p))/\gamma_n} \lambda_n'(x_1,\gamma_n v+g^{-1}(p))K(v)\mathrm{d}v. \quad (96)\end{aligned}$$

此时 λ_n 根据假设，10，以趋于 1 的概率有界。回顾一下，$\underline{t}_{\hat{\theta}} = g^{-1}(\underline{p}_{\hat{\theta}})$ $\bar{t}_{\hat{\theta}} = g^{-1}(\bar{p}_{\hat{\theta}})$。由于 g^{-1} 是递增的，并且 K 在零处达到最大值且满足$\lim_{|u|\to\infty} K(u) \to 0$，(95) 的阶为

$$\begin{aligned}&\gamma_n^{-1}\left[\sup_{p\in A_n} K\left(\frac{g^{-1}(\bar{p}_{\hat{\theta}})-g^{-1}(p)}{\gamma_n}\right) + \sup_{p\in A_n} K\left(\frac{g^{-1}(\underline{p}_{\hat{\theta}})-g^{-1}(p)}{\gamma_n}\right)\right] \\ &= \gamma_n^{-1}\left[K\left(\frac{g^{-1}(\bar{p}_{\hat{\theta}})-g^{-1}(\bar{p}_{\hat{\theta}}-a_n)}{\gamma_n}\right) + K\left(\frac{g^{-1}(\underline{p}_{\hat{\theta}})-g^{-1}(\underline{p}_{\hat{\theta}}+a_n)}{\gamma_n}\right)\right]. \quad (97)\end{aligned}$$

通过中值展开，(97) 中的第一项的阶为

$$\gamma_n^{-1}K\left(\left(\inf_{p\in[\underline{p}_{\hat{\theta}},\bar{p}_{\hat{\theta}}]}(g^{-1})'(p)\right)a_n/\gamma_n\right).$$

当 无穷$^{p\in[\underline{p}_{\hat{\theta}},\bar{p}_{\hat{\theta}}]}(g^{-1})'(p)$ > 0时，这就是 $o_P(1)$ 对于 $a_n/\gamma_n \to \infty$ 和高斯 K。第二项也同样适用于（97），因此（95）是 $o_P(1)$。最后，我们证明（96）收敛到$-\lambda_n'(x_1, g^{-1}(p))$。同样，我们可以写成

$$\begin{aligned}\lambda_n'(x_1,g^{-1}(p)) &= \frac{\lambda_n'(x_1,g^{-1}(p))}{\gamma_n} \\ \times\left\{\int_{-\infty}^{g^{-1}(p)-\bar{t}_{\hat{\theta}}} K(u/\gamma_n)\mathrm{d}u + \int_{g^{-1}(p)-\bar{t}_{\hat{\theta}}}^{p-\underline{t}_{\hat{\theta}}} K(u/\gamma_n)\mathrm{d}u + \int_{g^{-1}(p)-\underline{t}_{\hat{\theta}}}^{\infty} K(u/\gamma_n)\mathrm{d}u\right\}.\end{aligned}$$

在（96）中作变量替换（$u:=-\gamma_n v$），并且 K 关于零对称，因此得到

$$\sup_{x_1\in\mathcal{X}_1,p\in A_n}\left|\int_{(\underline{t}_{\hat\theta}-g^{-1}(p))/\gamma_n}^{(\bar t_{\hat\theta}-g^{-1}(p))/\gamma_n}\lambda_n'(x_1,\gamma_n v+g^{-1}(p))K(v)\mathrm{d}v-\lambda_n'(x_1,g^{-1}(p))\right|$$

$$\le \sup_{x_1\in\mathcal{X}_1,p\in A_n}\left|\int_{-\infty}^{g^{-1}(p)-\bar t_{\hat\theta}}\lambda_n'(x_1,g^{-1}(p))\gamma_n^{-1}K(u/\gamma_n)\mathrm{d}u\right| \tag{98}$$

$$+\sup_{x_1\in\mathcal{X}_1,p\in A_n}\left|\int_{g^{-1}(p)-\underline{t}_{\hat\theta}}^{\infty}\lambda_n'(x_1,g^{-1}(p))\gamma_n^{-1}K(u/\gamma_n)\mathrm{d}u\right| \tag{99}$$

$$+\sup_{x_1\in\mathcal{X}_1,p\in A_n}\left|\int_{g^{-1}(p)-\bar t_{\hat\theta}}^{g^{-1}(p)-\underline{t}_{\hat\theta}}[\lambda_n'(x_1,g^{-1}(p)-u)-\lambda_n'(x_1,g^{-1}(p))]\gamma_n^{-1}K(u/\gamma_n)\mathrm{d}u\right|. \tag{100}$$

