

2 Nonparametric regression

2.1 Introduction

The last chapter considered nonparametric estimation of densities and distributions. We now turn to the estimation of conditional expectations, commonly referred to as regressions. We start with a suitable definition of conditional expectation. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $D \in \mathbb{N}$.

Recall that if $X : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}^D, \mathcal{B}(\mathbb{R}^D) := \mathcal{B}^D)$ is a random vector, then $X^{-1}(\mathcal{B}^D) \subset \mathcal{F}$ is a σ -algebra and we write $X^{-1}(\mathcal{B}^D) := \sigma(X)$, the σ -algebra generated by X . Now, consider a random variable $Y : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}, \mathcal{B})$. It is legitimate to ask when Y is a random variable with respect to $\sigma(X)$.¹ The following theorem provides a useful characterization.

Theorem 1. *Let $X : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}^D, \mathcal{B}^D)$ be a random vector and $Y : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ be a random variable. Y is $\sigma(X) - \mathcal{B}$ measurable if, and only if, there exists $m : (\mathbb{R}^D, \mathcal{B}^D) \rightarrow (\mathbb{R}, \mathcal{B})$ such that $Y = m(X)$, and m is $\mathcal{B}^D - \mathcal{B}$ measurable.*

Proof. ($\Leftarrow$) We want to show that for every $B \in \mathcal{B}$ we have $Y^{-1}(B) \in \sigma(X)$. But $Y^{-1}(B) = X^{-1}(m^{-1}(B))$ and by measurability of m , $m^{-1}(B) \in \mathcal{B}^D$ and since X is a random vector $X^{-1}(m^{-1}(B)) \in \sigma(X)$. Thus, Y is $\sigma(X) - \mathcal{B}$ -measurable.

($\Rightarrow$) Suppose $Y^{-1}(B) \in \sigma(X)$ for all $B \in \mathcal{B}$. First, assume that Y is a simple function. Then, for $k \in \mathbb{N}$ we have $Y(\omega) = \sum_{i=1}^k a_i I_{A_i}(\omega)$ for a_i distinct and $\{A_i\}_{i=1}^k$ pairwise-disjoint. In this case, $Y^{-1}(\{a_i\}) = A_i$ and by assumption $A_i \in \sigma(X)$. Hence, there exists $B_i \in \mathcal{B}^D$ such that $X^{-1}(B_i) = A_i$ by definition of $\sigma(X)$. Hence,

$$Y(\omega) = \sum_{i=1}^k a_i I_{X^{-1}(B_i)}(\omega) = \sum_{i=1}^k a_i I_{X \in B_i}(\omega).$$

Define $m(x) = \sum_{i=1}^k a_i I_{B_i}(x)$, then $Y = m(X)$, m is $\mathcal{B}^D$ -measurable. Thus, the implication is proved for every Y simple that is $\sigma(X)$ -measurable.

If $Y : (\Omega, \mathcal{F}, P) \rightarrow [0, \infty)$ then, by Lemma 5, there exist $Y_n(\omega)$ simple such that

$$Y(\omega) = \lim_{n \rightarrow \infty} Y_n(\omega), \quad 0 \leq Y_n(\omega) \leq Y_{n+1}(\omega).$$

If each Y_n is $\sigma(X) - \mathcal{B}$ measurable and $Y_n := m_n(X)$ as in the proof when Y is a simple function, set $m(x) = \limsup_{n \rightarrow \infty} m_n(x)$ and note $Y = \lim_{n \rightarrow \infty} Y_n = \lim_{n \rightarrow \infty} m_n(X)$.

Given that $m(X) := (\limsup_{n \rightarrow \infty} m_n)(X) = \limsup_{n \rightarrow \infty} m_n(X)$, $m(x)$ is $\mathcal{B}^D - \mathcal{B}$ -measurable. For general Y , write $Y = Y^+ - Y^-$ which reduces to the preceding case. $\square$

¹More generally, if $\mathcal{G}$ is a σ -algebra and $\mathcal{G} \subset \mathcal{F}$, we say that X is $\mathcal{G} - \mathcal{B}$ measurable if for all $B \in \mathcal{B}$, $X^{-1}(B) \in \mathcal{G}$. If G denotes the collection of all such σ -algebras, then it can be easily shown that $\sigma(X) = \bigcap_{\mathcal{G} \in G} \mathcal{G}$.

Let $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ be the space of random variables $X : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ such that $\int X^2(\omega) dP < \infty$. Let X be a random vector taking values in $\mathbb{R}^D$ defined in the same probability space with $\sigma(X) \subset \mathcal{F}$. Then, $\mathcal{L}^2(\Omega, \sigma(X), P) \subset \mathcal{L}^2(\Omega, \mathcal{F}, P)$ is a (Hilbert) space with the same inner product $(\int X(\omega)X'(\omega)dP)$ and $\mathcal{L}^2(\Omega, \sigma(X), P)$ is a closed subspace of $\mathcal{L}^2(\Omega, \mathcal{F}, P)$. By the Projection Theorem, we next define conditional expectation.

Definition 1. Let $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$. The conditional expectation of Y given X is the unique element $\hat{Y} \in \mathcal{L}^2(\Omega, \sigma(X), P)$ such that

$$E((Y - \hat{Y})s) = 0, \text{ for all } s \in \mathcal{L}^2(\Omega, \sigma(X), P).$$

We write $\hat{Y} = E(Y|X)$ or $\hat{Y} = E(Y|\sigma(X))$.

Remark 1. 1. Again, by the Projection Theorem,

$$\|Y - E(Y|X)\|_2 := \left(\int_{\Omega} |Y - E(Y|X)|^2 dP \right)^{1/2} \leq \|Y - s\|_2 \text{ for all } s \in \mathcal{L}^2(\Omega, \sigma(X), P).$$

Hence,

$$E(Y|X) := \arg \inf_{s \in \mathcal{L}^2(\Omega, \sigma(X), P)} \|Y - s\| := \arg \inf_{m \in \mathcal{M}} \|Y - m(X)\|.$$

where $\mathcal{M}$ is the set of Borel measurable functions from $\mathbb{R}^D$ to $\mathbb{R}$.

2. Since $E(Y|X)$ is $\sigma(X)$ -measurable, by Theorem 1, there exists $m : \mathbb{R}^D \rightarrow \mathbb{R}$ which is Borel measurable such that $E(Y|X) = m(X)$ and m is unique. Hence, we can write $E[(Y - m(X))g(X)] = 0$, for all $g : \mathbb{R}^D \rightarrow \mathbb{R}$ Borel measurable such that $\int g^2 dP < \infty$.
3. The requirement that Y be square integrable is not necessary for the definition of conditional expectation. It is convenient here because the Projection Theorem can be used to obtain existence and uniqueness of $E(Y|X)$. However, $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ is sufficient to obtain existence and uniqueness of $E(Y|X)$.
4. We may write

$$Y = E(Y|X) + U = m(X) + U \tag{1}$$

where $U = Y - E(Y|X) = Y - m(X)$ and by definition of conditional expectation $E(Ug(X)) = 0$ for any g that is $\mathcal{B}^D - \mathcal{B}$ measurable. Note, in particular, that $E(UX) = 0$ and by definition $E(U|X) = 0$.

Let $Z = \begin{pmatrix} Y & X^T \end{pmatrix}^T$ be a $(1 + D)$ -dimensional random vector defined on $(\Omega, \mathcal{F}, P)$ having distribution F . Assume that we observe a random sample $\{Z_i = (Y_i \ X_i^T)^T\}_{i=1}^n$

where $X_i^T = (X_{i1} \cdots X_{iD})$ with $D \in \mathbb{N}$. We call Y the regressand and the components of X the regressors. Our objective is to estimate the regression $m(x)$ for some $x \in S \subset \mathbb{R}^D$. We write $U_i = Y_i - m(X_i)$ and call it the regression error. As $\{U_i\}_{i=1}^n$ is a sequence of measurable functions of $\{Z_i\}_{i=1}^n$, an IID sequence of random vectors, it is itself an IID sequence of (unobserved) random variables.

2.2 The Nadaraya-Watson estimator

Assuming that Z has a density denoted by $f_Z(z) := f(y, x)$ and the existence of $f_X(x) = \int f(y, x)dy$ with $f_X(x) \neq 0$, the conditional density of Y given $X = x$ exists and is denoted by $f_{Y|X=x}(y) = \frac{f(y, x)}{f_X(x)}$. Thus, we write

$$m(x) := E(Y|X = x) = \int y f_{Y|X=x}(y)dy = \frac{1}{f_X(x)} \int y f(y, x)dy. \quad (2)$$

Hence, it is natural to propose the following estimator for $m(x)$.

Definition 2. The Nadaraya-Watson estimator for $m(x)$ is defined as

$$\hat{m}(x) = \frac{1}{\hat{f}_X(x)} \int y \hat{f}(y, x)dy, \quad (3)$$

where $\hat{f}$ and $\hat{f}_X$ are Rosenblatt kernel density estimators. That is,

$$\hat{f}(y, x) = \frac{1}{n \det(G_n)} \sum_{i=1}^n \mathcal{K}(G_n^{-1}(Z_i - (y, x')')) \quad \text{and} \quad \hat{f}_X(x) = \frac{1}{n \det(H_n)} \sum_{i=1}^n K(H_n^{-1}(X_i - x))$$

where G_n and H_n are bandwidth matrices and $\mathcal{K}$ and K are $D+1$ and D dimensional kernels.

