

Kernel Density Estimation

Jiuzhou Miao

October 11, 2024



The first topic of this seminar is on *nonparametric estimation theories*, including *kernel density estimation*, *nonparametric local kernel and spline regression* and their asymptotic properties. Perhaps, more interesting topics (such as *minimax lower bounds* and *reproducing kernel Hilbert spaces*) will also be included. Throughout the whole topic, I will use the following notations:

For a function $\varphi(x), x \in \mathcal{D}$, use the notation $\|\varphi\|_\infty$ to denote $\sup_{x \in \mathcal{D}} |\varphi(x)|$. The notation $\varphi^{(p)}(\cdot)$ denotes the derivative of order p of the function $\varphi(\cdot)$ for some positive integer p and $\varphi^{(0)}(\cdot) = \varphi(\cdot)$. The notation $C^p(\mathcal{D})$ denotes the function space whose elements defined on \mathcal{D} have continuous derivatives of order p for some positive integer p and $C^0(\mathcal{D}) = C(\mathcal{D})$. For a stochastic function $\xi(x), x \in \mathcal{D}$, the notation $\mathcal{L}\{\xi(x), x \in \mathcal{D}\}$ denotes its distribution. For two positive sequences $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$, $a_n = o(b_n)$ means that $\lim_{n \rightarrow \infty} a_n/b_n = 0$ and $a_n = \mathcal{O}(b_n)$ or $a_n \lesssim b_n$ means that a_n/b_n is bounded by a positive constant. The notation $a_n \gtrsim b_n$ means that $b_n \lesssim a_n$ and $a_n \asymp b_n$ means that $a_n \lesssim b_n$ and $a_n \gtrsim b_n$. For a random variable sequence $\{X_n\}_{n=1}^\infty$, $X_n = o_p(1)$ [or $X_n = o_{a.s.}(1)$] means that X_n converges to zero in probability [or almost surely] as $n \rightarrow \infty$ and $X_n = \mathcal{O}_p(1)$ [or $X_n = \mathcal{O}_{a.s.}(1)$] means that X_n is bounded in probability [or almost surely]. For a non-stochastic function sequence $\{f_n(x), x \in \mathcal{D}\}_{n=1}^\infty$, the notation $f_n(x) = o(1)$ [or $f_n(x) = \mathcal{O}(1)$] means that $\sup_{x \in \mathcal{D}} |f_n(x)| = o(1)$ [or $\sup_{x \in \mathcal{D}} |f_n(x)| = \mathcal{O}(1)$]. For a stochastic function sequence, notations u_p , $u_{a.s.}$, \mathcal{U}_p and $\mathcal{U}_{a.s.}$ can be defined similarly.

1 Introduction

Let $\{X_i\}_{i=1}^n$ be independent random variables which have the same distribution with the random variable X . The distribution function and density function of X are $F(x) = \mathbb{P}(X \leq x)$ and $f(x) = F^{(1)}(x)$. We are interested in estimating the probability density function with nonparametric methods.

Note that the density function $f(\cdot)$ is the derivative of the distribution function $F(\cdot)$, one can obtain the approximation

$$f(x) \approx \{F(x+h) - F(x-h)\}/(2h) = \mathbb{P}(x-h < X \leq x+h)/(2h)$$

for some small h . A natural estimator of the distribution function $F(\cdot)$ is the well known empirical distribution function

$$\hat{F}_n(x) = n^{-1} \sum_{i=1}^n \mathbb{I}(X_i \leq x),$$

then one can use

$$\hat{F}_n(x+h) - \hat{F}_n(x-h) = n^{-1} \sum_{i=1}^n \mathbb{I}(x-h < X_i \leq x+h)$$

to estimate the probability $\mathbb{P}(x-h < X \leq x+h)$. Hence, one can use the estimator

$$\begin{aligned} \hat{f}^\dagger(x) &= (2nh)^{-1} \sum_{i=1}^n \mathbb{I}(x-h < X_i \leq x+h) \\ &= (nh)^{-1} \sum_{i=1}^n \mathbb{I}\left(-1 < h^{-1}(X_i - x) \leq 1\right)/2 \\ &= (nh)^{-1} \sum_{i=1}^n K^\dagger\left(h^{-1}(X_i - x)\right). \end{aligned}$$

to estimate $f(x)$.

