arXiv:2512.18627v1 [stat.ME] 21 Dec 2025

细网格点上一致推断的准确性

今井俊介 *[1]

[1]京都大学经济学研究生院

2025年12月23日

摘要

函数的一致置信带广泛应用于实证分析。人们提出了多种简单实现方法（其中最突出的是乘子自助法），并从理论上证明了其合理性。然而，对严格连续的指标集进行计算通常在计算上不可行，因此实践中会在有限评估网格上计算统计量，以确定临界值。本文量化了在有限网格点上实施乘子自助法时，评估网格需要多精细，才能得到有效的一致置信带。我们推导出一个明确的界，用以刻画由此产生的覆盖误差，并将离散化效应与网格上固有的高维自助法近似误差区分开来。该界为实践中选择网格大小提供了清晰的操作流程；我们还通过核密度估计示例展示了具体实现。

关键词：自助法；经验过程；高维中心极限定理；核平滑；非参数推断；一致置信带；

*本研究得到日本学术振兴会科学研究费助成项目编号 24KJ1472 的资助。

1 引言

一致置信带是一种有用的工具，可用于量化在连续点集上未知函数估计量的不确定性。关于非参数问题中一致推断的大量研究，至少可以追溯到早期的开创性工作，例如Bickel 和 Rosenblatt (1973年)以及相关论文。

如今，得益于 Chernozhukov 等人 （2014a，b）的开创性研究（以下简称 CCK）及其后续拓展，例如 Chernozhukov 等人 （2016年）；Chen 和 Kato （2020年），我们能够轻松验证一致高斯近似，并为各种可能属于非 Donsker 类的非参数估计量实施一致临界值。具体而言，CCK 在相当一般的条件下，建立了经验过程上确界与高斯过程上确界之间的近似结果。因此，采用近似高斯过程上确界的分位数作为临界值，便可构造出渐近覆盖率正确的置信带。更重要的是，在实际应用中，CCK 还表明，可以使用基于乘子自助法的临界值作为可行临界值，并获得渐近正确的覆盖率。CCK 理论已广泛应用于应用计量经济学和统计学，为一致推断中的解析（高斯）临界值和基于自助法的临界值提供依据，例如 Lee 等人 （2017）；Kato 和 Sasaki （2018年）；Fan 等人 （2022年）；Imai 等人（2025年）等。

然而，在实践中，无论是高斯过程在连续区间上的上确界，还是自助统计量在连续区间上的上确界，在计算上都不可行。因此，实际实现会用有限评估网格上的最大值替代连续区间上的上确界，并根据有限网格点上自助统计量的相应最大值计算临界值。由此自然引出一个问题：评估网格需要多精细，才能使所得程序在计算上可行的同时仍具备理论有效性？本研究将探讨这一问题。

本文对这一问题给出了定量解答。更具体地说，正如定理 1所示，我们量化了乘子自助统计量的网格最大值的条件 $(1-\alpha)$ 分位数（实践中使用的临界值）对学生化统计量在连续区间上的上确界的不可行 $(1-\alpha)$ 分位数的近似程度。定理 1 的关键特点是将近似误差分解为两个不同的误差来源。其一是离散化误差，即用网格上的最大值替代学生化统计量的连续域上确界所产生的误差。这一离散化部分包含两个要素。首先，离散化误差包括一项，用于刻画网格点之间出现尖锐局部尖峰的概率，而网格最大值可能会漏掉这些峰值。其次，即使网格之间没有出现尖锐尖峰，网格过粗仍可能仅仅因为网格间距过大而漏掉真正的峰值。因此，离散化误差还包含一项明确的网格条件，要求网格足够细，使得因稀疏造成的漏检可以忽略不计。另一项误差则反映了乘子自助法对细网格点上学生化统计量所构成的有限维向量之分布的近似精度。这一

项由针对高维最大值的高斯与乘子自助法近似理论控制，参见 Chernozhukov 等人 (2022 年）。这一分解直接转化为选择网格大小的实用方法；详细的实施流程见备注 2。

为了说明定理1中的界如何在实践中应用，第3节在标准正则条件下详细推导了一个核密度估计示例。该示例将界中的抽象要素转化为熟悉的量，并给出了一个简单且可实施的网格大小选择规则。

本文结构：本文其余部分安排如下。第2 节介绍一般框架，并给出基于网格的乘子自助法一致推断的主要覆盖误差界。第3节通过核密度估计量展开并说明该框架。第4 节包含证明。

2 主要结果

假设我们希望在紧区域 $\mathcal{X} := [x, x] \subset \mathbb{R}$ 上为某个函数 f 构造一致置信带，并且有一个针对 f 的估计量，其形式为独立变量之和 $\hat{f}_{h_n}(x) := n^{-1} \sum_{i=1}^{n} \psi_{h_n}(X_i,\ x)$，且

$$\hat{f}_{h_n}(x) - \mathbb{E}[\hat{f}_{h_n}(x)] = \frac{1}{n} \sum_{i=1}^{n} \{\psi_{h_n}(X_i, x) - \mathbb{E}[\psi_{h_n}(X_i, x)]\}, \tag{1}$$

其中 $\{X_i\}_{i=1}^{n}$ 是独立同分布观测的样本，h_n 表示一个可能依赖于 n的调节参数。

注 1. 为使论文简洁，并聚焦于离散化所导致的误差，我们将 $\hat{\mathbb{E}}[f_{h_n}(x)]$ 作为基准目标。然而，在实践中，关注参数通常是 $f(x)$，而非伪真目标 $\mathbb{E}[f_{h_n}(x)]$。在这种情况下，可以写作

$$\hat{f}_{h_n}(x) - f(x) := \frac{1}{n} \sum_{i=1}^{n} \{\psi_{h_n}(X_i, x) - \mathbb{E}[\psi_{h_n}(X_i, x)]\} + r_n(x),$$

其中 $r_n(x)$ 表示偏差项（或者更一般地说，线性化余项）。尽管如此，只要 $r_n(x)$ 相对于随机项，在 $\mathcal{X}$上满足一致的可忽略性条件，以下结果仍然成立（参见 Chernozhukov 等人，2023a 中的引理 1）。

假设 $\sigma_n^2(x) := n^{-1}\mathrm{Var}[\psi_{h_n}(X_i, x)]$，并令 $\hat{\sigma}_n^2(x)$ 为其估计量：

$$\hat{\sigma}_n^2(x) := \frac{1}{n} \left[\frac{1}{n} \sum_{i=1}^{n} \psi_{h_n}^2(X_i, x) - \left(\frac{1}{n} \sum_{i=1}^{n} \psi_{h_n}(X_i, x) \right)^2 \right]. \tag{2}$$

将标准化估计量和学生化估计量定义为

$$T_n(x) \coloneqq \frac{\hat{f}_{h_n}(x) - \mathbb{E}[\hat{f}_{h_n}(x)]}{\sigma_n(x)}, \quad \hat{T}_n(x) \coloneqq \frac{\hat{f}_{h_n}(x) - \mathbb{E}[\hat{f}_{h_n}(x)]}{\hat{\sigma}_n(x)},$$

分别地。对于每次自助迭代，我们独立于原始数据生成一组独立同分布的自助权重 $\{w_i^\star\}_{i=1}^n$，并计算

$$\hat{T}_n^\star(x) \coloneqq \frac{1}{n\hat{\sigma}_n(x)} \sum_{i=1}^n w_i^\star \{\psi_{h_n}(X_i, x) - \hat{f}_{h_n}(x)\}.$$

为简单起见，我们假设 $\{w_i^\star\}_{i=1}^n$ 服从标准正态分布。理想情况下，我们会基于自助过程在连续域上的上确界的条件（$1-\alpha$）分位数构造临界值，

$$\hat{c}_{n,1-\alpha}^{\star,\texttt{infeasible}} \coloneqq \inf\left\{t \in \mathbb{R} : \mathbb{P}^\star\left(\sup_{x\in\mathcal{X}} |\hat{T}_n^\star(x)| \le t \mid X_{1:n}\right) \ge 1-\alpha\right\},$$

其中，$\mathbb{P}^\star(\cdot \mid X_{1:n})$ 表示在给定原始样本 $\hat{}$ 的条件下、关于自助权重的概率。然而，通常无法通过计算求得 $\sup_{x\in\mathcal{X}} |T_n^\star(x)|$。因此，我们用有限网格上的最大值来近似连续域上确界。

为继续讨论，我们先介绍一些关于离散网格的记号。令 $\delta_n > 0$为网格尺寸。网格点 $\{x_j\}_{j=1}^p \subset \mathcal{X}$定义为

$$x_j \coloneqq \underline{x} + (j-1)\delta_n, \quad j = 1, \ldots, p-1, \quad x_p \coloneqq \overline{x}, \text{ with } p \coloneqq \left\lfloor \frac{\overline{x} - \underline{x}}{\delta_n} \right\rfloor + 2.$$

我们记作 $\mathcal{X}_{\delta_n} \coloneqq \{x_1, \ldots, x_p\}$。将最大间隔定义为

$$\Delta_n \coloneqq \max_{1\le j\le p-1} (x_{j+1} - x_j).$$

注意到 $x_1 = \underline{x}$和 $x_p = \overline{x}$，且最大间隔满足 $\Delta_n \le \delta_n$，因为对于$j \le p-2$，$x_{j+1} - x_j = \delta_n$，并且 $x_p - x_{p-1} \in (0, \delta_n]$。