Remark 2. 1. Let $h_{y,n}$ be the bandwidth associated with the first component of Z_i and $h_{1,n}, \dots, h_{D,n}$ be the bandwidths associated with the regressors. We note that

$$\begin{aligned} \int y \hat{f}(y, x)dy &= \frac{1}{n \det(G_n)} \int y \sum_{j=1}^n \mathcal{K}\left(G_n^{-1}\left(\begin{array}{c} Y_j - y \\ X_j - x \end{array}\right)\right) dy \\ &= \frac{1}{nh_{y,n}h_{1,n} \cdots h_{D,n}} \sum_{j=1}^n \int y \mathcal{K}\left(G_n^{-1}\left(\begin{array}{c} Y_j - y \\ X_j - x \end{array}\right)\right) dy \\ &= \frac{1}{n} \sum_{j=1}^n \int (Y_j - uh_{y,n})(h_{y,n} \cdots h_{D,n})^{-1} \mathcal{K}\left(u, H_n^{-1}(X_j - x)\right) h_{y,n} du \\ &= (nh_{1,n} \cdots h_{D,n})^{-1} \sum_{j=1}^n \int (Y_j - uh_{y,n}) \mathcal{K}\left(u, \frac{X_{j1} - x_1}{h_{1,n}}, \dots, \frac{X_{jD} - x_D}{h_{D,n}}\right) du \\ &= (nh_{1,n} \cdots h_{D,n})^{-1} \sum_{j=1}^n \left(Y_j \int \mathcal{K}\left(u, \frac{X_{j1} - x_1}{h_{1,n}}, \dots, \frac{X_{jD} - x_D}{h_{D,n}}\right) du \right. \\ &\quad \left. - h_{y,n} \int u \mathcal{K}\left(u, \frac{X_{j1} - x_1}{h_{1,n}}, \dots, \frac{X_{jD} - x_D}{h_{D,n}}\right) du \right). \end{aligned}$$

If $\int \mathcal{K}(u, x_1, \dots, x_D) du = K(x_1, \dots, x_D)$ and if $\int u \mathcal{K}(u, x_1, \dots, x_D) du = 0$, then

$$\int y \hat{f}(y, x) dy = (nh_{1,n} \dots h_{D,n})^{-1} \sum_{j=1}^n Y_j K(H_n^{-1}(X_j - x))$$

and the Nadaraya-Watson estimator can be written as

$$\hat{m}(x) = \frac{\sum_{j=1}^n K(H_n^{-1}(X_j - x)) Y_j}{\sum_{j=1}^n K(H_n^{-1}(X_j - x))} = \sum_{j=1}^n w_{jn}(x) Y_j,$$

where $w_{jn}(x; X_1, \dots, X_n) = \frac{K(H_n^{-1}(X_j - x))}{\sum_{i=1}^n K(H_n^{-1}(X_i - x))}$. Thus, $\hat{m}(x)$ can be written as a linear combination of the observations $Y_1, \dots, Y_n$ on the regressand Y with coefficients $w_{jn}(x; X_1, \dots, X_n)$. In addition, $\sum_{j=1}^n w_{jn}(x; X_1, \dots, X_n) = 1$. Hence, the Nadaraya-Watson estimator is a weighted average of the observations on the regressand.

2. In this context it is convenient to work with product kernels, where $\mathcal{K}(x_1, \dots, x_D) = \prod_{j=1}^D k(x_j)$ and k is a univariate kernel centered at zero.

The Nadaraya-Watson estimator can also be obtained from a different perspective that will motivate an entire new class of estimators. Hence, consider the following estimator.

Definition 3. The local constant estimator for $m(x)$ is denoted by $\hat{m}_0(x)$ and given by,

$$\begin{aligned} \hat{m}_0(x) &:= \arg \min_c \sum_{j=1}^n (Y_j - c)^2 K(H_n^{-1}(X_j - x)) := \arg \min_c (\mathbf{Y} - \mathbf{1}_n c)^T P(x) (\mathbf{Y} - \mathbf{1}_n c) \\ &:= \arg \min_c S_n(c) \end{aligned}$$

where $P(x) = \text{diag}\{K(H_n^{-1}(X_j - x))\}_{j=1}^n$, $\mathbf{1}_n = (1 \dots 1)^T$, $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ and c is a scalar.

Remark 3. Taking the first derivative of $S_n(c)$ with respect to c and finding $\hat{c} := \hat{m}_0(x)$ that sets it equal to zero gives $-2\mathbf{1}_n' P(x) \mathbf{Y} + 2\hat{m}_0(x) \mathbf{1}_n' P(x) \mathbf{1}_n = 0$ and

$$\hat{m}_0(x) = \frac{\sum_{j=1}^n K(H_n^{-1}(X_j - x)) Y_j}{\sum_{j=1}^n K(H_n^{-1}(X_j - x))}.$$

The second derivative of $S_n(c)$ with respect to c is $\mathbf{1}_n' P(x) \mathbf{1}_n = \sum_{i=1}^n K(H_n^{-1}(X_i - x))$. For a minimum, we need $\sum_{i=1}^n K(H_n^{-1}(X_i - x)) > 0$, which is true if $K(H_n^{-1}(X_i - x)) > 0$ for all i .

2.3 Local Polynomial Estimation

The optimization in the definition of the local constant estimator suggests a class of estimators that uses more flexible functions for the local approximation of $m(x)$. Taylor's Theorem suggests the following estimator, known in the literature as the local polynomial estimator.

Definition 4. Let $\alpha \in \mathbb{R}^D$ with $\alpha_i \in \{0, 1, \dots\}$, $|\alpha| = \sum_{i=1}^D \alpha_i$ and for $x \in \mathbb{R}^D$, $x^\alpha = x_1^{\alpha_1} \dots x_D^{\alpha_D}(x)$. In addition, let $\beta^\alpha := \partial^\alpha m(x) = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots \partial x_D^{\alpha_D}} m(x)$ and define the vector $\beta_{p,D} \in \mathbb{R}^c$ where $c = \binom{D+p}{p}$ such that

$$\beta_{p,D} = \left(\beta^{(0,\dots,0)} \quad \beta^{(0,1,\dots,0)} \quad \dots \quad \beta^{(0,0,\dots,1)} \quad \dots \right)^T.$$

The local polynomial estimator of order p for $m(x)$ is given by $\hat{m}_p(x) := \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} \hat{\beta}_{p,D}$, where

$$\hat{\beta}_{p,D} := \arg \min_{\beta} \sum_{i=1}^n \left(Y_i - \sum_{|\alpha| \leq p} \frac{1}{\alpha!} (X_i - x)^\alpha \beta^\alpha \right)^2 K(H_n^{-1}(X_i - x)) := \arg \min_{\beta} S_n(\beta). \quad (4)$$

Remark 4. 1. A much studied version of this estimator takes $p = 1$. $\hat{m}_1(x)$ is called the local linear estimator and in this case

$$\beta_{1,D} = \left(\beta^{(0,\dots,0)} \quad \beta^{(0,1,\dots,0)} \quad \dots \quad \beta^{(0,0,\dots,1)} \right)^T := \left(\beta^{(0,\dots,0)} \quad \beta_1^T \right)^T,$$

where $\beta_1 \in \mathbb{R}^D$. Then, if $\mathbf{X} := \begin{pmatrix} X_1 & \dots & X_n \end{pmatrix}^T$ and $\otimes$ denotes the Kronecker product

$$\begin{aligned} S_n(\beta) &:= \sum_{i=1}^n \left(Y_i - \sum_{|\alpha| \leq 1} \frac{1}{\alpha!} (X_i - x)^\alpha \beta^\alpha \right)^2 K(H_n^{-1}(X_i - x)) \\ &= \sum_{i=1}^n \left(Y_i - \beta^{(0,0,\dots,0)} - (X_i - x)^T \beta_1 \right)^2 K(H_n^{-1}(X_i - x)) \\ &= \left(\mathbf{Y} - \begin{pmatrix} 1_n & \mathbf{X} - (1_n \otimes x^T) \end{pmatrix} \begin{pmatrix} \beta^{(0,0,\dots,0)} \\ \beta_1 \end{pmatrix} \right)^T P(x) \\ &\quad \times \left(\mathbf{Y} - \begin{pmatrix} 1_n & \mathbf{X} - (1_n \otimes x^T) \end{pmatrix} \begin{pmatrix} \beta^{(0,0,\dots,0)} \\ \beta_1 \end{pmatrix} \right) \end{aligned}$$

If $R(x) = \begin{pmatrix} 1_n & \mathbf{X} - (1_n \otimes x^T) \end{pmatrix}$ the vector $\hat{\beta}_{1,D} := \begin{pmatrix} \hat{\beta}^{(0,\dots,0)} \\ \hat{\beta}_1 \end{pmatrix}$ that minimizes $S_n(\beta) = (\mathbf{Y} - R(x)\beta)^T P(x) (\mathbf{Y} - R(x)\beta)$ is given by

$$\hat{\beta}_{1,D} = \begin{pmatrix} \hat{\beta}^{(0,\dots,0)} \\ \hat{\beta}_1 \end{pmatrix} = \left(R(x)^T P(x) R(x) \right)^{-1} R(x)^T P(x) \mathbf{Y} \quad (5)$$

whenever $\det(R(x)^T P(x) R(x)) \neq 0$. Hence, the local polynomial estimator of order 1 for $m(x)$ is therefore given by

$$\hat{\beta}^{(0, \dots, 0)} := \hat{m}_1(x) = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} (R(x)^T P(x) R(x))^{-1} R(x)^T P(x) \mathbf{Y} := W(x) \mathbf{Y}.$$

where $W(x) := \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} (R(x)^T P(x) R(x))^{-1} R(x)^T P(x)$.