Definition 1.1. (Kernel function) The function $K(\cdot)$ is called as *kernel function* if $K(\cdot)$ is a symmetrical probability density function, i.e.,

- $K(-x) = K(x)$;
- $\int_{-\infty}^{\infty} K(x)dx = 1$.

Example 1.2. (Some kernel functions)

(1) Rectangular kernel:

$$K(x) = 2^{-1} \mathbb{I}(|x| \leq 1);$$

(2) Gaussian kernel:

$$K(x) = (2\pi)^{-1/2} \exp(-x^2/2);$$

(3) Epanechnikov kernel:

$$K(x) = 3(1-x^2) \mathbb{I}(|x| \leq 1)/4;$$

(4) Biweight kernel:

$$K(x) = 15(1-x^2)^2 \mathbb{I}(|x| \leq 1)/16.$$

Definition 1.3. (Kernel density estimation) The estimation

$$\hat{f}(x) = n^{-1} \sum_{i=1}^n K_h(X_i - x)$$

is called as *kernel density estimation*. In which h is a smooth parameter called as *bandwidth* and $K_h(\cdot) = K(\cdot/h)/h$ is a *rescaled kernel function* with $K(\cdot)$ by the bandwidth h .

In the following, we will establish the asymptotic properties of the estimator $\hat{f}(x)$, including derive the uniform convergence rate, pointwise and global asymptotic distribution. For convenience, assume that the support of X is $\mathcal{X} = [0, 1]$. The following assumption will be used throughout the whole notes.

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

Denote $\mathcal{I}_h = [h, 1 - h]$. We introduce the error decomposition of $\hat{f}(x) - f(x)$ by

$$\hat{f}(x) - f(x) = \underbrace{\hat{f}(x) - \mathbb{E}\{\hat{f}(x)\}}_{\text{Noise term}} + \underbrace{\mathbb{E}\{\hat{f}(x)\} - f(x)}_{\text{Bias term}}.$$

2 Uniform convergence

In this section, we will derive the uniform convergence rate of $\sup_{x \in \mathcal{I}_h} |\hat{f}(x) - f(x)|$ under some mild conditions. We start from the bias term.

Proposition 2.1. (*Uniform convergence of bias term*) Under Assumption (A), as $n \rightarrow \infty$, if $h = h_n \rightarrow 0$, one has that

$$\sup_{x \in \mathcal{I}_h} \left| \mathbb{E}\{\hat{f}(x)\} - f(x) - 2^{-1}f^{(2)}(x)h^2 \int_{-1}^1 v^2 K(v)dv \right| = o(h^2)$$

and

$$\sup_{x \in \mathcal{I}_h} \left| \mathbb{E}\{\hat{f}(x)\} - f(x) \right| = \mathcal{O}(h^2).$$

Proof. We only need to calculate the expectation of $\hat{f}(x)$.

$$\begin{aligned} \mathbb{E}\{\hat{f}(x)\} &= \mathbb{E}\{K_h(X_1 - x)\} \\ &= \int_{\mathcal{X}} h^{-1} K(h^{-1}(u - x)) f(u) du \\ &= \int_{-1}^1 K(v) f(x + hv) dv \\ &= \int_{-1}^1 K(v) \{f(x) + f^{(1)}(x)hv + 2^{-1}f^{(2)}(x)(hv)^2 + u(h^2)\} dv \\ &= f(x) + 2^{-1}f^{(2)}(x)h^2 \int_{-1}^1 v^2 K(v)dv + o(h^2), \end{aligned}$$

then one can obtain the first equation. The second equation can be obtained from

$$\begin{aligned} \sup_{x \in \mathcal{I}_h} \left| \mathbb{E}\{\hat{f}(x)\} - f(x) \right| &\leq \sup_{x \in \mathcal{I}_h} \left| \mathbb{E}\{\hat{f}(x)\} - f(x) - 2^{-1}f^{(2)}(x)h^2 \int_{-1}^1 v^2 K(v)dv \right| \\ &\quad + 2^{-1}h^2 \int_{-1}^1 v^2 K(v)dv \times \sup_{x \in \mathcal{I}_h} |f^{(2)}(x)| \\ &= \mathcal{O}(h^2). \end{aligned}$$

