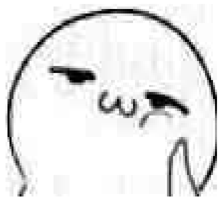
Nadaraya–Waston Regression

Jiuzhou Miao

November 29, 2024



In the previous topic, we discussed nonparametric density function estimation and its properties. Starting with this topic, we will talk about the topics on *nonparametric regression*.

1 Introduction

Let (X, Y) be a pair of real-valued random variables and satisfy the model

$$Y = m(X) + \sigma(X)\varepsilon,$$

in which $m : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to [0, \infty)$ are mean function and standard deviation function respectively. ε is the measurement error which satisfies $\mathbb{E}(\varepsilon|X = x) = 0$ and $\mathbb{E}(\varepsilon^2|X = x) = 1$. Later, more conditions will be attached to ε to derive some get some profound results. Let $\left\{(X_i, Y_i, \varepsilon_i)\right\}_{i=1}^n$ be i.i.d. copies of (X, Y, ε). i.e., they satisfies the model

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i.$$

We are interested in estimating the mean function

$$m(x) = \mathbb{E}(Y|X = x)$$

based on the data $\left\{(X_i, Y_i)\right\}_{i=1}^n$ with nonparametric methods.

We will use two methods to obtain the estimation of $m(\cdot)$. The first method is *nonparametric local kernel regression*, including *Nadaraya–Waston regression* and *local linear regression*. Later, the second method, *spline regression* will also be introduced.

The key idea of nonparametric local kernel regression is using a polynomial to approximate $m(\cdot)$ in the neighborhood of X_i. Assume that $m(\cdot)$ is smooth enough (has derivatives of order p on x), then one can approximate $m(z)$ by

$$\begin{aligned} m(z) &\approx m(x) + m^{(1)}(x)(z - x) + \cdots + m^{(p)}(x)(z - x)^p/p! \\ &= c_0 + c_1(z - x) + \cdots + c_p(z - x)^p, \end{aligned}$$

in which $c_j = c_j(x) = m^{(j)}(x)/j!$ for $1 \le j \le p$ and $c_0 = c_0(x) = m(x)$. Then one can estimate $m(\cdot)$ by using weighted least square method. i.e., solve

$$(\hat{c}_0, \dots, \hat{c}_p)^\mathsf{T} = \arg\min_{c_0,\dots,c_p \in \mathbb{R}} n^{-1} \sum_{i=1}^{n} K_h(X_i - x) \Big\{ Y_i - \sum_{j=0}^{p} c_j (X_i - x)^j \Big\}^2$$

with setting $\hat{m}(x) = \hat{c}_0$. When $p = 0$, the estimation $\hat{m}(\cdot)$ is called as *local constant estimation* or *Nadaraya–Waston estimation*, proposed by Nadaraya and Waston in 1964. When $p = 1$, the estimation $\hat{m}(\cdot)$ is called as *local linear estimation*. They will be introduced on this and next topics. For more general cases that $p \ge 2$, the estimation $\hat{m}(\cdot)$ is called as *local polynomial estimation*, which will be omitted in our seminar. For obtaining more details on local polynomial regression, one can read the monograph Fan and Gijbels (1996).

When $p = 0$, one can obtain the closed form of Nadaraya–Waston estimation, that is

$$\hat{m}(x) = \frac{n^{-1} \sum_{i=1}^{n} K_h(X_i - x) Y_i}{n^{-1} \sum_{i=1}^{n} K_h(X_i - x)}.$$

In the following, we will derive the asymptotic properties of $\hat{m}(x)$. Similar to the methods dealing with the kernel density estimation, We introduce the error decomposition of $\hat{m}(x) - m(x)$ by

$$\begin{aligned}
\hat{m}(x) - m(x) &= \frac{n^{-1} \sum_{i=1}^{n} K_h(X_i - x) Y_i}{n^{-1} \sum_{i=1}^{n} K_h(X_i - x)} - m(x) \\
&= \frac{n^{-1} \sum_{i=1}^{n} K_h(X_i - x) \big\{ Y_i - m(x) \big\}}{n^{-1} \sum_{i=1}^{n} K_h(X_i - x)} \\
&= \frac{\overbrace{n^{-1} \sum_{i=1}^{n} K_h(X_i - x) \big\{ m(X_i) - m(x) \big\}}^{\text{Bias term}} + \overbrace{n^{-1} \sum_{i=1}^{n} K_h(X_i - x) \sigma(X_i) \varepsilon_i}^{\text{Noise term}}}{\underbrace{n^{-1} \sum_{i=1}^{n} K_h(X_i - x)}_{\text{Density term}}} \\
&= \frac{B_n(x) + A_n(x)}{f_n(x)}.
\end{aligned}$$

Assume the support of X is $\mathcal{X} = [0, 1]$ and denote $\mathcal{I}_h = [h, 1 - h]$. The following assumptions will be used in Section 2–Section 4:

(A1) $m \in C^2(\mathcal{X})$, $f \in C^1(\mathcal{X})$, $\sigma \in C^1(\mathcal{X})$, $K \in C^2[-1, 1]$ is a symmetrical probability density function with support $[-1, 1]$, $\inf_{x \in \mathcal{X}} f(x) > 0$ and $\inf_{x \in \mathcal{X}} \sigma(x) > 0$.

(A2) ε satisfies that $\mathbb{E}(\varepsilon | X = x) = 0$ and $\mathbb{E}(\varepsilon^2 | X = x) = 1$. There exists some $\eta > 0$ such that $\sup_{x \in \mathcal{X}} \mathbb{E}\big(|\varepsilon|^{2+\eta} | X = x \big) < \infty$.

2 Uniform convergence

From the previous topic, we have established the uniform convergence of kernel density estimation. Note that the density term here has the same form as kernel density estimation, we present the following proposition without proof.

Proposition 2.1 (Uniform convergence of density term). *Under Assumption (A1), as $n \to \infty$, if $h = h_n$ satisfies that $h \to 0$ and $n^{-1/2}h^{-1/2}\log^{1/2} n \to 0$, one has that*

$$\sup_{x\in\mathcal{I}_h}\big|f_n(x) - f(x)\big| = o(h) + \mathcal{O}_{a.s.}(n^{-1/2}h^{-1/2}\log^{1/2} n).$$

In the following, we only need to deal with the bias term and the noise term. The main technical tool is still the Bernstein's inequality (Bosq (1998), Theorem 1.2). We firstly deal with the bias term.

Proposition 2.2 (Uniform convergence of bias term). *Under Assumption (A1), as $n \to \infty$, if $h = h_n$ satisfies that $h \to 0$ and $n^{-1/2}h^{-1/2}\log^{1/2} n \to 0$, one has that*

$$\sup_{x\in\mathcal{I}_h}\Big|B_n(x) - h^2 b(x)f(x)\Big| = o(h^2) + \mathcal{O}_{a.s.}(n^{-1/2}h^{1/2}\log^{1/2} n)$$

in which

$$b(x) = \big\{m^{(1)}(x)f^{(1)}(x)/f(x) + m^{(2)}(x)/2\big\}\int_{-1}^{1} v^2 K(v)\mathrm{d}v,$$

and

$$\sup_{x\in\mathcal{I}_h}\big|B_n(x)\big| = \mathcal{O}(h^2) + \mathcal{O}_{a.s.}(n^{-1/2}h^{1/2}\log^{1/2} n).$$

Proof. Let $\xi_{in}(x) = n^{-1}K_h(X_i - x)\big\{m(X_i) - m(x)\big\}$. One can easily obtain that

$$\begin{aligned}
\mathbb{E}\big\{\xi_{in}(x)\big\} &= \int_{\mathcal{X}} n^{-1}h^{-1}K\big(h^{-1}(u-x)\big)\big\{m(u) - m(x)\big\}f(u)\mathrm{d}u \\
&= \int_{-1}^{1} n^{-1}K(v)\big\{m(x+hv) - m(x)\big\}f(x+hv)\mathrm{d}v \\
&= \int_{-1}^{1} n^{-1}K(v)\big\{m^{(1)}(x)hv + 2^{-1}m^{(2)}(x)h^2v^2 + u(h^2)\big\} \\
&\qquad\quad \times \big\{f(x) + f^{(1)}(x)hv + u(h)\big\}\mathrm{d}v \\
&= n^{-1}h^2 b(x)f(x)\big\{1 + u(1)\big\}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\big\{\xi_{in}^2(x)\big\} &= \int_{\mathcal{X}} n^{-2}h^{-2}K^2\big(h^{-1}(u-x)\big)\big\{m(u) - m(x)\big\}^2 f(u)\mathrm{d}u \\
&= \int_{-1}^{1} n^{-2}h^{-1}K^2(v)\big\{m(x+hv) - m(x)\big\}^2 f(x+hv)\mathrm{d}v \\
&= \int_{-1}^{1} n^{-2}h^{-1}K^2(v)\big\{m^{(1)}(x)\big\}^2 f(x)h^2v^2\mathrm{d}v\big\{1 + u(1)\big\} \\
&= n^{-2}h\big\{m^{(1)}(x)\big\}^2 f(x)\int_{-1}^{1} v^2K^2(v)\mathrm{d}v\big\{1 + u(1)\big\} \asymp n^{-2}h.
\end{aligned}$$