2. Note that estimators for $\frac{\partial^1}{\partial x_1^{\alpha_1} \dots \partial x_D^{\alpha_D}} m(x) = \frac{\partial}{\partial x_d} m(x)$ for $d = 1, \dots, D$ can be obtained directly from equation (5) by selecting the appropriate element from the vector $\begin{pmatrix} \hat{\beta}^{(0, \dots, 0)} \\ \hat{\beta}_1 \end{pmatrix}$. Hence, we can define the estimator

$$\widehat{\frac{\partial}{\partial x_d} m(x)} := \begin{pmatrix} 0 & \dots & 1 & \dots & 0 \end{pmatrix} \begin{pmatrix} \hat{\beta}^{(0, \dots, 0)} \\ \hat{\beta}_1 \end{pmatrix},$$

where the number 1 appears on the $d+1$ position of the vector $\begin{pmatrix} 0 & \dots & 1 & \dots & 0 \end{pmatrix}$.

3. Let I_N denote the identity matrix of dimension N and note that $R(x) \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 1_n$.

Then, since $(R(x)^T P(x) R(x))^{-1} R(x)^T P(x) R(x) = I_{D+1}$ and

$$\begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} I_{D+1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 1$$

we have

$$\begin{aligned} 1 &= \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} (R(x)^T P(x) R(x))^{-1} R(x)^T P(x) R(x) \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} (R(x)^T P(x) R(x))^{-1} R(x)^T P(x) 1_n = W(x) 1_n \end{aligned}$$

Hence, $\hat{m}_1(x)$ can be expressed as a weighted average of the observations on Y . The conclusion easily generalizes to any polynomial estimator of order p .

3 Asymptotic properties of the Nadaraya-Watson estimator

3.1 The case where $D = 1$

We start with consistency.

Theorem 2. Let $\left\{ \begin{pmatrix} Y_i & X_i \end{pmatrix}^T \right\}_{i=1}^n$ be a sequence of IID random vectors distributed as $\begin{pmatrix} Y & X \end{pmatrix}'$ and assume that a) $|K(u)| \leq C$ for all $|u| \in \mathbb{R}$; b) $\int |K(u)| du < \infty$ and $\int K(u) du = 1$; c) $|u|K(u) \rightarrow 0$ as $u \rightarrow \infty$. In addition, assume that the marginal density of X , $f_X(x) > 0$ and x is a point of continuity of f_X and m with $|m(x)|, f(x) < C$. If $V(Y|X) = \sigma^2$, $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$ as $n \rightarrow \infty$,

$$\hat{m}(x) - m(x) \xrightarrow{p} 0 \text{ as } n \rightarrow \infty.$$

Proof. By definition

$$\begin{aligned} \hat{m}(x) &= \frac{\sum_{j=1}^n K\left(\frac{X_j - x}{h_n}\right) Y_j}{\sum_{j=1}^n K\left(\frac{X_j - x}{h_n}\right)} = \frac{1}{\frac{1}{nh_n} \sum_{j=1}^n K\left(\frac{X_j - x}{h_n}\right)} \frac{1}{nh_n} \sum_{j=1}^n K\left(\frac{X_j - x}{h_n}\right) Y_j \\ &= \frac{1}{\hat{f}_X(x)} \frac{1}{nh_n} \sum_{j=1}^n K\left(\frac{X_j - x}{h_n}\right) Y_j, \end{aligned}$$

where $\hat{f}_X(x) = \frac{1}{nh_n} \sum_{j=1}^n K\left(\frac{X_j - x}{h_n}\right)$. From chapter 1, under the conditions in this theorem $\hat{f}_X(x) \xrightarrow{p} f_X(x)$. Hence, it suffices to show that

$$\hat{g}(x) = \frac{1}{nh_n} \sum_{j=1}^n K\left(\frac{X_j - x}{h_n}\right) Y_j \xrightarrow{p} m(x)f_X(x).$$

Here, we show that $E(\hat{g}(x)) \rightarrow m(x)f_X(x)$ and $V(\hat{g}(x)) \rightarrow 0$ as $n \rightarrow \infty$. By the Law of Iterated Expectations and the IID assumption

$$\begin{aligned} E(\hat{g}(x)) &= \frac{1}{h_n} E\left(K\left(\frac{X - x}{h_n}\right) Y\right) = \frac{1}{h_n} E_X\left(K\left(\frac{X - x}{h_n}\right) E(Y|X)\right) \\ &= \frac{1}{h_n} E_X\left(K\left(\frac{X - x}{h_n}\right) m(X)\right) \\ &= \frac{1}{h_n} \int K\left(\frac{u - x}{h_n}\right) m(u) f_X(u) du = \int K(v) m(x + h_n v) f_X(x + h_n v) dv. \end{aligned}$$

Now, given that $K \in \mathcal{L}^1$ and provided $|m(x)|, f_X(x) < C$ and x is a point of continuity of f_X and m , we conclude by dominated convergence that

$$E(\hat{g}(x)) \rightarrow f_X(x)m(x) \int K(u) du = f_X(x)m(x).$$

provided $\int K(u) du = 1$. Now, given that $V(Y|X) = \sigma^2$, by the Law of Iterated Expectations

$$V(\hat{g}(x)) = \sigma^2 \frac{1}{nh_n^2} \int K^2\left(\frac{u - x}{h_n}\right) f_X(u) du + \frac{1}{nh_n^2} V\left(K\left(\frac{X - x}{h_n}\right) m(X)\right).$$

Note that $V\left(K\left(\frac{X-x}{h_n}\right)m(X)\right) = E\left(K^2\left(\frac{X-x}{h_n}\right)m^2(X)\right) - \left(E\left(K\left(\frac{X-x}{h_n}\right)m(X)\right)\right)^2$. Furthermore,

$$E\left(K^2\left(\frac{X-x}{h_n}\right)m^2(X)\right) = h_n \int K^2(u)m^2(x-uh)f_X(x-uh)du$$

and

$$E\left(K\left(\frac{X-x}{h_n}\right)m(X)\right) = h_n \int K(u)m(x-uh)f_X(x-uh)du.$$

Then,

$$\begin{aligned} V(\hat{g}(x)) &= \sigma^2 \frac{1}{nh_n} \int K^2(u)f_X(x-uh)du + \frac{1}{nh_n} \int K^2(u)m^2(x-uh)f_X(x-uh)du \\ &\quad + \frac{1}{n} \left(\int K(u)m(x-uh)f_X(x-uh)du \right)^2. \end{aligned}$$

Given the bounds on $|m|$ and f_X we have by dominated convergence that as $n \rightarrow \infty$ the integrals in the previous equation display converge. Hence, if $nh_n \rightarrow \infty$ we have that $V(\hat{g}(x)) \rightarrow 0$. $\square$

We now give sufficient conditions for asymptotic normality of $\hat{m}(x)$ under suitable centering and normalization.

Theorem 3. Let $\left\{ \begin{pmatrix} Y_i & X_i \end{pmatrix}^T \right\}_{i=1}^n$ be a sequence of IID random vectors distributed as $\begin{pmatrix} Y & X \end{pmatrix}'$ and assume that a) $|K(u)| \leq C$ for all $u \in \mathbb{R}$; b) $\int |K(u)|du < \infty$ and $\int K(u)du = 1$; c) $|u||K(u)| \rightarrow 0$ as $|u| \rightarrow \infty$. In addition, assume that the marginal density of X , $f_X(x) > 0$ and x is a point of continuity of f_X and m with $|m(x)|, f(x) < C$. If $V(Y|X) = \sigma^2$, $E(|Y - m(X)|^{2+\delta}|X) < C < \infty$ for some $\delta > 0$, $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$ as $n \rightarrow \infty$,

$$(nh_n)^{1/2}(\hat{m}(x) - E(\hat{m}(x)|X)) \xrightarrow{d} N\left(0, \frac{\sigma^2}{f_X(x)} \int K^2(u)du\right)$$

Proof. Since $\hat{m}(x) = \frac{(nh_n)^{-1} \sum_{j=1}^n K(h_n^{-1}(X_j - x))Y_j}{(nh_n)^{-1} \sum_{j=1}^n K(h_n^{-1}(X_j - x))}$ and the denominator converges in probability to $f_X(x) > 0$, we are mainly concerned with the convergence in distribution of the numerator, denoted by $\hat{g}(x)$. Let $\mathcal{X} = \{X_1, \dots, X_n\}$ and note that

$$\begin{aligned} E(\hat{g}(x)|\mathcal{X}) &= \frac{1}{nh_n} \sum_{j=1}^n K\left(\frac{X_j - x}{h_n}\right) E(Y_j|\mathcal{X}) \\ &= \frac{1}{nh_n} \sum_{j=1}^n K\left(\frac{X_j - x}{h_n}\right) m(X_j) \text{ by the fact that } \{(Y_i, X_i)\}_{i=1}^n \text{ is IID.} \end{aligned}$$

Then,

$$\hat{g}(x) - E(\hat{g}(x)|\mathcal{X}) = \frac{1}{nh_n} \sum_{j=1}^n K\left(\frac{X_j - x}{h_n}\right) (Y_j - m(X_j)).$$

Let $Z_{jn} = \frac{K\left(\frac{X_j - x}{h_n}\right)(Y_j - m(X_j))}{nh_n}$ and note that $E(Z_{jn}) = 0$ and

$$\begin{aligned} V(Z_{jn}) &= E(Z_{jn}^2) = \frac{\sigma^2}{n^2 h_n^2} E\left(K^2\left(\frac{X_j - x}{h_n}\right)\right) \\ &= \frac{\sigma^2}{n^2 h_n^2} \int K^2(h_n^{-1}(u - x)) f_X(u) du \end{aligned}$$

