arXiv:2603.03763v1 [math.ST] 4 Mar 2026

关于多指标模型的大带宽矩阵值核平滑估计量

Taku MORIYAMA 横滨市立大学数据科学学院

摘要

一般而言，采用较大带宽值的核平滑会导致过度平滑或欠拟合。然而，当包含无关变量时，相应的较大带宽值已知具有将这些变量收缩的作用。本研究针对多指标模型情形，考察了具有较大带宽矩阵元素的核条件密度估计量与回归估计量的渐近性质。研究表明，在不剔除无关变量的情况下，这些估计量的最优收敛速率并不取决于变量个数，而取决于有效维数。因此，核条件密度估计量与回归估计量被证明天然具备减轻维数灾难的能力。通过数值研究考察了有限样本表现，并讨论了带宽选择。最后，给出了关于波士顿房价数据的案例研究。

关键词：条件密度估计；核密度估计量；核平滑；Nadaraya-Watson 回归；均方误差

1 引言

对于随机序列 $X_1, \cdots, X_n$，点 x 处的核密度估计量由 $(nh)^{-1}\sum_{i=1}^{n} k(h^{-1}(x-X_i))$ 给出，其中 k 是核函数。k 需要是概率密度函数，而 h 对应于尺度参数。核密度估计量的渐近性质已得到充分研究，并且其对广泛分布类别的相合性已获认可。

Jones（1993）综述了带有大带宽的核密度估计量，即当 $h \to \infty$ 为 $n \to \infty$ 时的情形。这种情况称为过度平滑或欠拟合。研究证明，此时核密度估计量不再是非参数估计量，而可能是针对选作核函数的概率密度的相合估计量。文中还指出，如果底层分布和核函数均为正态分布，则方差校正核密度估计量收敛于正态分布。本研究将 Jones（1993）的研究扩展到核密度估计中的多变量情形，并探讨这些结果在核条件密度估计或回归估计中的应用。

多元核估计量受到维数灾难的影响（即维度越高，最优收敛速率越慢），已有多项研究致力于解决回归估计中的这一问题。如果存在与因变量无关的变量，通过（完全）剔除这些无关变量，估计量的收敛速率便取决于相关变量的数量。已有一些方法用于核回归估计。Lafferty 和 Wasserman（2008）提出了一种名为 RODEO 的有效算法。White 等 （2017）提出了 MEKRO，该方法通过考虑 $h = \infty$，在解释变量带宽的调和平均数约束下使测量误差最小化。Feng等 （2018）提出了一种称为非参数独立性筛选的方法，并研究了该方法在超高维情形下的性质。

Hall 等（2004）和 Hall 等（2007）分别考察了采用对角带宽矩阵的交叉验证核条件密度估计器和回归估计量的收缩性质。通过使与无关变量对应的带宽趋于无穷大，无需消除无关变量即可达到最优收敛速率。因此，不同于 RODEO 或 MEKRO 核

条件密度估计量和回归估计量无需额外的超参数（用于设定阈值或约束）即可达到最优收敛速率。这也意味着，未完全消除无关变量的非参数估计量不会受到变量设定错误（即模型设定错误）的影响。

本研究首先考虑条件密度估计和回归估计中全部或部分解释变量无关的情形。已有研究证明，条件密度的非参数级数估计量可达到最优极小极大收敛速率（Efromovich 2010）。我们将条件独立的情形扩展到部分线性相关的情形。Xia（2008）证明，在该情形下，采用标量矩阵的局部线性回归估计量具有最优收敛速率，并提出了相应的有效计算算法。

Conn and Li（2019）证明，当真实模型为多指标模型时，核回归估计量具有收缩性质（例如，Klock 等2021）。研究得出了带宽可能发散时核估计量渐近均方误差的收敛速度，结果表明其具有收缩性质。研究还表明，收缩性质所需的条件不同于带宽 h 假设收敛到零时的条件。研究证明，采用任意形式的正定带宽矩阵的核估计量均可达到最优收敛速率。

需要指出的是，本研究在不对底层分布作结构假设的情况下，考察具有大带宽矩阵的核平滑估计量的渐近性质。第2节从本研究的基础入手，考察多元核密度估计量的渐近性质。本研究进一步推广到所有解释变量均独立于结果变量时的条件密度或回归的核估计量。第3节和第4节分别研究部分解释变量独立的情形，以及相依关系遵循多指标模型的情形。一个重要结论是，在多指标模型情形下，最优带宽矩阵并非对角的。第5节给出了模拟研究以及对波士顿房价数据集的案例研究。第6节对本项研究作出结论。

2 回归或条件密度估计中的独立性情形

2.1 具有大带宽矩阵的密度估计量

设 $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ 的概率密度函数为 f。令 $\widehat{f}$ 为如下给出的多元核密度估计量

$$\widehat{f}(\boldsymbol{x}) := (n|\boldsymbol{H}|)^{-1} \sum_{i=1}^{n} k(\boldsymbol{H}^{-1}(\boldsymbol{x} - \boldsymbol{X}_i)),$$

其中 $k\colon\ \mathbb{R}^d \to \mathbb{R}$ 是核函数，尺度 $\boldsymbol{H}$ 是正则的 $d \times d$ 带宽矩阵。 $|\boldsymbol{H}|$ 表示矩阵 $\boldsymbol{H}$ 的行列式， $\|\boldsymbol{H}\|$ 表示绝对值。我们不假设 $\int k(\boldsymbol{z})\mathrm{d}\boldsymbol{z} = 1$，后文会发现这一假设并无必要。

注 1. 如果核 k 是球对称的，即存在 $\ell\colon\ \mathbb{R}^d \to \mathbb{R}$，使得 $k(\boldsymbol{w}) = \ell(\|\boldsymbol{w}\|_2)$，那么 $\widetilde{\ }$ $\widetilde{\ }$ $\widetilde{\ }$ 对于任意 $\boldsymbol{H}$ 和 $\boldsymbol{w}$，都存在一个对称矩阵 $\boldsymbol{H}$，使得 $k(\boldsymbol{H}^{-1}\boldsymbol{w}) = k(\boldsymbol{H}^{-1}\boldsymbol{w})$，其中 $\boldsymbol{H}$ 由极分解得到。

由 $k(\boldsymbol{w}) = \prod_{j=1}^{d} s(w_j^2)$ 给出的乘积核，其中 $s\colon\ \mathbb{R} \to \mathbb{R}$，是球对称的。单变量高斯核和埃潘尼奇尼科夫核满足该假设。

已知核密度估计量的所有最优带宽矩阵元素具有相同的阶（Wand and Jones 1995），这意味着存在一个常数矩阵 $\boldsymbol{C}$，使得 $\boldsymbol{H} = h\boldsymbol{C}$，即 $(\boldsymbol{H})_{ij} = O(h)$。

在本研究中， $\boldsymbol{H}$ 不一定是对称的，并且允许（即 $(\boldsymbol{H})_{ij} < 0$ （$\boldsymbol{H}$）。本研究允许带宽矩阵 $\boldsymbol{H}$ 中的元素发散，或者更具体地说，其绝对值趋于 $\widehat{\ }$ 无穷大，当 $n \to \infty$ 时。由于密度估计量 f 无法与真实密度 f 保持一致，除非

$\boldsymbol{H} \to \boldsymbol{O}$，考虑带宽矩阵 $\boldsymbol{H}$ s.t. $\boldsymbol{H} \nrightarrow \boldsymbol{O}$ 通常在概率密度估计中没有意义，如下所示。

引理1。 假设(i) $h \to \infty$ 当 $n \to \infty$ 时，(ii) $k(\mathbf{0})$ 存在，(iii) k 在原点附近二阶连续可微，(iv) $\nabla(k)(\mathbf{0}) = \mathbf{0}$，且(v) 2 f 阶矩存在。那么，

$$\|\boldsymbol{H}\|\widehat{f}(\boldsymbol{x}) - k(\mathbf{0}) = O(h^{-2}) + O_P(n^{-1/2}h^{-1}).$$

证明。 对于大带宽矩阵 $\boldsymbol{H}$，有如下展开式

$$\begin{aligned}\mathbb{E}[\widehat{f}(\boldsymbol{x})] =&\|\boldsymbol{H}\|^{-1}\mathbb{E}\left[k\left(\boldsymbol{H}^{-1}(\boldsymbol{x}-\boldsymbol{X}_i)\right)\right] = \|\boldsymbol{H}\|^{-1}\int k\left(\boldsymbol{H}^{-1}(\boldsymbol{x}-\boldsymbol{w})\right)f(\boldsymbol{w})\mathrm{d}\boldsymbol{w}\\ =&\|\boldsymbol{H}\|^{-1}\int [k(\mathbf{0}) + \{\boldsymbol{H}^{-1}(\boldsymbol{x}-\boldsymbol{w})\}'\{\nabla(k)(\mathbf{0})\} + O(h^{-2})]f(\boldsymbol{w})\mathrm{d}\boldsymbol{w}\\ \sim&\|\boldsymbol{H}\|^{-1}[k(\mathbf{0}) + (\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{H}^{-1}\{\nabla(k)(\mathbf{0})\} + O(h^{-2})];\end{aligned}$$

因此，

$$\mathbb{E}[\widehat{f}(\boldsymbol{x})] - \|\boldsymbol{H}\|^{-1}k(\mathbf{0}) = O(\|\boldsymbol{H}\|^{-1}h^{-2}).$$