Let $\check{\xi}_{in}(x) = \xi_{in}(x) - \mathbb{E}\big\{\xi_{in}(x)\big\}$, one has that $\mathbb{E}\big\{\check{\xi}_{in}(x)\big\} = 0$ and $\mathbb{E}\big\{\check{\xi}_{in}^2(x)\big\} \asymp n^{-2}h$. Note that

$$\big|\check{\xi}_{in}(x)\big| \le 2n^{-1}h^{-1}\|K\|_\infty\|m^{(1)}\|_\infty|X_i - x| \le 2n^{-1}\|K\|_\infty\|m^{(1)}\|_\infty,$$

then one has that for $k = 3, 4, \dots,$

$$\mathbb{E}\Big\{\big|\check{\xi}_{in}(x)\big|^k\Big\} = \mathbb{E}\Big\{\big|\check{\xi}_{in}(x)\big|^{k-2}\check{\xi}_{in}^2(x)\Big\} \le k!\big(2n^{-1}\|K\|_\infty\|m^{(1)}\|_\infty\big)^{k-2}\mathbb{E}\big\{\check{\xi}_{in}^2(x)\big\},$$

which implies that Cramér's conditions hold with $r = 2n^{-1}\|K\|_\infty\|m^{(1)}\|_\infty$. Then by the Bernstein's inequality, for some $\delta > 0$, one has that

$$\begin{aligned}
&\mathbb{P}\left\{\left|\sum_{i=1}^n \check{\xi}_{in}(x)\right| > \delta n^{-1/2}h^{1/2}\log^{1/2} n\right\}\\
&\le 2\exp\left\{\frac{-\delta^2 n^{-1}h\log n}{4\sum_{i=1}^n \mathbb{E}\big\{\check{\xi}_{in}^2(x)\big\} + 4n^{-1}\|K\|_\infty\|m^{(1)}\|_\infty \delta n^{-1/2}h^{1/2}\log^{1/2} n}\right\}\\
&= 2\exp\left\{\frac{-\delta^2\log n}{4nh^{-1}\sum_{i=1}^n \mathbb{E}\big\{\check{\xi}_{in}^2(x)\big\} + 4\|K\|_\infty\|m^{(1)}\|_\infty \delta n^{-1/2}h^{-1/2}\log^{1/2} n}\right\}\\
&\le 2n^{-10}.
\end{aligned}$$

In the following, we discretize $\mathcal{I}_h$ by $h = x_0 < x_1 < \cdots < x_{n^4-1} = 1-h$. One has that

$$\begin{aligned}
&\sum_{n=1}^\infty \mathbb{P}\left\{\max_{0\le l\le n^4-1}\left|\sum_{i=1}^n \check{\xi}_{in}(x_l)\right| > \delta n^{-1/2}h^{1/2}\log^{1/2} n\right\}\\
&\le \sum_{n=1}^\infty\sum_{l=0}^{n^4-1}\mathbb{P}\left\{\left|\sum_{i=1}^n \check{\xi}_{in}(x_l)\right| > \delta n^{-1/2}h^{1/2}\log^{1/2} n\right\}\\
&\le \sum_{n=1}^\infty 2n^{-6} < \infty,
\end{aligned}$$

By Borel-Cantelli's lemma, one obtains that

$$\max_{0\le l\le n^4-1}\left|\sum_{i=1}^n \check{\xi}_{in}(x_l)\right| = \mathcal{O}_{a.s.}(n^{-1/2}h^{1/2}\log^{1/2} n).$$

For $x \in [x_l, x_{l+1}]$, $l = 0, \dots, n^4-2$, one has that

$$\begin{aligned}
\big|\xi_{in}(x) - \xi_{in}(x_l)\big| &\le n^{-1}h^{-1}\Big\{\Big|K\big(h^{-1}(X_i - x)\big) - K\big(h^{-1}(X_i - x_l)\big)\Big|\big|m(X_i) - m(x)\big|\\
&\qquad + K\big(h^{-1}(X_i - x_l)\big)\big|m(x) - m(x_l)\big|\Big\}\\
&\le n^{-1}h^{-1}\big\{\|K^{(1)}\|_\infty h^{-1}|x - x_l|\|m^{(1)}\|_\infty|X_i - x|\\
&\qquad + \|K\|_\infty\|m^{(1)}\|_\infty|x - x_l|\big\}\\
&\lesssim n^{-5}h^{-1}
\end{aligned}$$

and

$$\max_{0\le l\le n^4-2}\sup_{x\in[x_l,x_{l+1}]}\left|\sum_{i=1}^n \check{\xi}_{in}(x) - \check{\xi}_{in}(x_l)\right| \lesssim 2n^{-4}h^{-1} = \mathcal{O}_p(n^{-4}h^{-1}).$$

Hence, one has that

$$
\begin{aligned}
\sup_{x\in\mathcal{I}_h}\Big|\sum_{i=1}^{n}\check{\xi}_{in}(x)\Big| &\le \max_{0\le l\le n^4-1}\Big|\sum_{i=1}^{n}\check{\xi}_{in}(x_l)\Big| + \max_{0\le l\le n^4-2}\sup_{x\in[x_l,x_{l+1}]}\Big|\sum_{i=1}^{n}\check{\xi}_{in}(x)-\check{\xi}_{in}(x_l)\Big| \\
&= \mathcal{O}_{a.s.}(n^{-1/2}h^{1/2}\log^{1/2}n) + \mathcal{O}_p(n^{-4}h^{-1}) \\
&= \mathcal{O}_{a.s.}(n^{-1/2}h^{1/2}\log^{1/2}n)
\end{aligned}
$$

and

$$
B_n(x) = h^2 b(x) f(x) + u(h^2) + \mathcal{U}_{a.s.}(n^{-1/2}h^{1/2}\log^{1/2}n)
$$

for $x\in\mathcal{I}_h$ uniformly, which can lead to the two results in Proposition 2.2. □

Next, we deal with the noise term.

Proposition 2.3 (Uniform convergence of noise term)**.** *Under Assumptions (A1) and (A2), as $n\to\infty$, if $h=h_n$ satisfies that $h\to 0$ and $n^{\alpha-1/2}h^{-1/2}\log^{1/2}n\to 0$ for $(2+\eta)^{-1}<\alpha<1/2$, one has that*

$$
\sup_{x\in\mathcal{I}_h}\big|A_n(x)\big| = \mathcal{O}_{a.s.}(n^{-1/2}h^{-1/2}\log^{1/2}n).
$$

Proof. Let $T_n = n^{\alpha}$, which satisfies that

$$
n^{1/2}h^{1/2}T_n^{-(1+\eta)}\to 0,\ \sum_{n=1}^{\infty}T_n^{-(2+\eta)}<\infty,\ T_n n^{-1/2}h^{-1/2}\log^{1/2}n\to 0.
$$

Denote

$$
\begin{aligned}
\mu_{in} &= \mathbb{E}\Big\{\varepsilon_i\mathbb{I}\big(|\varepsilon_i|\le T_n\big)\Big|X_i\Big\}, && \text{(Mean part)}\\
\varepsilon_{in,1} &= \varepsilon_i\mathbb{I}\big(|\varepsilon_i|>T_n\big), && \text{(Tail part)}\\
\varepsilon_{in,2} &= \varepsilon_i\mathbb{I}\big(|\varepsilon_i|\le T_n\big)-\mu_{in}. && \text{(Truncated part)}
\end{aligned}
$$

Then $A_n(x)$ can be divided into three parts

$$
A_n(x) = A_{n,0}(x) + A_{n,1}(x) + A_{n,2}(x)
$$

in which

$$
\begin{aligned}
A_{n,0}(x) &= n^{-1}\sum_{i=1}^{n}K_h(X_i-x)\sigma(X_i)\mu_{in},\\
A_{n,1}(x) &= n^{-1}\sum_{i=1}^{n}K_h(X_i-x)\sigma(X_i)\varepsilon_{in,1},\\
A_{n,2}(x) &= n^{-1}\sum_{i=1}^{n}K_h(X_i-x)\sigma(X_i)\varepsilon_{in,2}.
\end{aligned}
$$

For the mean part, note that

$$
\begin{aligned}
|\mu_{in}| &= \Big|\mathbb{E}\Big\{\varepsilon_i\mathbb{I}\big(|\varepsilon_i|\le T_n\big)\Big|X_i\Big\}\Big|\\
&= \Big|\mathbb{E}\big(\varepsilon_i|X_i\big)-\mathbb{E}\Big\{\varepsilon_i\mathbb{I}\big(|\varepsilon_i|>T_n\big)\Big|X_i\Big\}\Big|\\
&\le \mathbb{E}\Big\{|\varepsilon_i|\mathbb{I}\big(|\varepsilon_i|>T_n\big)\Big|X_i\Big\}\\
&\le T_n^{-(1+\eta)}\sup_{x\in\mathcal{X}}\mathbb{E}\big(|\varepsilon|^{2+\eta}|X=x\big),
\end{aligned}
$$

one has that

$$\sup_{x\in\mathcal{I}_h}\big|A_{n,0}(x)\big| \lesssim T_n^{-(1+\eta)}\sup_{x\in\mathcal{I}_h}\big|f_n(x)\big| = \mathcal{O}_p(T_n^{-(1+\eta)}) = o_p(n^{-1/2}h^{-1/2}).$$

For the tail part, note that

$$\sum_{n=1}^{\infty}\mathbb{P}\big(|\varepsilon_n| > T_n\big) \le \mathbb{E}\big(|\varepsilon|^{2+\eta}\big)\sum_{n=1}^{\infty}T_n^{-(2+\eta)} < \infty,$$

one has that

$$\mathbb{P}\Big[\limsup_{n\to\infty}\big\{|\varepsilon_n| > T_n\big\}\Big] = \mathbb{P}\Big[\limsup_{N\to\infty}\big\{|\varepsilon_N| > T_N\big\}\Big] = \mathbb{P}\Big[\cap_{N=1}^{\infty}\cup_{n=N}^{\infty}\big\{|\varepsilon_n| > T_n\big\}\Big] = 0,$$

which means that

$$\begin{aligned}
&\mathbb{P}\big(\exists N \ge 1, \forall n \ge N, |\varepsilon_n| \le T_n\big) = 1,\\
&\mathbb{P}\big(\exists N \ge 1, \forall n \ge N, |\varepsilon_i| \le T_n, 1 \le i \le n\big) = 1,\\
&\mathbb{P}\big(\exists N \ge 1, \forall n \ge N, \varepsilon_{in,1} = 0, 1 \le i \le n\big) = 1.
\end{aligned}$$

Hence, one has that for large n, $A_{n,1}(x) = 0$, a.s. uniformly.