Now, let $s_n^2 = \sum_{j=1}^n E(Z_{jn}^2) = \frac{\sigma^2}{nh_n} \frac{1}{h_n} \int K^2(h_n^{-1}(u - x)) f_X(u) du$ and

$$\begin{aligned} X_{jn} = \frac{Z_{jn}}{s_n} &= \frac{Z_{jn}}{(nh_n)^{-1/2} \left(\frac{\sigma^2}{h_n} \int K^2(h_n^{-1}(\alpha - x)) f_X(\alpha) d\alpha \right)^{1/2}} \\ &= \frac{K(h_n^{-1}(X_j - x))(Y_j - m(X_j))}{(nh_n)^{1/2} \left(\frac{\sigma^2}{h_n} \int K^2(h_n^{-1}(\alpha - x)) f_X(\alpha) d\alpha \right)^{1/2}} \end{aligned}$$

By Liapounov's CLT $\sum_{j=1}^n X_{jn} \xrightarrow{d} N(0, 1)$ provided that $\lim_{n \rightarrow \infty} \sum_{j=1}^n E(|X_{jn}|^{2+\delta}) = 0$ for some $\delta > 0$. Note that $|X_{jn}| = \frac{|K(h_n^{-1}(X_j - x))||Y_j - m(X_j)|}{(nh_n)^{1/2} c(n)^{1/2}}$ with $c(n) = \frac{\sigma^2}{h_n} \int K^2(h_n^{-1}(\alpha - x)) f_X(\alpha) d\alpha$ and $|X_{jn}|^{2+\delta} = \frac{|K(h_n^{-1}(X_j - x))|^{2+\delta} |Y_j - m(X_j)|^{2+\delta}}{(nh_n)^{1+\delta/2} c(n)^{1+\delta/2}}$. $c(n)$ is nonstochastic, therefore

$$E(|X_{jn}|^{2+\delta}) = (nh_n c(n))^{-1-\delta/2} E(|K(h_n^{-1}(X_j - x))|^{2+\delta} |Y_j - m(X_j)|^{2+\delta})$$

and

$$\sum_{j=1}^n E(|X_{jn}|^{2+\delta}) = (nh_n c(n))^{-1-\delta/2} \sum_{j=1}^n E(|K(h_n^{-1}(X_j - x))|^{2+\delta} |Y_j - m(X_j)|^{2+\delta}).$$

Now, if $E(|Y_j - m(X_j)|^{2+\delta} |X_j) < C$

$$\begin{aligned} E(|K(h_n^{-1}(X_j - x))|^{2+\delta} |Y_j - m(X_j)|^{2+\delta}) &= E(|K(h_n^{-1}(X_j - x))|^{2+\delta} E(|Y_j - m(X_j)|^{2+\delta} |X_j)) \\ &\leq C \int \left| K\left(\frac{\alpha - x}{h_n}\right) \right|^{2+\delta} f_X(\alpha) d\alpha. \end{aligned}$$

Consequently,

$$\begin{aligned} \sum_{j=1}^n E(|X_{jn}|^{2+\delta}) &\leq (nh_n c(n))^{-1-\delta/2} nC \int \left| K\left(\frac{\alpha - x}{h_n}\right) \right|^{2+\delta} f_X(\alpha) d\alpha \\ &= (nh_n)^{-\delta/2} (c(n))^{-1-\delta/2} C \int |K(u)|^{2+\delta} f_X(x + h_n u) du \\ &\leq (nh_n)^{-\delta/2} (c(n))^{-1-\delta/2} C C^{1+\delta} \int |K(u)| f_X(x + h_n u) du. \end{aligned}$$

By dominated convergence, as $n \rightarrow \infty$, $c(n) \rightarrow \sigma^2 f_X(x) \int K^2(u) du$ and $\int |K(u)| f_X(x + h_n u) du \rightarrow f_X(x) \int |K(u)| du < C$. Since $nh_n \rightarrow \infty$ we have that $\lim_{n \rightarrow \infty} \sum_{j=1}^n E(|X_{jn}|^{2+\delta}) = 0$. Hence,

$$\frac{\sum_{j=1}^n K(h_n^{-1}(X_j - x))(Y_j - m(X_j))}{(nh_n)^{1/2} (\sigma^2 h_n^{-1} \int K^2(h_n^{-1}(\alpha - x)) f(\alpha) d\alpha)^{1/2}} = \frac{(nh_n)^{1/2} (\hat{g}(x) - E(\hat{g}(x)|\mathcal{X}))}{c(n)^{1/2}} \xrightarrow{d} N(0, 1),$$

which implies that $(nh_n)^{1/2} (\hat{g}(x) - E(\hat{g}(x)|\mathcal{X})) \xrightarrow{d} N(0, \sigma^2 f_X(x) \int K^2(u) du)$.

Now, since

$$\hat{m}(x) - E(\hat{m}(x)|\mathcal{X}) = \frac{\hat{g}(x) - E(\hat{g}(x)|\mathcal{X})}{\hat{f}_X(x)},$$

we have that

$$(nh_n)^{1/2} (\hat{m}(x) - E(\hat{m}(x)|\mathcal{X})) = \frac{(nh_n)^{1/2} (\hat{g}(x) - E(\hat{g}(x)|\mathcal{X}))}{\hat{f}_X(x)} \xrightarrow{d} \frac{1}{f_X(x)} Z_n,$$

where $Z_n \sim N(0, \sigma^2 f_X(x) \int K^2(u) du)$. Consequently,

$$(nh_n)^{1/2} (\hat{m}(x) - E(\hat{m}(x)|\mathcal{X})) \xrightarrow{d} N\left(0, \frac{\sigma^2}{f_X(x)} \int K^2(u) du\right).$$

□

Remark 5. We write

$$\sqrt{nh_n}(\hat{m}(x) - m(x)) = \sqrt{nh_n}(\hat{m}(x) - E(\hat{m}(x)|\mathcal{X})) + \sqrt{nh_n}(E(\hat{m}(x)|\mathcal{X}) - m(x))$$

and observe that

$$\begin{aligned} E(\hat{m}(x)|\mathcal{X}) - m(x) &= \frac{1}{\sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right)} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) (m(X_i) - m(x)) \\ &= \frac{1}{\hat{f}_X(x)} \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) (m(X_i) - m(x)). \end{aligned}$$

Hence, to establish the asymptotic distribution of $\sqrt{nh_n}(\hat{m}(x) - m(x))$ it is necessary to investigate the order (in probability) of the conditional bias of the Nadaraya-Watson estimator, which will be done in subsection 3.3. First we state our results for consistency and asymptotic normality for the case where $D > 1$.

3.2 The case where $D > 1$

Theorems 2 and 7 extend easily to the case where $D > 1$. Hence, we state the next two theorems without proof.

Theorem 4. Let $\left\{ \begin{pmatrix} Y_i & X_i^T \end{pmatrix}^T \right\}_{i=1}^n$ be a sequence of IID random vectors distributed as $\begin{pmatrix} Y & X \end{pmatrix}'$ and assume that a) $|K(u)| \leq C$ for all $u \in \mathbb{R}^D$; b) $\int |K(u)|du < \infty$ and $\int K(u)du = 1$; c) $\|u\|_1 |K(u)| \rightarrow 0$ as $\|u\|_1 \rightarrow \infty$. In addition, assume that the marginal density of X , $f_X(x) > 0$ and x is a point of continuity of f_X and m with $|m(x)|, f(x) < C$. If $V(Y|X) = \sigma^2$, $h_{d,n} > 0$ for all n and $h_{d,n} \rightarrow 0$ as $n \rightarrow \infty$ and $n \det(H_n) \rightarrow \infty$ as $n \rightarrow \infty$,

$$\hat{m}(x) - m(x) \xrightarrow{p} 0 \text{ as } n \rightarrow \infty.$$

Theorem 5. Let $\left\{ \begin{pmatrix} Y_i & X_i^T \end{pmatrix}^T \right\}_{i=1}^n$ be a sequence of IID random vectors distributed as $\begin{pmatrix} Y & X \end{pmatrix}'$ and assume that a) $|K(u)| \leq C$ for all $u \in \mathbb{R}$; b) $\int |K(u)|du < \infty$ and $\int K(u)du = 1$; c) $\|u\|_1 |K(u)| \rightarrow 0$ as $\|u\|_1 \rightarrow \infty$. In addition, assume that the marginal density of X , $f_X(x) > 0$ and x is a point of continuity of f_X and m with $|m(x)|, f(x) < C$. If $V(Y|X) = \sigma^2$, $E(|Y - m(X)|^{2+\delta}|X) < C < \infty$ for some $\delta > 0$, $h_{d,n} \rightarrow 0$ and $n \det(H_n) \rightarrow \infty$ as $n \rightarrow \infty$,

$$(n \det(H_n))^{1/2}(\hat{m}(x) - E(\hat{m}(x)|X)) \xrightarrow{d} N\left(0, \frac{\sigma^2}{f_X(x)} \int K^2(u)du\right).$$

Remark 6. As in the case of density estimation, the multivariate Nadaraya-Watson estimator converges in distribution at a rate that is slower than the parametric rate. In addition, the speed of convergence decays exponentially with D . Hence, the curse of dimensionality manifests itself again in the case of the Nadaraya-Watson estimator.