由于

$$\mathbb{E}\left[\|\boldsymbol{H}\|^{-2}k^2\left(\boldsymbol{H}^{-1}(\boldsymbol{x}-\boldsymbol{X}_i)\right)\right] = \|\boldsymbol{H}\|^{-2}\int [k^2(\mathbf{0}) + 2k(\mathbf{0})\{\boldsymbol{H}^{-1}(\boldsymbol{x}-\boldsymbol{w})\}'\{\nabla(k)(\mathbf{0})\} + O(h^{-2})]f(\boldsymbol{w})\mathrm{d}\boldsymbol{w},$$

$$\begin{aligned}\mathbb{E}[\widehat{f}^2(\boldsymbol{x})] - (\mathbb{E}[\widehat{f}(\boldsymbol{x})])^2 =&\frac{1}{n}\|\boldsymbol{H}\|^{-2}\int [k^2(\mathbf{0}) + 2k(\mathbf{0})\{\boldsymbol{H}^{-1}(\boldsymbol{x}-\boldsymbol{w})\}'\{\nabla(k)(\mathbf{0})\} + O(h^{-2})]f(\boldsymbol{w})\mathrm{d}\boldsymbol{w}\\ &-\frac{1}{n}\|\boldsymbol{H}\|^{-2}\{\int [k(\mathbf{0}) + \{\boldsymbol{H}^{-1}(\boldsymbol{x}-\boldsymbol{w})\}'\{\nabla(k)(\mathbf{0})\} + O(h^{-2})]f(\boldsymbol{w})\mathrm{d}\boldsymbol{w}\}^2\\ \sim&\frac{1}{n}\|\boldsymbol{H}\|^{-2}\{\nabla(k)(\mathbf{0})\}\boldsymbol{H}^{-1}\mathbb{V}[\boldsymbol{X}]\boldsymbol{H}^{-1}\{\nabla(k)(\mathbf{0})\}'.\end{aligned}$$

引理得证。 □

注2。 将该假设与引理*1*中通常的小带宽矩阵情形下的假设进行比较，可以看出，此处的假设针对的是核函数，而非底层分布。不过，这并不构成限制，常用的核函数（例如高斯核或 *Epanechnikov* 型核）均满足该假设。关于矩的假设与底层分布的重尾性有关。

注释3。 在假设核 k 为概率密度函数的条件下，具有如下大带宽矩阵的核密度估计量的期望

$$\int \|\boldsymbol{H}\|^{-1}k\left(\boldsymbol{H}^{-1}(\boldsymbol{x}-\boldsymbol{w})\right)f(\boldsymbol{w})\mathrm{d}\boldsymbol{w}$$

可视为卷积。因此，它也是一个概率密度函数，对应于随机变量 $\boldsymbol{X} + \boldsymbol{H}\boldsymbol{K}$。 $\boldsymbol{K}$ 是与 $\boldsymbol{X}$相互独立的随机变量，其概率密度函数为 k。由于缩放（放大） $\boldsymbol{H}$，当 n 足够大时，该随机变量由 $\boldsymbol{H}\boldsymbol{K}$主导，而 $\boldsymbol{H}\boldsymbol{K}$在 $\mathbb{R}^d$ 上渐近均匀分布

在继续讨论条件密度或回归的核估计量之前，我们先说明以下几点。

注4. 除核密度估计量外，后文引入的条件密度或回归的核估计量也是渐近正态的，这可由*Slutsky*定理证明。

注5. 引理*1*意味着 $\widehat{f}(\boldsymbol{x}) \xrightarrow{p} f(\boldsymbol{x})$，当 $f(\boldsymbol{x}) \neq k(\mathbf{0})$时。然而，只要 $f(\boldsymbol{x}) - k(\mathbf{0}) = o(1)$，相合性仍然成立。例子包括分布污染这一情形，这在随后关于条件密度或回归估计的讨论中十分重要。

注 6. 从均方误差等逐点误差（即逐点估计）的角度来看，模型不必在全局上趋近于极限模型。这意味着，在这种情况下，相合性要求的不是$f(\boldsymbol{x}) \equiv k(\mathbf{0})$，而只要求 $f(\boldsymbol{x}) = k(\mathbf{0})$。

根据该引理，接下来我们讨论大带宽矩阵下的条件密度核估计量或回归核估计量。

2.2 回归估计中的独立情形

设 d 维独立同分布随机变量由 $\boldsymbol{X}_i := (\boldsymbol{X}'_{1,i}, \boldsymbol{X}'_{2,i})'$ 组成，且 d_2 维随机变量 $\boldsymbol{X}_{2,i}$ 的概率密度函数为 f_2。令 f_2 为如下给出的 d_2 维核密度估计量:

$$\widehat{f}_2(\boldsymbol{x}_2) := (n\|\boldsymbol{H}_{22}\|)^{-1} \sum_{i=1}^{n} k_2(\boldsymbol{H}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{X}_{2,i})),$$

其中 $k_2 : \mathbb{R}^{d_2} \to \mathbb{R}$ 是核函数， $\boldsymbol{H}_{22}$ 是 $d_2 \times d_2$ 带宽矩阵，$(\boldsymbol{H}_{22})_{ij} \in \mathbb{R}$ 和 $(\boldsymbol{H}_{22})_{ij} = O(h_{22})$。核回归估计量定义如下:

$$\widehat{m}(\boldsymbol{x}_2) := \widehat{f}_2(\boldsymbol{x}_2)^{-1}(n\|\boldsymbol{H}_{22}\|)^{-1} \sum_{i=1}^{n} \boldsymbol{X}_{1,i} k_2(\boldsymbol{H}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{X}_{2,i})).$$

那么，以下　得到了关于大带宽矩阵核回归估计量的结果　.

定理1. 假设(i) $h_{22} \to \infty$ 当 $n \to \infty$时，(ii) $k_2(\mathbf{0}) \neq 0$ 存在；(iii) k_2 在原点附近二阶连续可微；(iv) $\nabla(k_2)(\mathbf{0}) = \mathbf{0}$；且(v) 2*nd moments of* f_1 和 f_2 存在。那么，

$$\mathbb{E}[\{\widehat{m}(\boldsymbol{x}_2) - \mathbb{E}[\boldsymbol{X}_{1,i}]\}^2] = O(h_2^{-4} + n^{-1}).$$

证明. $(n\|\boldsymbol{H}_{22}\|)^{-1} \sum_{i=1}^{n} \boldsymbol{X}_{1,i} k_2(\boldsymbol{H}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{X}_{2,i}))$的期望为

$$\begin{aligned}\|\boldsymbol{H}_{22}\|^{-1} \int \boldsymbol{w}_1 k_2\left(\boldsymbol{H}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{w}_2)\right) f(\boldsymbol{w})\mathrm{d}\boldsymbol{w} =& \|\boldsymbol{H}_{22}\|^{-1} \int \boldsymbol{w}_1[k_2(\mathbf{0}) + O(h_{22}^{-2})] f(\boldsymbol{w})\mathrm{d}\boldsymbol{w} \\ \sim& \|\boldsymbol{H}_{22}\|^{-1}\mathbb{E}[\boldsymbol{X}_{1,i}][k_2(\mathbf{0}) + O(h_{22}^{-2})].\end{aligned}$$

方差为

$$\begin{aligned}&\frac{1}{n}\|\boldsymbol{H}_{22}\|^{-2} \int \boldsymbol{w}_1\boldsymbol{w}_1'[k_2^2(\mathbf{0}) + O(h_{22}^{-2})] f(\boldsymbol{w})\mathrm{d}\boldsymbol{w} \\ &\quad - \frac{1}{n}\|\boldsymbol{H}_{22}\|^{-2}\{\int \boldsymbol{w}_1[k_2(\mathbf{0}) + O(h_{22}^{-2})] f(\boldsymbol{w})\mathrm{d}\boldsymbol{w}\}\{\int \boldsymbol{w}_1[k_2(\mathbf{0}) + O(h_{22}^{-2})] f(\boldsymbol{w})\mathrm{d}\boldsymbol{w}\}' \\ \sim& \frac{1}{n}\|\boldsymbol{H}_{22}\|^{-2} k_2^2(\mathbf{0})\mathbb{V}[\boldsymbol{X}_{1,i}].\end{aligned}$$

那么，我们有

$$\mathbb{E}[\{\|\boldsymbol{H}_{22}\|^{-1}\mathbb{E}[\boldsymbol{X}_{1,i}]k_2(\mathbf{0}) - (n\|\boldsymbol{H}_{22}\|)^{-1} \sum_{i=1}^{n} \boldsymbol{X}_{1,i} k_2(\boldsymbol{H}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{X}_{2,i}))\}^2] = O(\|\boldsymbol{H}_{22}\|^{-2}(h_{22}^{-4} + n^{-1})).$$

将 H¨older 不等式应用于渐近均方误差（AMSE），并结合引理1关于 $f_2(\boldsymbol{x}_2)$ 的结果，可得结论。 □

推论1。 假设定理*1*的条件成立。如果 $h_{22}^{-4} = O(n^{-1})$ 或 $o(n^{-1})$，则采用最优带宽时的均方误差为 n^{-1} 阶。

推论1表明，核回归估计量收敛于无条件期望 $\mathbb{E}[\boldsymbol{X}_{1,i}]$，其收敛速率为 $\sqrt{n}$。核估计量称为局部常数估计量；在通常的带宽条件（$h_{22} \to 0$）下，它已知是无偏的，因此只要带宽矩阵足够大（严格来说，想象为 $h_{22} = O(1)$），即可达到 $\sqrt{n}$-相合性。

从证明中可以看出，分子和分母的核函数与带宽矩阵必须相同，这与使用较小（即通常的）带宽矩阵时不同。 f_2

2.3 条件密度估计中的独立情形

Set $\widehat{f}$ 是核　　密度函数，定义在 $\boldsymbol{X}_i$ 上，核函数为 $k : \mathbb{R}^d \to \mathbb{R}$，带宽矩　　x

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_{11} & \boldsymbol{H}_{12} \\ \boldsymbol{H}_{21} & \boldsymbol{H}_{22} \end{pmatrix},$$

其中 $\boldsymbol{H}_{k\ell}$ 是 $d_k \times d_\ell$ 矩阵，且（$\boldsymbol{H}_{k\ell}$）$_{ij} \in \mathbb{R}$。假设每个块中的所有元素都具有相同的阶，更具体地说，（$\boldsymbol{H}_{k\ell}$）$_{ij} = O$（$h_{k\ell}$）。于是，我们得到以下结果。