For the truncated part, one has that $\mathbb{E}(\varepsilon_{in,2}|X_i) = 0$ and

$$\begin{aligned}
\mathbb{E}(\varepsilon_{in,2}^2|X_i) &= \mathbb{E}\Big\{\varepsilon_i^2\mathbb{I}\big(|\varepsilon_i| \le T_n\big)\Big|X_i\Big\} - \mu_{in}^2\\
&= 1 + \mathcal{U}_p(T_n^{-\eta} + T_n^{-2(1+\eta)}),
\end{aligned}$$

and for $k = 3, 4, \dots,$

$$\mathbb{E}\big(|\varepsilon_{in,2}|^k\big|X_i\big) = \mathbb{E}\big(|\varepsilon_{in,2}|^{k-2}\varepsilon_{in,2}^2\big|X_i\big) \le (2T_n)^{k-2}\mathbb{E}(\varepsilon_{in,2}^2|X_i).$$

Let $\nu_{in}(x) = n^{-1}K_h(X_i - x)\sigma(X_i)\varepsilon_{in,2}$, one has that $\mathbb{E}\big\{\nu_{in}(x)\big\} = 0$,

$$\begin{aligned}
\mathbb{E}\big\{\nu_{in}^2(x)\big\} &= \mathbb{E}\big\{n^{-2}K_h^2(X_i - x)\sigma^2(X_i)\varepsilon_{in,2}^2\big\}\\
&= \mathbb{E}\big\{n^{-2}K_h^2(X_i - x)\sigma^2(X_i)\big\} \times \big\{1 + u(1)\big\}\\
&= n^{-2}h^{-1}\sigma^2(x)f(x)\int_{-1}^{1}K^2(v)\mathrm{d}v\big\{1 + u(1)\big\},
\end{aligned}$$

and for $k = 3, 4, \dots,$

$$\mathbb{E}\Big\{\big|\nu_{in}(x)\big|^k\Big\} \le \big(2n^{-1}h^{-1}\|K\|_\infty\|\sigma\|_\infty T_n\big)^{k-2}\mathbb{E}\big\{\nu_{in}^2(x)\big\},$$

which implies that Cramér's conditions hold with $r = 2n^{-1}h^{-1}\|K\|_\infty\|\sigma\|_\infty T_n$. Then by the Bernstein's inequality, for some $\delta > 0$, one has that

$$\begin{aligned}
&\mathbb{P}\left\{\left|\sum_{i=1}^{n}\nu_{in}(x)\right| > \delta n^{-1/2}h^{-1/2}\log^{1/2} n\right\}\\
&\le 2\exp\left\{\frac{-\delta^2 n^{-1}h^{-1}\log n}{4\sum_{i=1}^{n}\mathbb{E}\big\{\nu_{in}^2(x)\big\} + 4n^{-1}h^{-1}\|K\|_\infty\|\sigma\|_\infty T_n\delta n^{-1/2}h^{-1/2}\log^{1/2} n}\right\}\\
&= 2\exp\left\{\frac{-\delta^2\log n}{4nh\sum_{i=1}^{n}\mathbb{E}\big\{\check{\xi}_{in}^2(x)\big\} + 4\|K\|_\infty\|\sigma\|_\infty\delta T_n n^{-1/2}h^{-1/2}\log^{1/2} n}\right\}\\
&\le 2n^{-10}.
\end{aligned}$$

In the following, we discretize $\mathcal{I}_h$ by $h = x_0 < x_1 < \cdots < x_{n^4-1} = 1-h$. One has that

$$\begin{aligned}
&\sum_{n=1}^{\infty} \mathbb{P}\left\{ \max_{0\le l\le n^4-1} \left| \sum_{i=1}^{n} \nu_{in}(x_l) \right| > \delta n^{-1/2} h^{-1/2} \log^{1/2} n \right\} \\
\le &\sum_{n=1}^{\infty} \sum_{l=0}^{n^4-1} \mathbb{P}\left\{ \left| \sum_{i=1}^{n} \nu_{in}(x_l) \right| > \delta n^{-1/2} h^{-1/2} \log^{1/2} n \right\} \\
\le &\sum_{n=1}^{\infty} 2n^{-6} < \infty,
\end{aligned}$$

By Borel-Cantelli's lemma, one obtains that

$$\max_{0\le l\le n^4-1} \left| \sum_{i=1}^{n} \nu_{in}(x_l) \right| = \mathcal{O}_{a.s.}(n^{-1/2} h^{-1/2} \log^{1/2} n).$$

For $x \in [x_l, x_{l+1}]$, $l = 0, \ldots, n^4-2$, one has that

$$\begin{aligned}
\left| \nu_{in}(x) - \nu_{in}(x_l) \right| &\le n^{-1} h^{-1} \left| K\big(h^{-1}(X_i - x)\big) - K\big(h^{-1}(X_i - x_l)\big) \right| \sigma(X_i) |\varepsilon_{in,2}| \\
&\le 2n^{-1} h^{-1} \|K^{(1)}\|_\infty h^{-1} |x - x_l| \|\sigma\|_\infty T_n \\
&\lesssim n^{-5} h^{-2} T_n
\end{aligned}$$

and

$$\max_{0\le l\le n^4-2} \sup_{x\in[x_l,x_{l+1}]} \left| \sum_{i=1}^{n} \nu_{in}(x) - \nu_{in}(x_l) \right| \lesssim n^{-4} h^{-2} T_n.$$

Hence, one has that

$$\begin{aligned}
\sup_{x\in\mathcal{I}_h} \left| A_{n,2}(x) \right| &= \sup_{x\in\mathcal{I}_h} \left| \sum_{i=1}^{n} \nu_{in}(x) \right| \\
&\le \max_{0\le l\le n^4-1} \left| \sum_{i=1}^{n} \nu_{in}(x_l) \right| + \max_{0\le l\le n^4-2} \sup_{x\in[x_l,x_{l+1}]} \left| \sum_{i=1}^{n} \nu_{in}(x) - \nu_{in}(x_l) \right| \\
&= \mathcal{O}_{a.s.}(n^{-1/2} h^{-1/2} \log^{1/2} n) + \mathcal{O}_p(n^{-4} h^{-2} T_n) \\
&= \mathcal{O}_{a.s.}(n^{-1/2} h^{-1/2} \log^{1/2} n).
\end{aligned}$$

Put three parts together, one has that

$$\sup_{x\in\mathcal{I}_h} \left| A_n(x) \right| = \mathcal{O}_{a.s.}(n^{-1/2} h^{-1/2} \log^{1/2} n).$$

□

By Proposition 2.1–2.3, one can obtain the following theorem.

Theorem 2.4 (Uniform convergence of Nadaraya–Waston estimation)**.** *Under the assumptions of Proposition 2.3, one has that*

$$\sup_{x\in\mathcal{I}_h} \left| \hat{m}(x) - m(x) - h^2 b(x) \right| = o(h^2) + \mathcal{O}_{a.s.}(n^{-1/2} h^{-1/2} \log^{1/2} n),$$

in which $b(x)$ is given in Proposition 2.2, and

$$\sup_{x\in\mathcal{I}_h} \left| \hat{m}(x) - m(x) \right| = \mathcal{O}(h^2) + \mathcal{O}_{a.s.}(n^{-1/2} h^{-1/2} \log^{1/2} n).$$

Remark. Furthermore, assume that $h \asymp n^{-1/5} \log^{1/5} n$, one has that

$$\sup_{x \in \mathcal{I}_h} \left| \hat{m}(x) - m(x) \right| = \mathcal{O}_{a.s.}(n^{-2/5} \log^{2/5} n),$$

which is the optimal convergence rate of Nadaraya–Waston estimation. In such case, α and η need to satisfy that $(2+\eta)^{-1} < \alpha < 2/5$ which implies that $\eta > 1/2$.

3 Boundary behavior

In this section, we examine the convergence rate of Nadaraya–Waston estimation at the boundary points of $\mathcal{X}$, i.e., $x \in [0, h]$ or $x \in [1-h, 1]$. We only focus on the case that $x \in [1-h, 1]$. The case that $x \in [0, h]$ had been given in the slides written by Prof. Lijian Yang.