由于根据假设 10，$\mathcal{X}_1$ 有界，因此 $\sup_{x_1\in\mathcal{X}_1,\ p\in A_n}|\lambda_n'(x_1,g^{-1}(p))|$ 以趋于 1 的概率有界。因此，由于 g^{-1} 递增，$K\ge 0$ 和 $\bar t_{\hat\theta}=g^{-1}(\bar p_{\hat\theta})$，（98）的阶为 $\int_{-\infty}^{(g^{-1}(\bar p_{\hat\theta}-a_n)-g^{-1}(p))/\gamma_n}K(v)\mathrm{d}v$。这里，

$$(g^{-1}(\bar p_{\hat\theta}-a_n)-g^{-1}(p))/\gamma_n\le(\inf_{p\in[0,1]}(g^{-1})'(p))(-a_n/\gamma_n)\le(1/\|g'\|_\infty)(-a_n/\gamma_n)$$

通过中值展开。由于下确界为正，（98）对于 $a_n/\gamma_n\to\infty$ 为 $o_P(1)$，类似的论证也适用于（99）。项（100）可以用类似于（80）的方法处理。首先注意，对 $p\in A_n$ 取上确界等同于对 $t\in[g^{-1}(\underline{p}_{\hat\theta}+a_n)\ g^{-1}(\bar p_{\hat\theta}-a_n)]$ 取上确界：因为 g^{-1} 递增，所以 T_n。取一个 $\gamma_n\ll\rho_n<(1/\|g'\|_\infty)a_n$（例如 $\rho_n:=(1/\|g'\|_\infty)\kappa_2 n^{-(\alpha+\beta)/2}$，其中某个 $0<\kappa_2<\kappa_1$）。那么（100）有界于

$$\sup_{x_1\in\mathcal{X}_1,t\in T_n}\left|\int_{[t-\bar t_{\hat\theta},t-\underline{t}_{\hat\theta}]\cap\{|u|\le\rho_n\}}[\lambda_n'(x_1,t-u)-\lambda_n'(x_1,t)]\gamma_n^{-1}K(u/\gamma_n)\mathrm{d}u\right| \tag{101}$$

$$\sup_{x_1\in\mathcal{X}_1,t\in T_n}\left|\int_{[t-\bar t_{\hat\theta},t-\underline{t}_{\hat\theta}]\cap\{|u|>\rho_n\}}[\lambda_n'(x_1,t-u)-\lambda_n'(x_1,t))]\gamma_n^{-1}K(u/\gamma_n)\mathrm{d}u\right|. \tag{102}$$

在此，（101）的界为

$$\sup_{x_1\in\mathcal{X}_1,t\in T_n}\sup_{|u|\le\rho_n}|\lambda_n'(x_1,t-u)-\lambda_n'(x_1,t)|\sup_{x_1\in\mathcal{X}_1,t\in T_n}\int_{-\infty}^{\infty}\gamma_n^{-1}K(u/\gamma_n)\mathrm{d}u$$

$$\le\sup_{x_1\in\mathcal{X}_1,t\in T_n}\sup_{|u|\le\rho_n}|\lambda_n'(x_1,t-u)-\lambda_n'(x_1,t)|.$$