引理2. 假设(i) $\boldsymbol{H}_{11}$ 是正则的， $h_{11} \to 0$ 且 $h_{22} \to \infty$ 满足 $n \to \infty$，(ii)

$$\boldsymbol{H}_{12} = \boldsymbol{O} \quad or \quad h_{12} = O(h_{11})$$

and

$$\boldsymbol{H}_{21} = \boldsymbol{O} \quad or \quad h_{21} = O(h_{11}),$$

(iii) $k_2(\mathbf{0})$ 存在，其中

$$k_2(\boldsymbol{z}_2) = \int k(\boldsymbol{z}_1, \boldsymbol{z}_2)\mathrm{d}\boldsymbol{z}_1,$$

(iv) k 在 $\{(\boldsymbol{z}_1,\mathbf{0})|\boldsymbol{z}_1 \in \mathbb{R}^{d_1}\}$上二阶连续可微，(v) $\nabla(k)(\boldsymbol{z}_1,\mathbf{0}) = \mathbf{0}$ 对任意 $\boldsymbol{z}_1$均成立，(vi)$\int \boldsymbol{z}_1 k(\boldsymbol{z}_1,\mathbf{0})\mathrm{d}\boldsymbol{z}_1 = \mathbf{0}$，(vii) f 关于变量 $\boldsymbol{x}_1$ 偏连续可微，对任意 $\boldsymbol{x}_2$均成立。于是，

$$\|\boldsymbol{H}_{22}\|\widehat{f}(\boldsymbol{x}) - k_2(\mathbf{0})f_1(\boldsymbol{x}_1) = O_P(n^{-1/2}\|\boldsymbol{H}_{11}\|^{-1/2}) + O(h_{11}^2 + h_{22}^{-2}).$$

其中 f_1 是$\boldsymbol{X}_{1,i}$的概率密度函数。

证明. 在本证明中，设 $\boldsymbol{H}_{12} \neq \boldsymbol{O}$ 和 $\boldsymbol{H}_{21} \neq \boldsymbol{O}$。 $\boldsymbol{H}$ 的正则性并不意味着 $\boldsymbol{H}_{11}$ 可逆；不过，为便于记号，令 $\boldsymbol{\Xi}^{-1} := \boldsymbol{H}_{11}^{-1} + \boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12}\boldsymbol{\Upsilon}^{-1}\boldsymbol{H}_{21}\boldsymbol{H}_{11}^{-1}$且 $\boldsymbol{\Upsilon} := \boldsymbol{H}_{22} - \boldsymbol{H}_{21}\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12}$。于是有

$$\boldsymbol{H}^{-1} = \begin{pmatrix} \boldsymbol{\Xi}^{-1} & -\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12}\boldsymbol{\Upsilon}^{-1} \\ -\boldsymbol{\Upsilon}^{-1}\boldsymbol{H}_{21}\boldsymbol{H}_{11}^{-1} & \boldsymbol{\Upsilon}^{-1} \end{pmatrix}.$$

对于该分块矩阵的行列式，有

$$|\boldsymbol{H}|^{-1} = |\boldsymbol{H}_{11}|^{-1}|\boldsymbol{H}_{22} - \boldsymbol{H}_{21}\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12}|^{-1} = |\boldsymbol{H}_{11}|^{-1}|\boldsymbol{\Upsilon}|^{-1}$$

由西尔维斯特行列式定理可得

$$\begin{aligned} |\boldsymbol{\Xi}|^{-1} =& |\boldsymbol{\Xi}^{-1}| = |\boldsymbol{H}_{11}|^{-2}|\boldsymbol{H}_{11} + \boldsymbol{H}_{12}\boldsymbol{\Upsilon}^{-1}\boldsymbol{H}_{21}| \\ =& |\boldsymbol{H}_{11}|^{-1}|\boldsymbol{\Upsilon}|^{-1}|\boldsymbol{\Upsilon} + \boldsymbol{H}_{21}\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12}| = |\boldsymbol{H}_{11}|^{-1}|\boldsymbol{\Upsilon}|^{-1}|\boldsymbol{H}_{22}|. \end{aligned}$$

将它们结合起来，我们得到 $\|\boldsymbol{H}\|\|\boldsymbol{\Xi}^{-1}\| = \|\boldsymbol{H}_{22}\|$。

$(\boldsymbol{\Xi}^{-1})_{ij} = O(h_{11}^{-1})$、$(\boldsymbol{\Upsilon}^{-1})_{ij} = O(h_{22}^{-1})$、$(-\boldsymbol{H}_{11}^{-1}\boldsymbol{H}_{12}\boldsymbol{\Upsilon}^{-1})_{ij} = O(h_{22}^{-1})$、$(-\boldsymbol{\Upsilon}^{-1}\boldsymbol{H}_{21}\boldsymbol{H}_{11}^{-1}\boldsymbol{\Xi})_{ij} = O(h_{11}h_{22}^{-1})$，即 $\boldsymbol{H}^{-1}$ 中元素的顺序形式为

$$\begin{pmatrix} h_{11}^{-1} & h_{22}^{-1} \\ h_{22}^{-1} & h_{22}^{-1} \end{pmatrix}.$$

通过变量替换 $\boldsymbol{\Xi}^{-1}(\boldsymbol{x}_1 - \boldsymbol{w}_1) = \boldsymbol{z}_1$ 和 $(\boldsymbol{x}_2 - \boldsymbol{w}_2) = \boldsymbol{z}_2$, $\mathbb{E}[\widehat{f}(\boldsymbol{x})] = \|\boldsymbol{H}\|^{-1}\int k\left(\boldsymbol{H}^{-1}(\boldsymbol{x}-\boldsymbol{w})\right) f(\boldsymbol{w})\mathrm{d}\boldsymbol{w}$可表示为

$$\begin{aligned}
\mathbb{E}[\widehat{f}(\boldsymbol{x})] =& \|\boldsymbol{H}\|^{-1}\|\boldsymbol{\Xi}\| \int [k(\boldsymbol{z}_1, \boldsymbol{0}) + \{(-\boldsymbol{z}_2'(\boldsymbol{\Upsilon}^{-1})'\boldsymbol{H}_{12}'(\boldsymbol{H}_{11}^{-1})', \\
& - \boldsymbol{z}_1'\boldsymbol{\Xi}'(\boldsymbol{H}_{11}^{-1})'\boldsymbol{H}_{21}'(\boldsymbol{\Upsilon}^{-1})' + \boldsymbol{z}_2'(\boldsymbol{\Upsilon}^{-1})')'\nabla(k)(\boldsymbol{z}_1, \boldsymbol{0})\} + O(h_{22}^{-2})] \\
& \times \{f(\boldsymbol{x}_1, \boldsymbol{x}_2 - \boldsymbol{z}_2) - \boldsymbol{z}_1'\boldsymbol{\Xi}'\nabla_1(f)(\boldsymbol{x}_1, \boldsymbol{x}_2 - \boldsymbol{z}_2) + O(h_{11}^2)\}\mathrm{d}\boldsymbol{z}_1\mathrm{d}\boldsymbol{z}_2 \\
=& \|\boldsymbol{H}_{22}\|^{-1}|\{k_2(\boldsymbol{0})f_1(\boldsymbol{x}_1) + O(h_{11}^2 + h_{22}^{-2})\}, \\
\mathbb{E}[\|\boldsymbol{H}\|^{-2}k^2\left(\boldsymbol{H}^{-1}(\boldsymbol{x} - \boldsymbol{X}_i)\right)] \sim& \|\boldsymbol{H}\|^{-2}\|\boldsymbol{\Xi}\| \int k^2(\boldsymbol{z}_1, \boldsymbol{0})f(\boldsymbol{x}_1, \boldsymbol{x}_2 - \boldsymbol{w}_2)\mathrm{d}\boldsymbol{z}_1\mathrm{d}\boldsymbol{w}_2
\end{aligned}$$

where ∇_1 是作用于前 d_1 个变量的向量微分算子。因此，我们有 e

$$\mathbb{E}[\widehat{f}^2(\boldsymbol{x})] - (\mathbb{E}[\widehat{f}(\boldsymbol{x})])^2 \sim \frac{1}{n}\|\boldsymbol{H}\|^{-1}\|\boldsymbol{H}_{22}\|^{-1}\left[f_1(\boldsymbol{x}_1)\int k^2(\boldsymbol{z}_1, \boldsymbol{0})\mathrm{d}\boldsymbol{z}_1 - \|\boldsymbol{\Xi}\|\{k_2(\boldsymbol{0})f_1(\boldsymbol{x}_1)\}^2\right].$$

对于其余情形 $\boldsymbol{H}_{12} = \boldsymbol{O}$ 或 $\boldsymbol{H}_{21} = \boldsymbol{O}$，仍有 $(\boldsymbol{\Xi}^{-1})_{ij} = O(h_{11}^{-1})$ 和 $(\boldsymbol{\Upsilon}^{-1})_{ij} = O(h_{22}^{-1})$，上述证明同理可得。 □

引理2表明，核条件密度估计量收敛于无条件密度 f_1，其收敛速度可能取决于 $d = d_1 + d_2$，而非 d_1。

斯卢茨基定理指出，条件密度估计量 $\widehat{f}(\boldsymbol{x})\widehat{f_2}(\boldsymbol{x}_2))^{-1}$ 依概率收敛于 $f_1(\boldsymbol{x}_1)$，前提是 $f(\boldsymbol{x})$和 $f_2(\boldsymbol{x}_2)$均为一致估计量，且共用相应的带宽矩阵 $\boldsymbol{H}_{22}$，并且 $\widehat{f}$ 的核函数满足 $\int k(\boldsymbol{z}_1, \boldsymbol{0})\mathrm{d}\boldsymbol{z}_1 = k_2(\boldsymbol{0})$，其中 k_2 是 $\widehat{f}_2$的核函数。