Proposition 3.1 (Boundary behavior of density term and bias term)**.** *Under Assumption (A1), as $n \to \infty$, if $h = h_n$ satisfies that $h \to 0$ and $n^{-1/2}h^{-1/2} \log^{1/2} n \to 0$, one has that*

$$f_n(x) = f(x) \int_{-1}^{(1-x)/h} K(v) \mathrm{d}v + \mathcal{U}(h) + \mathcal{U}_{a.s.}(n^{-1/2}h^{-1/2} \log^{1/2} n),$$
$$B_n(x) = hm^{(1)}(x) f(x) \int_{-1}^{(1-x)/h} vK(v) \mathrm{d}v + \mathcal{U}(h^2) + \mathcal{U}_{a.s.}(n^{-1/2}h^{1/2} \log^{1/2} n)$$

for all $x \in [1-h, 1]$ uniformly.

Proposition 3.2 (Boundary behavior of noise term)**.** *Under Assumptions (A1) and (A2), as $n \to \infty$, if $h = h_n$ satisfies that $h \to 0$ and $n^{\alpha-1/2}h^{-1/2} \log^{1/2} n \to 0$ for $(2+\eta)^{-1} < \alpha < 1/2$, one has that*

$$A_n(x) = \mathcal{U}_{a.s.}(n^{-1/2}h^{-1/2} \log^{1/2} n)$$

for all $x \in [1-h, 1]$ uniformly.

The proofs of Proposition 3.1 and 3.2 can be obtained by imitating the proofs of Proposition 2.2 and 2.3. **You need to be careful with the range of integrals after the variable substitution.** By using Proposition 3.1 and 3.2, one can obtain the boundary behavior of Nadaraya–Waston estimation.

Theorem 3.3 (Boundary behavior of Nadaraya–Waston estimation)**.** *Under the assumptions of Proposition 3.2, one has that*

$$\hat{m}(x) = m(x) + \mathcal{U}(h) + \mathcal{U}_{a.s.}(n^{-1/2}h^{-1/2} \log^{1/2} n)$$

for all $x \in [1-h, 1]$ uniformly.

4 Pointwise asymptotic distribution

We only need to derive the asymptotic normality of the noise term by using Lindeberg–Feller's CLT. Then the asymptotic normality of $\hat{m}(x) - m(x)$ can be obtained by using Slutsky's theorem.

Proposition 4.1 (Asymptotic normality of noise term). *Under Assumptions (A1) and (A2), as $n \to \infty$, if $h = h_n$ satisfies that $h \to 0$ and $nh \to \infty$, one has that for a fixed $x \in \mathcal{I}_h$,*

$$\sqrt{nh} A_n(x) \xrightarrow{d} \mathcal{N}\big(0, \sigma^2(x) f(x) \int_{-1}^{1} K^2(v) \mathrm{d}v\big).$$

Proof. We first calculate the expectation and variance of the noise part. One has that $\mathbb{E}\{A_n(x)\} = 0$ and

$$\begin{aligned}
\operatorname{Var}\{A_n(x)\} &= n^{-1} \operatorname{Var}\{K_h(X_1 - x)\sigma(X_1)\varepsilon_1\} \\
&= n^{-1} \mathbb{E}\{K_h^2(X_1 - x)\sigma^2(X_1)\} \\
&= n^{-1} h^{-1} \sigma^2(x) f(x) \int_{-1}^{1} K^2(v) \mathrm{d}v \{1 + u(1)\} \\
&\asymp n^{-1} h^{-1}.
\end{aligned}$$

Denote $\zeta_{in}(x) = n^{-1} K_h(X_i - x)\sigma(X_i)\varepsilon_i$. In the following, we will check Lyapunov's condition, i.e.,

$$\lim_{n \to \infty} V_n^{-(2+\eta)}(x) \sum_{i=1}^{n} \mathbb{E}\big(|\xi_{in}(x)|^{2+\eta}\big) = 0,$$

in which $V_n^2(x) = \operatorname{Var}\{A_n(x)\}$. Note that

$$\begin{aligned}
\mathbb{E}\big(|\xi_{in}(x)|^{2+\eta}\big) &= \mathbb{E}\Big\{\mathbb{E}\big(|\xi_{in}(x)|^{2+\eta} \big| X_i\big)\Big\} \\
&= (nh)^{-(2+\eta)} \mathbb{E}\Big\{K^{2+\eta}\big(h^{-1}(X_i - x)\big)\sigma^{2+\eta}(X_i)\mathbb{E}\big(|\varepsilon_i|^{2+\eta} \big| X_i\big)\Big\} \\
&\le (nh)^{-(2+\eta)} \sup_{x \in \mathcal{X}} \mathbb{E}\big(|\varepsilon|^{2+\eta} | X = x\big) \mathbb{E}\Big\{K^{2+\eta}\big(h^{-1}(X_i - x)\big)\sigma^{2+\eta}(X_i)\Big\} \\
&\le n^{-(2+\eta)} h^{-(1+\eta)} \sup_{x \in \mathcal{X}} \mathbb{E}\big(|\varepsilon|^{2+\eta} | X = x\big) \|\sigma\|_\infty^{2+\eta} \|f\|_\infty \int_{-1}^{1} K^{2+\eta}(v) \mathrm{d}v \\
&\lesssim n^{-(2+\eta)} h^{-(1+\eta)},
\end{aligned}$$

one has that

$$V_n^{-(2+\eta)}(x) \sum_{i=1}^{n} \mathbb{E}\big(|\xi_{in}(x)|^{2+\eta}\big) \lesssim (nh)^{-(1+\eta)/2},$$

implies that Lyapunov's condition holds. Hence, one has that

$$V_n^{-1}(x) A_n(x) \xrightarrow{d} \mathcal{N}(0, 1)$$

and

$$\begin{aligned}
\frac{\sqrt{nh} A_n(x)}{\sqrt{f(x)\sigma^2(x) \int_{-1}^{1} K^2(v) \mathrm{d}v}} &= \frac{A_n(x)}{\sqrt{n^{-1} h^{-1} f(x) \sigma^2(x) \int_{-1}^{1} K^2(v) \mathrm{d}v}} \\
&= V_n^{-1}(x) A_n(x) \times \frac{V_n(x)}{\sqrt{n^{-1} h^{-1} \sigma^2(x) f(x) \int_{-1}^{1} K^2(v) \mathrm{d}v}} \\
&\xrightarrow{d} \mathcal{N}(0, 1)
\end{aligned}$$

by using Slutsky's theorem. □

Theorem 4.2 (Asymptotic normality of Nadaraya–Waston estimation)**.** *Under Assumptions (A1) and (A2), as $n \to \infty$, if $h = h_n$ satisfies that $n^{-1/2}h^{-1/2}\log^{1/2} n \to 0$ and $nh^5 = \mathcal{O}(1)$, one has that for a fixed $x \in \mathcal{I}_h$,*

$$\sqrt{nh}\Big\{\hat{m}(x) - m(x) - h^2 b(x)\Big\} \xrightarrow{d} \mathcal{N}\Big(0, \sigma^2(x) f^{-1}(x) \int_{-1}^{1} K^2(v)\mathrm{d}v\Big),$$

in which $b(x)$ is given in Proposition 2.2. Furthermore, assume that $nh^5 \to 0$, one has that for a fixed $x \in \mathcal{I}_h$,

$$\sqrt{nh}\big\{\hat{m}(x) - m(x)\big\} \xrightarrow{d} \mathcal{N}\Big(0, \sigma^2(x) f^{-1}(x) \int_{-1}^{1} K^2(v)\mathrm{d}v\Big).$$

5 Uniform asymptotic distribution

In this section, we will derive the uniform asymptotic distribution of $\hat{m}(x)-m(x)$ over the whole $x \in \mathcal{I}_h$ under some mild conditions. By using the uniform asymptotic distribution, one can construct $100\%(1-\alpha)$ simultaneous confidence band for the mean function $m(x), x \in \mathcal{I}_h$. The research on constructing simultaneous confidence band for the mean function based on nonparametric local kernel regression estimation can refer to Johnston (1982), Härdle (1989), Eubank and Speckman (1993), Xia (1998), Cai et al. (2019), Cai et al. (2021), etc. The proofs given in this section mainly refer to Härdle (1989) and Cai et al. (2021).

In the following, we mainly focus on the uniform asymptotic distribution of the noise term $A_n(x)$. The technical tool is the strong approximation theorem proposed by Tusnády (1977).

Lemma 5.1 (Tusnády (1977), Theorem 1)**.** *Let $\{\boldsymbol{\eta}_i\}_{i=1}^n$ be i.i.d. random vectors on the 2-dimensional unit square with $\mathbb{P}(\boldsymbol{\eta}_i \le \mathbf{t}) = \lambda(\mathbf{t})$, $\mathbf{0} \le \mathbf{t} \le \mathbf{1}$, where $\mathbf{t} = (t_1, t_2)$ and $\mathbf{1} = (1,1)$ are 2-dimensional vectors, $\lambda(\mathbf{t}) = t_1 t_2$. The empirical distribution function $F_n(\mathbf{t})$ based on sample $(\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n)$ is defined as $F_n(\mathbf{t}) = n^{-1}\sum_{i=1}^n \mathbb{I}(\boldsymbol{\eta}_i \le \mathbf{t})$ for $\mathbf{0} \le \mathbf{t} \le \mathbf{1}$. The 2-dimensional Brownian bridge $B(\mathbf{t})$ is defined by $B(\mathbf{t}) = W(\mathbf{t}) - \lambda(\mathbf{t})W(\mathbf{1})$ for $\mathbf{0} \le \mathbf{t} \le \mathbf{1}$, where $W(\mathbf{t})$ is a 2-dimensional Wiener process. Then there is a version $B_n(\mathbf{t})$ of $B(\mathbf{t})$ such that*

$$\mathbb{P}\bigg[\sup_{\mathbf{0}\le\mathbf{t}\le\mathbf{1}} \Big| n\big\{F_n(\mathbf{t}) - \lambda(\mathbf{t})\big\} - \sqrt{n}B_n(\mathbf{t})\Big| > (C\log n + x)\log n\bigg] < ke^{-\lambda x}$$

holds for all x, where C, K, λ are positive constants.