3.3 Bias of the Nadaraya Watson Estimator

Recall that a sequence of random variables $\{Z_n\}_{n \in \mathbb{N}}$ is said to be bounded in probability if for all $\epsilon > 0$ there exists $C_\epsilon > 0$ such that for all $n \geq N_\epsilon$

$$P(|Z_n| > C_\epsilon) \leq \epsilon$$

and we write $Z_n = O_p(1)$. If $\{a_n\}_{n \in \mathbb{N}}$ is a nonstochastic sequence sequence of positive real numbers, we say that $Z_n = O_p(a_n)$ if $\frac{Z_n}{a_n} = O_p(1)$. It is a direct consequence of Markov's Inequality that: a) if $E(|Z_n|) = O(a_n)$ then $Z_n = O_p(a_n)$ and b) $E(Z_n^2) = O(a_n)$ then $Z_n = O_p(a_n^{1/2})$.

Theorem 6. Let $\left\{ \begin{pmatrix} Y_i & X_i \end{pmatrix}^T \right\}_{i=1}^n$ be a sequence of IID random vectors distributed as $\begin{pmatrix} Y & X \end{pmatrix}^T$ and assume that a) $|K(u)| \leq C$ for all $u \in \mathbb{R}$; b) $\int |K(u)|du < \infty$ and $\int K(u)du = 1$; c) $|u| |K(u)| \rightarrow 0$ as $|u| \rightarrow \infty$; d) $\int u K(u)du = 0$, $\int u^2 K(u)du := \mu_2$. In addition, assume that the marginal density of X , $f_X(x) > 0$, x is a point of continuity of

f_X and m with $|m(x)|, f(x) < C$ and $f, m \in \mathcal{C}_b^2$. If $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$ as $n \rightarrow \infty$, then for $I_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) (m(X_i) - m(x))$ we have

$$\begin{aligned} a) E(I_n(x)) &= \frac{h_n^2}{2} \mu_2 f_X(x) \left(m^{(2)}(x) + 2 \frac{m^{(1)}(x) f_X^{(1)}(x)}{f_X(x)} \right) + o(h_n^2) \text{ and } V(I_n(x)) = O(h_n/n), \\ b) I_n - \frac{h_n^2}{2} \mu^2 f_X(x) \left(m^{(2)}(x) + 2 \frac{m^{(1)}(x) f_X^{(1)}(x)}{f_X(x)} \right) &= O_p \left(h_n^2 + \frac{h_n^{1/2}}{n^{1/2}} \right). \end{aligned}$$

Proof. a) First, note that

$$\begin{aligned} E(I_n(x)) &= \int K(u)(m(x+hu) - m(x))f_X(x+hu)du \\ &= \int K(u)(m(x+hu)f_X(x+hu) - m(x)f_X(x) + m(x)f_X(x) - m(x)f_X(x+hu))du \\ &= \int K(u)(m(x+hu)f_X(x+hu) - m(x)f_X(x))du \\ &\quad - m(x) \int K(u)(f_X(x+hu) - f_X(x))du \\ &= \int K(u) \left([mf_X]^{(1)}(x)h_nu + \frac{1}{2}[mf_X]^{(2)}(x + \lambda h_nu)h_n^2u^2 \right) du \\ &\quad - m(x) \int K(u) \left(f_X^{(1)}(x)h_nu + \frac{1}{2}f_X^{(2)}(x + \lambda h_nu)h_n^2u^2 \right) du \\ &\text{for } \lambda \in (0, 1), \text{ since } m, f \in \mathcal{C}_b^2 \\ &= \frac{h_n^2}{2} \left(\int K(u)[mf_X]^{(2)}(x + \lambda h_nu)u^2du - m(x) \int K(u)f_X^{(2)}(x + \lambda h_nu)u^2du \right) \\ &\text{since } \int uK(u)du = 0. \end{aligned}$$

Also, since $m, f \in \mathcal{C}_b^2$ and $\int u^2|K(u)|du = \mu_2$, by Lebesgue's Dominated Convergence Theorem we have

$$\begin{aligned} \frac{1}{h_n^2} E(I_n(x)) &= \frac{1}{2} \left(\mu_2 [mf_X]^{(2)}(x) + o(1) - m(x)(\mu_2 f_X^{(2)}(x) + o(1)) \right) \\ &= \frac{1}{2} \mu_2 f_X(x) \left(m^{(2)}(x) + 2 \frac{m^{(1)}(x) f_X^{(1)}(x)}{f_X(x)} \right) + o(1). \end{aligned}$$

Second, given that $V(I_n) = E(I_n^2) - E(I_n)^2$ we investigate $E(I_n^2)$.

$$\begin{aligned} E(I_n^2) &= \frac{1}{n^2 h_n^2} \left(\sum_{i=1}^n E \left(K^2 \left(\frac{X_i - x}{h_n} \right) (m(X_i) - m(x))^2 \right) \right. \\ &\quad \left. + \sum_{i=1}^n \sum_{j=1, j \neq i}^n E \left(K \left(\frac{X_i - x}{h_n} \right) (m(X_i) - m(x)) K \left(\frac{X_j - x}{h_n} \right) (m(X_j) - m(x)) \right) \right) \end{aligned}$$

where

$$E \left(K^2 \left(\frac{X_i - x}{h_n} \right) (m(X_i) - m(x))^2 \right) = h_n \int K^2(u) (m(x + hu) - m(x))^2 f_X(x + hu) du.$$

By the fact that the sequence $\{X_i\}_{i \in \mathbb{N}}$ is IID

$$\begin{aligned} E \left(K \left(\frac{X_i - x}{h_n} \right) (m(X_i) - m(x)) K \left(\frac{X_j - x}{h_n} \right) (m(X_j) - m(x)) \right) \\ = h_n^2 \left(\int K(u) (m(x + hu) - m(x)) f_X(x + hu) du \right)^2. \end{aligned}$$

Consequently,

$$\begin{aligned} E(I_n^2) &= \frac{1}{nh_n} \int K^2(u) (m(x + hu) - m(x))^2 f_X(x + hu) du \\ &\quad + \left(1 - \frac{1}{n} \right) \left(\int K(u) (m(x + hu) - m(x)) f_X(x + hu) du \right)^2, \end{aligned}$$

and

$$\begin{aligned} V(I_n) &= \frac{1}{nh_n} \int K^2(u) (m(x + hu) - m(x))^2 f_X(x + hu) du \\ &\quad - \frac{1}{n} \left(\int K(u) (m(x + hu) - m(x)) f_X(x + hu) du \right)^2 = I_{n1} - I_{n2} \end{aligned}$$

where $I_{n2} = \frac{h_n^4}{n} \left(\frac{1}{2} \mu_2 f_X(x) \left(m^{(2)}(x) + 2 \frac{m^{(1)}(x) f_X^{(1)}(x)}{f_X(x)} \right) + o(1) \right)^2$. By the Mean Value Theorem, for $\lambda \in (0, 1)$, $m(x + hu) - m(x) = m^{(1)}(x + \lambda h_n u) h_n u$. Hence,

$$I_{n1} = \frac{h_n^2}{nh_n} \int u^2 K^2(u) (m^{(1)}(x + \lambda h_n u))^2 f_X(x + h_n u) du,$$

and given that $m, f \in \mathcal{C}_b^2$, by Lebesgue's Dominated Convergence Theorem we have

$$\int u^2 K^2(u) (m^{(1)}(x + \lambda h_n u))^2 f_X(x + h_n u) du \rightarrow (m^{(1)}(x))^2 f_X(x) \int u^2 K^2(u) du$$

and $\frac{nh_n}{h_n^2} V(I_n) = O(1)$.

b) Let $B(x) = \frac{1}{2} \mu_2 f_X(x) \left(m^{(2)}(x) + 2 \frac{m^{(1)}(x) f_X^{(1)}(x)}{f_X(x)} \right)$ and $\mathcal{I}_n = I_n - h_n^2 B(x)$. Then, $\mathcal{I}_n^2 = (I_n - E(I_n) + E(I_n) - h_n^2 B(x))^2$ and

$$\begin{aligned} E(\mathcal{I}_n^2) &= V(I_n) - (E(I_n) - h_n^2 B(x))^2 \\ &= O \left(\frac{h_n^2}{nh_n} \right) + o(h_n^4). \end{aligned}$$

Thus, by Markov's Inequality $\mathcal{I}_n = O_p \left(\left(\frac{h_n^2}{nh_n} + h_n^4 \right)^{1/2} \right) = O_p \left(\frac{h_n^{1/2}}{n^{1/2}} + h_n^2 \right)$. $\square$

Remark 7. Note that,

$$\begin{aligned}
E(\hat{m}(x)|\chi) &= m(x) + \frac{1}{\hat{f}_X(x)} I_n = m(x) + \frac{I_n}{\hat{f}_X(x)} - \frac{I_n}{f_X(x)} + \frac{I_n}{f_X(x)} \\
&= m(x) + \left(\frac{1}{\hat{f}_X(x)} - \frac{1}{f_X(x)} \right) I_n + \frac{I_n}{f_X(x)} \\
&= m(x) + (1 + o_p(1)) \frac{I_n}{f_X(x)} \\
&= m(x) + \frac{1}{f_X(x)} (1 + o_p(1)) (I_n - h_n^2 B(x) + h_n^2 B(x)) \\
&= m(x) + \frac{h_n^2 B(x)}{f_X(x)} + o_p(h_n^2) + (1 + o_p(1)) \frac{\mathcal{I}_n}{f_X(x)}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
\sqrt{nh_n} (\hat{m}(x) - E(\hat{m}(x)|\chi)) &= \sqrt{nh_n} \left(\hat{m}(x) - \left(m(x) + \frac{h_n^2 B(x)}{f_X(x)} + o_p(h_n^2) \right) \right) \\
&\quad + \sqrt{nh_n} \frac{\mathcal{I}_n}{f_X(x)}.
\end{aligned}$$

Since $\mathcal{I}_n = O_p \left(\frac{h_n}{(nh_n)^{1/2}} + h_n^2 \right)$ we have

$$\begin{aligned}
\sqrt{nh_n} (\hat{m}(x) - E(\hat{m}(x)|\chi)) &= \sqrt{nh_n} \left(\hat{m}(x) - \left(m(x) + \frac{h_n^2 B(x)}{f_X(x)} + o_p(h_n^2) \right) \right) \\
&\quad + \sqrt{nh_n} \left(\frac{h_n^{1/2}}{n^{1/2}} + h_n^2 \right) \frac{1}{f_X(x)} O_p(1).
\end{aligned}$$

If $nh_n^5 \rightarrow 0$ then,

$$\begin{aligned}
\sqrt{nh_n} (\hat{m}(x) - E(\hat{m}(x)|\chi)) &= \sqrt{nh_n} \left(\hat{m}(x) - \left(m(x) + \frac{h_n^2 B(x)}{f_X(x)} + o_p(h_n^2) \right) \right) \\
&\quad + o_p(1).
\end{aligned}$$

We immediately have the following Theorem.