根据假设10和11，$\lambda_n'(x_1,\cdot)$关于 x_1 均匀连续，因此在紧集 $[g^{-1}(\underline{p}_{\hat\theta})\ g^{-1}(\bar p_{\hat\theta})]\supset[g^{-1}(\underline{p}_{\hat\theta}+a_n)\ g^{-1}(\bar p_{\hat\theta}-a_n)]$上一致连续。因此，对于足够小的 ρ_n，$|u|\le\rho_n$ 意味着 $t-u$，$t\in[g^{-1}(\underline{p}_{\hat\theta})\ g^{-1}(\bar p_{\hat\theta})]$ 对所有$t\in T_n$都成立。（注意，ρ_n 足够小当且仅当 $|u|\le\rho_n$ 意味着

$$|u|\le\min\left\{g^{-1}(\underline{p}_{\hat\theta}+a_n)-g^{-1}(\underline{p}_{\hat\theta}),g^{-1}(\bar p_{\hat\theta})-g^{-1}(\bar p_{\hat\theta}-a_n)\right\}.$$

通过中值展开，右侧小于或等于 $\inf_{p\in[0,1]}(g^{-1})'(p)a_n \le (1/\|g'\|_\infty)a_n$。但这样一来，$\rho_n < (1/\|g'\|_\infty)a_n$ 就足够小了。) 因此，

$$\sup_{x\in\mathcal{X}_1,t\in T_n}\sup_{|u|\le\rho_n}|\lambda_n'(x_1,t-u)-\lambda_n'(x_1,t)| = o_P(1),$$

因此，(101) 为 $o_P(1)$。根据假设10和11，当 $\hat{\theta}_2 \xrightarrow{P} \theta_{0,2} \neq 0$ 时，$2\sup_{x\in\mathcal{X}_1,t\in[\underline{t}_{\hat{\theta}},\bar{t}_{\hat{\theta}}]}|\lambda_n(x_1,t)|$ 以趋于 1 的概率有界，因此类似于 (82)，可以证明 (102) 为 $o_P(1)$。于是，(100) 为 $o_P(1)$，这表明 (96) 按所需方式收敛。因此，(88) 对于 $k=1$ 成立。对于 $k=2$ 的情形，通过写成以下形式可得

$$\mathbb{E}[m(X)\mid\theta^\intercal X=t] = (1/v(\theta,t))\int_{\mathcal{X}_2} m\left(\frac{t-\theta_2x_2}{\theta_1},x_2\right)\Psi\left(\frac{t-\theta_2x_2}{\theta_1},x_2\right)\mathrm{d}x_2$$
$$v(\theta,t) := \int_{\mathcal{X}_2}\Psi\left(\frac{t-\theta_2x_2}{\theta_1},x_2\right)\mathrm{d}x_2$$

而不是 (91) 和 (92)。根据假设 10，这些同样表现良好。由此可得，所需的收敛性 $\hat{}\dot{q}_{\mu,k}$（条件 (68)，部分 (85)）成立。

条件 (68)，部分 (86)。按照与 (85) 相同的思路，令 $Y_j := 1$ 对所有 $j\in[n]$。由此可得条件 (68) 成立。

剩余的一致收敛结果。证明

$$\sup_{p\in A_n}|\hat{\mu}(\hat{\theta},p)-\mu(\hat{\theta},p)| \xrightarrow{P} 0, \tag{103}$$
$$\sup_{p\in A_n}|\hat{\mu}_2(\hat{\theta},p)-\mu_2(\hat{\theta},p)| \xrightarrow{P} 0 \tag{104}$$

$\hat{}$完成了命题 7 的证明。由于 $h(\theta,\cdot)$ 根据假设 6，以趋于 1 的概率远离零，因此（103）可由（73）和（74）推出。同理，（104）可由 $\|\hat{}q_{\mu_2}-q_{\mu_2}\|_{A_n}\xrightarrow{P}0$ 推出，如（73）所示。■

参考文献

Alberto Abadie 和 Guido W. Imbens。《平均处理效应匹配估计量的大样本性质》。《计量经济学》，74(1):235–267，2006年1月。ISSN 0012-9682，1468-0262。doi: 10.1111/j.1468-0262.2006.00655.x。URLhttp://doi.wiley.com/10.1111/j.1468-0262.2006.00655.x。

Alberto Abadie 和 Guido W. Imbens。基于估计倾向得分的匹配。《计量经济学》，84(2):781–807，2016。ISSN 0012-9682。doi: 10.3982/ECTA11293。URLhttps://www.econometricsociety.org/doi/10.3982/ECTA11293。