结合引理1和引理2，可得以下推论。

定理2. </style}假设满足引理2中关于$(\boldsymbol{x})$的条件。那么，

$$\mathbb{E}[\{\widehat{f}_2(\boldsymbol{x}_2)^{-1}\widehat{f}(\boldsymbol{x}) - f_1(\boldsymbol{x}_1)\}^2] = O(h_{11}^4 + h_{22}^{-4} + n^{-1}\|\boldsymbol{H}_{11}\|^{-1}).$$

证明. 在引理2的条件下，引理1中关于 $f_2(\boldsymbol{x}_2)$的假设成立。渐近均方误差渐近等于

$$\{k_2(\boldsymbol{0})\}^{-2}\mathbb{E}[\{\|\boldsymbol{H}_{22}\|(\widehat{f}(\boldsymbol{x}) - \|\boldsymbol{H}_{22}\|^{-1}k_2(\boldsymbol{0})f_1(\boldsymbol{x}_1)) - \|\boldsymbol{H}_{22}\|^2 f_1(\boldsymbol{x}_1)(\widehat{f}_2(\boldsymbol{x}_2) - k_2(\boldsymbol{0}))\}^2].$$

应用引理1和引理2以及H¨older不等式，可得所需结果。 □

在独立条件下，条件密度估计就是边际密度估计。以下结果表明，其最优收敛速率与核（边际）密度估计量的最优收敛速率一致。

推论2。 假设定理*2*的条件成立。若 $h_{22}^{-1} = O(h_{11})$ 或 $o(h_{11})$，使渐近均方误差最小可得$h_{11} = O(n^{-(d_1+4)^{-1}})$，且采用最优带宽时的均方误差阶为 $n^{-4(d_1+4)^{-1}}$。

注7。 在假设核 k 是概率密度函数的条件下，采用分块带宽矩阵的核密度估计量的期望可视为随机变量 $\boldsymbol{X} + \boldsymbol{H}\boldsymbol{K}$ 的概率密度函数。 $\boldsymbol{K} := (\boldsymbol{K}_1', \boldsymbol{K}_2')'$ 是一个与 $\boldsymbol{X}$ 独立的随机变量，其概率密度函数为 k。由于放大与缩小 $\boldsymbol{H}$，当 n 足够大时，该随机变量由$(\boldsymbol{X}_1', (\boldsymbol{H}_{22}\boldsymbol{K}_2)')'$主导，其中最后一个 d_2 变量渐近服从 $\mathbb{R}^{d_2}$ 上的均匀分布。

注8。令

$$\boldsymbol{H}=\begin{pmatrix} h_{11}\boldsymbol{C}_{11} & h_{11}\boldsymbol{C}_{12} \\ h_{11}\boldsymbol{C}_{21} & h_{11}\boldsymbol{C}_{22} \end{pmatrix},$$

其中， $\boldsymbol{C}_{11}$、 $\boldsymbol{C}_{12}$、 $\boldsymbol{C}_{21}$ 和 $\boldsymbol{C}_{22}$ 是大小为 $d_1\times d_1$、 $d_1\times d_2$、 $d_2\times d_1$和 $d_2\times d_2$的常数矩阵，满足 $\boldsymbol{C}_{11}\neq\boldsymbol{O}$ 为正则矩阵且 $\boldsymbol{C}_{22}\neq\boldsymbol{O}$。

$$\boldsymbol{E}_n:=\begin{pmatrix} \boldsymbol{E}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & h_{11}^{-1}h_{22}\boldsymbol{E}_{22} \end{pmatrix},$$

其中， $\boldsymbol{E}_{11}$ 和 $\boldsymbol{E}_{22}$ 分别是大小为 $d_1\times d_1$ 和 $d_2\times d_2$的单位矩阵。于是，

$$\boldsymbol{E}_n\boldsymbol{H}=\begin{pmatrix} h_{11}\boldsymbol{C}_{11} & h_{11}\boldsymbol{C}_{12} \\ h_{22}\boldsymbol{C}_{21} & h_{22}\boldsymbol{C}_{22} \end{pmatrix},$$

并且，当 $\boldsymbol{C}_{21}=\boldsymbol{O}$时， $\boldsymbol{E}_n\boldsymbol{H}$ 满足带宽矩阵的假设（即引理*2*中的(i)–(ii) ）。由于

$$\begin{aligned}&(n\|\boldsymbol{E}_n\boldsymbol{H}\|)^{-1}\sum_{i=1}^{n}k((\boldsymbol{E}_n\boldsymbol{H})^{-1}(\boldsymbol{x}-\boldsymbol{X}_i))\\ =&\|\boldsymbol{E}_n\|^{-1}(n\|\boldsymbol{H}\|)^{-1}\sum_{i=1}^{n}k(\boldsymbol{H}^{-1}((\boldsymbol{x}_1-\boldsymbol{X}_{1,i})',h_{22}^{-1}h_{11}(\boldsymbol{x}_2-\boldsymbol{X}_{2,i})')'),\end{aligned}$$

带有（部分）大带宽矩阵 $\boldsymbol{E}_n\boldsymbol{H}$ 的$\widehat{f}$，等于随机变量$(\boldsymbol{X}_{1,i}',h_{22}^{-1}h_{11}\boldsymbol{X}_{2,i}')'$对应的小带宽矩阵 $\boldsymbol{H}$ 核密度估计量的 $\|\boldsymbol{E}_n\|^{-1}$ 倍。

核密度估计量的期望可视为随机变量的概率密度函数

$$\boldsymbol{X}_i+\boldsymbol{E}_n\boldsymbol{H}\boldsymbol{K}$$

并且是其 $\|\boldsymbol{E}_n\|^{-1}$ 倍

$$(\boldsymbol{X}_{1,i}',h_{22}^{-1}h_{11}\boldsymbol{X}_{2,i}')'+\boldsymbol{H}\boldsymbol{K},$$

其中， $\boldsymbol{K}$ 是与（$\boldsymbol{X}_{1,\ i}'$、$h_{22}^{-1}h_{11}\boldsymbol{X}_{2,\ i}'$）′独立的随机变量，对每个 n 均成立。该随机变量受（$\boldsymbol{X}_1'$、$\boldsymbol{0}'$）′支配，其中最后一个 d_3 变量渐近缩小。

注9。带宽矩阵假设

$$\boldsymbol{H}_{12}=\boldsymbol{O}\quad or\quad h_{12}=O(h_{11})$$

and

$$\boldsymbol{H}_{21}=\boldsymbol{O}\quad or\quad h_{21}=O(h_{11})$$

其中 $h_{11}\to 0$ 和 $h_{22}\to\infty$ 分别为 $n\to\infty$，足以使

$$\widehat{f}_2(\boldsymbol{x}_2)^{-1}\widehat{f}(\boldsymbol{x})\xrightarrow{p}f_1(\boldsymbol{x}_1).$$

对于 h_2 $\backsim 1\to\infty$ f_2 的带宽需要经过合理设计（参见附录：补充结果 *).*

3 回归中的条件独立性或条件密度估计

本节证明，对于所有可能的带宽矩阵，核估计量都能达到最优收敛速率。这表明，通过渐近地最小化平均积分平方误差来进行带宽选择的方法，无需指定依赖结构，也能选出理论上的最优带宽矩阵。下一节将研究带宽选择方法的数值性质。

3.1 回归估计中条件独立的情形

设 d 维随机变量为 $\boldsymbol{X}_i := (\boldsymbol{X}'_{1,\,i},\ \boldsymbol{X}'_{2,\,i},\ \boldsymbol{X}'_{3,\,i})'$，$\boldsymbol{Z}_i = (\boldsymbol{X}'_{2,\,i},\ \boldsymbol{X}'_{3,\,i})'$，且 d_2+d_3 维随机变量 $\boldsymbol{Z}_i$ 的概率密度函数 $\widehat{}$ 为 f_{23}。令 f_{23} 为 $\boldsymbol{Z}_i$ 的核密度估计量，其定义如下：

$$\widehat{f}_{23}(\boldsymbol{z}) := (n\|\boldsymbol{\Omega}\|)^{-1}\sum_{i=1}^{n} k_{23}(\boldsymbol{\Omega}^{-1}(\boldsymbol{z}-\boldsymbol{Z}_i)),$$

k_{23}：$\mathbb{R}^{d_2+d_3}\to\mathbb{R}$ 且带宽假设为（d_2+d_3）$\times$（d_2+d_3）矩阵 x

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{H}_{22} & \boldsymbol{H}_{23} \\ \boldsymbol{H}_{32} & \boldsymbol{H}_{33} \end{pmatrix},$$

其中，每个块中的所有元素都假设为同阶。令

$$\boldsymbol{\Omega}^{-1} := \begin{pmatrix} \boldsymbol{\Xi}^{-1} & -\boldsymbol{H}_{22}^{-1}\boldsymbol{H}_{23}\boldsymbol{\Upsilon}^{-1} \\ -\boldsymbol{\Upsilon}^{-1}\boldsymbol{H}_{32}\boldsymbol{H}_{22}^{-1} & \boldsymbol{\Upsilon}^{-1} \end{pmatrix},$$

其中 $\boldsymbol{\Upsilon} = \boldsymbol{H}_{33}-\boldsymbol{H}_{32}\boldsymbol{H}_{22}^{-1}\boldsymbol{H}_{23}$ 且 $\boldsymbol{\Xi}^{-1} = \boldsymbol{H}_{22}^{-1}+\boldsymbol{H}_{22}^{-1}\boldsymbol{H}_{23}\boldsymbol{\Upsilon}^{-1}\boldsymbol{H}_{32}\boldsymbol{H}_{22}^{-1}$。以下关于核回归估计量的结果

$$\widehat{m}(\boldsymbol{z}) := \widehat{f}_{23}(\boldsymbol{z})^{-1}(n\|\boldsymbol{\Omega}\|)^{-1}\sum_{i=1}^{n}\boldsymbol{X}_{1,i}k_{23}(\boldsymbol{\Omega}^{-1}(\boldsymbol{z}-\boldsymbol{Z}_i)),$$