□

Then we deal with the noise term. The following useful Bernstein's inequality is a quite powerful weapon in estimating the uniform convergence rate of the summation of independent random variables.

Lemma 2.2. (*Bosq (1998), Theorem 1.2*) Let $\{\eta_i\}_{i=1}^n$ be independent random variables with mean zero. If there exists $r > 0$ such that (Cramér's conditions)

$$\mathbb{E}(|\eta_i|^k) \leq r^{k-2} k! \mathbb{E}(\eta_i^2) < \infty, 1 \leq i \leq n, k \geq 3,$$

then for any $t > 0$,

$$\mathbb{P}\left\{\left|\sum_{i=1}^n \eta_i\right| > t\right\} \leq 2 \exp\left\{-\frac{t^2}{4 \sum_{i=1}^n \mathbb{E}(\eta_i^2) + 2rt}\right\}.$$

Proposition 2.3. (*Uniform convergence of noise term*) Under Assumption (A), as $n \rightarrow \infty$, if $h = h_n$ satisfies that $h \rightarrow 0$ and $n^{-1/2} h^{-1/2} \log^{1/2} n \rightarrow 0$, one has that

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

Proof. Let $\xi_{in}(x) = n^{-1} K_h(X_i - x)$ and $\check{\xi}_{in}(x) = \xi_{in}(x) - \mathbb{E}\{\xi_{in}(x)\}$, then one has that

$$\hat{f}(x) - \mathbb{E}\{\hat{f}(x)\} = \sum_{i=1}^n \check{\xi}_{in}(x).$$

One can obtain that $\mathbb{E}\{\xi_{in}(x)\} \asymp n^{-1}$ and

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

Then, one has that $\mathbb{E}\{\check{\xi}_{in}(x)\} = 0$, $\mathbb{E}\{\check{\xi}_{in}^2(x)\} \asymp n^{-2} h^{-1}$. Note that

$$|\check{\xi}_{in}(x)| \leq |\xi_{in}(x)| + \mathbb{E}\{|\xi_{in}(x)|\} \leq 2\|K\|_{\infty} n^{-1} h^{-1},$$

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

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

which implies that Cramér's conditions hold with $r = 2\|K\|_\infty n^{-1}h^{-1}$. Then by the Bernstein's inequality in Lemma 2.2, for some $\delta > 0$, one has that

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

In the following, we discretize \mathcal{I}_h by $h = x_0 < x_1 < \dots < x_{n^4-1} = 1 - h$. By the subadditivity of probability, one has that

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

By Borel-Cantelli's lemma, one obtains that

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

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

$$|\xi_{in}(x) - \xi_{in}(x_l)| \leq n^{-1}h^{-2}\|K^{(1)}\|_\infty |x - x_l| \leq n^{-5}h^{-2}\|K^{(1)}\|_\infty(1 - 2h)$$

and

$$\begin{aligned} \max_{0 \leq l \leq n^4-2} \sup_{x \in [x_l, x_{l+1}]} \left|\sum_{i=1}^n \xi_{in}(x) - \xi_{in}(x_l)\right| & \leq 2n^{-4}h^{-2}\|K^{(1)}\|_\infty(1 - 2h) \\ & = \mathcal{O}_p(n^{-4}h^{-2}). \end{aligned}$$

Hence, one has that

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

□

Put Proposition 2.1 and 2.3 together, one can obtain the following theorem.