Theorem 7. Let $\left\{ \begin{pmatrix} Y_i & X_i \end{pmatrix}^T \right\}_{i=1}^n$ be a sequence of IID random vectors distributed as $\begin{pmatrix} Y & X \end{pmatrix}^T$ and assume that a) $|K(u)| \leq C$ for all $u \in \mathbb{R}$; b) $\int |K(u)| du < \infty$ and $\int K(u) du = 1$; c) $|u| |K(u)| \rightarrow 0$ as $|u| \rightarrow \infty$; d) $\int u K(u) du = 0$, $\int u^2 K(u) du = \mu_2$. In addition, assume that the marginal density of X , $f_X(x) > 0$ and x is a point of continuity of f_X and m with $|m(x)|, f(x) < C$. If $V(Y|X) = \sigma^2$, $E(|Y - m(X)|^{2+\delta}|X) < C < \infty$ for some $\delta > 0$, $h_n \rightarrow 0$, $nh_n \rightarrow \infty$ and $nh_n^5 \rightarrow 0$ as $n \rightarrow \infty$,

$$\sqrt{nh_n} \left(\hat{m}(x) - \left(m(x) + \frac{h_n^2 B(x)}{f_X(x)} + o_p(h_n^2) \right) \right) \xrightarrow{d} N \left(0, \frac{\sigma^2}{f_X(x)} \int K^2(\psi) d\psi \right).$$

4 Asymptotic properties of the Local Linear Estimator

4.1 Algebraic results

Before we study the asymptotic behavior of the local linear estimator we give some useful algebraic facts. We start by noting that since

$$R(x) = \begin{pmatrix} 1 & X_{11} - x_1 & \cdots & X_{1D} - x_D \\ 1 & X_{21} - x_1 & \cdots & X_{2D} - x_D \\ \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} - x_1 & \cdots & X_{nD} - x_D \end{pmatrix} \text{ and } P(x) = \text{diag} \left\{ K \left(H_n^{-1}(X_t - x) \right) \right\}_{t=1}^n$$

matrix multiplication gives

$$\begin{aligned} \frac{1}{n \det(H_n)} R(x)^T P(x) R(x) &:= S(x) = \begin{pmatrix} s_0(x) & s_1(x) & \cdots & s_D(x) \\ s_1(x) & s_{(1,1)}(x) & \cdots & s_{(1,D)}(x) \\ \vdots & \vdots & \ddots & \vdots \\ s_D(x) & s_{(D,1)}(x) & \cdots & s_{(D,D)}(x) \end{pmatrix} \\ &= \begin{pmatrix} s_0(x) & S_1(x)^T \\ S_1(x) & S_2(x) \end{pmatrix} \end{aligned}$$

where

$$s_0(x) = \frac{1}{n \det(H_n)} \sum_{t=1}^n K \left(H_n^{-1}(X_t - x) \right), \quad s_j(x) = \frac{1}{n \det(H_n)} \sum_{t=1}^n K \left(H_n^{-1}(X_t - x) \right) (X_{tj} - x_j)$$

and

$$s_{(i,j)}(x) = \frac{1}{n \det(H_n)} \sum_{t=1}^n K \left(H_n^{-1}(X_t - x) \right) (X_{ti} - x_i) (X_{tj} - x_j) \text{ for } i, j = 1, \dots, D.$$

Note that $s_{(i,j)}(x) = s_{(j,i)}(x)$. Furthermore,

$$\frac{1}{n \det(H_n)} R^T(x) P(x) \mathbf{Y} := \tau(x) = \begin{pmatrix} \tau_0(x) \\ \tau_1(x) \\ \vdots \\ \tau_D(x) \end{pmatrix}$$

where $\tau_0(x) = \frac{1}{n \det(H_n)} \sum_{t=1}^n K \left(H_n^{-1}(X_t - x) \right) Y_t$, $\tau_j(x) = \frac{1}{n \det(H_n)} \sum_{t=1}^n K \left(H_n^{-1}(X_t - x) \right) (X_{tj} - x_j) Y_t$.

Hence, if $S(x)$ is invertible we may write the local linear estimator as

$$\hat{\beta}_{1,D}(x) = S(x)^{-1} \tau(x).$$

Lemma 1. Assume that all partial derivatives of order 1 of $m(x)$ exist for $x \in \mathbb{R}^D$. Denote these partial derivatives by $\frac{\partial}{\partial_d} m(x)$ for $d = 1, \dots, D$ and define

$$m^{(1)}(x) := \left(\frac{\partial}{\partial_1} m(x) \quad \frac{\partial}{\partial_2} m(x) \quad \cdots \quad \frac{\partial}{\partial_D} m(x) \right)^T.$$

In addition, define $\mathbf{m} := \left(m(X_1) \quad m(X_2) \quad \cdots \quad m(X_n) \right)^T$, $\mathbf{U} := \left(U_1 \quad U_2 \quad \cdots \quad U_n \right)^T$ where $U_t = Y_t - m(X_t)$. Also, put

$$G_n = \text{diag}\{1, \text{diag}\{1/h_d^2\}_{d=1}^D\}$$

and

$$N_n = \text{diag}\{\sqrt{n \det(H_n)}, \text{diag}\{\sqrt{n \det(H_n)} h_d\}_{d=1}^D\},$$

both $D+1$ -diagonal matrices. Then, letting $\beta_{1,D}(x) = \left(m(x) \quad \cdots \quad m^{(1)}(x)^T \right)^T$

$$\begin{aligned} & N_n \left(\hat{\beta}_{1,D}(x) - \beta_{1,D}(x) - (G_n S(x))^{-1} \frac{1}{n \det(H_n)} G_n R^T(x) P(x) (\mathbf{m} - R(x) \beta(x)) \right) \\ &= N_n (G_n S(x))^{-1} \frac{1}{n \det(H_n)} G_n R^T(x) P(x) \mathbf{U} \end{aligned}$$

Proof. We can immediately write

$$\begin{aligned} \hat{\beta}_{1,D}(x) - \beta_{1,D}(x) &= S(x)^{-1} \frac{1}{n \det(H_n)} R^T(x) P(x) \mathbf{Y} - S(x)^{-1} \frac{1}{n \det(H_n)} R^T(x) P(x) R(x) \beta_{1,D}(x) \\ &= (G_n S(x))^{-1} \frac{1}{n \det(H_n)} G_n R^T(x) P(x) (\mathbf{m} + \mathbf{U} - R(x) \beta_{1,D}(x)) \\ &= (G_n S(x))^{-1} \frac{1}{n \det(H_n)} G_n R^T(x) P(x) (\mathbf{m} - R(x) \beta_{1,D}(x)) \\ &\quad + (G_n S(x))^{-1} \frac{1}{n \det(H_n)} G_n R^T(x) P(x) \mathbf{U}. \end{aligned}$$

Consequently,

$$\begin{aligned} & N_n \left(\hat{\beta}_{1,D}(x) - \beta_{1,D}(x) - (G_n S(x))^{-1} \frac{1}{n \det(H_n)} G_n R^T(x) P(x) (\mathbf{m} - R(x) \beta_{1,D}(x)) \right) \\ &= N_n (G_n S(x))^{-1} \frac{1}{n \det(H_n)} G_n R^T(x) P(x) \mathbf{U}. \end{aligned}$$

□

4.2 The case where $D = 1$

If $D = 1$, we have from the previous subsection that

$$R(x) = \begin{pmatrix} 1 & X_1 - x \\ 1 & X_2 - x \\ \vdots & \\ 1 & X_n - x \end{pmatrix}, \quad P(x) = \text{diag} \left\{ K \left(h_n^{-1}(X_t - x) \right) \right\}_{t=1}^n$$

and matrix multiplication gives

$$\begin{aligned} \frac{1}{nh_n} R(x)^T P(x) R(x) &:= S(x) = \begin{pmatrix} s_0(x) & s_1(x) \\ s_1(x) & s_{(1,1)}(x) \end{pmatrix} \\ &= \begin{pmatrix} s_0(x) & S_1(x)^T \\ S_1(x) & S_2(x) \end{pmatrix} \end{aligned}$$

where

$$\begin{aligned} s_0(x) &= \frac{1}{nh_n} \sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right), \quad s_1(x) = \frac{1}{nh_n} \sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right) (X_t - x), \quad \text{and} \\ s_{(1,1)}(x) &:= s_2(x) = \frac{1}{nh_n} \sum_{t=1}^n K\left(\frac{X_t - x}{h_n}\right) (X_t - x)^2. \end{aligned}$$

Furthermore, $G_n = \text{diag}\{1, 1/h_n^2\}$, $N_n = \text{diag}\{\sqrt{nh_n}, \sqrt{nh_n}h_n\}$ and

$$G_n S(x) = \begin{pmatrix} s_0(x) & s_1(x) \\ s_1(x)/h_n^2 & s_{(1,1)}(x)/h_n^2 \end{pmatrix}.$$

From chapter 1 we know that $s_0(x) = f_X(x) + o_p(1)$. The next lemma gives the order of the remaining terms.