也是引理2的推论。

定理 3. 假设(i) $\boldsymbol{H}_{22}$ 是正则的，$h_{22}\to 0$ 且 $h_{33}\to\infty$，(ii)

$$\boldsymbol{H}_{23}=\boldsymbol{O} \quad and \quad \boldsymbol{H}_{32}=\boldsymbol{O}$$

or

$$h_{23}=O(h_{22}) \quad and \quad h_{32}=O(h_{22}),$$

(iii) $k_3(\boldsymbol{0})\neq 0$ 存在，其中

$$k_3(\boldsymbol{z}_3) = \int k_{23}(\boldsymbol{z}_2,\boldsymbol{z}_3)\mathrm{d}\boldsymbol{w}_2,$$

(iv) k_{23} 在 $\{(\boldsymbol{z}_2,\boldsymbol{0})|\boldsymbol{z}_2\in\mathbb{R}^{d_2}\}$ 上二阶连续可微，(v) $\int \boldsymbol{z}_2 k_{23}(\boldsymbol{z}_2,\boldsymbol{0})\mathrm{d}\boldsymbol{z}_2 = \boldsymbol{0}$，(vi) 的一阶矩存在，(vii) f 关于变量 $\boldsymbol{x}_2$ 偏连续可微，对任意 $\boldsymbol{x}_3$ 均成立。则有

$$\mathbb{E}[\{\widehat{m}(\boldsymbol{z})-m(\boldsymbol{x}_2)\}^2] = O(h_{22}^4+h_{33}^{-4}+n^{-1}\|\boldsymbol{H}_{22}\|^{-1}),$$

其中

$$m(\boldsymbol{x}_2) := \int \boldsymbol{w}_1\frac{f_{12}(\boldsymbol{w}_1,\boldsymbol{x}_2)}{f_2(\boldsymbol{x}_2)}\mathrm{d}\boldsymbol{w}_1,$$

f_{12} 和 f_2 分别是$(\boldsymbol{X}_{1,i},\boldsymbol{X}_{2,i})$ 及 $\boldsymbol{X}_{2,i}$ 的概率密度函数。

证明。 通过与引理2的证明类似地进行变量替换 $\boldsymbol{\Xi}^{-1}(\boldsymbol{x}_2-\boldsymbol{w}_2) = \boldsymbol{z}_2$，

$$\begin{aligned}
\|\boldsymbol{\Omega}\|^{-1}\mathbb{E}[\boldsymbol{X}_{1,i}k_{23}(\boldsymbol{\Omega}^{-1}(\boldsymbol{z}-\boldsymbol{Z}_i))] =& \|\boldsymbol{\Omega}\|^{-1}\|\boldsymbol{\Xi}\|\int \boldsymbol{w}_1[k_{23}(\boldsymbol{z}_2,\boldsymbol{0})+\{(-\boldsymbol{w}_3'(\boldsymbol{\Upsilon}^{-1})'\boldsymbol{H}_{23}'(\boldsymbol{H}_{22}^{-1})', \\
& -\boldsymbol{z}_2'\boldsymbol{\Xi}'(\boldsymbol{H}_{22}^{-1})'\boldsymbol{H}_{32}'(\boldsymbol{\Upsilon}^{-1})'+\boldsymbol{w}_3'(\boldsymbol{\Upsilon}^{-1})')'\nabla(k_{23})(\boldsymbol{z}_2,\boldsymbol{0})\}+O(h_{33}^{-2})\}] \\
& \times\{f(\boldsymbol{w}_1,\boldsymbol{x}_2,\boldsymbol{w}_3)-\boldsymbol{z}_2'\boldsymbol{\Xi}'\nabla_2(f)(\boldsymbol{w}_1,\boldsymbol{x}_2,\boldsymbol{w}_3)+O(h_{22}^2)\}\mathrm{d}\boldsymbol{w}_1\mathrm{d}\boldsymbol{z}_2\mathrm{d}\boldsymbol{w}_3 \\
=& \|\boldsymbol{H}_{33}\|^{-1}\{k_3(\boldsymbol{0})\int \boldsymbol{w}_1 f(\boldsymbol{w}_1,\boldsymbol{x}_2,\boldsymbol{w}_3)\mathrm{d}\boldsymbol{w}_1\mathrm{d}\boldsymbol{w}_3+O(h_{22}^2+h_{33}^{-2})\}.
\end{aligned}$$

同理，我们有

$$\|\boldsymbol{\Omega}\|^{-2}\mathbb{E}[\boldsymbol{X}_{1,i}\boldsymbol{X}_{1,i}'k_{23}^2(\boldsymbol{\Omega}^{-1}(\boldsymbol{z}-\boldsymbol{Z}_i))] \sim |\boldsymbol{\Omega}\|^{-2}\|\boldsymbol{\Xi}\| \int \boldsymbol{w}_1\boldsymbol{w}_1'k_{23}^2(\boldsymbol{z}_2,\boldsymbol{0})f(\boldsymbol{w}_1,\boldsymbol{x}_2,\boldsymbol{w}_3)\mathrm{d}\boldsymbol{w}_1\mathrm{d}\boldsymbol{z}_2\mathrm{d}\boldsymbol{w}_3$$

且（$n\|\boldsymbol{\Omega}\|$）$^{-1}\sum_{i=1}^n \boldsymbol{X}_{1,\ i}k_{23}$（$\boldsymbol{\Omega}^{-1}$（$\boldsymbol{z}-\boldsymbol{Z}_i$））的渐近方差为

$$\begin{aligned}&\frac{1}{n}|\boldsymbol{\Omega}\|^{-2}\|\boldsymbol{\Xi}\| \Bigg[\int \boldsymbol{w}_1\boldsymbol{w}_1'k_{23}^2(\boldsymbol{z}_2,\boldsymbol{0})f(\boldsymbol{w}_1,\boldsymbol{x}_2,\boldsymbol{w}_3)\mathrm{d}\boldsymbol{w}_1\mathrm{d}\boldsymbol{z}_2\mathrm{d}\boldsymbol{w}_3 \\ &\quad - \|\boldsymbol{\Xi}\|k_3^2(\boldsymbol{0})\{\int \boldsymbol{w}_1 f(\boldsymbol{w}_1,\boldsymbol{x}_2,\boldsymbol{w}_3)\mathrm{d}\boldsymbol{w}_1\mathrm{d}\boldsymbol{w}_3\}\{\int \boldsymbol{w}_1 f(\boldsymbol{w}_1,\boldsymbol{x}_2,\boldsymbol{w}_3)\mathrm{d}\boldsymbol{w}_1\mathrm{d}\boldsymbol{w}_3\}' \Bigg].\end{aligned}$$

因此，应用引理2于 f_{23}（$\boldsymbol{z}$）并结合 H¨older 不等式，可得到渐近均方误差的阶。

□

定理 3 表明，核回归估计量收敛于部分条件期望 m（$\boldsymbol{x}_2$），其中收敛速度不仅取决于 d_2+d_3，还取决于 d_2。

推论 3。假设满足定理 *3* 的条件。如果 $h_{33}^{-1}=O$（h_{22}）或 o（h_{22}），则最小化渐近均方误差得到$h_{22}=O$（$n^{-(d_2+4)\ -1}$），且采用最优带宽时的均方误差的阶为 $n^{-4(d_2+4)\ -1}$。

3.2 条件密度估计中的条件独立情形

$\widehat{f}$ 是 $\boldsymbol{X}_i$ 的核密度函数，其核函数为 $k:\ \mathbb{R}^d\to\mathbb{R}$，且带宽矩阵设为

$$\boldsymbol{H}=\begin{pmatrix}\boldsymbol{H}_{11} & \boldsymbol{H}_{12} & \boldsymbol{H}_{13}\\ \boldsymbol{H}_{21} & \boldsymbol{H}_{22} & \boldsymbol{H}_{23}\\ \boldsymbol{H}_{31} & \boldsymbol{H}_{32} & \boldsymbol{H}_{33}\end{pmatrix}=:\begin{pmatrix}\boldsymbol{\Psi} & \boldsymbol{H}_{12,3}\\ \boldsymbol{H}_{3,12} & \boldsymbol{H}_{33}\end{pmatrix},$$

假设每个块中的所有元素都具有相同的阶。

让我们考虑 以下情形：给定 $\boldsymbol{X}_{2,\ i}=\boldsymbol{x}_2$ 时，$\boldsymbol{X}_{1,\ i}$ 和 $\boldsymbol{X}_{3,\ i}$ 条件独立

$$\mathbb{P}[\boldsymbol{X}_{1,i}=\boldsymbol{x}_1,\boldsymbol{Z}_i=\boldsymbol{z}]=\mathbb{P}[\boldsymbol{X}_{1,i}=\boldsymbol{x}_1|\boldsymbol{X}_{2,i}=\boldsymbol{x}_2]\mathbb{P}[\boldsymbol{Z}_i=\boldsymbol{z}].$$

即 $\mathbb{P}[\boldsymbol{X}_{1,i}=\boldsymbol{x}_1|\boldsymbol{Z}_i=\boldsymbol{z}]=\mathbb{P}[\boldsymbol{X}_{1,i}=\boldsymbol{x}_1|\boldsymbol{X}_{2,i}=\boldsymbol{x}_2]$。

假设引理2中关于 $f(\boldsymbol{x})$ 和 $f_{23}(\boldsymbol{z})$ 的条件均成立，其中 $\frown f_{23}(\boldsymbol{z})$ 的带宽矩阵为

$$\boldsymbol{\Omega}:=\begin{pmatrix}\boldsymbol{H}_{22} & \boldsymbol{H}_{23}\\ \boldsymbol{H}_{32} & \boldsymbol{H}_{33}\end{pmatrix}.$$

那么，关于核条件密度估计量（$\widehat{f}_{23}(\boldsymbol{z})$）$^{-1}\widehat{f}(\boldsymbol{x})$，我们有如下推论。