Theorem 2.4. (*Uniform convergence of kernel density estimation*) Under the assumptions of Proposition 2.3, one has that

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

and

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

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

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

which is the optimal convergence rate of kernel density estimation.

3 Pointwise asymptotic distribution

In this section, we will derive the pointwise asymptotic distribution of $\hat{f}(x) - f(x)$ for any $x \in \mathcal{I}_h$ under some mild conditions. The following Lindeberg–Feller’s CLT is a powerful tool to obtain the asymptotic normality of $\hat{f}(x) - \mathbb{E}\{\hat{f}(x)\}$.

Lemma 3.1. (*Lindeberg–Feller’s CLT*) Let $\{\eta_{in}\}_{i=1}^{k_n}$, $n \geq 1$ be independent random variables with mean zero, $\mathbb{E}(\eta_{in}^2) = \sigma_{in}^2$. Denote $S_n = \sum_{i=1}^{k_n} \eta_{in}$ and $V_n^2 = \sum_{i=1}^{k_n} \sigma_{in}^2$. Then $V_n^{-1} S_n \xrightarrow{d} \mathcal{N}(0, 1)$ if either of the following conditions holds:

- (*Lindeberg’s condition*) for any $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} V_n^{-2} \sum_{i=1}^{k_n} \mathbb{E} \left\{ \eta_{in}^2 \mathbb{I}(|\eta_{in}| \geq \varepsilon V_n) \right\} = 0.$$

- (*Lyapunov’s condition*) for some $\delta > 0$,

$$\lim_{n \rightarrow \infty} V_n^{-(2+\delta)} \sum_{i=1}^{k_n} \mathbb{E}(|\eta_{in}|^{2+\delta}) = 0.$$

Proposition 3.2. (*Asymptotic normality of noise term*) Under Assumption (A), as $n \rightarrow \infty$, if $h = h_n$ satisfies that $nh \rightarrow \infty$, one has that for a fixed $x \in \mathcal{I}_h$,

$$\sqrt{nh} \left[\hat{f}(x) - \mathbb{E}\{\hat{f}(x)\} \right] \xrightarrow{d} \mathcal{N}(0, f(x) \int_{-1}^1 K^2(v) dv).$$

Proof. Note that

$$\mathbb{E}\{\check{\xi}_{in}^2(x)\} = n^{-2} h^{-1} f(x) \int_{-1}^1 K^2(v) dv \{1 + \mathcal{U}(h^2)\},$$

then one has that

$$V_n^2(x) = n^{-1}h^{-1}f(x) \int_{-1}^1 K^2(v)dv \{1 + \mathcal{U}(h^2)\} \asymp n^{-1}h^{-1}.$$

and $V_n^{-(2+\delta)}(x) \asymp (nh)^{1+\delta/2}$. Use the inequality $(a+b)^p \leq 2^{p-1}(a^p + b^p)$, $a, b > 0$ and $p \geq 1$, one has that

$$\mathbb{E}\left\{|\check{\xi}_{in}(x)|^{2+\delta}\right\} \leq 2^{1+\delta} \left[\mathbb{E}\left\{|\xi_{in}(x)|^{2+\delta}\right\} + \left| \mathbb{E}\{\xi_{in}(x)\} \right|^{2+\delta} \right].$$

Note that

$$\begin{aligned} \mathbb{E}\left\{|\xi_{in}(x)|^{2+\delta}\right\} &= \int_{\mathcal{X}} n^{-(2+\delta)} h^{-(2+\delta)} K^{2+\delta}(h^{-1}(u-x)) f(u) du \\ &= \int_{-1}^1 n^{-(2+\delta)} h^{-(1+\delta)} K^{2+\delta}(v) f(x+hv) dv \\ &\leq n^{-(2+\delta)} h^{-(1+\delta)} \sup_{x \in \mathcal{X}} f(x) \int_{-1}^1 K^{2+\delta}(v) dv \\ &= \mathcal{U}(n^{-(2+\delta)} h^{-(1+\delta)}) \end{aligned}$$

and $\left| \mathbb{E}\{\xi_{in}(x)\} \right|^{2+\delta} = \mathcal{U}(n^{-(2+\delta)})$, then

$$\mathbb{E}\left\{|\check{\xi}_{in}(x)|^{2+\delta}\right\} \leq 2^{1+\delta} \left[\mathbb{E}\left\{|\xi_{in}(x)|^{2+\delta}\right\} + \left| \mathbb{E}\{\xi_{in}(x)\} \right|^{2+\delta} \right] = \mathcal{U}(n^{-(2+\delta)} h^{-(1+\delta)}).$$