在本研究中，我们使用以下可行自助法临界值，它基于 $\sup_{x\in\mathcal{X}} |T_n^\star(x)|$，用于近似$\hat{}$（不可行）临界值。

$$\hat{c}_{n,1-\alpha}^{\star} \coloneqq \inf\left\{t \in \mathbb{R} : \mathbb{P}^\star\left(\max_{1\le j\le p} |\hat{T}_n^\star(x_j)| \le t \mid X_{1:n}\right) \ge 1-\alpha\right\},$$

并研究这种近似的质量，给出使所得临界值能够实现有效一致推断的条件。为此，我们引入以下假设。

假设1。

(i) $\{X_i\}_{i=1}^n$ 是来自独立同分布观测的样本。

(ii) 存在某个常数 $B_n \geq 1$，使得 $\|Y_{i,h_n}(x_j)\|_{\psi_1} \leq B_n$ 和 $\mathbb{E}[Y_{i,h_n}^4(x_j)] \leq B_n^2$ 对所有$1 \leq i \leq n$ 和 $1 \leq j \leq p$均成立，其中

$$Y_{i,h_n}(x) := \frac{\psi_{h_n}(X_i, x) - \mathbb{E}[\psi_{h_n}(X_i, x)]}{\sqrt{\text{Var}[\psi_{h_n}(X_i, x)]}},$$

且 $\|\xi\|_{\psi_1} := \text{无穷}\{C > 0 : \mathbb{E}[\exp(|\xi|/C)] \leq 2\}$。

（iii）估计量 $f_{h_n}(x)$ 在方程（1）中对 $\mathcal{X}_{\delta_n}$一致地具有线性形式。

（iv）$\inf_{1 \leq j \leq p} \text{Var}[\psi_{h_n}(X_i, x_j)] \geq c > 0$。

（v）存在一个确定性序列 L_n 和 $\varepsilon_n \downarrow 0$，使得

$$\mathbb{P}\left(\sup_{\{(x,y):|x-y| \leq \Delta_n} \frac{|\hat{T}_n(x) - \hat{T}_n(y)|}{|x-y|} > L_n \right) \leq \varepsilon_n.$$

有必要对两个假设（ii）和（v）作进一步解释。假设1（ii）依据Chernozhukov 等人（2022年）的结果提出，用于控制高维高斯近似和乘子自助法近似误差。详见Chernozhukov 等人 （2022年）中的条件E和M。假设1（v）是用于控制离散化误差的高层次条件。我们将在后文详细讨论假设1（v），并在第1节的核密度估计示例中，为假设3.1（v）提供原始的充分条件。

以下定理给出了一个显式界，用于刻画以基于网格的自助法最大值 $\max_{1 \leq j \leq p} |T_n^\star(x_j)|$计算得到的可行自助法临界值 $\hat{c}_{n,1-\alpha}^\star$近似上确界统计量 $\sup_{x \in \mathcal{X}} |T_n(x)|$ 的 $(1-\alpha)$-分位数时产生的近似误差。

定理 1. 在假设 1下，令 $r := 2\left\{n^{-1} B_n^2 \log^3(np)\right\}^{1/4}$，则有

$$\begin{aligned} &\left| \mathbb{P}\left(\sup_{x \in \mathcal{X}} |\hat{T}_n(x)| \leq \hat{c}_{n,1-\alpha}^\star \right) - (1-\alpha) \right| \\ &\quad \leq 3\Big(\varepsilon_n + 1\{L_n \Delta_n / 2 > r\} \Big) + O\left(\left(\frac{B_n^2 \log^5(np)}{n} \right)^{1/4} \right). \end{aligned}$$

该定理的证明见第4.1节。

我们首先详细解释定理1右侧的三个项，然后介绍有效实施的工作流程（备注 2）。

定理1 右侧的前两项量化了由不可行的连续域上确界替换为有限网格上的最大值所产生的离散化误差。最后一项对应网格上的高维高斯近似误差和乘子自助法近似误差，其收敛速率由Chernozhukov 等人 (2022年)的结果给出。

ˆ 第一项反映了 $T_n(x)$ 在相邻点之间出现快速局部变化的概率，因此网格最大值可能会漏掉发生在网格点之间的尖峰。要理解这一项为何出现，请注意，网格间距至多为 Δ_n。因此，对任意位置 $x \in \mathcal{X}$，都存在一个ˆ网格点 x_j，其与该位置的距离不超过 Δ_n。如果 $T_n($ $x)$ 在如此短的距离内变化不大，ˆ ˆ那么 $T_n(x)$ 的值就可以由 $T_n(x_j)$ 很好地近似，因此连续域上确界 ˆ ˆ $\sup_{x\in\mathcal{X}} |T_n(x)|$ 也可以由网格最大值 $\max_{1\le j\le p} |T_n(x_j)|$ 很好地近似。更确切地说，量

$$\sup_{\{(x,y):|x-y|\le\Delta_n} \frac{|\hat{T}_n(x) - \hat{T}_n(y)|}{|x-y|}$$

ˆ 衡量在距离不超过 Δ_n 的范围内，x 每变化一个单位时，T_n 的最大变化量。当这一局部 ˆ 斜率不超过 L_n 时，从任意 x 移动到附近的网格点，T_n 的变化最多为 $L_n\Delta_n/2$。因此，网格点之间出现的任何峰值都不会比观测到的网格最大值高出超过 $L_n\Delta_n/2$；第一项是这种有利情形未能成立的概率。

第二项是确定性的网格设计条件，用以补充第一项。如上文 ˆ 所述，当 T_n 在距离不超过 Δ_n 的范围内的局部斜率不超过 L_n 时，从任意位置 $x \in \mathcal{X}$ 移动到附近的网格点 x_j，T_n 的变化最多为 $L_n\Delta_n/2$。因此，在此事件下，有

$$\sup_{x\in\mathcal{X}} |\hat{T}_n(x)| \le \max_{1\le j\le p} |\hat{T}_n(x_j)| + L_n\Delta_n/2,$$

因此，连续域上确界与网格最大值之间的差异受最坏情况下的网格间波动 $L_n\Delta_n/2$ 控制。指示函数 $1\{L_n\Delta_n/2 > r\}$ 用于检验这种最坏情况下的波动相对于容差水平 r 是否足够小；该容差水平源于高维高斯近似误差，以及高维高斯近似和自助法近似证明中使用的反集中不等式。若 $L_n\Delta_n/2 \le r$，则以网格最大值替代连续域上确界所带来的离散化误差，必然不超过 r（至多再加上第一项所涵盖事件发生的概率）。在实践中，这一项为网格选取提供了清晰的指导：网格间距应足够精细，使得 $L_n\Delta_n/2$ 不超过 r。

最后一项反映了：在将连续域上确界化简为有限维问题后，基于网格上最大值的乘子自助法在多大程度上准确地近似目标统计量的分布。该收敛速率继承自Chernozhukov 等人(2022年）提出的高维高斯近似与乘子自助法理论。在（Chernozhukov 等人，2023b，第 2375 页）中，他们推测该界“在协方差不受限制时近乎最优”。Chernozhukov 等人(2022年）的一个关键特点是，无需高维向量ˆ ˆ $(T_n(x_1)\dots,T_n(x_p))$ 的协方差矩阵一致可逆。这种稳健性在非参数应用中尤为重要，因为此时网格会变得稠密，对

估计量的邻近点进行评估时通常高度相关，因此随着 p 增大，所得有限维协方差矩阵可能趋于奇异。基于此，采用Chernozhukov 等人(2022年）给出的协方差不受限制时的收敛速率，作为自助法近似误差的基准，是合理的选择。对于实际应用，最好能刻画Chernozhukov 等人(2022年）所述高斯近似与乘子自助法近似误差中的常数。不过，建立这样的结果在技术上似乎颇具挑战，因此留待未来研究。

备注2（有效实施流程）。 定理1为网格所需的精细程度提供了定量判据。以下流程总结了如何以透明的方式选择网格大小。

1. **选择一个越界概率容差水平 ε_n，用于衡量局部变化：** 数量 ε_n界定了这样一个事件的概率：学生化统计量在与网格间距相当的距离上变化过于剧烈，以至于网格最大值可能遗漏网格之间一个狭窄的尖峰。如果目标仅是一阶有效性，取 $\varepsilon_n = o(1)$即可。如果希望使离散化相对于内在的高维近似误差（定理 1中的最后一项）可以忽略，一个自然的选择是令 ε_n 渐近地小于该项。

2. **确定相应的局部变化界 L_n：** 一旦确定了 ε_n，假设 1（v）便规定了相应的界 L_n，用于描述学生化统计量在局部（距离不超过网格间距）变化的速度。在应用中， L_n 可根据估计量和数据生成过程（DGP）的性质得到。

3. **确定 B_n 以及由此确定的阈值 r：** 量 B_n 由假设 1(ii) 中归一化加数的尾部性质和矩性质决定。一旦确定了 B_n，对于给定的网格点数量 p，定理 1 中的阈值 $r = 2\{n^{-1}B_n^2\log^3(np)\}^{1/4}$ 也就确定了。

4. **选择网格尺寸 Δ_n 以消除指示函数项：** 给定 L_n和 r，只要网格足够精细，使局部变化界所隐含的最坏情况网格间偏差低于阈值，定理 1 中的指示函数项就会消失，即 $L_n\Delta_n/2 \le r$。由此可得一条明确规则：逐渐减小 Δ_n（或增大 p），直至满足条件 $L_n\Delta_n/2 \le r$。消除指示函数项后，覆盖误差便简化为越界概率界 ε_n 加上定理 1中的最后一项。

- **注意（不要不必要地将网格设得过密）：** 尽管细化网格可以降低离散化误差，但也会增大 p，进而通过 $\log(np)$ 影响高维自助法近似误差。因此，网格应仅细化到足以满足网格条件 $L_n\Delta_n/2 \le r$，同时使 p 保持在高维近似仍然有效的范围内，即 $n^{-1}B_n^2\log^5(np) \to 0$。

下一节中，我们将通过一个核密度估计示例，具体讨论实现方法。关于一阶有效性，请参见命题 1 以及评注4–6，其中给出了明确的网格规则和实用指导。至于更强的目标——使离散化误差相对于高维近似误差可以忽略——尽管目前还无法给出完全以常数为级别的方案，我们在评注 6.