定理 4. 假设满足引理*2*中关于 $f(\boldsymbol{x})$ 和 $f_{23}(\boldsymbol{z})$ 的条件，即(i) $\boldsymbol{\Psi}$ 是正则的，$(\boldsymbol{\Psi})_{ij}=O($ $h_{11})$，$h_{11}\to 0$ 和 $h_{33}\to\infty$ 当 $n\to\infty$ 时，(ii)

$$\boldsymbol{H}_{12,3}=\boldsymbol{O}\quad and\quad \boldsymbol{H}_{3,12}=\boldsymbol{O}$$

or

$$h_{12,3}=O(h_{22})\quad and\quad h_{3,12}=O(h_{22}),$$

(iii) $k_3(\boldsymbol{0})\neq 0$ 存在，其中

$$k_3(\boldsymbol{z}_3)=\int k(\boldsymbol{w}_1,\boldsymbol{w}_2,\boldsymbol{z})\mathrm{d}\boldsymbol{w}_1\mathrm{d}\boldsymbol{w}_2,$$

(iv) k 在 $\{(z_1, z_2, 0)|(z_1, z_2) \in \mathbb{R}^{d_1+d_2}\}$上二阶连续可微，*(v)* 对任意 (z_1, z_2)，$\nabla_3 k(z_1, z_2, 0) = \mathbf{0}$，*(vi)* $\int (z_1', z_2')' k(z_1, z_2, 0) dz_1 dz_2 = \mathbf{0}$，*(vii)* 对任意 x_3，f 关于变量 (x_1, x_2) 偏连续可微。于是，

$$\mathbb{E}[\{\widehat{f}_{23}(\boldsymbol{z})^{-1}\widehat{f}(\boldsymbol{x}) - f_2(\boldsymbol{x}_2)^{-1} f_{12}(\boldsymbol{x}_1, \boldsymbol{x}_2)\}^2] = O(h_{11}^4 + h_{33}^{-4} + n^{-1}\|\boldsymbol{\Psi}\|^{-1}),$$

其中，f_{12} 和 f_2 分别是（$\boldsymbol{X}_{1,\,i}$,$\boldsymbol{X}_{2,\,i}$）以及$\boldsymbol{X}_{2,\,i}$ 的概率密度函数。

证明。 由于渐近均方误差渐近表示为

$$\{k_3(\mathbf{0})\}^{-2}\mathbb{E}[\{\|\boldsymbol{H}_{33}\|(\widehat{f}(\boldsymbol{x}) - \|\boldsymbol{H}_{33}\|^{-1}k_3(\mathbf{0})f_{12}(\boldsymbol{x}_1, \boldsymbol{x}_2)) - \|\boldsymbol{H}_{33}\|^2 f_{12}(\boldsymbol{x}_1, \boldsymbol{x}_2)(\widehat{f}_3(\boldsymbol{x}_3) - k_3(\mathbf{0}))\}^2],$$

应用引理2和H¨older不等式，我们得到该结果。 □

定理 4 表明，核条件密度估计量收敛于部分条件密度$f_2(\boldsymbol{x}_2)^{-1}f_{12}(\boldsymbol{x}_1,\ \boldsymbol{x}_2)$，其中收敛速度可能取决于 $d = d_1 + d_2 + d_3$，而非 $d_1 + d_2$。

推论 4。 设定理 *4* 的条件成立。如果$h_{33}^{-1} = O$（h_{11}）或 o（h_{11}），则最小化渐近均方误差得到$h_{11} = O$（$n^{-(d_1+d_2+4)^{-1}}$），且采用最优带宽时，均方误差的阶为 $n^{-4(d_1+d_2+4)^{-1}}$。

4 多指标模型在回归或条件密度估计中的情形

考虑多指标模型的情形，即存在一个 $d_2 \times (d_2 + d_3)$ 矩阵 $\boldsymbol{A}$，使得秩$\boldsymbol{A} = d_2$且

$$\mathbb{P}[\boldsymbol{X}_{1,i} = \boldsymbol{x}_1, \boldsymbol{Z}_i = \boldsymbol{z}_i] = \mathbb{P}[\boldsymbol{X}_{1,i} = \boldsymbol{x}_1|\boldsymbol{A}\boldsymbol{Z}_i = \boldsymbol{A}\boldsymbol{z}_i]\mathbb{P}[\boldsymbol{Z}_i = \boldsymbol{z}_i],$$

其中 $\boldsymbol{A} = (\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_{d_2})'$，但 $\boldsymbol{A}$ 未知。于是，$\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_{d_2} \in \mathbb{R}^{d_2+d_3}$ 线性无关。我们可以选取任意一组线性无关向量，使其也属于 $\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_{d_2}$的张成空间。

对于任意一组线性无关的 $\boldsymbol{b}_1,\ \cdots,\ \boldsymbol{b}_{d_3} \in \mathbb{R}^{d_2+d_3}$，若它们也与 $\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_{d_2}$相互独立，其中 $\boldsymbol{B} := (\boldsymbol{b}_1,\ \cdots,\ \boldsymbol{b}_{d_3})'$，则始终有

$$\mathbb{P}[\boldsymbol{X}_{1,i} = \boldsymbol{x}_1|\boldsymbol{Z}_i = \boldsymbol{z}_i] \equiv \mathbb{P}[\boldsymbol{X}_{1,i} = \boldsymbol{x}_1|\boldsymbol{A}\boldsymbol{Z}_i = \boldsymbol{A}\boldsymbol{z}_i, \boldsymbol{B}\boldsymbol{Z}_i = \boldsymbol{B}\boldsymbol{z}_i].$$

令 $\boldsymbol{A}\boldsymbol{Z}_i =: \boldsymbol{Y}_{2,\,i}$，$\boldsymbol{B}\boldsymbol{Z}_i =: \boldsymbol{Y}_{3,\,i}$，则该情形可归结为条件独立的情形

$$\mathbb{P}[\boldsymbol{X}_{1,i} = \boldsymbol{x}_1|\boldsymbol{Y}_{2,i} = \boldsymbol{y}_2, \boldsymbol{Y}_{3,i} = \boldsymbol{y}_3] = \mathbb{P}[\boldsymbol{X}_{1,i} = \boldsymbol{x}_1|\boldsymbol{Y}_{2,i} = \boldsymbol{y}_2]$$

（见第 2.3 节）。

设 d 维随机变量为 $\boldsymbol{Y}_i := (\boldsymbol{Y}_{1,\,i}',\ \boldsymbol{Y}_{2,\,i}',\ \boldsymbol{Y}_{3,\,i}')'$，且 $\boldsymbol{D} := (\boldsymbol{A}',\ \boldsymbol{B}')'$，其中$\boldsymbol{Y}_{1,\,i} = \boldsymbol{X}_{1,\,i}$。$(d_2 + d_3)$ 维随机变量 $(\boldsymbol{Y}_{2,\,i}',\ \boldsymbol{Y}_{3,\,i}')' =$的概率密度函数 g_{23} 满足

$$g_{23}(\boldsymbol{w}) := \|\boldsymbol{D}\|^{-1} f_{23}(\boldsymbol{D}^{-1}\boldsymbol{w}) = \|\boldsymbol{D}\|^{-1} f_{23}(\boldsymbol{z}),$$

其中，f_{23} 是$(\boldsymbol{X}_{2,i},\boldsymbol{X}_{3,i})$的概率密度函数，且 $\boldsymbol{w} := \boldsymbol{D}\boldsymbol{z}$。

4.1 多指标模型回归情形

设 k　　采用带宽矩阵 $\boldsymbol{\Omega}$ 的 $\boldsymbol{W}$ 核密度估计量，在点 $\boldsymbol{w} := \boldsymbol{D}\boldsymbol{z}$ be

$$\widehat{g}_{23}(\boldsymbol{w}) := (n\|\boldsymbol{\Omega}\|)^{-1} \sum_{i=1}^{n} k_{23}(\boldsymbol{\Omega}^{-1}(\boldsymbol{w} - \boldsymbol{W}_i)).$$

那么，以下结论成立:

$$\widehat{g}_{23}(\boldsymbol{w}) = (n\|\boldsymbol{D}^{-1}\boldsymbol{\Omega}\|)^{-1}\sum_{i=1}^{n}\|\boldsymbol{D}\|^{-1}k_{23}(\boldsymbol{\Omega}^{-1}\boldsymbol{D}(\boldsymbol{z}-\boldsymbol{Z}_i))$$

可视为估计量 $\widehat{f}_{23}(\boldsymbol{z})$，其带宽矩阵为 $\boldsymbol{D}^{-1}\boldsymbol{\Omega}$，而非 $\boldsymbol{\Omega}$，核函数为$\|\boldsymbol{D}\|^{-1}k_{23}$，其中 $\boldsymbol{D}$ 是正则的。这意味着 $\widehat{f}_{23}(\boldsymbol{z})$ 实际上具有 $\widehat{g}_{23}(\boldsymbol{w})$ 的形式，即条件相依的密度估计量；也就是说，在引理2的假设下，有

$$\|\boldsymbol{D}^{-1}\boldsymbol{\Omega}\|\widehat{f}_{23}(\boldsymbol{z}) - k_{23}(\boldsymbol{0})g_2(\boldsymbol{A}\boldsymbol{z}) = O_P(n^{-1/2}h_{22}^{-d_2/2}) + O(h_{11}^2 + h_{22}^{-2}),$$

其中 g_2 是 $\boldsymbol{Y}_{2,i}$ 的概率密度函数。

此外,

$$\begin{aligned}\widehat{r}(\boldsymbol{w}) :=& \{\widehat{g}_{23}(\boldsymbol{w})\}^{-1}(n\|\boldsymbol{\Omega}\|)^{-1}\sum_{i=1}^{n}\boldsymbol{Y}_{1,i}k_{23}(\boldsymbol{\Omega}^{-1}(\boldsymbol{w}-\boldsymbol{W}_i))\\ =& \left\{(n\|\boldsymbol{D}^{-1}\boldsymbol{\Omega}\|)^{-1}\sum_{i=1}^{n}k_{23}(\boldsymbol{\Omega}^{-1}\boldsymbol{D}(\boldsymbol{z}-\boldsymbol{Z}_i))\right\}^{-1}(n\|\boldsymbol{D}^{-1}\boldsymbol{\Omega}\|)^{-1}\sum_{i=1}^{n}\boldsymbol{X}_{1,i}k_{23}(\boldsymbol{\Omega}^{-1}\boldsymbol{D}(\boldsymbol{z}-\boldsymbol{Z}_i))\end{aligned}$$