Lemma 2. Assume $f_X \in \mathcal{C}_b^2$, $K \geq 0$, $\int K(u)du = 1$, $\int uK(u)du = 0$, $\int u^2K(u)du = \mu_2$, $\int u^2K^2(u)du < \infty$ and $\int u^4K^2(u)du < \infty$. If $h_n \rightarrow 0$, $nh_n^3 \rightarrow \infty$ we have

$$G_n S(x) = \begin{pmatrix} f_X(x) & 0 \\ \mu_2 f_X^{(1)}(x) & \mu_2 f_X(x) \end{pmatrix} + \begin{pmatrix} o_p(1) & O_p(h_n^2) \\ o_p(1) & o_p(1) \end{pmatrix} = Q + \begin{pmatrix} o_p(1) & O_p(h_n^2) \\ o_p(1) & o_p(1) \end{pmatrix}.$$

Proof. Note that $E(s_1(x)) = h_n \int uK(u)f_X(x + h_n u)du$ and

$$V(s_1(x)) = \frac{h_n}{n} \int K^2(u)u^2 f_X(x + h_n u)du - \frac{h_n^2}{n} \left(\int K(u)u f_X(x + h_n u)du \right)^2.$$

Hence, $V(s_1(x)) = O\left(\frac{h_n}{n}\right)$. Also, for $\lambda \in (0, 1)$

$$\begin{aligned} \frac{1}{h_n^2} E(s_1(x)) &= \frac{1}{h_n} \int uK(u)f_X(x + h_n u)du = \frac{1}{h_n} \int uK(u)[f_X(x) + h_n u f_X^{(1)}(x + \lambda h_n u)]du \\ &= \int u^2 K(u) f_X^{(1)}(x + \lambda h_n u)du \rightarrow f_X^{(1)}(x) \int u^2 K(u)du := \mu_2 f_X^{(1)}(x) \text{ as } n \rightarrow \infty. \end{aligned}$$

Now,

$$\begin{aligned} V\left(\frac{1}{h_n^2} s_1(x)\right) &= \frac{1}{nh_n^3} \int K^2(u)u^2 f_X(x + h_n u)du - \frac{1}{nh_n^2} \left(\int K(u)u f_X(x + h_n u)du \right)^2 \\ &\rightarrow 0 \text{ if } nh_n^3 \rightarrow \infty. \end{aligned}$$

Hence, $s_1(x) = O_p(h_n^2)$ and $h_n^{-2} s_1(x) - \mu_2 f_X^{(1)}(x) = o_p(1)$.

Lastly, $E(s_{(1,1)}(x)) = h_n^2 \int K(u)u^2 f_X(x + h_n u) du$ and

$$V(s_{(1,1)}(x)) = \frac{h_n^3}{n} \int K^2(u)u^4 f_X(x + h_n u) du - \frac{h_n^4}{n} \left(\int K(u)u^2 f_X(x + h_n u) du \right)^2.$$

Hence, $\frac{1}{h_n^2} E(s_{(1,1)}(x)) = \int K(u)u^2 f_X(x + h_n u) du \rightarrow f_X(x)\mu_2$ as $n \rightarrow \infty$ and

$$V(h_n^{-2} s_{(1,1)}(x)) = \frac{1}{nh_n} \int K^2(u)u^4 f_X(x + h_n u) du - \frac{1}{n} \left(\int K(u)u^2 f_X(x + h_n u) du \right)^2.$$

Then, $h_n^{-2} s_{(1,1)}(x) - f_X(x)\mu_2 = o_p(1)$. □

Remark 8. 1. If $f_X(x) > 0$ and $\mu_2 > 0$, Q is invertible.

2. It follows immediately from Slutsky Theorem and invertibility of Q that $(G_n S(x))^{-1} - Q^{-1} = o_p(1)$.

3. Recall the formula for the inverse of a partitioned matrix: if A_{11} , A_{22} and the Schur complements $A_{11} - A_{12}A_{22}^{-1}A_{21}$ and $A_{22} - A_{21}A_{11}^{-1}A_{12}$ are invertible, then

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & -A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \\ -A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{pmatrix}.$$

Applied to Q , this gives

$$Q^{-1} = \begin{pmatrix} \frac{1}{f_X(x)} & 0 \\ -\frac{f_X^{(1)}(x)}{f_X(x)^2} & \frac{1}{\mu_2 f_X(x)} \end{pmatrix}.$$

Applied to $G_n S(x)$, with $A_{11} = s_0(x)$, $A_{12} = s_1(x)$, $A_{21} = s_1(x)/h_n^2$ and $A_{22} = s_{(1,1)}(x)/h_n^2$ (all conditions hold with probability approaching one), and using Lemma 2,

$$\begin{aligned} A_{11} - A_{12}A_{22}^{-1}A_{21} &= s_0(x) - \frac{s_1(x)^2}{s_{(1,1)}(x)} = s_0(x) + O_p(h_n^2) = f_X(x) + o_p(1), \\ A_{22} - A_{21}A_{11}^{-1}A_{12} &= \frac{s_{(1,1)}(x)}{h_n^2} - \frac{s_1(x)}{h_n^2} \frac{s_1(x)}{s_0(x)} = \mu_2 f_X(x) + o_p(1), \\ A_{11}^{-1}A_{12} &= \frac{s_1(x)}{s_0(x)} = O_p(h_n^2), \quad A_{22}^{-1}A_{21} = \frac{s_1(x)}{s_{(1,1)}(x)} = \frac{f_X^{(1)}(x)}{f_X(x)} + o_p(1). \end{aligned}$$

Therefore,

$$(G_n S(x))^{-1} = Q^{-1} + \begin{pmatrix} o_p(1) & O_p(h_n^2) \\ o_p(1) & o_p(1) \end{pmatrix}.$$

The $(1, 1)$ error term is only $o_p(1)$ because the Lemma 2 gives $s_0(x) - f_X(x) = o_p(1)$, not $O_p(h_n^2)$.

4. Note that if $m \in \mathcal{C}_b^2$, the i^{th} element of $\mathbf{m} - R(x)\beta_{1,1}(x)$ is given by $m(X_t) - m(x) - m^{(1)}(x)(X_t - x) = \frac{1}{2}m^{(2)}(x + \lambda(X_t - x))(X_t - x)^2$

Theorem 8. Let $\{(Y_i, X_i)^T\}_{i=1}^n$ be a sequence of IID random vectors distributed as $(Y, X)^T$, with $m(x) := E(Y|X = x)$, $U := Y - m(X)$ and $\beta_{1,1}(x) := (m(x), m^{(1)}(x))^T$. Assume that

- a) K is symmetric about zero, has compact support and $|K(u)| \leq C$ for all $u \in \mathbb{R}$;
- b) $\int K(u) du = 1$, $\mu_2 := \int u^2 K(u) du > 0$ and put $\nu_2 := \int u^2 K^2(u) du$;
- c) $m, f_X \in \mathcal{C}_b^2$ and $f_X(x) > 0$, where f_X is the marginal density of X ;
- d) $V(Y|X) = \sigma^2$ and $E(|U|^{2+\delta} | X) \leq C < \infty$ for some $\delta > 0$;
- e) $h_n \rightarrow 0$, $nh_n^3 \rightarrow \infty$ and $nh_n^5 = O(1)$ as $n \rightarrow \infty$.

Then, with $N_n = \text{diag}\{\sqrt{nh_n}, \sqrt{nh_n}h_n\}$,

$$N_n \left(\hat{\beta}_{1,1}(x) - \beta_{1,1}(x) - \begin{pmatrix} \frac{1}{2}h_n^2\mu_2 m^{(2)}(x) \\ 0 \end{pmatrix} \right) \xrightarrow{d} N \left(\mathbf{0}, \frac{\sigma^2}{f_X(x)} \begin{pmatrix} \int K^2(u) du & 0 \\ 0 & \frac{\nu_2}{\mu_2^2} \end{pmatrix} \right).$$

Proof. Write the last equality in Lemma 1 (with $D = 1$) as

$$T_n := N_n \left(\hat{\beta}_{1,1}(x) - \beta_{1,1}(x) - B_n \right) = N_n (G_n S(x))^{-1} G_n v_n,$$

where $B_n := (G_n S(x))^{-1} \frac{1}{nh_n} G_n R(x)^T P(x) (\mathbf{m} - R(x)\beta_{1,1}(x))$ and

$$v_n := \frac{1}{nh_n} R(x)^T P(x) \mathbf{U} = \begin{pmatrix} \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) U_i \\ \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) (X_i - x) U_i \end{pmatrix} := \begin{pmatrix} v_{n1} \\ v_{n2} \end{pmatrix}.$$

Step 1: moments of v_n . By assumption, $E(v_n) = \mathbf{0}$, and the change of variables $X_i = x + h_n u$ gives

$$\begin{aligned} V(v_{n1}) &= \frac{\sigma^2}{nh_n} \int K^2(u) f_X(x + h_n u) du, & V(v_{n2}) &= \frac{\sigma^2 h_n}{n} \int u^2 K^2(u) f_X(x + h_n u) du, \\ E(v_{n1} v_{n2}) &= \frac{\sigma^2}{n} \int u K^2(u) f_X(x + h_n u) du. \end{aligned}$$

Hence $\sqrt{nh_n} v_{n1} = O_p(1)$ and $v_{n2} = O_p((h_n/n)^{1/2})$, that is, $\sqrt{nh_n} v_{n2} = O_p(h_n)$ and $\sqrt{nh_n} v_{n2}/h_n = O_p(1)$.