Hence, one has that

$$V_n^{-(2+\delta)}(x) \sum_{i=1}^n \mathbb{E}\left\{|\check{\xi}_{in}(x)|^{2+\delta}\right\} = \mathcal{O}(n^{-\delta/2} h^{-\delta/2}) = o(1),$$

and Lyapunov's condition holds. By using Lemma 3.1, one has that

$$V_n^{-1}(x) \sum_{i=1}^n \check{\xi}_{in}(x) \xrightarrow{d} \mathcal{N}(0, 1)$$

and by using Slutsky's theorem, one has that

$$\begin{aligned} \frac{\sqrt{nh} \left[\hat{f}(x) - \mathbb{E}\{\hat{f}(x)\} \right]}{\sqrt{f(x) \int_{-1}^1 K^2(v) dv}} &= \frac{\sum_{i=1}^n \check{\xi}_{in}(x)}{\sqrt{n^{-1}h^{-1}f(x) \int_{-1}^1 K^2(v) dv}} \\ &= V_n^{-1}(x) \sum_{i=1}^n \check{\xi}_{in}(x) \times \frac{V_n(x)}{\sqrt{n^{-1}h^{-1}f(x) \int_{-1}^1 K^2(v) dv}} \\ &\xrightarrow{d} \mathcal{N}(0, 1). \end{aligned}$$

□

Note that if $nh^5 = \mathcal{O}(1)$, then

$$\sqrt{nh} \sup_{x \in \mathcal{I}_h} \left| \mathbb{E}\{\hat{f}(x)\} - f(x) - 2^{-1}f^{(2)}(x)h^2 \int_{-1}^1 v^2 K(v)dv \right| = o(n^{1/2}h^{5/2}) = o(1),$$

by using Slutsky's theorem, one can obtain the asymptotic normality of kernel density estimation.

Theorem 3.3. (*Asymptotic normality of kernel density estimation*) Under Assumption (A), as $n \rightarrow \infty$, if $h = h_n$ satisfies that $nh \rightarrow \infty$ and $nh^5 = \mathcal{O}(1)$, one has that for a fixed $x \in \mathcal{I}_h$,

$$\sqrt{nh} \left\{ \hat{f}(x) - f(x) - 2^{-1}f^{(2)}(x)h^2 \int_{-1}^1 v^2 K(v)dv \right\} \xrightarrow{d} \mathcal{N}\left(0, f(x) \int_{-1}^1 K^2(v)dv\right).$$

Furthermore, assume that $nh^5 \rightarrow 0$, one has that for a fixed $x \in \mathcal{I}_h$,

$$\sqrt{nh} \{ \hat{f}(x) - f(x) \} \xrightarrow{d} \mathcal{N}\left(0, f(x) \int_{-1}^1 K^2(v)dv\right).$$

4 Uniform asymptotic distribution

In this section, we will derive the uniform asymptotic distribution of $\hat{f}(x) - f(x)$ over the whole $x \in \mathcal{I}_h$ under some mild conditions, i.e., derive the asymptotic distribution of $\sup_{x \in \mathcal{I}_h} |\hat{f}(x) - f(x)|$. By using the uniform asymptotic distribution, one can construct $100\%(1 - \alpha)$ simultaneous confidence band for the density function $f(x)$, $x \in \mathcal{I}_h$, which is more difficult in techniques but more useful in practice. The research on constructing simultaneous confidence band for the density function is firstly studied by [Bickel and Rosenblatt \(1973\)](#), while the proofs given in this section mainly refer to the notes written by Prof. Lijian Yang in 2011 which is a simplified version of [Bickel and Rosenblatt \(1973\)](#).