可视为估计量 $\widehat{m}(\boldsymbol{z})$，其带宽矩阵为 $\boldsymbol{D}^{-1}\boldsymbol{\Omega}$，而非 $\boldsymbol{\Omega}$，核函数为k_{23}。于是，在定理 3的假设下，有

$$\mathbb{E}[\{\widehat{m}(\boldsymbol{z}) - r(\boldsymbol{A}\boldsymbol{z})\}^2] = O(h_{22}^4 + h_{33}^{-4} + n^{-1}h_{22}^{-d_2}),$$

其中

$$r(\boldsymbol{A}\boldsymbol{z}) := \int \boldsymbol{w}_1\frac{g_{12}(\boldsymbol{w}_1, \boldsymbol{A}\boldsymbol{z})}{g_2(\boldsymbol{A}\boldsymbol{z})}\mathrm{d}\boldsymbol{w}_1,$$

g_{12} 是 $(\boldsymbol{Y}_{1,i},\boldsymbol{Y}_{2,i})$ 的概率密度函数。

由此可知，对于多指标模型，回归估计量的最优收敛速率取决于 d_2，而非 (d_2+d_3)。

注记10. 存在 $\boldsymbol{D}^{-1}\boldsymbol{\Omega}$，使得$\boldsymbol{\Omega}$满足相应的假设（例如定理 *3* 中的(i)–(iii)），这可通过逆向计算得出。该矩阵的形式为

$$\boldsymbol{D}^{-1}\boldsymbol{\Omega} = \boldsymbol{D}^{-1}\begin{pmatrix} h_{22}\boldsymbol{C}_{22} & h_{22}\boldsymbol{C}_{23} \\ h_{22}\boldsymbol{C}_{32} & h_{33}\boldsymbol{C}_{33}\end{pmatrix},$$

其中，$\boldsymbol{C}_{22}$、$\boldsymbol{C}_{23}$、$\boldsymbol{C}_{32}$ 和 $\boldsymbol{C}_{33}$ 是大小分别为 $d_2\times d_2$、$d_2\times d_3$、$d_3\times d_2$和 $d_3\times d_3$的常数矩阵，满足 $\boldsymbol{C}_{22}\neq\boldsymbol{O}$ 为正则且满足 $\boldsymbol{C}_{33}\neq\boldsymbol{O}$。最优 $\boldsymbol{D}^{-1}\boldsymbol{\Omega}$ 的元素排列形式为

$$\begin{pmatrix} h_{22}^{-1} & h_{33}^{-1} \\ h_{22}^{-1} & h_{33}^{-1}\end{pmatrix}.$$

带宽矩阵为 $\boldsymbol{D}^{-1}\boldsymbol{\Omega}$ 的核密度估计量 $\widehat{f}_{23}$ 的期望可见为随机变量概率密度函数的 $\|\boldsymbol{D}\|^{-1}$ 倍

$$\boldsymbol{W} + \begin{pmatrix} h_{22}\boldsymbol{C}_{22} & h_{22}\boldsymbol{C}_{23} \\ h_{22}\boldsymbol{C}_{32} & h_{33}\boldsymbol{C}_{33}\end{pmatrix}\boldsymbol{K}.$$

$\boldsymbol{K}:=(\boldsymbol{K}_2', \boldsymbol{K}_3')'$是与 $\boldsymbol{W}$ 独立的随机变量，其概率密度函数为 k_{23}。由于当 n 足够大时的放大与收缩矩阵，该随机变量由 $(\boldsymbol{Y}_2', h_{33}(\boldsymbol{C}_{33}\boldsymbol{K}_3)')'$支配，其中最后一个 d_3 变量渐近服从在 $\mathbb{R}^{d_3}$ 上的均匀分布。

注 11。 设核 k_{23} 为球对称的，即存在 ℓ_{23}： $\mathbb{R}^{d_2+d_3} \to \mathbb{R}$，使得 $k_{23}(\boldsymbol{w}) = \ell_{23}(\|\boldsymbol{w}\|_2)$。那么，对于任意正交矩阵 $\boldsymbol{R}$，有

$$\ell_{23}(\|\boldsymbol{w}\|_2) \equiv \ell_{23}(\|\boldsymbol{R}\boldsymbol{w}\|_2),$$

这意味着不$\boldsymbol{\Omega}$ 而是$\boldsymbol{\Omega}\boldsymbol{R}^{-1}$ 必须满足相应的假设（例如定理 *3* 中的(i)–(iii)）。矩阵的形式为

$$\boldsymbol{D}^{-1}\boldsymbol{\Omega}\boldsymbol{R}^{-1} = \boldsymbol{D}^{-1}\begin{pmatrix} h_{22}\boldsymbol{C}_{22} & h_{22}\boldsymbol{C}_{23} \\ h_{22}\boldsymbol{C}_{32} & h_{33}\boldsymbol{C}_{33} \end{pmatrix}\boldsymbol{R}^{-1};$$

然而，这也 对于任意正交矩阵 $\boldsymbol{R}$，理论上都不可能是分块对角的（更不用说对角的） .

核密度估计量的期望 $\boldsymbol{\Omega}\boldsymbol{R}^{-1}$ 对应的值是随机变量 $\widehat{\boldsymbol{W}} + \boldsymbol{\Omega}\boldsymbol{R}^{-1}\boldsymbol{K}$的概率密度函数的 $\|\boldsymbol{D}\|^{-1}$ 倍。 $\boldsymbol{K} := (\boldsymbol{K}_2', \boldsymbol{K}_3')'$ 是与 $\boldsymbol{W}$ 相互独立的随机变量，其概率密度函数为 ℓ_{23}。因此，对于球对称的 k_{23}， $\boldsymbol{R}^{-1}\boldsymbol{K} = \boldsymbol{K}$ 在分布上成立。当 n 足够大时，该随机变量由$(\boldsymbol{Y}_2', (\boldsymbol{H}_{33}\boldsymbol{K}_3)')'$主导，其中最后一个 d_3变量渐近服从 $\mathbb{R}^{d_3}$上的均匀分布。

4.2 条件密度估计中的多指标模型情形

让我们考虑与第 4.1 节相同的设定。我们也使用带宽矩阵的记号

$$\begin{pmatrix} \boldsymbol{\Psi} & \boldsymbol{H}_{12,3} \\ \boldsymbol{H}_{3,12} & \boldsymbol{H}_{33} \end{pmatrix} := \boldsymbol{H}$$

并考虑引理2的假设，其中 $(\boldsymbol{\Psi})_{ij} = O(h_{11})$、$(\boldsymbol{H}_{3,12})_{ij} = O(h_{33})$ 以及 $(\boldsymbol{H}_{33})_{ij} = O(h_{33})$。$\boldsymbol{H}_{12,3} = \boldsymbol{O}$ 或 $(\boldsymbol{H}_{12,3})_{ij} = O(h_{33})$。令 $\widehat{g}$ 为 $\boldsymbol{Y}_i$ 在点 $\boldsymbol{y} := \boldsymbol{D}\boldsymbol{x}$ 处的核密度估计量，由下式给出

$$\widehat{g}(\boldsymbol{y}) := (n\|\boldsymbol{H}\|)^{-1}\sum_{i=1}^{n} k(\boldsymbol{H}^{-1}(\boldsymbol{y} - \boldsymbol{Y}_i)),$$

其中

$$\widetilde{\boldsymbol{D}} := \begin{pmatrix} \boldsymbol{E}_{11} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{D} \end{pmatrix}.$$

于是,

$$\widehat{g}(\boldsymbol{y}) = (n\|\widetilde{\boldsymbol{D}}^{-1}\boldsymbol{H}\|)^{-1}\sum_{i=1}^{n} \|\boldsymbol{D}\|^{-1}k(\boldsymbol{H}^{-1}\widetilde{\boldsymbol{D}}(\boldsymbol{x} - \boldsymbol{X}_i))$$

可视为带有带宽矩阵 $\widetilde{\boldsymbol{D}}^{-1}\boldsymbol{H}$ 与核函数 $\|\boldsymbol{D}\|^{-1}k$的估计量 $\widehat{f}(\boldsymbol{x})$。在定理 4 的假设下，成立

$$\|\widetilde{\boldsymbol{D}}^{-1}\boldsymbol{H}\|\widehat{f}(\boldsymbol{x}) - k(\boldsymbol{0})g(\widetilde{\boldsymbol{D}}\boldsymbol{x}) = O_P(n^{-1/2}h_{11}^{-(d_1+d_2)/2}) + O(h_{11}^2 + h_{22}^{-2}),$$

其中 g 表示 $\boldsymbol{Y}$的概率密度函数。由于

$$\widehat{g}_{23}(\boldsymbol{w}) := (n\|\boldsymbol{D}^{-1}\boldsymbol{\Omega}\|)^{-1}\sum_{i=1}^{n} \|\boldsymbol{D}\|^{-1}k_{23}(\boldsymbol{\Omega}^{-1}\boldsymbol{D}(\boldsymbol{z} - \boldsymbol{Z}_i))$$