3 示例：核密度估计器

作为示例，我们考虑如何为核密度估计器的期望构造一致置信带。

3.1 理论分析

本节以核密度估计器（KDE）为例，说明如何确定 B_n 和 L_n，以及它们如何进而指导网格间距 Δ_n的合理选择。KDE定义为

$$\hat{f}_{h_n}(x) \coloneqq \frac{1}{n}\sum_{i=1}^{n} K_{i,h_n}(x), \text{ with } K_{i,h_n}(x) \coloneqq \frac{1}{h_n} K\left(\frac{X_i - x}{h_n}\right),$$

其中， K 是核函数， $h_n > 0$ 是带宽，二者满足稍后引入的假设。为简化记号，定义

$$Y_{i,h_n}(x) \coloneqq \frac{1}{\sqrt{\mathrm{Var}[K_{i,h_n}(x)]}} \left\{K_{i,h_n}(x) - \mathbb{E}[K_{i,h_n}(x)]\right\}.$$

假设2。

(i) $\{X_i\}_{i=1}^n$ 是来自密度函数为 f 的分布的独立同分布样本，且密度函数 f 连续可微，$\sup_{x\in\mathcal{X}} |f'(x)| < \infty$，并且 $0 < \inf_{x\in\mathcal{X}} f(x) < \sup_{x\in\mathcal{X}} f(x) < \infty$，其中 $f'(x) \coloneqq \partial f(x)/\partial x$。

(ii) $\sigma_n(x)$ 在 $\mathcal{X}$ 上对所有 $n \geq 1$ 连续可微，且

$$0 < \inf_{x\in\mathcal{X}} (nh_n)^{1/2}\sigma_n(x) \leq \sup_{x\in\mathcal{X}} (nh_n)^{1/2}\sigma_n(x) < \infty, \qquad \sup_{x\in\mathcal{X}} (nh_n)^{1/2}|\partial_x \sigma_n(x)| < \infty.$$

(iii) 带宽 $h_n > 0$ 满足 $h_n \to 0$ 和 $nh_n^3 \to \infty$。

(iv) 核函数 $K : \mathbb{R} \to \mathbb{R}$ 二阶连续可微，$\int_{\mathbb{R}} K(u)du = 1$，对于 $v = 0, 1, 2$，$\sup_{u\in\mathbb{R}} |K^{(v)}(u)| < \infty$，且

$$\int_{\mathbb{R}} |K(u)|du < \infty, \int_{\mathbb{R}} |u|K^2(u)du < \infty, \int_{\mathbb{R}} K^4(u)du < \infty,$$
$$\int_{\mathbb{R}} |K'(u)|du < \infty, \int_{\mathbb{R}} \{K'(u)\}^2 du < \infty, \int_{\mathbb{R}} |K''(u)|du < \infty.$$

这类假设在相关文献中很常见（例如参见 Wasserman, 2006;Li 和 Racine, 2007;Tsybakov, 2009）。假设 $nh_n^3 \to \infty$ 中的条件 2(iii) 比 KDE 分析中最常见的条件 $nh_n \to \infty$ 更强，但这一限制在实践中并不过分严格，因为常用的带宽选择方法（例如最优均方积分误差速率（$h_n \asymp n^{-1/5}$）满足 $nh_n^3 \to \infty$。

以下引理给出了核密度估计示例中网格选择所需的显式量。具体而言，引理 1(ii) 给出了 B_n 在假设 1(ii) 中的一个有效取值，而引理 1(iii) 则通过构造 $L_n(\varepsilon_n)$验证了假设 1(v)。稍后讨论实现问题时，我们将用到引理 1(i)。

引理 1. 在假设 2下，以下结论成立。

(i) For all $x \in \mathcal{X}$, $\mathrm{Var}[K_{i,h_n}(x)] = h_n^{-1} f(x) \int K^2(u)du + o(h_n^{-1})$.

(ii) $\max_{1\le j\le p} \mathbb{E}[Y_{i,h_n}^4(x_j)]$ 和 $\max_{1\le j\le p} \|Y_{i,h_n}(x_j)\|_{\psi_1}$ 有界，且

$$\max_{1\le j\le p} \mathbb{E}[Y_{i,h_n}^4(x_j)] \le \max_{1\le j\le p} \frac{16 \sup_{x\in\mathcal{X}} |f(x)|}{h_n^3 \mathrm{Var}[K_{i,h_n}(x_j)]^2} \int_{\mathbb{R}} K^4(u)du,$$
$$\max_{1\le j\le p} \|Y_{i,h_n}(x_j)\|_{\psi_1} \le \max_{1\le j\le p} \frac{1}{\log 2} \cdot \frac{2 \sup_{u\in\mathbb{R}} |K(u)|}{h_n \sqrt{\mathrm{Var}[K_{i,h_n}(x_j)]}}.$$

(iii) 对于给定的 $\varepsilon_n \in (0,1)$，定义 $L_n(\varepsilon_n)$ 如下。则假设 1(v) 成立。

$$L_n(\varepsilon_n) := 2\left(\sup_{x\in\mathcal{X}} \frac{1}{\sigma_n(x)} A_1\left(\frac{\varepsilon_n}{2}\right) + \sup_{x\in\mathcal{X}} \frac{\partial \sigma_n(x)}{\partial x} \frac{1}{\sigma_n^2(x)} A_0\left(\frac{\varepsilon_n}{2}\right)\right),$$

with

$$A_0(\varepsilon_n) := \sqrt{\frac{2v_0 \log(2m_n/\varepsilon_n)}{n}} + \frac{M_0 \log(2m_n/\varepsilon_n)}{3n} + \frac{\sup_{u\in\mathbb{R}} |K'(u)|}{n^{1/2} h_n^{1/2}},$$
$$A_1(\varepsilon_n) := \sqrt{\frac{2v_1 \log(2m_n/\varepsilon_n)}{n}} + \frac{M_1 \log(2m_n/\varepsilon_n)}{3n} + \frac{\sup_{u\in\mathbb{R}} |K''(u)|}{n^{1/2} h_n^{3/2}},$$

and

$$m_n := \lfloor n^{1/2} h_n^{-3/2} (\overline{x} - \underline{x}) \rfloor + 2,$$
$$M_0 := 2h_n^{-1} \sup_{u\in\mathbb{R}} |K(u)|, \quad v_0 := h_n^{-1} \sup_{x\in\mathcal{X}} |f(x)| \int_{\mathbb{R}} \{K(u)\}^2 du,$$
$$M_1 := 2h_n^{-2} \sup_{u\in\mathbb{R}} |K'(u)|, \quad v_1 := h_n^{-3} \sup_{x\in\mathcal{X}} |f(x)| \int_{\mathbb{R}} \{K'(u)\}^2 du.$$

该引理的证明见第4.2节。

注 3。 忽略对数因子后，我们可以看到

$$\sup_{x\in\mathcal{X}} \sigma_n^{-1}(x) \asymp (nh_n)^{1/2}, \;\; A_1(\varepsilon_n) \asymp (nh_n^3)^{-1/2},$$
$$\sup_{x\in\mathcal{X}} \sigma_n^{-2}(x)\{\partial\sigma_n(x)/\partial x\} \asymp (nh_n)^{1/2}, \;\; A_0(\varepsilon_n) \asymp (nh_n)^{-1/2}.$$

这意味着，在核密度估计的情形下，与第一项相比， $L_n(\varepsilon_n)$ 中的第二项可以忽略。因此我们可以使用

$$\tilde{L}_n(\varepsilon_n) := 2\sup_{x\in\mathcal{X}} \sigma_n^{-1}(x) A_1(\varepsilon_n/2).$$

虽然在核密度估计示例中（如上所示），第二项的阶较小，但其大小可能取决于估计量，在更一般的情形下也未必可以忽略。因此，我们给出包含这两项的一般界。

根据引理 1 和注 3，我们可以进一步确定网格大小。对于每个网格尺寸$\delta \in (0, \;\; x - x]$，令 $\mathcal{X}_\delta \equiv \{x_1, \ldots, x_{p(\delta)}\}$ 为第 2 节中定义的等距网格。根据引理 1(ii)，我们可以取

$$B_n = \sup_{x\in\mathcal{X}} \Big(B_{\psi,n}(x) \vee B_{4,n}(x) \Big),$$

其中

$$B_{\psi,n}(x) = \frac{1}{\log 2} \cdot \frac{2\sup_{u\in\mathbb{R}} |K(u)|}{h_n \sqrt{\mathrm{Var}[K_{i,h_n}(x)]}}, \qquad B_{4,n}(x) = \sqrt{\frac{16\sup_{t\in\mathcal{X}} |f(t)|}{h_n^3 \mathrm{Var}[K_{i,h_n}(x)]^2} \int_{\mathbb{R}} K^4(u)\, du}.$$