Kenneth S. Alexander。按集合索引的加权经验过程的增长率与样本模数。《概率论及相关领域》，75:379–423，1987年。

N. Balakrishnan 和 Shanti S. Gupta。2 来自指数分布和右截断指数分布的次序统计量的高阶矩及其在寿命试验问题中的应用。载于次序统计量：应用，第17卷，《统计学手册》，第25–59页。爱思唯尔，1998年。doi: https://doi.org/10.1016/S0169-7161(98)17004-9。URLhttps://www.sciencedirect.com/science/article/pii/S0169716198170049。ISSN: 0169-7161。

Richard Kwasi Bannor、Gupta Amarnath Krishna Kumar、Helena Oppong-Kyeremeh 和 CamillusAbaw- iera Wongnaa。《现代水稻品种在印度东部的贫困状况中的采用与影响》。《*Rice Science*》，27(1):56–66，2020年1月。ISSN 16726308。doi: 10.1016/j.rsci.2019.12.006。URLhttps://linkinghub.elsevier.com/retrieve/pii/S1672630819301106。

Herman J. Bierens。《高级计量经济学专题：横截面与时间序列模型的估计、检验和设定》。剑桥大学出版社，第1版，1994年7月。ISBN 978-0-521-56511-0 978-0-521-41900-0 978-0-511-59927-9。doi：10.1017/CBO9780511599279。URL https://www.cambridge.org/core/product/identifier/9780511599279/type/book。Patrick Billingsley。《概率与测度》。威利概率与数理统计丛书。威利，纽约，第3版，1995年。ISBN 978-0-471-00710-4。Elenka Brenna。《我该照顾妈妈还是孩子? 意大利三明治世代的抑郁负担》。《健康政策》，125(3):415–423，2021年3月。ISSN 01688510。doi：10.1016/j.healthpol.2020.11.014。URL https://linking.elsevier.com/retrieve/pii/S0168851020302979。Kristen Capogrossi 和 Wen You。《学校营养计划对低收入儿童体重的影响：处理效应分析》。《健康经济学》，26(8):980–1000，2017年。Youngmin Cho。《非标准工作时间安排对劳动者健康的影响：工作与家庭冲突的中介作用》。《国际社会福利杂志》，27(1):74–87，2018年1月。ISSN 1369-6866, 1468-2397。doi：10.1111/ ijsw.12269。URL https://onlinelibrary.wiley.com/doi/10.1111/ijsw.12269。William G. Cochran。《分析性研究中的匹配》。《美国公共卫生与国民健康杂志》，43(6_Pt_1):684–691，1953年6月。ISSN 0002-9572。doi：10.2105/AJPH.43.6_Pt_1.684。URL https://ajph.aphapublications.org/doi/full/10.2105/AJPH.43.6_Pt_1.684。William G. Cochran 和 Donald B. Rubin。《观察性研究中的偏差控制：综述》。《Sankhy ¯a：印度统计学杂志，A辑（1961—2002）》，35(4):417–446，1973年。ISSN 0581572X。URL http://www.jstor.org/stable/25049893。Francisco Cribari-Neto、N. Garcia 和 Klaus L. P. Vasconcellos。《关于二项随机变量逆矩的注记》。《巴西计量经济学评论》，20:269–277，2000年。Rajeev Dehejia 和 Sadek Wahba。《非实验因果研究的倾向得分匹配方法》。技术报告 w6829，美国国家经济研究局，马萨诸塞州剑桥，1998年12月。URL http://www.nber.org/papers/w6829.pdf。

Rajeev H.Dejia 和 Sadek Wahba。《非实验因果研究的倾向得分匹配方法》。《经济学与统计学评论》，84(1):151–161，2002年。ISSN 00346535, 15309142。URLhttp://www.jstor.org/stable/3211745。

P. M. Densen、P. Padget、B. Webster、C. S. Nicol 和 C. Rich。《心血管梅毒研究》II：疗效评估中的方法学问题。《美国梅毒、淋病与性病杂志》，36(1):64–76，1952年1月。ISSN 0096-6738。