Step 2: decomposition of T_n . Let $E_n := (G_n S(x))^{-1} - Q^{-1}$. By Remark 8.3, $E_n = \begin{pmatrix} o_p(1) & O_p(h_n^2) \\ o_p(1) & o_p(1) \end{pmatrix}$, and $G_n v_n = (v_{n1}, v_{n2}/h_n^2)^T$. Then

$$T_n = N_n E_n G_n v_n + N_n Q^{-1} G_n v_n := T_{n1} + T_{n2},$$

with

$$T_{n1} = \begin{pmatrix} o_p(1)\sqrt{nh_n}v_{n1} + O_p(1)\sqrt{nh_n}v_{n2} \\ o_p(1)h_n\sqrt{nh_n}v_{n1} + o_p(1)\sqrt{nh_n}v_{n2}/h_n \end{pmatrix} = o_p(1)$$

by Step 1. Furthermore,

$$T_{n2} = \begin{pmatrix} \frac{1}{f_X(x)}\sqrt{nh_n}v_{n1} \\ -h_n\frac{f_X^{(1)}(x)}{f_X^2(x)}\sqrt{nh_n}v_{n1} + \frac{1}{\mu_2 f_X(x)}\frac{\sqrt{nh_n}}{h_n}v_{n2} \end{pmatrix} = T_{n2}^* + \begin{pmatrix} 0 \\ o_p(1) \end{pmatrix}, \quad T_{n2}^* := \begin{pmatrix} \frac{1}{f_X(x)}\sqrt{nh_n}v_{n1} \\ \frac{1}{\mu_2 f_X(x)}\frac{\sqrt{nh_n}}{h_n}v_{n2} \end{pmatrix},$$

since $h_n\sqrt{nh_n}v_{n1} = O_p(h_n) = o_p(1)$. Hence $T_n = T_{n2}^* + o_p(1)$.

Step 3: asymptotic normality of T_{n2}^ .* Let

$$\Sigma := \frac{\sigma^2}{f_X(x)} \begin{pmatrix} \int K^2(u) du & 0 \\ 0 & \frac{\nu_2}{\mu_2^2} \end{pmatrix}.$$

By the Cramér–Wold device it suffices to show that $a^T T_{n2}^* \xrightarrow{d} N(0, a^T \Sigma a)$ for every $a = (a_1, a_2)^T \in \mathbb{R}^2$, $a \neq \mathbf{0}$. Write $a^T T_{n2}^* = \sum_{i=1}^n \xi_{ni}$, where

$$\xi_{ni} := \frac{1}{\sqrt{nh_n}} K\left(\frac{X_i - x}{h_n}\right) \left(\frac{a_1}{f_X(x)} + \frac{a_2}{\mu_2 f_X(x)} \frac{X_i - x}{h_n} \right) U_i.$$

For each n the ξ_{ni} are i.i.d. with $E(\xi_{ni}) = 0$. By Step 1,

$$\begin{aligned} V(a^T T_{n2}^*) &= \frac{a_1^2}{f_X^2(x)} nh_n V(v_{n1}) + \frac{2a_1 a_2}{\mu_2 f_X^2(x)} n E(v_{n1} v_{n2}) + \frac{a_2^2}{\mu_2^2 f_X^2(x)} \frac{n}{h_n} V(v_{n2}) \\ &\rightarrow \frac{a_1^2}{f_X^2(x)} \sigma^2 f_X(x) \int K^2(u) du + 0 + \frac{a_2^2}{\mu_2^2 f_X^2(x)} \sigma^2 f_X(x) \nu_2 = a^T \Sigma a, \end{aligned}$$

where the cross term vanishes because $\int u K^2(u) du = 0$ by the symmetry of K . Moreover, by assumption,

$$\begin{aligned} \sum_{i=1}^n E|\xi_{ni}|^{2+\delta} &\leq \frac{Cn h_n}{(nh_n)^{1+\delta/2}} \int |K(u)|^{2+\delta} \left| \frac{a_1}{f_X(x)} + \frac{a_2 u}{\mu_2 f_X(x)} \right|^{2+\delta} f_X(x + h_n u) du \\ &= O((nh_n)^{-\delta/2}) \rightarrow 0, \end{aligned}$$

so Liapunov's condition holds. Since $a^T \Sigma a > 0$, Liapunov's CLT gives $a^T T_{n2}^* \xrightarrow{d} N(0, a^T \Sigma a)$. Therefore $T_{n2}^* \xrightarrow{d} N(\mathbf{0}, \Sigma)$ and, by Step 2, $T_n \xrightarrow{d} N(\mathbf{0}, \Sigma)$.

Step 4: the bias term. By Remark 8.4, the i^{th} element of $\mathbf{m} - R(x)\beta_{1,1}(x)$ is $\frac{1}{2}g(X_i)(X_i - x)^2$, where

$$g(X_i) := \frac{2[m(X_i) - m(x) - m^{(1)}(x)(X_i - x)]}{(X_i - x)^2} = m^{(2)}(x + \lambda_i(X_i - x)), \quad \lambda_i \in (0, 1).$$

Thus $|g| \leq \sup |m^{(2)}| < \infty$ and $g(x + h_n u) \rightarrow m^{(2)}(x)$ for every u as $h_n \rightarrow 0$. Hence $\frac{1}{nh_n} G_n R(x)^T P(x) (\mathbf{m} - R(x)\beta_{1,1}(x)) = \frac{1}{2}(z_{n1}, z_{n2})^T$ with

$$z_{n1} = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) g(X_i)(X_i - x)^2, \quad z_{n2} = \frac{1}{nh_n^3} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) g(X_i)(X_i - x)^3.$$

By dominated convergence,

$$\begin{aligned} E(z_{n1}) &= h_n^2 \int u^2 K(u) g(x + h_n u) f_X(x + h_n u) du = h_n^2 (\mu_2 m^{(2)}(x) f_X(x) + o(1)), \\ E(z_{n2}) &= h_n \int u^3 K(u) g(x + h_n u) f_X(x + h_n u) du = h_n \left(m^{(2)}(x) f_X(x) \int u^3 K(u) du + o(1) \right) \\ &= o(h_n), \end{aligned}$$

since $\int u^3 K(u) du = 0$. Also,

$$V(z_{n1}) \leq \frac{1}{nh_n^2} E \left[K^2 \left(\frac{X_i - x}{h_n} \right) g^2(X_i) (X_i - x)^4 \right] = O \left(\frac{h_n^3}{n} \right) = o(h_n^4), \quad V(z_{n2}) = O \left(\frac{h_n}{n} \right) = o(h_n^2),$$

because $nh_n \rightarrow \infty$. Consequently, $z_{n1} = h_n^2 (\mu_2 m^{(2)}(x) f_X(x) + o_p(1))$ and $z_{n2} = o_p(h_n)$.

Now write

$$N_n B_n = \frac{1}{2} N_n E_n \begin{pmatrix} z_{n1} \\ z_{n2} \end{pmatrix} + \frac{1}{2} N_n Q^{-1} \begin{pmatrix} z_{n1} \\ z_{n2} \end{pmatrix} := B_{n1} + B_{n2}.$$

Using the orders of E_n , z_{n1} and z_{n2} ,

$$B_{n1} = \frac{1}{2} \begin{pmatrix} \sqrt{nh_n} [o_p(1)z_{n1} + O_p(h_n^2)z_{n2}] \\ h_n \sqrt{nh_n} [o_p(1)z_{n1} + o_p(1)z_{n2}] \end{pmatrix} = \begin{pmatrix} o_p(\sqrt{nh_n^5}) \\ o_p(\sqrt{nh_n^5}) \end{pmatrix} = o_p(1),$$

and

$$B_{n2} = \frac{1}{2} \begin{pmatrix} \sqrt{nh_n} \frac{z_{n1}}{f_X(x)} \\ h_n \sqrt{nh_n} \left[-\frac{f_X^{(1)}(x)}{f_X^2(x)} z_{n1} + \frac{1}{\mu_2 f_X(x)} z_{n2} \right] \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \sqrt{nh_n} h_n^2 \mu_2 m^{(2)}(x) + o_p(\sqrt{nh_n^5}) \\ O_p(\sqrt{nh_n^5}) + o_p(\sqrt{nh_n^5}) \end{pmatrix}.$$

Since $nh_n^5 = O(1)$, letting $b_n := (\frac{1}{2} h_n^2 \mu_2 m^{(2)}(x), 0)^T$ we obtain $N_n B_n = N_n b_n + o_p(1)$.

Conclusion. Combining Steps 3 and 4,

$$N_n \left(\hat{\beta}_{1,1}(x) - \beta_{1,1}(x) - b_n \right) = T_n + N_n (B_n - b_n) = T_n + o_p(1) \xrightarrow{d} N(\mathbf{0}, \Sigma),$$

that is,

$$\begin{aligned} \sqrt{nh_n} \left(\hat{m}(x) - m(x) - \frac{1}{2} h_n^2 \mu_2 m^{(2)}(x) \right) &\xrightarrow{d} N \left(0, \frac{\sigma^2}{f_X(x)} \int K^2(u) du \right), \\ \sqrt{nh_n^3} \left(\hat{m}^{(1)}(x) - m^{(1)}(x) \right) &\xrightarrow{d} N \left(0, \frac{\sigma^2}{f_X(x)} \frac{\nu_2}{\mu_2^2} \right), \end{aligned}$$

jointly, with asymptotically independent components. □