Recall Proposition 2.1, one knows that under Assumption (A)

$$\sup_{x \in \mathcal{I}_h} \left| \mathbb{E}\{\hat{f}(x)\} - f(x) \right| = \mathcal{O}(h^2),$$

if $h = h_n \rightarrow 0$ as $n \rightarrow \infty$. Then we only need to derive the asymptotic distribution of $\sup_{x \in \mathcal{I}_h} |\hat{f}(x) - \mathbb{E}\{\hat{f}(x)\}|$. The technical tool required for this section is the strong approximation theorem proposed by [Komlós et al. \(1975\)](#)

Lemma 4.1. ([Komlós et al. \(1975\)](#), Theorem 3) For a fixed n , let $\{\eta_i\}_{i=1}^n$ be i.i.d. random variables with

$$\mathbb{P}(\eta_1 \leq t) = \begin{cases} 0, & t < 0 \\ t, & 0 \leq t \leq 1 \\ 1, & t > 1 \end{cases},$$

let $F_n(t)$ be the empirical distribution function based on the sample $\{\eta_i\}_{i=1}^n$ and let $B_n(t)$ be a Brownian bridge. There is a version of $F_n(t)$ and $B_n(t)$ such that

$$\mathbb{P} \left[\sup_{t \in [0,1]} \left| n\{F_n(t) - t\} - \sqrt{n}B_n(t) \right| > C \log n + x \right] < ke^{-\lambda x}$$

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

Denote the empirical process $Z_n(t) = \sqrt{n}\{\hat{F}_n(t) - F(t)\}$. Here, $\hat{F}_n(t)$ and $F(t)$ are empirical distribution function and distribution function based on $\{X_i\}_{i=1}^n$. By using Lemma 4.1, one can easily obtain that

$$\sup_{t \in [0,1]} |Z_n(t) - B_n(F(t))| = \mathcal{O}_{a.s.}(n^{-1/2} \log n).$$

Define the standardized stochastic process

$$\begin{aligned} \Xi_n(x) &= n^{1/2} h^{1/2} f^{-1/2}(x) [\hat{f}(x) - \mathbb{E}\{\hat{f}(x)\}] \\ &= n^{1/2} h^{1/2} f^{-1/2}(x) \int_{\mathcal{X}} K_h(u - x) d\{\hat{F}_n(u) - F(u)\} \\ &= h^{1/2} f^{-1/2}(x) \int_{\mathcal{X}} K_h(u - x) dZ_n(u) \\ &= h^{-1/2} f^{-1/2}(x) \int_{\mathcal{X}} K(h^{-1}(u - x)) dZ_n(u) \end{aligned}$$

and stochastic processes

$$\begin{aligned} \Xi_{n,0}(x) &= h^{-1/2} f^{-1/2}(x) \int_{\mathcal{X}} K(h^{-1}(u - x)) dB_n(F(u)) \\ \Xi_{n,1}(x) &= h^{-1/2} f^{-1/2}(x) \int_{\mathcal{X}} K(h^{-1}(u - x)) dW_n(F(u)) \\ \Xi_{n,2}(x) &= h^{-1/2} f^{-1/2}(x) \int_{\mathcal{X}} K(h^{-1}(u - x)) f^{1/2}(u) dW_n(u) \\ \Xi_{n,3}(x) &= h^{-1/2} \int_{\mathcal{X}} K(h^{-1}(u - x)) dW_n(u). \end{aligned}$$

We will prove that $\Xi_n(x)$ can be uniformly approximated by $\Xi_{n,3}(x)$ over $x \in \mathcal{I}_h$ in Proposition 4.2–4.5 and $\Xi_{n,3}(x)$ is asymptotically Gumbel in Proposition 4.7.