Recall the variance of the noise term $A_n(x)$

$$\mathrm{Var}\big\{A_n(x)\big\} = n^{-1}h^{-1}\sigma^2(x) f(x) \int_{-1}^{1} K^2(v)\mathrm{d}v\big\{1 + u(1)\big\}.$$

Denote the empirical process $Z_n(u,\varepsilon) = \sqrt{n}\big\{\hat{F}_n(u,\varepsilon) - F(u,\varepsilon)\big\}$. Here, $\hat{F}_n(u,\varepsilon)$ and $F(u,\varepsilon)$ are empirical distribution function and distribution function based on $\big\{(X_i,\varepsilon_i)\big\}_{i=1}^n$. We introduce the Rosenblatt quantile transformation for (X,ε):

$$\mathcal{T}(X,\varepsilon) = \big(F_X(X), F_{\varepsilon|X}(\varepsilon|X)\big),$$

in which $F_X(\cdot)$ and $F_{\varepsilon|X}(\cdot|\cdot)$ are distribution function of X and conditional distribution function of $\varepsilon|X$ respectively. The Jacobian determinant of $\mathcal{T}(X,\varepsilon)$ is $f(X,\varepsilon)$, in which $f(x,\varepsilon)$ is the joint density function of (X,ε). By using Lemma 5.1, one has that

$$\sup_{(u,\varepsilon)\in\mathcal{X}\times\mathcal{E}}\Big|Z_n(u,\varepsilon)-B_n\big(\mathcal{T}(u,\varepsilon)\big)\Big|=\mathcal{O}_{a.s.}(n^{-1/2}\log^2 n).$$

in which $\mathcal{E}$ is the range of ε. Define the standardized stochastic process

$$\begin{aligned}\Xi_n(x)&=n^{1/2}h^{1/2}\sigma^{-1}(x)f^{-1/2}(x)A_n(x)\\&=n^{1/2}h^{1/2}\sigma^{-1}(x)f^{-1/2}(x)\int_{\mathcal{X}}\int_{\mathcal{E}}K_h(u-x)\sigma(u)\varepsilon\mathrm{d}\big\{\hat{F}_n(u,\varepsilon)-F(u,\varepsilon)\big\}\\&=h^{-1/2}\sigma^{-1}(x)f^{-1/2}(x)\int_{\mathcal{X}}\int_{\mathcal{E}}K\big(h^{-1}(u-x)\big)\sigma(u)\varepsilon\mathrm{d}Z_n(u,\varepsilon).\end{aligned}$$

Let $\{\kappa_n\}_{n=1}^{\infty}$ be a divergent positive sequence. Denote $s_n(x)=\int_{\mathcal{E}_n}\varepsilon^2 f(x,\varepsilon)\mathrm{d}\varepsilon$ in which $\mathcal{E}_n=\big\{\varepsilon:|\varepsilon|\le\kappa_n\big\}$. Denote $W_n^*(t_1,t_2)$ as a sequence of Wiener processes satisfying $B_n(t_1,t_2)=W_n^*(t_1,t_2)-t_1t_2W_n^*(1,1)$ and $W_n(u)$ is a two-sided Wiener process on $(-\infty,\infty)$.

Lemma 5.2. *Under Assumptions (A1) and (A2), as $n\to\infty$, one has that*

$$\sup_{x\in\mathcal{X}}\big|s_n(x)/f(x)-1\big|=\mathcal{O}(\kappa_n^{-\eta}).$$

Proof. Note that $s_n(x)/f(x)=\mathbb{E}\Big\{\varepsilon^2\mathbb{I}\big(|\varepsilon|\le\kappa_n\big)\Big|X=x\Big\}$, one has that

$$\sup_{x\in\mathcal{X}}\big|s_n(x)/f(x)-1\big|\le\kappa_n^{-\eta}\sup_{x\in\mathcal{X}}\mathbb{E}\big(|\varepsilon|^{2+\eta}|X=x\big)=\mathcal{O}(\kappa_n^{-\eta}).$$

□

Define the following stochastic processes

$$\begin{aligned}\Xi_{n,0}(x)&=h^{-1/2}\sigma^{-1}(x)f^{-1/2}(x)\int_{\mathcal{X}}\int_{\mathcal{E}_n}K\big(h^{-1}(u-x)\big)\sigma(u)\varepsilon\mathrm{d}Z_n(u,\varepsilon),\\\Xi_{n,1}(x)&=h^{-1/2}\sigma^{-1}(x)s_n^{-1/2}(x)\int_{\mathcal{X}}\int_{\mathcal{E}_n}K\big(h^{-1}(u-x)\big)\sigma(u)\varepsilon\mathrm{d}Z_n(u,\varepsilon),\\\Xi_{n,2}(x)&=h^{-1/2}\sigma^{-1}(x)s_n^{-1/2}(x)\int_{\mathcal{X}}\int_{\mathcal{E}_n}K\big(h^{-1}(u-x)\big)\sigma(u)\varepsilon\mathrm{d}B_n\big(\mathcal{T}(u,\varepsilon)\big),\\\Xi_{n,3}(x)&=h^{-1/2}\sigma^{-1}(x)s_n^{-1/2}(x)\int_{\mathcal{X}}\int_{\mathcal{E}_n}K\big(h^{-1}(u-x)\big)\sigma(u)\varepsilon\mathrm{d}W_n^*\big(\mathcal{T}(u,\varepsilon)\big),\\\Xi_{n,4}(x)&=h^{-1/2}\sigma^{-1}(x)s_n^{-1/2}(x)\int_{\mathcal{X}}K\big(h^{-1}(u-x)\big)\sigma(u)s_n^{1/2}(u)\mathrm{d}W_n(u),\\\Xi_{n,5}(x)&=h^{-1/2}\int_{\mathcal{X}}K\big(h^{-1}(u-x)\big)\mathrm{d}W_n(u).\end{aligned}$$

In the previous topic, we have established the uniform asymptotic distribution of $\Xi_{n,5}(x), x\in\mathcal{I}_h$. We present it as the following proposition.

Proposition 5.3. *Under Assumption (A1), as $n \to \infty$, if $h = h_n$ satisfies that $h \to 0$, one has that*

$$a_h\Big(\sup_{x\in\mathcal{I}_h}\big|\Xi_{n,5}(x)/\Lambda\big| - b_h\Big) \xrightarrow{d} Z,$$

in which the random variable Z satisfies $\mathbb{P}(Z \le z) = \exp\big\{-2\exp(-z)\big\}$ for all $z \in \mathbb{R}$ and

$$a_h = (2\log h^{-1})^{1/2},\ b_h = a_h + 2^{-1}a_h^{-1}\log\Big(\frac{C}{2\pi^2}\Big),$$
$$C = 2^{-1}\Lambda^{-1}\int_{-1}^{-1}\big\{K^{(1)}(v)\big\}^2\mathrm{d}v,\ \Lambda = \sqrt{\int_{-1}^{1}K^2(v)\mathrm{d}v}.$$

In the following, we will prove that $\Xi_n(x)$ can be uniformly approximated by $\Xi_{n,5}(x)$ over $x \in \mathcal{I}_h$ in Proposition 5.4–5.9.

Proposition 5.4. *Under Assumptions (A1) and (A2), as $n \to \infty$, if $h = h_n$ satisfies that $h \to 0$ and $\kappa_n^{-\eta}h^{-2}\log n = \mathcal{O}(1)$, one has that*

$$\sup_{x\in\mathcal{I}_h}\big|\Xi_n(x) - \Xi_{n,0}(x)\big| = o_p(\log^{-1/2} n).$$

Proof. We have that

$$\begin{aligned}
&\log^{1/2} n\big\{\Xi_n(x) - \Xi_{n,0}(x)\big\}\\
&= h^{-1/2}\log^{1/2} n\sigma^{-1}(x)f^{-1/2}(x)\int_{\mathcal{X}}\int_{\mathcal{E}\setminus\mathcal{E}_n}K\big(h^{-1}(u-x)\big)\sigma(u)\varepsilon\mathrm{d}Z_n(u,\varepsilon)\\
&= n^{1/2}h^{1/2}\log^{1/2} n\sigma^{-1}(x)f^{-1/2}(x)\sum_{i=1}^n n^{-1}\Big[K_h(X_i-x)\sigma(X_i)\varepsilon_i\mathbb{I}\big(|\varepsilon_i| > \kappa_n\big)\\
&\qquad - \mathbb{E}\Big\{K_h(X_i-x)\sigma(X_i)\varepsilon_i\mathbb{I}\big(|\varepsilon_i| > \kappa_n\big)\Big\}\Big]\\
&:= \sum_{i=1}^n \zeta_{in}(x).
\end{aligned}$$