将 $p = p(\delta)$ 和 B_n 代入定理 r 1 中 的 定义可得

$$r(\delta) = 2\left\{ \frac{B_n^2 \log^3\big(n\, p(\delta)\big)}{n} \right\}^{1/4}.$$

下面的命题说明如何选择 δ_n ，以确保一致推断的一阶有效性。

命题1。 假设 2成立。设 $\{\varepsilon_n\}_{n\geq 1}$ 为满足$\varepsilon_n \downarrow 0$的确定性序列，并将 $L_n(\varepsilon_n)$ 定义如备注 3。取

$$\delta_n := \sup\left\{ \delta \in (0, \overline{x} - \underline{x}] : \; \delta \leq 2r(\delta)/\tilde{L}_n(\varepsilon_n) \right\},$$

并令 Δ_n 表示所得网格的最大间隔$\mathcal{X}_{\delta_n}$ （如第 2 节所定义）。则

$$\frac{\tilde{L}_n(\varepsilon_n)\Delta_n}{2} \leq r(\delta_n),$$

˜因此定理 $1\{L_n(\varepsilon_n)\Delta_n/2 > r\}$ 中的指标项 1 对于这种网格选择为零。

˜证明。 由于网格是等间距的， $\Delta_n \leq \delta_n$，并且根据构造有 $L_n(\varepsilon_n)\delta_n/2 \leq r(\delta_n)$。因此

˜ ˜$L_n(\varepsilon_n)\Delta_n/2 \leq L_n(\varepsilon_n)\delta_n/2 \leq r(\delta_n)$。 □

现在我们讨论实施命题1的实际问题。具体而言，我们重点介绍备注4 和备注5中应用时出现的若干瓶颈。此外，备注6 提供了一个简单的实施程序。

备注4（实现问题1：未知量）。命题1中的量涉及未知的分布量，例如 $\sup_{x\in\mathcal{X}} f(x)$、$\sup_{x\in\mathcal{X}} \sigma_n^{-1}(x)$ 和 $\sup_{x\in\mathcal{X}}\{\mathsf{Var}[K_{i,h_n}(x)]\}^{-1}$。一种可行的实施方法是用代入估计量替换这些未知对象。注意，引理1(i)表明 $\mathsf{Var}[K_{i,h_n}(x)] = h_n^{-1} f(x) \int K^2(u)du + o(h_n^{-1})$。因此，我们可以用 $\sup_{x\in\mathcal{X}}\{\mathsf{Var}[K_{i,h_n}(x)]\}^{-1} \approx h_n\{\int_{\mathbb{R}} K^2(u)du \cdot \inf_{x\in\mathcal{X}} f(x)\}^{-1}$ 进行近似。类似地，由于$\sigma_n(x)^2 = n^{-1}\mathsf{Var}[K_{i,h_n}(x)]$,，我们有 $\sup_{x\in\mathcal{X}} \sigma_n(x)^{-1} \approx (nh_n\{\int_{\mathbb{R}} K^2(u)du \cdot \inf_{x\in\mathcal{X}} f(x)\}^{-1})^{1/2}$。因此，只需估计 $\sup_{x\in\mathcal{X}} f(x)$ 和 $\inf_{x\in\mathcal{X}} f(x)$。尽管有许多可行方法，但可以使用直方图估计量的最大值和最小值，并采用标准软件包提供的默认分箱规则。

备注5（实现问题2： δ_n的实现）。命题1中的网格规则通过约束˜ $\delta \leq 2r(\delta)/L_n(\varepsilon_n)$进行定义，其中 $r(\delta)$仅通过项 $\log\{np(\delta)\}$依赖于 δ，且 $p(\delta) = \lfloor (x - x)/\delta \rfloor + 2$。由于下取整算子的存在，$p(\delta)$是 δ的阶梯函数，因此通常难以方便地求得最大的可容许 δ 的闭式解。

在实践中，这一问题可以按如下方式轻松处理。令 $|\mathcal{X}| := x - x$，并考虑连续代理 $\tilde{p}(\delta) := |\mathcal{X}|/\delta$（或 $|\mathcal{X}|/\delta + 2$），该代理消除了下取整算子。将 $\tilde{p}(\delta)$代入 $r(\delta)$，得到关于 δ的光滑隐式一维方程，例如可以通过二分法进行数值求解。令 δ 表示等式 $\delta = 2\tilde{r}(\delta)/L_n(\varepsilon_n)$的一个数值解，其中 $\tilde{r}(\delta)$定义为 $r(\delta)$，并将 $p(\delta)$替换为 $\tilde{p}(\delta)$。随后，我们将网格点的整数数量定义为

$$\hat{p} := \left\lceil \frac{|\mathcal{X}|}{\tilde{\delta}} \right\rceil + 2 \qquad \text{and set} \qquad \hat{\delta} := \frac{|\mathcal{X}|}{\hat{p} - 1}.$$

这种舍入会产生稍微更细的网格（网格尺寸更小），因此对于满足原不等式而言是保守的。

备注6 (实现问题 3：简单实现)。本备注讨论一种基于阶数的网格选择方法。虽然不指定常数不利于实践中的透明度，但这大大简化了实现，并提供了一条易于使用的规则。在全文中，我们忽略对数因子（例如涉及 $\log(2m_n/\varepsilon_n)$ 和 $\log(np_n)$ 的因子），并着重关注 n 和 h_n 中的主导多项式速率。

(i) **仅要求有效性**。为确保一阶有效性，只需选择网格，使定理1 中的指标项消失，即 $L_n(\varepsilon_n)\Delta_n/2 \lesssim r$。注意， $L_n(\varepsilon_n)$ 和 $r(\delta)$ 满足

$L_n(\varepsilon_n) \asymp h_n^{-1}$和 $r(\delta) \asymp (nh_n)^{-1/4}$ (因为 $B_n^2 \asymp h_n^{-1}$)。因此，一种自然的基于阶数的选择是

$$\Delta_n = \frac{2r}{L_n(\varepsilon_n)} c_\Delta n^{-\gamma} = c_\Delta n^{-1/4-\gamma} h_n^{3/4},$$

或者等价地

$$p_n = \left\lceil \frac{|\mathcal{X}|}{\Delta_n} \right\rceil = \left\lceil c_\Delta^{-1} |\mathcal{X}| n^{1/4+\gamma} h_n^{-3/4} \right\rceil,$$

其中常数 $c_\Delta \in (0, 1)$ 和 $\gamma > 0$ 都很小。实际应用中，γ 应当非常小。

(ii) **离散化误差小于高斯近似误差**。在 (i) 中采用相同的网格选择时，定理 1 中的指标项最终会消失。此外，选择$\varepsilon_n = o(\rho_n)$ 可使稀有尖峰概率项可以忽略，其中

$$\rho_n := \left(\frac{B_n^2 \log^5(np_n)}{n} \right)^{1/4}.$$

在核密度估计示例中，(i) 中的 p_n 关于 n 至多按多项式增长，因此 $\log(np_n) = O(\log n)$，且$\rho_n \asymp (nh_n)^{-1/4}$，相差至多对数因子。因此，一旦选定网格以确保有效性，整体精度便由高维近似项而非离散化决定。

4 证明

4.1 定理1的证明

首先，注意到

$$\begin{aligned}
\rho_n^\star &:= \sup_{t\in\mathbb{R}} \left| \mathbb{P}^\star \left(\max_{1\le j\le p} |\hat{T}_n^\star(x_j)| \le t \mid X_{1:n} \right) - \mathbb{P} \left(\sup_{x\in\mathcal{X}} |\hat{T}_n(x)| \le t \right) \right| \\
&\le \sup_{t\in\mathbb{R}} \left| \mathbb{P}^\star \left(\max_{1\le j\le p} |\hat{T}_n^\star(x_j)| \le t \mid X_{1:n} \right) - \mathbb{P} \left(\max_{1\le j\le p} |\hat{T}_n(x_j)| \le t \right) \right| \\
&\quad + \sup_{t\in\mathbb{R}} \left| \mathbb{P} \left(\max_{1\le j\le p} |\hat{T}_n(x_j)| \le t \right) - \mathbb{P} \left(\sup_{x\in\mathcal{X}} |\hat{T}_n(x)| \le t \right) \right| \\
&=: (\rho^\star.I) + (\rho^\star.II).
\end{aligned}$$

由于 $(\rho^\star.I) \le 1$，当 $n^{-1}B_n^2\log^5(np) > 1$ 时，界 $(\rho^\star.I) \le C\{n^{-1}B_n^2\log^5(np)\}^{1/4}$ 是显然的，因此不失一般性，我们可以假设 $n^{-1}B_n^2\log^5(np) \le 1$ 。在以下推导中，我们经常会用到这个不等式。