Proposition 4.2. *Under Assumption (A), as $n \rightarrow \infty$, if $h = h_n \rightarrow 0$, one has that*

$$\sup_{x \in \mathcal{I}_h} |\Xi_n(x) - \Xi_{n,0}(x)| = \mathcal{O}_{a.s.}(n^{-1/2} h^{-1/2} \log n).$$

Proof. Note that

$$\begin{aligned} \Xi_n(x) - \Xi_{n,0}(x) &= h^{-1/2} f^{-1/2}(x) \int_{\mathcal{X}} K(h^{-1}(u - x)) d\{Z_n(u) - B_n(F(u))\} \\ &= h^{-1/2} f^{-1/2}(x) K(h^{-1}(u - x)) \left\{ Z_n(u) - B_n(F(u)) \right\} \Big|_{\mathcal{X}} \\ &\quad - h^{-1/2} f^{-1/2}(x) \int_{\mathcal{X}} \left\{ Z_n(u) - B_n(F(u)) \right\} d\{K(h^{-1}(u - x))\} \\ &= -h^{-3/2} f^{-1/2}(x) \int_{\mathcal{X}} \left\{ Z_n(u) - B_n(F(u)) \right\} K^{(1)}(h^{-1}(u - x)) du, \end{aligned}$$

one has that

$$\begin{aligned} \sup_{x \in \mathcal{I}_h} |\Xi_n(x) - \Xi_{n,0}(x)| &\lesssim h^{-3/2} \times \mathcal{O}_{a.s.}(n^{-1/2} \log n) \times h \int_{-1}^1 |K^{(1)}(v)| dv \\ &= \mathcal{O}_{a.s.}(n^{-1/2} h^{-1/2} \log n). \end{aligned}$$

□

Proposition 4.3. *Under Assumption (A), as $n \rightarrow \infty$, if $h = h_n \rightarrow 0$, one has that*

$$\sup_{x \in \mathcal{I}_h} |\Xi_{n,0}(x) - \Xi_{n,1}(x)| = \mathcal{O}_p(h^{1/2}).$$

Proof. By the definition of Brownian bridge, one has that $B_n(F(u)) - W_n(F(u)) = -F(u)W_n(1)$. Hence, one has that

$$\begin{aligned} \Xi_{n,0}(x) - \Xi_{n,1}(x) &= -W_n(1)h^{-1/2}f^{-1/2}(x) \int_{\mathcal{X}} K(h^{-1}(u-x))f(u)du \\ &= -W_n(1)h^{1/2}f^{-1/2}(x) \int_{-1}^1 K(v)f(x+hv)dv \\ &= -W_n(1)h^{1/2}f^{-1/2}(x) \int_{-1}^1 K(v)f(x+hv)dv \end{aligned}$$

and

$$\begin{aligned} \sup_{x \in \mathcal{I}_h} |\Xi_{n,0}(x) - \Xi_{n,1}(x)| &\lesssim |W_n(1)| \times h^{1/2} \int_{-1}^1 K(v)dv \\ &= \mathcal{O}_p(h^{1/2}). \end{aligned}$$

□

Proposition 4.4. *Under Assumption (A), as $n \rightarrow \infty$, if $h = h_n \rightarrow 0$, one has that*

$$\mathcal{L}\{\Xi_{n,1}(x), x \in \mathcal{I}_h\} = \mathcal{L}\{\Xi_{n,2}(x), x \in \mathcal{I}_h\}.$$

Proof. Note that $\Xi_{n,1}(x)$ and $\Xi_{n,2}(x)$ are two Gaussian process with zero mean function, we only need to verify that they have same covariance function. By Itô's isometric, one has that

$$\begin{aligned} \mathbb{E}\{\Xi_{n,1}(x)\Xi_{n,1}(x')\} &= h^{-1}f^{-1/2}(x)f^{-1/2}(x') \int_{\mathcal{X}} K(h^{-1}(u-x))K(h^{-1}(u-x'))dF(u) \\ &= h^{-1}f^{-1/2}(x)f^{-1/2}(x') \int_{\mathcal{X}} K(h^{-1}(u-x))K(h^{-1}(u-x'))f(u)du \end{aligned}$$

and

$$\mathbb{E}\{\Xi_{n,2}(x)\Xi_{n,2}(x')\} = h^{-1}f^{-1/2}(x)f^{-1/2}(x') \int_{\mathcal{X}} K(h^{-1}(u-x))K(h^{-1}(u-x'))f(u)du$$