Eva Christine Erhardt。《自营就业之外的小额信贷：来自保加利亚企业的证据》。《劳动经济学》，47:75–95，2017年8月。ISSN 09275371。doi: 10.1016/j.labeco.2017.04.009。URL https://linkinghub.elsevier.com/retrieve/pii/S0927537116302561。Markus Frölich。《匹配估计量与最优带宽选择》。《统计与计算》，15(3):197–215，2005年7月。ISSN 0960-3174, 1573-1375。doi: 10.1007/s11222-005-1309-6。URL http://link.springer.com/10.1007/s11222-005-1309-6。Jinyong Hahn。《倾向得分在平均处理效应高效半参数估计中的作用》。《计量经济学》，66(2):315，1998年3月。ISSN 00129682。doi: 10.2307/2998560。URL https://www.jstor.org/stable/2998560?origin=crossref。James J. Heckman、Hidehiko Ichimura 和 Petra Todd。《匹配作为计量经济学评估估计量》。《经济研究评论》，65(2):261–294，1998年。ISSN 00346527, 1467937X。URL http://www.jstor.org/stable/2566973。Jonas Heese、Mozaffar Khan 和 Karthik Ramanna。《美国证券交易委员会是否被俘获？来自评论信审查的证据》。《会计与经济学杂志》，64(1):98–122，2017年8月。ISSN 01654101。doi: 10.1016/j.jacceco.2017.06.002。URL https://linkinghub.elsevier.com/retrieve/pii/S0165410117300344。Martin Huber、Michael Lechner 和 Andreas Steinmayr。《倾向得分半径匹配与偏差调整：调节参数及有限样本行为》。《实证经济学》，49(1):1–31，2015a年8月。ISSN 0377-7332, 1435-8921。doi: 10.1007/s00181-014-0847-1。URL http://link.springer.com/10.1007/s00181-014-0847-1。Martin Huber、Michael Lechner 和 Conny Wunsch。《工作场所健康促进与员工劳动力市场表现》。《卫生经济学杂志》，43:170–189，2015b年9月。ISSN 01676296。doi: 10.1016/j.jhealeco.2015. 07.002。URL https://linkinghub.elsevier.com/retrieve/pii/S0167629615000776。Jonathan Izudi、Imelda K. Tamwesigire 和 Francis Bajunirwe。《完成痰涂片监测是否会影响乌干达东部农村成年结核病患者的治疗成功率和治愈率？倾向得分匹配分析》。《PLOS ONE》，14(12):e0226919，2019年12月。ISSN 1932-6203。doi: 10.1371/journal.pone. 0226919。URL https://dx.plos.org/10.1371/journal.pone.0226919。Shao-Hsun Keng 和 S heng-Jang Sheu。《兴奋剂及其与香烟合用对死亡率的影响：以槟榔为例》。《欧洲卫生经济学杂志》，14(4):677–695，2013年8月。ISSN 1618-7598, 1618-7601。doi: 10.1007/s10198-012-0415-6。URL http://link.springer.com/10.1007/s10198-012-0415-6。Yuvaraj Krishnamoorthy 和 Tanveer Rehman。《产前保健就诊对儿童免疫接种的影响：使用全国代表性调查进行的倾向得分匹配分析》。《家庭医学》，39(4):603–609，2022年7月。ISSN 1460-2229。doi: 10.1093/fampra/cmab124。URL https://academic.oup.com/fampra/article/39/4/603/6375553。Ying-Ying Lee。《接受处理者多值处理效应的高效倾向得分回归估计量》。《计量经济学杂志》，204(2):207–222，2018年6月。ISSN 03044076。doi: 10.1016/j.jeconom.2018.02.002。URL https://linkinghub.elsevier.com/retrieve/pii/S0304407618300290。Jerzy Neyman。《概率论在农业试验中的应用》博士论文，华沙大学，1924年。

Anouk Patel-Campillo 和 V.B. Salas García。《打破贫困循环？有条件现金转移与高等教育 attainment》。《国际教育发展杂志》，92:102612，2022年7月。ISSN 07380593。doi: 10.1016/j.ijedudev.2022.102612。URLhttps://linkinghub.elsevier.com/retrieve/pii/S0738059322000621。

Herbert Robbins。关于斯特林公式的一点说明。《美国数学月刊》，62(1):26，1955年1月。ISSN 00029890。doi: 10.2307/2308012。URLhttp://www.jstor.org/stable/2308012?origin=crossref。