首先，根据 $(\rho^\star.I)$，有

$$
\begin{aligned}
(\rho^\star.I) \leq & \sup_{t\in\mathbb{R}}\left|\mathbb{P}^\star\left(\max_{1\leq j\leq p}|T_n^\star(x_j)|\leq t \mid X_{1:n}\right)-\mathbb{P}\left(\max_{1\leq j\leq p}|T_n(x_j)|\leq t\right)\right| \\
& +\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\max_{1\leq j\leq p}|\hat{T}_n(x_j)|\leq t\right)-\mathbb{P}\left(\max_{1\leq j\leq p}|T_n(x_j)|\leq t\right)\right| \\
& +\sup_{t\in\mathbb{R}}\left|\mathbb{P}^\star\left(\max_{1\leq j\leq p}|\hat{T}_n^\star(x_j)|\leq t \mid X_{1:n}\right)-\mathbb{P}^\star\left(\max_{1\leq j\leq p}|T_n^\star(x_j)|\leq t \mid X_{1:n}\right)\right| \\
:= & (\rho^\star.I.I)+(\rho^\star.I.II)+(\rho^\star.I.III),
\end{aligned}
$$

其中 $T_n^\star(x) := n^{-1}\sigma_n^{-1}\sum_{i=1}^n w_i^\star\{\psi_{h_n}(X_i, x) - \hat{f}_{h_n}(x)\}$。根据 Chernozhukov 等人(2022年)中的引理 4.6，我们可以看到

$$
(\rho^\star.I.I) = O_p\left(\left(\frac{B_n^2\log^5(np)}{n}\right)^{1/4}\right). \tag{3}
$$

此外，根据（Le Cam，1986，第 2 章）中的引理 15，我们有

$$
(\rho^\star.I.II) \leq \sup_{t\in\mathbb{R}}\mathbb{P}\left(t<\max_{1\leq j\leq p}|T_n(x_j)|\leq t+r\right)+\sup_{t\in\mathbb{R}}\mathbb{P}\left(\max_{1\leq j\leq p}\left|\hat{T}_n(x_j)-T_n(x_j)\right|>r\right). \tag{4}
$$

其中 $r := 2\left(n^{-1}B_n^2\log^3(pn)\right)^{1/4}$。关于公式（4）中的第一项，注意到

$$
\begin{aligned}
&\sup_{t\in\mathbb{R}}\mathbb{P}\left(t<\max_{1\leq j\leq p}|T_n(x_j)|\leq t+r\right) \\
&\leq \sup_{t\in\mathbb{R}}\mathbb{P}\left(t<\max_{1\leq j\leq p}|G_n(x_j)|\leq t+r\right)+\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\max_{1\leq j\leq p}|T_n(x_j)|\leq t\right)-\mathbb{P}\left(\max_{1\leq j\leq p}|G_n(x_j)|\leq t\right)\right|.
\end{aligned}
$$

Nazarov 不等式给出

$$
\sup_{t\in\mathbb{R}}\mathbb{P}\left(t<\max_{1\leq j\leq p}|G_n(x_j)|\leq t+r\right)\leq r(\sqrt{2\log p}+2)=O\left(\left(\frac{B_n^2\log^5(np)}{n}\right)^{1/4}\right),
$$

以及 Ch ernozhukov 等人 （2022年）指出，存在一个普适常数 C 使得

$$
\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\max_{1\leq j\leq p}|T_n(x_j)|\leq t\right)-\mathbb{P}\left(\max_{1\leq j\leq p}|G_n(x_j)|\leq t\right)\right|\leq C\left(\frac{B_n^2\log^5(np)}{n}\right)^{1/4}. \tag{5}
$$

因此

$$
\sup_{t\in\mathbb{R}}\mathbb{P}\left(t<\max_{1\leq j\leq p}|T_n(x_j)|\leq t+r\right)=O\left(\left(\frac{B_n^2\log^5(np)}{n}\right)^{1/4}\right). \tag{6}
$$

接下来，我们评估公式（4）中的第二项。由于

$$
\max_{1\leq j\leq p}\left|\hat{T}_n(x_j)-T_n(x_j)\right|\leq\max_{1\leq j\leq p}|T_n(x_j)|\max_{1\leq j\leq p}\frac{|\hat{\sigma}_n^2(x_j)-\sigma_n^2(x_j)|}{\hat{\sigma}_n(x_j)|\hat{\sigma}_n(x_j)+\sigma_n(x_j)|},
$$

我们可以看到

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le j\le p}\left|\hat{T}_n(x_j)-T_n(x_j)\right|>r\right)\\
&\le \mathbb{P}\left(\max_{1\le j\le p}|T_n(x_j)|\max_{1\le j\le p}\frac{|\hat{\sigma}_n^2(x_j)-\sigma_n^2(x_j)|}{\hat{\sigma}_n(x_j)|\hat{\sigma}_n(x_j)+\sigma_n(x_j)|}>r\right)\\
&\le \mathbb{P}\left(\max_{1\le j\le p}|T_n(x_j)|\cdot C\sqrt{\frac{B_n^2\log(pn)}{n}}>r\right)+\mathbb{P}\left(\max_{1\le j\le p}\frac{|\hat{\sigma}_n^2(x_j)-\sigma_n^2(x_j)|}{\hat{\sigma}_n(x_j)|\hat{\sigma}_n(x_j)+\sigma_n(x_j)|}>C\sqrt{\frac{B_n^2\log(pn)}{n}}\right).
\end{aligned}$$

由于方差估计量 $\hat{\sigma}_n^2(x)$ 按式（2）定义，在假设1（ii）下，Chernozhukov 等人（2022年）的引理 4.2 给出

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le j\le p}\frac{|\hat{\sigma}_n^2(x_j)-\sigma_n^2(x_j)|}{\hat{\sigma}_n(x_j)|\hat{\sigma}_n(x_j)+\sigma_n(x_j)|}>C\sqrt{\frac{B_n^2\log(pn)}{n}}\right)\\
&\quad\le\frac{1}{n}+3\left(\frac{B_n^2\log^3(pn)}{n}\right)^{1/2}=O\left(\left(\frac{B_n^2\log^5(np)}{n}\right)^{1/4}\right).
\end{aligned}$$

此外，还成立：

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le j\le p}|T_n(x_j)|\cdot C\sqrt{\frac{B_n^2\log(pn)}{n}}>r\right)\\
&=1-\mathbb{P}\left(\max_{1\le j\le p}|T_n(x_j)|\cdot C\sqrt{\frac{B_n^2\log(pn)}{n}}\le r\right)\\
&\le \mathbb{P}\left(\max_{1\le j\le p}|G_n(x_j)|\cdot C\sqrt{\frac{B_n^2\log(pn)}{n}}>r\right)+\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\max_{1\le j\le p}|T_n(x_j)|\le t\right)-\mathbb{P}\left(\max_{1\le j\le p}|G_n(x_j)|\le t\right)\right|.
\end{aligned}$$

Since，由并集界和高斯尾部的标准性质可得 e

$$\mathbb{P}\left(\max_{1\le j\le p}|G_n(x_j)|\cdot C\sqrt{\frac{B_n^2\log(pn)}{n}}>r\right)\le 2p\exp\left(-\frac{r^2}{2}\cdot\left(C\sqrt{\frac{B_n^2\log(pn)}{n}}\right)^{-2}\right).$$

由 $r=2\left(n^{-1}B_n^2\log^3(pn)\right)^{1/4}$ 和 $B_n^2\log^5(np)/\le 1$，我们有

$$2p\exp\left(-\frac{r^2}{2}\cdot\left(C\sqrt{\frac{B_n^2\log(pn)}{n}}\right)^{-2}\right)=O\left(\left(\frac{B_n^2\log^5(np)}{n}\right)^{1/4}\right).$$

结合公式（5），我们有

$$\mathbb{P}\left(\max_{1\le j\le p}|T_n(x_j)|\cdot C\sqrt{\frac{B_n^2\log(pn)}{n}}>r\right)=O\left(\left(\frac{B_n^2\log^5(np)}{n}\right)^{1/4}\right).$$

综上

$$\mathbb{P}\left(\max_{1\leq j\leq p}\left|\hat{T}_n(x_j)-T_n(x_j)\right|>r\right)=O\left(\left(\frac{B_n^2\log^5(np)}{n}\right)^{1/4}\right). \tag{7}$$

将计算结果代入等式（6）和等式（7），再代入等式（4，可得

$$(\rho^\star.I.II)=O\left(\left(\frac{B_n^2\log^5(np)}{n}\right)^{1/4}\right). \tag{8}$$

同理，可得

$$(\rho^\star.I.III)=O_p\left(\left(\frac{B_n^2\log^5(np)}{n}\right)^{1/4}\right). \tag{9}$$

由等式（3）、等式（8）和等式（9，可得

$$(\rho^\star.I)\leq(\rho^\star.I.I)+(\rho^\star.I.II)++(\rho^\star.I.III)=O_p\left(\left(\frac{B_n^2\log^5(np)}{n}\right)^{1/4}\right). \tag{10}$$