可以看作是具有带宽矩阵 $\boldsymbol{D}^{-1}\boldsymbol{\Omega}$ 和核函数 $\|\boldsymbol{D}\|^{-1}k$的估计量 $\widehat{f}_{23}(\boldsymbol{z})$，$\widehat{g}_{23}(\boldsymbol{D}\boldsymbol{z})^{-1}\widehat{g}(\widetilde{\boldsymbol{D}}\boldsymbol{x})$可以看作是具有带宽矩阵 $\widetilde{\boldsymbol{D}}^{-1}\boldsymbol{H}$ 和$\widetilde{\ }$核函数 k的估计量 $\widehat{f}_{23}(\boldsymbol{z})^{-1}\widehat{f}(\boldsymbol{x})$。若 $\boldsymbol{D}^{-1}\boldsymbol{H}$ 满足定理 4中的相应假设，则先前的所有结果均成立，其中这样的 $\boldsymbol{H}$ 可通过反向计算容易得到。此处亦请注意注释3。

因此，我们看到

$$\mathbb{E}[\{\widehat{f}_{23}(\boldsymbol{z})^{-1}\widehat{f}(\boldsymbol{x}) - g_2(\boldsymbol{A}\boldsymbol{z})^{-1}g_{12}(\boldsymbol{x}_1, \boldsymbol{A}\boldsymbol{z})\}^2] = O(h_{11}^4 + h_{33}^{-4} + n^{-1}h_{11}^{-d_1-d_2}).$$

由此可知，对于多指标模型，条件密度估计量的最优收敛速率取决于 d_2，而非 $(d_2 + d_3)$。

5 模拟与案例研究

本节报告条件密度和回归核估计量的数值性质。回归估计量的带宽矩阵 $\boldsymbol{H}$ 通过几种方法确定：使用优化算法进行留一最小二乘交叉验证（LSCV），或使用 R 中 np 软件包的 npregbw 函数（Li and Racine 2004），以及 MEKRO（White 等 2017）。平均积分平方误差（MISE）

$$\mathbb{E}_{\boldsymbol{T}}[\mathbb{E}[\{\widehat{m}(\boldsymbol{X})-m(\boldsymbol{X})\}^2|\boldsymbol{T}]$$

此外，还测量了 $\boldsymbol{T}$，其中 $\boldsymbol{T}$ 表示服从 $\boldsymbol{X}$分布的训练数据。在核条件密度估计中，带宽矩阵通过似然交叉验证和最小二乘交叉验证等方法确定（Hall 等2004）。np软件包中的npcdensbw函数用于计算带宽矩阵。所有核函数均为高斯核，且呈球对称，因此所有带宽矩阵都假定为对称矩阵（见注释1）。

以下两种设定属于条件独立情形，见White 等 （2017）。第一种情形为 $d_1=1$、 $d_2=2$、 $d_3=1$

$$Y=\sin(2\pi X_1)+\sin(\pi X_2)+0.5\epsilon_0,$$

其中（X_1， X_2， X_3）是取值于 $[0,1]^3$上的三元均匀随机变量。相互独立变量ϵ_0 服从标准正态分布。

第二种情形为 $d_1=1$、 $d_2=2$、 $d_3=8$

$$Y=\sin\{2\pi(X_1+X_2)/(1+X_3)\}+\epsilon_1,$$

其中，(X_1， $\cdots$， X_{10})是在 $[0,1]^{10}$ 上的均匀随机变量，且 ϵ_1服从正态分布，参数为$\mathbb{V}[\epsilon]=\mathbb{V}[2\pi$(\(X_1+X_2\)/(\(1+X_3\))]/3。该模型是多指标模型，其中

$$\boldsymbol{A}=\begin{pmatrix}1&1&0\\0&0&1\end{pmatrix}$$

是满足 $\mathbb{E}[\boldsymbol{X}_{1,i}|\boldsymbol{Z}=\boldsymbol{z}]=m(\boldsymbol{A}\boldsymbol{z})$的一个例子。

表1–5和表6–10分别对应第一种情况和第二种情况。表1–4展示了回归估计量的带宽取值和MISE。暴力搜索得到的带宽矩阵 $\boldsymbol{H}$应为

$$\boldsymbol{H}_s:=\begin{pmatrix}h&0&0\\0&h&0\\0&0&h\end{pmatrix},\boldsymbol{H}_d:=\begin{pmatrix}h_1&0&0\\0&h_2&0\\0&0&h_3\end{pmatrix},\boldsymbol{H}_c:=\begin{pmatrix}h_1&h_4&h_5\\h_4&h_2&h_6\\h_5&h_6&h_3\end{pmatrix},$$

其中自由度为1，3，6。MEKRO和npregbw搜索得到的最优矩阵形式为 $\boldsymbol{H}_d$；然而，其元素应为正数，这使其与暴力搜索得到的带宽矩阵有所不同。表5展示了带宽矩阵的取值

$$\begin{pmatrix}h_0&0&0&0\\0&h_1&0&0\\0&0&h_2&0\\0&0&0&h_3\end{pmatrix}$$

这些取值是由npcdensbw选出的条件密度估计量的带宽矩阵取值。

第三种情况是

$$Y=2(X_1+X_2)+X_3+X_4+\epsilon,$$

这来自 Conn and Li（2019）。(X_1， $\cdots$， X_9) 是 $[0,1]^9$, 上的均匀随机变量，而 ϵ 服从标准正态分布。该模型为单指标模型，阶数分别为 $d_1=1$、 $d_2=1$ 和 $d_3=8$。表 11–16 展示了带宽值和平均积分平方误差（MISE）。

表 1：MISE结果

n	标量	sd	对角	sd	对称	sd	npregbw	sd	MEKRO	sd
$Y =$ 正弦$(2\pi X_1) +$正弦$(\pi X_2) + 0.5\epsilon_0$										
100	0.180	0.130	0.093	0.047	0.165	0.084	0.087	0.031	0.553	0.060
300	0.098	0.098	0.047	0.055	0.105	0.067	0.033	0.006	0.554	0.044
500	0.091	0.133	0.036	0.055	0.084	0.038	0.023	0.004	0.549	0.041
1000	0.050	0.076	0.015	0.053	0.048	0.050	0.014	0.003	0.370	0.257
$Y =$ 正弦$\{2\pi(X_1 + X_2)/(1 + X_3)\} + \epsilon_1$										
100	0.439	0.063	0.324	0.178	0.498	0.212	0.210	0.092	0.495	0.037
300	0.372	0.047	0.139	0.129	0.152	0.166	0.079	0.010	0.501	0.022
500	0.344	0.057	0.101	0.113	0.093	0.117	0.053	0.006	0.492	0.018
1000	0.303	0.058	0.044	0.061	0.059	0.060	0.035	0.003	0.491	0.011
$Y = 2(X_1 + X_2) + X_3 + X_4 + \epsilon$										
100	0.381	0.117	0.422	0.236	0.698	0.328	0.477	0.246	0.812	0.098
300	0.265	0.101	0.155	0.096	0.186	0.093	0.158	0.053	0.785	0.067
500	0.211	0.068	0.103	0.062	0.102	0.034	0.097	0.030	0.793	0.052
1000	0.176	0.081	0.066	0.046	0.035	0.018	0.059	0.017	0.788	0.053

6 结论

非参数估计量受到维数灾难的影响；然而，已有证明表明，核估计量仅取决于因变量的相关变量数量。

本研究未提出带宽选择算法；然而，现有方法，例如允许较大带宽值的最小二乘交叉验证，仍然适用。假设真实密度 f 的核函数及其海森矩阵的所有元素均平方可积，这通常是核估计中的假设。则平均积分平方误差（MISE）的收敛速率与渐近均方误差相同。这意味着，在含有无关变量的核估计中，使 MISE 最小化的带宽矩阵的相应元素趋于无穷大。

可以通过对数变换等数据变换，放宽对底层分布矩可积性的严格假设。变换后的密度估计量的形式为

$$(n\|\boldsymbol{H}\|)^{-1}\sum_{i=1}^{n}(\prod_{j=1}^{d}x_j)^{-1}k(\boldsymbol{H}^{-1}(\ln\boldsymbol{x}-\ln\boldsymbol{X}_i)),$$

其中 ln 表示逐元素变换。矩条件 $\mathbb{E}[X_{i,k}^2] < \infty$ 变为 $\mathbb{E}[(\ln X_{i,\ k})^2] < \infty$，且不改变收敛速率。进一步的研究及其数值性质留待未来工作。

参考文献

Conn, D. 和 Li, G.（2019）。高维非参数回归中 Nadaraya–Watson 核估计量的神谕性质。《斯堪的纳维亚统计学杂志》，**46**（3），735–764。

Efromovich, S.（2010）。条件密度估计中的降维与自适应。《美国统计学会杂志》，**105**（490），761–774。

Feng, Y.、Wu, Y. 和 Stefanski, L.A.（2018）。通过偏好平滑带宽进行非参数独立性筛选。《统计规划与推断杂志》，**197**，1–14。

Hall, P.、Racine, J. 和 Li, Q.（2004）。交叉验证与条件概率密度的估计。《美国统计学会杂志》，**99**（468），1015–1026。

Hall, P., Li, Q., 和 Racine, J. S.（2007年）。存在无关回归变量时回归函数的非参数估计。《经济学与统计学评论》，**89**（4），784–789。

Jones, M. (1993年). 带宽较大时的核密度估计. 《澳大利亚统计学杂志》,**35**(3), 319–326.

Klock, T., Lanteri, A. 和 Vigogna, S. (2021年). 基于响应条件最小二乘法估计多指标模型. 《电子统计学杂志》,**15**(1), 589 – 629. Lafferty, J. 和 Wasserman, L. (2008). RODEO：稀疏贪心非参数回归. 《统计年鉴》, **36**(1), 28 – 63. Li, Q. 和 Racine, J. (2004). 交叉验证局部线性非参数回归 . 《中华统计学志》 , 第 485–512 页. Wand, M. P. 和 Jones, M. C. (1995). 《核平滑》. 施普林格美国公司. White, K. R., Stefanski, L. A. 和 Wu, Y. (2017). 使用测量误差选择似然进行核回归中的变量选择. 《美国统计学会杂志》, **112**(520), 1587–1597. Xia, Y. (2008). 多指标模型与降维. 《美国统计学会杂志》,**103**(484), 1631–1640.