By using Theorem 13.5 in Billingsley (1999), we need to show that $\sum_{i=1}^n \zeta_{in}(x) \xrightarrow{p} 0$ for a fixed $x \in \mathcal{I}_h$ and check the tightness of $\sum_{i=1}^n \zeta_{in}(x)$ on $\mathcal{I}_h$, i.e., for any $x \in [x_1, x_2] \subset \mathcal{I}_h$,

$$\mathbb{E}\Big[\Big\{\sum_{i=1}^n \zeta_{in}(x) - \sum_{i=1}^n \zeta_{in}(x_1)\Big\}\Big\{\sum_{i=1}^n \zeta_{in}(x_2) - \sum_{i=1}^n \zeta_{in}(x)\Big\}\Big] \lesssim |x_1 - x_2|^2.$$

By Markov's inequality, $\forall\delta > 0$, one has that

$$\mathbb{P}\Big\{\Big|\sum_{i=1}^n \zeta_{in}(x)\Big| > \delta\Big\} \le \delta^{-2}\mathbb{E}\Big\{\Big|\sum_{i=1}^n \zeta_{in}(x)\Big|^2\Big\} = \delta^{-2}n\mathrm{Var}\big\{\zeta_{1n}(x)\big\}$$

and

$$\begin{aligned}
\mathrm{Var}\big\{\zeta_{1n}(x)\big\} &\le n^{-1}h\log n\sigma^{-2}(x)f^{-1}(x)\mathbb{E}\Big\{K_h^2(X_1-x)\sigma^2(X_1)\varepsilon_1^2\mathbb{I}\big(|\varepsilon_1| > \kappa_n\big)\Big\}\\
&= n^{-1}h\log n\sigma^{-2}(x)f^{-1}(x)\mathbb{E}\Big[K_h^2(X_1-x)\sigma^2(X_1)\mathbb{E}\Big\{\varepsilon_1^2\mathbb{I}\big(|\varepsilon_1| > \kappa_n\big)\Big|X_1\Big\}\Big]\\
&\lesssim \kappa_n^{-\eta}n^{-1}h\log n\mathbb{E}\big[K_h^2(X_1-x)\sigma^2(X_1)\big]\\
&= \mathcal{U}(\kappa_n^{-\eta}n^{-1}\log n),
\end{aligned}$$

then $n\mathrm{Var}\big\{\zeta_{1n}(x)\big\} = \mathcal{U}(\kappa_n^{-\eta}\log n) = u(1)$ and $\sum_{i=1}^n \zeta_{in}(x) \xrightarrow{p} 0$ for a fixed $x \in \mathcal{I}_h$.

Note that

$$\mathbb{E}\Bigg[\bigg\{\sum_{i=1}^n \zeta_{in}(x) - \sum_{i=1}^n \zeta_{in}(x_1)\bigg\}^2\Bigg] = \sum_{i=1}^n \mathbb{E}\big\{\zeta_{in}(x) - \zeta_{in}(x_1)\big\}^2 = n\mathbb{E}\big\{\zeta_{1n}(x) - \zeta_{1n}(x_1)\big\}^2$$

and

$$\zeta_{1n}(x) - \zeta_{1n}(x_1) = n^{-1/2}h^{1/2}\log^{1/2} n\big\{\mathrm{I}_n(x) + \mathrm{II}_n(x)\big\},$$

in which

$$\begin{aligned}\mathrm{I}_n(x) = {}&\big\{\sigma^{-1}(x)f^{-1/2}(x) - \sigma^{-1}(x_1)f^{-1/2}(x_1)\big\}\Big[K_h(X_1 - x)\sigma(X_1)\varepsilon_1\mathbb{I}\big(|\varepsilon_1| > \kappa_n\big)\\ &- \mathbb{E}\Big\{K_h(X_1 - x)\sigma(X_1)\varepsilon_1\mathbb{I}\big(|\varepsilon_1| > \kappa_n\big)\Big\}\Big],\\ \mathrm{II}_n(x) = {}&\sigma^{-1}(x_1)f^{-1/2}(x_1)\big\{K_h(X_1 - x) - K_h(X_1 - x_1)\big\}\sigma(X_1)\varepsilon_1\mathbb{I}\big(|\varepsilon_1| > \kappa_n\big)\\ &- \sigma^{-1}(x_1)f^{-1/2}(x_1)\mathbb{E}\Big[\big\{K_h(X_1 - x) - K_h(X_1 - x_1)\big\}\sigma(X_1)\varepsilon_1\mathbb{I}\big(|\varepsilon_1| > \kappa_n\big)\Big].\end{aligned}$$

One has that

$$\begin{aligned}\mathbb{E}\big\{\mathrm{I}_n^2(x)\big\} &= \big\{\sigma^{-1}(x)f^{-1/2}(x) - \sigma^{-1}(x_1)f^{-1/2}(x_1)\big\}^2\\ &\quad\times \mathrm{Var}\Big\{K_h(X_1 - x)\sigma(X_1)\varepsilon_1\mathbb{I}\big(|\varepsilon_1| > \kappa_n\big)\Big\}\\ &\lesssim |x - x_1|^2\mathbb{E}\Big\{K_h^2(X_1 - x)\sigma^2(X_1)\varepsilon_1^2\mathbb{I}\big(|\varepsilon_1| > \kappa_n\big)\Big\}\\ &\lesssim \kappa_n^{-\eta}h^{-1}|x - x_1|^2\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\big\{\mathrm{II}_n^2(x)\big\} &= \sigma^{-2}(x_1)f^{-1}(x_1)\mathrm{Var}\Big[\big\{K_h(X_1 - x) - K_h(X_1 - x_1)\big\}\sigma(X_1)\varepsilon_1\mathbb{I}\big(|\varepsilon_1| > \kappa_n\big)\Big]\\ &\lesssim \mathbb{E}\Big[\big\{K_h(X_1 - x) - K_h(X_1 - x_1)\big\}^2\sigma^2(X_1)\varepsilon_1^2\mathbb{I}\big(|\varepsilon_1| > \kappa_n\big)\Big]\\ &\lesssim \kappa_n^{-\eta}\mathbb{E}\Big[\big\{K_h(X_1 - x) - K_h(X_1 - x_1)\big\}^2\sigma^2(X_1)\Big]\\ &\lesssim \kappa_n^{-\eta}h^{-3}|x - x_1|^2,\end{aligned}$$

then by using Minkowski's inequality

$$\begin{aligned}\sqrt{\mathbb{E}\Bigg[\bigg\{\sum_{i=1}^n \zeta_{in}(x) - \sum_{i=1}^n \zeta_{in}(x_1)\bigg\}^2\Bigg]} &= n^{1/2}\Big[\mathbb{E}\big\{\zeta_{1n}(x) - \zeta_{1n}(x_1)\big\}^2\Big]^{1/2}\\ &= h^{1/2}\log^{1/2} n\Big[\mathbb{E}\big\{\mathrm{I}_n(x) + \mathrm{II}_n(x)\big\}^2\Big]^{1/2}\\ &\le h^{1/2}\log^{1/2} n\Big[\mathbb{E}\big\{\mathrm{I}_n^2(x)\big\}\Big]^{1/2}\\ &\quad + h^{1/2}\log^{1/2} n\Big[\mathbb{E}\big\{\mathrm{II}_n^2(x)\big\}\Big]^{1/2}\\ &\lesssim \kappa_n^{-\eta/2}h^{-1}\log^{1/2} n|x - x_1|.\end{aligned}$$

Similarly, one can obtain that

$$\sqrt{\mathbb{E}\Big[\Big\{\sum_{i=1}^{n}\zeta_{in}(x_2)-\sum_{i=1}^{n}\zeta_{in}(x)\Big\}^2\Big]}\lesssim \kappa_n^{-\eta/2}h^{-1}\log^{1/2}n|x_2-x|.$$

Hence, by using Cauchy–Schwarz's inequality

$$\begin{aligned}
&\mathbb{E}\Big[\Big\{\sum_{i=1}^{n}\zeta_{in}(x)-\sum_{i=1}^{n}\zeta_{in}(x_1)\Big\}\Big\{\sum_{i=1}^{n}\zeta_{in}(x_2)-\sum_{i=1}^{n}\zeta_{in}(x)\Big\}\Big]\\
&\leq \sqrt{\mathbb{E}\Big[\Big\{\sum_{i=1}^{n}\zeta_{in}(x)-\sum_{i=1}^{n}\zeta_{in}(x_1)\Big\}^2\Big]}\sqrt{\mathbb{E}\Big[\Big\{\sum_{i=1}^{n}\zeta_{in}(x_2)-\sum_{i=1}^{n}\zeta_{in}(x)\Big\}^2\Big]}\\
&\leq \kappa_n^{-\eta}h^{-2}\log n|x-x_1||x_2-x|\\
&\lesssim |x_1-x_2|^2.
\end{aligned}$$

□

Proposition 5.5. *Under Assumptions (A1) and (A2), as $n\to\infty$, one has that*

$$\sup_{x\in\mathcal{I}_h}\big|\Xi_{n,0}(x)/\Xi_{n,1}(x)-1\big|=\mathcal{O}_p(\kappa_n^{-\eta}).$$

Proposition 5.5 can be obtained by using Lemma 5.2.