关于 $(\rho^\star.II)$，根据（Le Cam，1986，第 2 章）中的引理 15，可得

$$\begin{aligned}
(\rho^\star.II)&\leq\sup_{t\in\mathbb{R}}\mathbb{P}\left(t<\max_{1\leq j\leq p}|\hat{T}_n(x_j)|\leq t+r\right)+\mathbb{P}\left(\sup_{x\in\mathcal{X}}|\hat{T}_n(x)|-\max_{1\leq j\leq p}|\hat{T}_n(x_j)|>r\right)\\
&\leq\sup_{t\in\mathbb{R}}\mathbb{P}\left(t<\max_{1\leq j\leq p}|G_n(x_j)|\leq t+r\right)\\
&\quad+\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\max_{1\leq j\leq p}|\hat{T}_n(x_j)|\leq t\right)-\mathbb{P}\left(\max_{1\leq j\leq p}|G_n(x_j)|\leq t\right)\right|\\
&\quad+\mathbb{P}\left(\sup_{x\in\mathcal{X}}|\hat{T}_n(x)|-\max_{1\leq j\leq p}|\hat{T}_n(x_j)|>r\right)\\
&=:(\rho^\star.II.I)+(\rho^\star.II.II)+(\rho^\star.II.III).
\end{aligned}$$

Nazarov 不等式给出

$$(\rho^\star.II.I)\leq r(\sqrt{2\log p}+2)=O\left(\left(\frac{B_n^2\log^5(np)}{n}\right)^{1/4}\right).$$

此外，根据公式（8）和公式（5），我们有

$$\begin{aligned}
(\rho^\star.II.II)&=\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\max_{1\leq j\leq p}|\hat{T}_n(x_j)|\leq t\right)-\mathbb{P}\left(\max_{1\leq j\leq p}|G_n(x_j)|\leq t\right)\right|\\
&\leq(\rho^\star.I.II)+\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\max_{1\leq j\leq p}|T_n(x_j)|\leq t\right)-\mathbb{P}\left(\max_{1\leq j\leq p}|G_n(x_j)|\leq t\right)\right|\\
&=O\left(\left(\frac{B_n^2\log^5(np)}{n}\right)^{1/4}\right).
\end{aligned}$$

最后，就（$\rho^\star.III$）而言，注意到

$$0 \le \sup_{x\in\mathcal{X}} |\hat{T}_n(x)| - \max_{1\le j\le p} |\hat{T}_n(x_j)| \le \sup_{\{(x,y):|x-y|\le \sup_{x\in\mathcal{X}} \min_{1\le j\le p}|x-x_j|\}} |\hat{T}_n(x) - \hat{T}_n(y)|.$$

回顾一下，我们将最大间隔定义为 $\Delta_n := \max_{1\le j\le p_n-1}(x_{j+1} - x_j)$。那么，有

$$\begin{aligned}
&\sup_{\{(x,y):|x-y|\le \sup_{x\in\mathcal{X}} \min_{1\le j\le p}|x-x_j|\}} |\hat{T}_n(x) - \hat{T}_n(y)| \\
&\le \sup_{x\in\mathcal{X}} \min_{1\le j\le p} |x - x_j| \cdot \sup_{\{(x,y):|x-y|\le\Delta_n\}} \frac{|\hat{T}_n(x) - \hat{T}_n(y)|}{|x-y|}.
\end{aligned}$$

这意味着

$$(\rho^\star.III) \le \mathbb{P}\left(\sup_{x\in\mathcal{X}} \min_{1\le j\le p} |x - x_j| \cdot \sup_{\{(x,y):|x-y|\le\Delta_n\}} \frac{|\hat{T}_n(x) - \hat{T}_n(y)|}{|x-y|} > r \right).$$

在假设 Assumption 1下，有

$$\mathbb{P}\left(\sup_{\{(x,y):|x-y|\le\Delta_n} \frac{|\hat{T}_n(x) - \hat{T}_n(y)|}{|x-y|} > L_n \right) \to 0.$$

因此

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\{(x,y):|x-y|\le\Delta_n} |\hat{T}_n(x) - \hat{T}_n(y)| > r \right) \\
&\le \mathbb{P}\left(\sup_{\{(x,y):|x-y|\le\Delta_n} \frac{|\hat{T}_n(x) - \hat{T}_n(y)|}{|x-y|} > L_n \right) + 1\{L_n \sup_{x\in\mathcal{X}} \min_{1\le j\le p} |x - x_j| > r\} \\
&\le \mathbb{P}\left(\sup_{\{(x,y):|x-y|\le\Delta_n} \frac{|\hat{T}_n(x) - \hat{T}_n(y)|}{|x-y|} > L_n \right) + 1\{L_n\Delta_n/2 > r\},
\end{aligned}$$

其中，最后一个不等式由 $\sup_{x\in\mathcal{X}} \min_{1\le j\le p} |x - x_j| \le \Delta_n/2$ 得出。综上，我们有

$$\begin{aligned}
&(\rho^\star.II) \\
&\le \mathbb{P}\left(\sup_{\{(x,y):|x-y|\le\Delta_n} \frac{|\hat{T}_n(x) - \hat{T}_n(y)|}{|x-y|} > L_n \right) + 1\{L_n\Delta_n/2 > r\} + O\left(\left(\frac{B_n^2 \log^5(np)}{n} \right)^{1/4} \right).
\end{aligned} \tag{11}$$

由公式（10）和公式（11），我们有

$$\rho_n^\star \le \mathbb{P}\left(\sup_{\{(x,y):|x-y|\le\Delta_n} \frac{|\hat{T}_n(x) - \hat{T}_n(y)|}{|x-y|} > L_n \right) + 1\{L_n\Delta_n/2 > r\} + O_p\left(\left(\frac{B_n^2 \log^5(np)}{n} \right)^{1/4} \right).$$

与Kato 和 Sasaki (2018年）中定理 2 的证明中的步骤 3 相同，我们有

$$\begin{aligned}
&\left| \mathbb{P}\left(\sup_{x\in\mathcal{X}} |\hat{T}_n(x)| \le \hat{c}^\star_{n,1-\alpha} \right) - (1-\alpha) \right| \\
&\le 3\left\{ \mathbb{P}\left(\sup_{\{(x,y):|x-y|\le\Delta_n} \frac{|\hat{T}_n(x) - \hat{T}_n(y)|}{|x-y|} > L_n \right) + 1\{L_n\Delta_n/2 > r\} \right\} + O\left(\left(\frac{B_n^2 \log^5(np)}{n} \right)^{1/4} \right).
\end{aligned}$$

证明至此完成。

4.2 引理1的证明

引理的证明*1(i)*. 由于

$$\mathbb{E}[K_{i,h_n}^2(x)] = \frac{1}{h_n}\int K^2(u)f(x+uh_n)du, \quad \mathbb{E}[K_{i,h_n}(x)] = \int K(u)f(x+uh_n)du,$$

，在假设2下，有

$$\begin{aligned}
&\sup_{x\in\mathcal{X}}\left|\text{Var}[K_{i,h_n}(x)] - \frac{f(x)}{h_n}\int K^2(u)du\right| \\
&= \sup_{x\in\mathcal{X}}\left|\mathbb{E}[K_{i,h_n}^2(x)] - \mathbb{E}[K_{i,h_n}(x)]^2 - \frac{1}{h_n}\int f(x)K^2(u)du\right| \\
&\le \sup_{x\in\mathcal{X}}\left|\frac{1}{h_n}\int K^2(u)\{f(x+uh_n)-f(x)\}du\right| + \sup_{x\in\mathcal{X}}\left|\left(\int K(u)f(x+uh_n)du\right)^2\right| \\
&\le \sup_{x\in\mathcal{X}}|f'(x)|\left|\int |u|K^2(u)du\right| + \sup_{x\in\mathcal{X}}|f(x)|^2\left(\int |K(u)|du\right)^2 \\
&= o(h_n^{-1}).
\end{aligned}$$

□

引理*1(ii)* 的证明。 首先，由 c_r 不等式和詹森不等式可得

$$\mathbb{E}[Y_{i,h_n}^4(x)] \le \frac{16}{h_n^4\text{Var}[K_{i,h_n}(x)]^2}\int_{\mathcal{X}} K^4\left(\frac{t-x}{h_n}\right)f(t)dt \le \frac{16\sup_{x\in\mathcal{X}}|f(x)|}{h_n^3\text{Var}[K_{i,h_n}(x)]^2}\int_{\mathbb{R}} K^4(u)du.$$

由于 $\sup_{u\in\mathbb{R}}|K(u)| < \infty$，根据三角不等式可得

$$|Y_{i,h_n}(x)| = \frac{|K_{i,h_n}(x) - \mathbb{E}[K_{i,h_n}(x)]|}{\sqrt{\text{Var}[K_{i,h_n}(x)]}} \le \frac{2\sup_{u\in\mathbb{R}}|K(u)|}{h_n\sqrt{\text{Var}[K_{i,h_n}(x)]}}.$$

因此，根据 ψ_1-范数的定义可得

$$\|Y_{i,h_n}(x)\|_{\psi_1} \le \frac{1}{\log 2}\cdot\frac{2\sup_{u\in\mathbb{R}}|K(u)|}{h_n\sqrt{\text{Var}[K_{i,h_n}(x)]}}.$$

□

在证明引理1(iii) 之前，我们先给出一些补充的辅助结果。这些引理的证明见第 4.2.1 节至4.2.3节。

引理 2。 在假设 2成立时，有

$$\frac{\partial}{\partial x}\mathbb{E}[\hat{f}_{h_n}(x)] = \mathbb{E}\left[\frac{\partial}{\partial x}\hat{f}_{h_n}(x)\right], \quad \frac{\partial}{\partial x}\mathbb{E}\left[\frac{\partial}{\partial x}\hat{f}_{h_n}(x)\right] = \mathbb{E}\left[\frac{\partial^2}{\partial^2 x}\hat{f}_{h_n}(x)\right].$$

引理3。 在假设2下，

(i) 上确界$_{x\in\mathcal{X}}\left|\frac{\hat{\sigma}_n(x)}{\sigma_n(x)}-1\right| = o_p(1)$. (ii)
上确界$_{x\in\mathcal{X}}\left|\frac{\partial}{\partial x}\hat{\sigma}_n(x) - \frac{\partial}{\partial x}\sigma_n(x)\right| = o_p(1)$.