Paul R. Rosenbaum 和 Donald B. Rubin。倾向得分在因果效应观察性研究中的核心作用。《生物计量学》，70(1):41–55，1983年。ISSN 0006-3444, 1464-3510。doi: 10.1093/biomet/70.1.41。URLhttps://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/70.1.41。

Paul R. Rosenbaum 和 Donald B. Rubin。利用纳入倾向得分的多变量匹配抽样方法构建对照组。《美国统计学家》，39(1)：33–38，1985年2月。ISSN 0003-1305, 1537-2731。doi: 10.1080/00031305.1985.10479383。URL http://www.tandfonline.com/doi/abs/10.1080/00031305.1985.10479383。Donald B. Rubin。随机研究和非随机研究中处理效应的估计。《教育心理学杂志》，66(5)：688–701，1974年。ISSN 0022-0663。doi: 10.1037/h0037350。URL http://content.apa.org/journals/edu/66/5/688。Donald B. Rubin。《因果效应的匹配抽样》。剑桥大学出版社，2006年。doi: 10.1017/ CBO9780511810725。Luca Salmasi 和 Luca Pieroni。移民政策与出生体重：意大利法律中的正外部性。《健康经济学杂志》，43：128–139，2015年9月。ISSN 01676296。doi: 10.1016/j.jhealeco.2015.06.009。URL https://linkinghub.elsevier.com/retrieve/pii/S0167629615000752。E. Schuster 和 S. Yakowitz。对非参数回归理论的贡献及其在系统识别中的应用。《统计年刊》，7(1)：139–149，1979年。ISSN 00905364。URL http://www.jstor.org/stable/2958838。出版方：数理统计学会。Ramalingam Shanmugam 和 Barry C. Arnold。基于样本最小值条件分布的特征刻画。Sankhy ¯a：《印度统计学杂志》，A辑（1961–2002），50(3)：452–459，1988年。ISSN 0581572X。URL http://www.jstor.org/stable/25050713。Chung-Hua Shen 和 Yuan Chang。雄心与良知：企业社会责任是否有回报？匹配方法的应用。《商业伦理学杂志》，88(S1)：133–153，2009年4月。ISSN 0167-4544, 1573-0697。doi: 10.1007/s10551-008-9826-9。URL http://link.springer.com/10.1007/s10551-008-9826-9。Galen R. Shorack 和 Jon A. Wellner。《统计学应用中的经验过程》。应用数学经典丛书第59卷。工业与应用数学学会，费城，2009年。ISBN 978-0-89871-684-9。OCLC：ocn421147764。Elizabeth A. Stuart。因果推断的匹配方法：综述与展望。《统计科学》，25(1)：1–21，2010年。doi: 10.1214/09-STS313。URL https://doi.org/10.1214/09-STS313。Aad W. van der Vaart。《渐近统计学》。剑桥统计与概率数学丛书。剑桥大学出版社，1998年。ISBN 1-107-26372-7。Aad W. van der Vaart 和 Jon A. Wellner。《弱收敛与经验过程》。施普林格统计学丛书。施普林格纽约，纽约，NY，1996年。ISBN 978-1-4757-2547-6 978-1-4757-2545-2。doi: 10.1007/978-1-4757-2545-2。URL http://link.springer.com/10.1007/978-1-4757-2545-2。Nerina Vecchio、Debbie Davies 和 Nicholas Rohde。医疗保健服务获取不足对急诊就诊的影响：身体健康状况与心理健康状况的比较。《PLOS ONE》，13(8)：1–15，2018年8月。ISSN 1932-6203。doi: 10.1371/journal.pone.0202559。URL https://dx.plos.org/10.1371/journal.pone.0202559。

王伟、陆莉莉、Mohammedhamid Mohammedosman Kelifa、于燕、何安琪、曹娜、郑思、闫文军和杨银梅。COVID-19疫情期间遭受职场暴力的中国医务工作者的心理健康问题：一项采用倾向评分匹配分析的横断面研究。《风险管理与医疗保健政策》，第13卷：2827–2833，2020年12月。ISSN 1179-1594。doi: 10.2147/RMHP.S279170。