Proposition 5.6. *Under Assumptions (A1) and (A2), as $n\to\infty$, if $h=h_n$ satisfies that $h\to 0$ and $\kappa_n n^{-1/2}h^{-1/2}\log^{5/2}n\to 0$, one has that*

$$\sup_{x\in\mathcal{I}_h}\big|\Xi_{n,1}(x)-\Xi_{n,2}(x)\big|=\mathcal{O}_p(\kappa_n n^{-1/2}h^{-1/2}\log^2 n)=o_p(\log^{-1/2}n).$$

Proof. For two functions $f(x,y)$ and $g(x,y)$ both defined on $[-B,B]\times[-A,A]$, the generalized integration by parts formula is

$$\begin{aligned}
\int_{-A}^{A}\int_{-B}^{B}f(x,y)\mathrm{d}g(x,y)=&\int_{-A}^{A}\int_{-B}^{B}g(x,y)\mathrm{d}f(x,y)\\
&+\int_{-A}^{A}f(B,y)\mathrm{d}g(B,y)-\int_{-A}^{A}f(-B,y)\mathrm{d}g(-B,y)\\
&-\int_{-B}^{B}g(x,A)\mathrm{d}f(x,A)+\int_{-B}^{B}g(x,-A)\mathrm{d}f(x,-A).
\end{aligned}$$

Then one has that

$$\begin{aligned}
&\Xi_{n,1}(x)-\Xi_{n,2}(x)\\
&=h^{-1/2}\sigma^{-1}(x)s_n^{-1/2}(x)\int_{\mathcal{X}}\int_{\mathcal{E}_n}K\big(h^{-1}(u-x)\big)\sigma(u)\varepsilon\mathrm{d}\Big\{Z_n(u,\varepsilon)-B_n\big(\mathcal{T}(u,\varepsilon)\big)\Big\}\\
&=h^{-1/2}\sigma^{-1}(x)s_n^{-1/2}(x)\int_{-1}^{1}\int_{\mathcal{E}_n}K(v)\sigma(x+hv)\varepsilon\mathrm{d}\Big\{Z_n(x+hv,\varepsilon)-B_n\big(\mathcal{T}(x+hv,\varepsilon)\big)\Big\}\\
&=h^{-1/2}\sigma^{-1}(x)s_n^{-1/2}(x)\int_{-1}^{1}\int_{\mathcal{E}_n}Z_n(x+hv,\varepsilon)-B_n\big(\mathcal{T}(x+hv,\varepsilon)\big)\mathrm{d}\big\{K(v)\sigma(x+hv)\varepsilon\big\}\\
&\quad+h^{-1/2}\sigma^{-1}(x)s_n^{-1/2}(x)\int_{-1}^{1}K(v)\sigma(x+hv)\kappa_n\mathrm{d}\Big\{Z_n(x+hv,\kappa_n)-B_n\big(\mathcal{T}(x+hv,\kappa_n)\big)\Big\}\\
&\quad+h^{-1/2}\sigma^{-1}(x)s_n^{-1/2}(x)\int_{-1}^{1}K(v)\sigma(x+hv)\kappa_n\mathrm{d}\Big\{Z_n(x+hv,-\kappa_n)-B_n\big(\mathcal{T}(x+hv,-\kappa_n)\big)\Big\}.
\end{aligned}$$

Note that

$$
\begin{aligned}
&\Big|\int_{-1}^{1}\int_{\mathcal{E}_n} Z_n(x+hv,\varepsilon)-B_n\big(\mathcal{T}(x+hv,\varepsilon)\big)\mathrm{d}\big\{K(v)\sigma(x+hv)\varepsilon\big\}\Big| \\
&\leq \mathcal{O}_{a.s.}(\kappa_n n^{-1/2}\log^2 n)\int_{-1}^{1}\big|K^{(1)}(v)\big|\sigma(x+hv)+K(v)\big|\sigma^{(1)}(x+hv)\big|\mathrm{d}v \\
&= \mathcal{O}_{a.s.}(\kappa_n n^{-1/2}\log^2 n),
\end{aligned}
$$

then the first term is bounded by $\mathcal{O}_{a.s.}(\kappa_n n^{-1/2}h^{-1/2}\log^2 n)$. By integration by parts, the second and third term are both bounded by $\mathcal{O}_{a.s.}(\kappa_n n^{-1/2}h^{-1/2}\log^2 n)$. Hence,

$$
\sup_{x\in\mathcal{I}_h}\big|\Xi_{n,1}(x)-\Xi_{n,2}(x)\big| = \mathcal{O}_p(\kappa_n n^{-1/2}h^{-1/2}\log^2 n) = o_p(\log^{-1/2} n).
$$

□

Proposition 5.7. *Under Assumptions (A1) and (A2), as $n\to\infty$, if $h=h_n$ satisfies that $h\log n\to 0$, one has that*

$$
\sup_{x\in\mathcal{I}_h}\big|\Xi_{n,2}(x)-\Xi_{n,3}(x)\big| = \mathcal{O}_p(h^{1/2}) = o_p(\log^{-1/2} n).
$$

Proof. Note that

$$
\begin{aligned}
&\Xi_{n,2}(x)-\Xi_{n,3}(x) \\
&= h^{-1/2}\sigma^{-1}(x)s_n^{-1/2}(x)\int_{\mathcal{X}}\int_{\mathcal{E}_n} K\big(h^{-1}(u-x)\big)\sigma(u)\varepsilon\mathrm{d}\Big\{B_n\big(\mathcal{T}(u,\varepsilon)\big)-W_n^*\big(\mathcal{T}(u,\varepsilon)\big)\Big\} \\
&= -W_n^*(1,1)\times h^{-1/2}\sigma^{-1}(x)s_n^{-1/2}(x)\int_{\mathcal{X}}\int_{\mathcal{E}_n} K\big(h^{-1}(u-x)\big)\sigma(u)\varepsilon f(u,\varepsilon)\mathrm{d}u\mathrm{d}\varepsilon \\
&= -W_n^*(1,1)\times h^{1/2}\sigma^{-1}(x)s_n^{-1/2}(x)\int_{-1}^{1}\int_{\mathcal{E}_n} K(v)\sigma(x+hv)\varepsilon f(x+hv,\varepsilon)\mathrm{d}v\mathrm{d}\varepsilon
\end{aligned}
$$

and

$$
\begin{aligned}
&\Big|\int_{-1}^{1}\int_{\mathcal{E}_n} K(v)\sigma(x+hv)\varepsilon f(x+hv,\varepsilon)\mathrm{d}v\mathrm{d}\varepsilon\Big| \\
&= \Big|\int_{-1}^{1} K(v)\sigma(x+hv)\mathbb{E}\Big\{\varepsilon\mathbb{I}\big(|\varepsilon|\leq\kappa_n\big)\Big|X=x+hv\Big\}f(x+hv)\mathrm{d}v\Big| \\
&\leq \int_{-1}^{1} K(v)\sigma(x+hv)\mathbb{E}\Big\{|\varepsilon|\mathbb{I}\big(|\varepsilon|\leq\kappa_n\big)\Big|X=x+hv\Big\}f(x+hv)\mathrm{d}v \\
&\leq \int_{-1}^{1} K(v)\sigma(x+hv)\mathbb{E}\big\{|\varepsilon|\big|X=x+hv\big\}f(x+hv)\mathrm{d}v \\
&\leq \int_{-1}^{1} K(v)\sigma(x+hv)f(x+hv)\mathrm{d}v \\
&= \mathcal{U}(1),
\end{aligned}
$$

in which the inequality of the penultimate line can be obtained by

$$
\sup_{x\in\mathcal{I}_h, v\in[-1,1]}\mathbb{E}\big\{|\varepsilon|\big|X=x+hv\big\} \leq \sup_{x\in\mathcal{I}_h, v\in[-1,1]}\sqrt{\mathbb{E}\big\{|\varepsilon|^2\big|X=x+hv\big\}} = 1,
$$

then one has that

$$\sup_{x\in\mathcal{I}_h}\big|\Xi_{n,2}(x)-\Xi_{n,3}(x)\big| \lesssim \big|W_n^*(1,1)\big| \times h^{1/2}\times \mathcal{O}(1)=\mathcal{O}_p(h^{1/2})=o_p(\log^{-1/2} n).$$

□

Proposition 5.8. *Under Assumption (A1), as $n\to\infty$, one has that*

$$\mathcal{L}\big\{\Xi_{n,3}(x), x\in\mathcal{I}_h\big\}=\mathcal{L}\big\{\Xi_{n,4}(x), x\in\mathcal{I}_h\big\}.$$

Proof. Note that $\Xi_{n,3}(x)$ and $\Xi_{n,4}(x)$ are two Gaussian process with zero mean function, we only need to verify that they have same covariance function. By Itô's Isometry, one has that

$$\begin{aligned}
\mathbb{E}\big\{\Xi_{n,3}(x)\Xi_{n,3}(x')\big\} &= h^{-1}\sigma^{-1}(x)\sigma^{-1}(x')s_n^{-1/2}(x)s_n^{-1/2}(x')\\
&\quad\times\int_{\mathcal{X}}\int_{\mathcal{E}_n} K\big(h^{-1}(u-x)\big)K\big(h^{-1}(u-x')\big)\sigma^2(u)\varepsilon^2\mathrm{d}\mathcal{T}(u,\varepsilon)\\
&= h^{-1}\sigma^{-1}(x)\sigma^{-1}(x')s_n^{-1/2}(x)s_n^{-1/2}(x')\\
&\quad\times\int_{\mathcal{X}}\int_{\mathcal{E}_n} K\big(h^{-1}(u-x)\big)K\big(h^{-1}(u-x')\big)\sigma^2(u)\varepsilon^2 f(u,\varepsilon)\mathrm{d}u\mathrm{d}\varepsilon\\
&= h^{-1}\sigma^{-1}(x)\sigma^{-1}(x')s_n^{-1/2}(x)s_n^{-1/2}(x')\\
&\quad\times\int_{\mathcal{X}} K\big(h^{-1}(u-x)\big)K\big(h^{-1}(u-x')\big)\sigma^2(u)s_n(u)\mathrm{d}u\\
&= \mathbb{E}\big\{\Xi_{n,4}(x)\Xi_{n,4}(x')\big\}.
\end{aligned}$$

□

The following Assumption (A3) involves the boundedness of the derivative of $s_n(x)$:

(A3) $s_n(x)$ satisfies that $\max_{n\geq 1}\sup_{x\in\mathcal{X}}\big|s_n^{(1)}(x)\big|<\infty$.