引理 4. 在假设2下,

(i) 对任意 $\varepsilon \in (0,1)$，均有

$$\mathbb{P}\left(\sup_{x\in\mathcal{X}}\left|\hat{f}_{h_n}(x) - \mathbb{E}\left[\hat{f}_{h_n}(x)\right]\right| > A_0(\varepsilon)\right) \le \varepsilon,$$

(ii) 对任意 $\varepsilon \in (0,1)$，均有

$$\mathbb{P}\left(\sup_{x\in\mathcal{X}}\left|\frac{\partial}{\partial x}\hat{f}_{h_n}(x) - \mathbb{E}\left[\frac{\partial}{\partial x}\hat{f}_{h_n}(x)\right]\right| > A_1(\varepsilon)\right) \le \varepsilon.$$

引理 *1(iii)* 的证明。 由中值定理可知

$$\sup_{\{(x,y):|x-y|\le\Delta_n} \frac{|\hat{T}_n(x) - \hat{T}_n(y)|}{|x-y|} \le \sup_{x\in\mathcal{X}}\left|\frac{\partial}{\partial x}\hat{T}_n(x)\right|.$$

这意味着

$$\mathbb{P}\left(\sup_{\{(x,y):|x-y|\le\Delta_n} \frac{|\hat{T}_n(x) - \hat{T}_n(y)|}{|x-y|} > L_n\right) \le \mathbb{P}\left(\sup_{x\in\mathcal{X}}\left|\frac{\partial}{\partial x}\hat{T}_n(x)\right| > L_n\right).$$

注意到

$$\frac{\partial}{\partial x}\hat{T}_n(x) := \frac{1}{\hat{\sigma}_n(x)}\left(\frac{\partial}{\partial x}\hat{f}_{h_n}(x) - \frac{\partial}{\partial x}\mathbb{E}[\hat{f}_{h_n}(x)]\right) - \frac{\hat{f}_{h_n}(x) - \mathbb{E}[\hat{f}_{h_n}(x)]}{\hat{\sigma}_n^2(x)}\partial_x\hat{\sigma}_n(x).$$

由引理 引理2，我们可以处理

$$\frac{\partial}{\partial x}\hat{T}_n(x) = \frac{1}{\hat{\sigma}_n(x)}\left(\frac{\partial}{\partial x}\hat{f}_{h_n}(x) - \mathbb{E}\left[\frac{\partial}{\partial x}\hat{f}_{h_n}(x)\right]\right) - \frac{\hat{f}_{h_n}(x) - \mathbb{E}[\hat{f}_{h_n}(x)]}{\hat{\sigma}_n^2(x)}\partial_x\hat{\sigma}_n(x).$$

由引理3，可得

$$\begin{aligned}
&\sup_{x\in\mathcal{X}}\left|\frac{\partial}{\partial x}\hat{T}_n(x)\right| \\
&\le \{1+o_p(1)\}\sup_{x\in\mathcal{X}}\frac{1}{\sigma_n(x)}\left|\frac{\partial}{\partial x}\hat{f}_{h_n}(x) - \mathbb{E}\left[\frac{\partial}{\partial x}\hat{f}_{h_n}(x)\right]\right| \\
&\qquad + \{1+o_p(1)\}\sup_{x\in\mathcal{X}}\frac{\partial\sigma_n(x)}{\partial x}\frac{1}{\sigma_n^2(x)}\left|\hat{f}_{h_n}(x) - \mathbb{E}\left[\hat{f}_{h_n}(x)\right]\right| \\
&\le 2\left(\sup_{x\in\mathcal{X}}\frac{1}{\sigma_n(x)}\left|\frac{\partial}{\partial x}\hat{f}_{h_n}(x) - \mathbb{E}\left[\frac{\partial}{\partial x}\hat{f}_{h_n}(x)\right]\right| + \sup_{x\in\mathcal{X}}\frac{\partial\sigma_n(x)}{\partial x}\frac{1}{\sigma_n^2(x)}\left|\hat{f}_{h_n}(x) - \mathbb{E}\left[\hat{f}_{h_n}(x)\right]\right|\right),
\end{aligned}$$

以趋近于 1 的概率。取

$$L_n = 2\left(\sup_{x\in\mathcal{X}}\frac{1}{\sigma_n(x)}A_1\left(\frac{\varepsilon}{2}\right)+\sup_{x\in\mathcal{X}}\frac{\partial\sigma_n(x)}{\partial x}\frac{1}{\sigma_n^2(x)}A_0\left(\frac{\varepsilon}{2}\right)\right),$$

于是有

$$\left\{\sup_{x\in\mathcal{X}}\left|\frac{\partial}{\partial x}\hat{f}_{h_n}(x)-\mathbb{E}\left[\frac{\partial}{\partial x}\hat{f}_{h_n}(x)\right]\right|\le A_1\left(\frac{\varepsilon}{2}\right),\right.$$
$$\left.\sup_{x\in\mathcal{X}}\left|\hat{f}_{h_n}(x)-\mathbb{E}\left[\hat{f}_{h_n}(x)\right]\right|\le A_0\left(\frac{\varepsilon}{2}\right)\right\}\implies\sup_{x\in\mathcal{X}}\left|\frac{\partial}{\partial x}\hat{T}_n(x)\right|\le L_n,$$

以趋近于 1 的概率。因此，并集界和引理 4 给出

$$\begin{aligned}&\mathbb{P}\left(\sup_{x\in\mathcal{X}}\left|\frac{\partial}{\partial x}\hat{T}_n(x)\right|>L_n\right)\\&\le\mathbb{P}\left(\sup_{x\in\mathcal{X}}\left|\frac{\partial}{\partial x}\hat{f}_{h_n}(x)-\mathbb{E}\left[\frac{\partial}{\partial x}\hat{f}_{h_n}(x)\right]\right|\le A_1\left(\frac{\varepsilon}{2}\right)\right)\\&\quad+\mathbb{P}\left(\sup_{x\in\mathcal{X}}\left|\hat{f}_{h_n}(x)-\mathbb{E}\left[\hat{f}_{h_n}(x)\right]\right|\le A_0\left(\frac{\varepsilon}{2}\right)\right)+o(1)\le\varepsilon+o(1).\end{aligned}$$

这意味着我们可以取

$$L_n = 2\left(\sup_{x\in\mathcal{X}}\frac{1}{\sigma_n(x)}A_1\left(\frac{\varepsilon}{2}\right)+\sup_{x\in\mathcal{X}}\frac{\partial\sigma_n(x)}{\partial x}\frac{1}{\sigma_n^2(x)}A_0\left(\frac{\varepsilon}{2}\right)\right),$$

使得

$$\mathbb{P}\left(\sup_{\{(x,y):|x-y|\le\Delta_n}\frac{|\hat{T}_n(x)-\hat{T}_n(y)|}{|x-y|}>L_n\right)\le\varepsilon+o(1).$$

□

4.2.1 引理 2 的证明

引理2（*i*）。注意到

$$\mathbb{E}[\hat{f}_{h_n}(x)]=\int\frac{1}{h_n}K\left(\frac{t-x}{h_n}\right)f(t)dt,$$

被积函数的导数为

$$\frac{\partial}{\partial x}\frac{1}{h_n}K\left(\frac{t-x}{h_n}\right)f(t)=-\frac{1}{h_n^2}K'\left(\frac{t-x}{h_n}\right)f(t).$$

那么，在假设2下，有

$$\int\left|\frac{\partial}{\partial x}\frac{1}{h_n}K\left(\frac{t-x}{h_n}\right)f(t)\right|dt\le\frac{\sup_{x\in\mathcal{X}}|f(x)|}{h_n}\int|K'(u)|\,du<\infty.$$

因此，控制收敛定理完成了证明。 □

引理证明2（*ii*）。与引理 2（i）的证明类似，

$$\mathbb{E}\left[\frac{\partial}{\partial x}\hat{f}_{h_n}(x)\right] = \int -\frac{1}{h_n^2}K'\left(\frac{t-x}{h_n}\right)f(t),$$

被积函数的导数为

$$\frac{\partial}{\partial x}\left\{-\frac{1}{h_n^2}K'\left(\frac{t-x}{h_n}\right)f(t)\right\} = \frac{2}{h_n^3}K''\left(\frac{t-x}{h_n}\right)f(t).$$

那么，在假设2下，有

$$\int\left|\frac{\partial}{\partial x}\left\{-\frac{1}{h_n^2}K'\left(\frac{t-x}{h_n}\right)f(t)\right\}\right|dt \leq \frac{\sup_{x\in\mathcal{X}}|f(x)|}{h_n^2}\int|K''(u)|\,du < \infty.$$

因此，控制收敛定理完成了证明。 □

4.2.2 引理3的证明3

引理3的证明。与 Imai 等人（2025）补充材料中引理 6 和 7 的证明类似，在假设2下，我们可以利用 van der Vaart 和 Wellner（1996）的引理 2.6.15、2.6.16 和 2.6.18，证明 $\sup_{x\in\mathcal{X}}\left|\frac{\hat{\sigma}_n(x)}{\sigma_n(x)} - 1\right|$ 和 $\sup_{x\in\mathcal{X}}\left|\frac{\partial}{\partial x}\hat{\sigma}_n(x) - \frac{\partial}{\partial x}\sigma_n(x)\right|$ 是由 VC 型类索引的经验过程的上确界（关于 VC 型类的定义，见 Chernozhukov 等人（2014b）的定义 2.1）。随后，我们可以利用 Chernozhukov 等人（2014b）的推论 5.1证明这些结论成立。

□

4.2.3 引理4

首先证明引理4(ii)，然后证明引理4(i)。
引理*4(ii)* 的证明。定义 $D_n(x) := \frac{\partial}{\partial x}\hat{f}_{h_n}(x) - \mathbb{E}\left[\frac{\partial}{\partial x}\hat{f}_{h_n}(x)\right]$。然后

$$D_n(x) = \frac{1}{n}\sum_{i=1}^{n}\xi_i(x), \text{ with } \xi_i(x) := -\frac{1}{h_n^2}K'_{i,h_n}(x) - \mathbb{E}\left[-\frac{1}{h_n^2}K'_{i,h_n}(x)\right].$$

对于每个 $x \in \mathcal{X}$，我们可以看到 $|\xi_i(x)| \leq 2h_n^{-2}\sup_{u\in\mathbb{R}}|K'(u)| =: M_1$ 且

$$\begin{aligned}\text{Var}[\xi_i(x)] &\leq \mathbb{E}\left[\left\{\frac{1}{h_n^2}K'_{i,h_n}(x)\right\}^2\right] \\ &= \frac{1}{h_n^4}\int(K')^2\left(\frac{t-x}{h_n}\right)f(t)dt \leq \frac{\sup_{x\in\mathcal{X}}|f(x)|}{h_n^3}\int_{\mathbb{R}}\{K'(u)\}^2du =: v_1.\end{aligned}$$

那么，对于任意 $t > 0$，由伯恩斯坦不等式可得，

$$\mathbb{P}\left(|D_n(x)| > \sqrt{\frac{2v_1t}{n}} + \frac{M_1t}{3n}\right) \leq 2\exp(-t). \tag{12}$$

定义 $\eta_n := n^{-1/2}h_n^{3/2}$，以及 $\mathcal{X}$ 上的网格点 $\{x_k\}_{k=1}^{m_{1,n}}$，如下所示:

$$x_k := \underline{x} + (k-1)\eta_n,\ \ k = 1, \ldots, m_{1,n} - 1\ \ x_{m_{1,n}} := \overline{x},\ \ m_{1,n} := \left\lfloor \frac{|\overline{x} - \underline{x}|}{\eta_n} \right\rfloor + 2.$$

因此，由并集界和公式（12）可得

$$\mathbb{P}\left(\max_{1 \le k \le m_{1,n}} |D_n(x_k)| > \sqrt{\frac{2v_1 t}{n}} + \frac{M_1 t}{3n} \right) \le 2m_{1,n} \exp(-t).$$

取 $t_n(\varepsilon) := \log(2m_{1,n}/\varepsilon)$，则由于 $2m_{1,n}\exp(-t_n(\varepsilon)) = \varepsilon$,

$$\mathbb{P}\left(\max_{1 \le k \le m_{1,n}} |D_n(x_k)| > \sqrt{\frac{2v_1 t_n(\varepsilon)}{n}} + \frac{M_1 t_n(\varepsilon)}{3n} \right) \le \varepsilon. \tag{13}$$

对于任意 $x \in \mathcal{X}$，存在 $k \in \{1, \ldots, m_{1,n}\}$ 使得 $|x - x_k| \le \eta_n/2 = n^{-1/2}h_n^{3/2}/2$。利用引理引理 2，可知 $|\partial D_n(x)/\partial x| \le 2h_n^{-3}$上确界$_{u\in\mathbb{R}} |K''(u)|$。因此，由中值定理可得

$$\begin{aligned} &|D_n(x) - D_n(x_k)| \\ &\le |x - x_k| \sup_{x \in \mathcal{X}} \left| \frac{\partial}{\partial x} D_n(x) \right| \le \frac{h_n^{3/2}}{2n^{1/2}} \cdot \frac{2\sup_{u\in\mathbb{R}} |K''(u)|}{h_n^3} = \frac{\sup_{u\in\mathbb{R}} |K''(u)|}{n^{1/2}h_n^{3/2}}. \end{aligned}$$

因此

$$\sup_{x \in \mathcal{X}} |D_n(x)| \le \max_{1 \le k \le m_{1,n}} |D_n(x_k)| + \frac{\sup_{u\in\mathbb{R}} |K''(u)|}{n^{1/2}h_n^{3/2}}. \tag{14}$$

由公式（13）和公式（14），可得

$$\mathbb{P}\left(\sup_{x \in \mathcal{X}} |D_n(x)| > \sqrt{\frac{2v_1 \log(2m_{1,n}/\varepsilon)}{n}} + \frac{M_1 \log(2m_{1,n}/\varepsilon)}{3n} + \frac{\sup_{u\in\mathbb{R}} |K''(u)|}{n^{1/2}h_n^{3/2}} \right) \le \varepsilon.$$

□

引理 4(i) 的证明。将 K'' 替换为 K'、将 K' 替换为 $\hat{}\hat{}K$，并相应调整缩放后，证明与 (ii) 部分完全相同。具体而言，定义 $D_{0,n}(x) := f_{h_n}(x) - \mathbb{E}[f_{h_n}(x)] = n^{-1}\sum_{i=1}^n \xi_{0,i}(x)$，其中 $\xi_{0,i}(x) := K_{i,h_n}(x) - \mathbb{E}[K_{i,h_n}(x)]$。于是$|\xi_{0,i}(x)| \le M_0$ 且 $\mathrm{Var}[\xi_{0,i}(x)] \le v_0$。在辅助网格上应用伯恩斯坦不等式，网格尺寸为 η_n,$0 := n^{-1/2}h_n^{3/2}$，并利用中值定理以及上界 $\sup_{x\in\mathcal{X}} |\partial D_n$,$0(x)/\partial x| \le 2h_n^{-2}\sup_{u\in\mathbb{R}} |K'(u)|$，即可得到所要结论。

□

参考文献

Bickel, P. J. 和 Rosenblatt, M.（1973年）。关于密度函数估计量偏差的若干整体度量。《统计年鉴》，第 1071–1095 页。2

Chen, X. 和 Kato, K.（2020年）。刀切乘子自助法：对U-过程上确界的有限样本近似及其应用。《概率论及相关领域》，176：1097–1163。2

Chernozhukov, V.、Chetverikov, D. 和 Kato, K.（2014a）。反集中与诚实的自适应置信带。《统计年鉴》，42(5)：1787。2

Chernozhukov, V.、Chetverikov, D. 和 Kato, K.（2014b）。经验过程上确界的高斯近似。《统计年鉴》，42(4)：1564–1597。2，20

Chernozhukov, V.、Chetverikov, D. 和 Kato, K.（2016年）。复杂度不断增加的经验过程上确界的经验自助法和乘子自助法，以及相关的高斯耦合。随机过程及其应用，126(12)：3632–3651。2

Chernozhukov, V., Chetverikov, D., Kato, K., 和 Koike, Y.（2022年）。高维下改进的中心极限定理与自助法近似。统计学年刊，50(5)：2562–2586。3，5，6，7，13，14

Chernozhukov, V., Chetverikov, D., Kato, K., 和 Koike, Y.（2023a）。高维数据自助法。统计学及其应用年度评论，10(1)：427–449。3

Chernozhukov, V., Chetverikov, D., 和 Koike, Y.（2023b）。高维下近乎最优的中心极限定理与自助法近似。应用概率年刊，33(3)：2374–2425。6

Fan, Q., Hsu, Y.-C., Lieli, R. P., 和 Zhang, Y.（2022年）。利用高维数据估计条件平均处理效应。商业与经济统计学杂志，40(1)：313–327。2

Imai, S.、Qin, L. 和 Yanagi, T.（2025年）。双重差分中组-时间条件平均处理效应的双重稳健一致置信带。《商业与经济统计学杂志》，即将出版。2，20

Kato, K. 和 Sasaki, Y.（2018年）。未知误差分布下反卷积的一致置信带。《计量经济学杂志》，207(1)：129–161。2，16

Le Cam, L.（1986）。统计决策理论中的渐近方法。*Springer* 统计学丛书。13，15

Lee, S., Okui, R. 和 Whang, Y.-J.（2017）。条件平均处理效应函数的双重稳健一致置信带。《应用计量经济学杂志》，32(7):1207–1225。2Li, Q. 和 Racine, J. S.（2007）。非参数计量经济学：理论与实践。普林斯顿大学出版社。9Tsybakov, A. B.（2009）。非参数估计导论。纽约施普林格出版社。9van der Vaart, A. 和 Wellner, J. A.（1996）。弱收敛与经验过程：统计学应用。施普林格。20Wasserman, L.（2006）。非参数统计学大全。施普林格科学与商业媒体公司。9