Proposition 5.9. *Under Assumptions (A1)–(A3), as $n\to\infty$, if $h=h_n$ satisfies that $h\log n\to 0$ and $\kappa_n^{-\eta}h^{-1/2}\log^{1/2} n\to 0$, one has that*

$$\sup_{x\in\mathcal{I}_h}\big|\Xi_{n,4}(x)-\Xi_{n,5}(x)\big| = \mathcal{O}_p(h^{1/2}+\kappa_n^{-\eta}h^{-1/2})=o_p(\log^{-1/2} n).$$

Proof. By integration by parts, one has that

$$
\begin{aligned}
\Xi_{n,4}(x) - \Xi_{n,5}(x) &= h^{-1/2}\int_{\mathcal{X}} K\big(h^{-1}(u-x)\big)\big\{\sigma^{-1}(x)s_n^{-1/2}(x)\sigma(u)s_n^{1/2}(u)-1\big\}\mathrm{d}W_n(u)\\
&= -h^{-1/2}\int_{\mathcal{X}} W_n(u)\mathrm{d}\Big[K\big(h^{-1}(u-x)\big)\big\{\sigma^{-1}(x)s_n^{-1/2}(x)\sigma(u)s_n^{1/2}(u)-1\big\}\Big]\\
&= -h^{-3/2}\int_{\mathcal{X}} W_n(u)K^{(1)}\big(h^{-1}(u-x)\big)\big\{\sigma^{-1}(x)s_n^{-1/2}(x)\sigma(u)s_n^{1/2}(u)-1\big\}\mathrm{d}u\\
&\quad - h^{-1/2}\int_{\mathcal{X}} W_n(u)K\big(h^{-1}(u-x)\big)\sigma^{-1}(x)s_n^{-1/2}(x)\\
&\quad \times\big\{\sigma^{(1)}(u)s_n^{1/2}(u)+2^{-1}\sigma(u)s_n^{(1)}(u)s_n^{-1/2}(u)\big\}\mathrm{d}u\\
&= -h^{-1/2}\int_{-1}^{1} W_n(x+hv)K^{(1)}(v)\\
&\quad \times\big\{\sigma^{-1}(x)s_n^{-1/2}(x)\sigma(x+hv)s_n^{1/2}(x+hv)-1\big\}\mathrm{d}v\\
&\quad - h^{1/2}\int_{-1}^{1} W_n(x+hv)K(v)\sigma^{-1}(x)s_n^{-1/2}(x)\\
&\quad \times\big\{\sigma^{(1)}(x+hv)s_n^{1/2}(x+hv)\\
&\qquad + 2^{-1}\sigma(x+hv)s_n^{(1)}(x+hv)s_n^{-1/2}(x+hv)\big\}\mathrm{d}v.
\end{aligned}
$$

By using Lemma 5.2 and Assumption (A1), one has that

$$
\begin{aligned}
\sup_{x\in\mathcal{I}_h, v\in[-1,1]}\big|s_n(x+hv)/f(x+hv)-1\big| &= \mathcal{O}(\kappa_n^{-\eta}),\\
\sup_{x\in\mathcal{I}_h, v\in[-1,1]}\big|f(x+hv)/f(x)-1\big| &= \mathcal{O}(h),\\
\sup_{x\in\mathcal{I}_h}\big|s_n(x)/f(x)-1\big| &= \mathcal{O}(\kappa_n^{-\eta}),\\
\sup_{x\in\mathcal{I}_h, v\in[-1,1]}\big|\sigma(x+hv)/\sigma(x)-1\big| &= \mathcal{O}(h).
\end{aligned}
$$

then one has that

$$
\begin{aligned}
\sup_{x\in\mathcal{I}_h, v\in[-1,1]}\big|s_n^{-1/2}(x)s_n^{1/2}(x+hv)-1\big| &\lesssim \sup_{x\in\mathcal{I}_h, v\in[-1,1]}\big|s_n^{-1}(x)s_n(x+hv)-1\big|\\
&= \mathcal{O}(h+\kappa_n^{-\eta})
\end{aligned}
$$

and

$$
\sup_{x\in\mathcal{I}_h, v\in[-1,1]}\big|\sigma^{-1}(x)s_n^{-1/2}(x)\sigma(x+hv)s_n^{1/2}(x+hv)-1\big| = \mathcal{O}(h+\kappa_n^{-\eta}).
$$

Hence, the first term is bounded by

$$
\begin{aligned}
&h^{-1/2}\sup_{x\in\mathcal{I}_h}\Big|\int_{-1}^{1} W_n(x+hv)K^{(1)}(v)\big\{\sigma^{-1}(x)s_n^{-1/2}(x)\sigma(x+hv)s_n^{1/2}(x+hv)-1\big\}\mathrm{d}v\Big|\\
&\lesssim h^{-1/2}\sup_{x\in\mathcal{I}_h, v\in[-1,1]}\big|W_n(x+hv)\big|\times\mathcal{O}(h+\kappa_n^{-\eta})\\
&= \mathcal{O}_p(h^{1/2}+h^{-1/2}\kappa_n^{-\eta}).
\end{aligned}
$$

The second term is bounded by $\mathcal{O}_p(h^{1/2})$ trivially. Hence,

$$
\sup_{x\in\mathcal{I}_h}\big|\Xi_{n,4}(x)-\Xi_{n,5}(x)\big| = \mathcal{O}_p(h^{1/2}+\kappa_n^{-\eta}h^{-1/2}) = o_p(\log^{-1/2} n).
$$

□

By putting Proposition 5.3–5.9 together, one can obtain the uniform asymptotic distribution of noise term.

Proposition 5.10 (Uniform asymptotic distribution of noise term). *Under Assumptions (A1)–(A3), as* $n \to \infty$*, if* $h = h_n$ *satisfies that* $h \log n \to 0$, $\kappa_n^{-\eta} h^{-2} \log n = \mathcal{O}(1)$, $\kappa_n n^{-1/2} h^{-1/2} \log^{5/2} n \to 0$ *and* $\log h^{-1} = \mathcal{O}(\log n)$*, one has that*

$$a_h \Big\{ \sqrt{nh} \Lambda^{-1} \sup_{x \in \mathcal{I}_h} \sigma^{-1}(x) f^{-1/2}(x) \big| A_n(x) \big| - b_h \Big\} \xrightarrow{d} Z.$$

Put Proposition 5.10 and Proposition 2.1, 2.2 together, one can obtain the uniform asymptotic distribution of Nadaraya–Waston estimation.

Theorem 5.11 (Uniform asymptotic distribution of Nadaraya–Waston estimation). *Under Assumptions (A1)–(A3), as* $n \to \infty$*, if* $h = h_n$ *satisfies that* $\kappa_n^{-\eta} h^{-2} \log n = \mathcal{O}(1)$, $\kappa_n n^{-1/2} h^{-1/2} \log^{5/2} n \to 0$, $nh^5 \log n \to 0$ *and* $\log h^{-1} = \mathcal{O}(\log n)$*, one has that*

$$a_h \Big\{ \sqrt{nh} \Lambda^{-1} \sup_{x \in \mathcal{I}_h} \sigma^{-1}(x) f^{1/2}(x) \big| \hat{m}(x) - m(x) \big| - b_h \Big\} \xrightarrow{d} Z.$$

Bibliography

Billingsley, P. (1999). Convergence of probability measures. *Wiley Series in Probability and Statistics.*

Bosq, D. (1998). *Nonparametric statistics for stochastic processes: estimation and prediction.* Springer, New York.

Cai, L., Gu, L., Wang, Q., and Wang, S. (2021). Simultaneous confidence bands for nonparametric regression with missing covariate data. *Ann Inst Stat Math*, 73(6):1249–1279.

Cai, L., Liu, R., Wang, S., and Yang, L. (2019). Simultaneous confidence bands for mean and variance functions based on deterministic design. *Stat Sin*, 29(1):505–525.

Eubank, R. L. and Speckman, P. L. (1993). Confidence bands in nonparametric regression. *J Am Stat Assoc*, 88(424):1287–1301.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* CRC Press, London.

Härdle, W. (1989). Asymptotic maximal deviation of M-smoothers. *J Multivar Anal*, 29(2):163–179.

Johnston, G. J. (1982). Probabilities of maximal deviations for nonparametric regression function estimates. *J Multivar Anal*, 12(3):402–414.

Tusnády, G. (1977). A remark on the approximation of the sample df in the multidimensional case. *Period Math Hung*, 8(1):53–55.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J R Stat Soc B*, 60(4):797–811.