□

Proposition 4.5. *Under Assumption (A), as $n \rightarrow \infty$, if $h = h_n \rightarrow 0$, one has that*

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

Proof. Note that

$$\begin{aligned}
 \Xi_{n,2}(x) - \Xi_{n,3}(x) &= h^{-1/2} \int_{\mathcal{X}} K(h^{-1}(u-x)) \{f^{1/2}(u)f^{-1/2}(x) - 1\} dW_n(u) \\
 &= -h^{-1/2} \int_{\mathcal{X}} W_n(u) d\left[K(h^{-1}(u-x)) \{f^{1/2}(u)f^{-1/2}(x) - 1\}\right] \\
 &= -h^{-3/2} \int_{\mathcal{X}} W_n(u) K^{(1)}(h^{-1}(u-x)) \{f^{1/2}(u)f^{-1/2}(x) - 1\} du \\
 &\quad - 2^{-1} h^{-1/2} \int_{\mathcal{X}} W_n(u) K(h^{-1}(u-x)) f^{-1/2}(u) f^{-1/2}(x) f^{(1)}(u) du \\
 &= -h^{-1/2} \int_{-1}^1 W_n(x+hv) K^{(1)}(v) \{f^{1/2}(x+hv)f^{-1/2}(x) - 1\} dv \\
 &\quad - 2^{-1} h^{1/2} \int_{-1}^1 W_n(x+hv) K(v) f^{-1/2}(x+hv) f^{-1/2}(x) f^{(1)}(x+hv) dv,
 \end{aligned}$$

and

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

then one has that

$$\begin{aligned}
 &h^{-1/2} \sup_{x \in \mathcal{I}_h} \left| \int_{-1}^1 W_n(x+hv) K^{(1)}(v) \{f^{1/2}(x+hv)f^{-1/2}(x) - 1\} dv \right| \\
 &\leq h^{-1/2} \sup_{x \in \mathcal{I}_h, v \in [-1,1]} |f^{1/2}(x+hv)f^{-1/2}(x) - 1| \sup_{x \in \mathcal{I}_h, v \in [-1,1]} |W_n(x+hv)| \int_{-1}^1 |K^{(1)}(v)| dv \\
 &= \mathcal{O}_p(h^{1/2}),
 \end{aligned}$$

and

$$\begin{aligned}
 &h^{1/2} \sup_{x \in \mathcal{I}_h} \left| \int_{-1}^1 W_n(x+hv) K(v) f^{-1/2}(x+hv) f^{-1/2}(x) f^{(1)}(x+hv) dv \right| \\
 &\lesssim h^{1/2} \sup_{x \in \mathcal{I}_h, v \in [-1,1]} |W_n(x+hv)| \\
 &= \mathcal{O}_p(h^{1/2}).
 \end{aligned}$$

□

The following is a reformulation of Theorem 11.1.5 and Theorem 12.3.5 of [Leadbetter et al. \(1983\)](#)

Lemma 4.6. ([Leadbetter et al. \(1983\)](#), Theorem 11.1.5 and Theorem 12.3.5) *If the Gaussian process $\zeta(s), 0 \leq t \leq T$ is stationary with mean zero and variance one, and correlation function satisfying for some constant $C > 0$,*

$$\mathbb{E}\{\zeta(s)\zeta(s+t)\} = 1 - C|t|^\alpha + o(|t|^\alpha)$$

as $t \rightarrow 0$, then as $T \rightarrow \infty$,

$$a_T \left(\sup_{s \in [0, T]} |\zeta(s)| - b_T \right) \xrightarrow{d} Z,$$

where the random variable Z satisfies $\mathbb{P}(Z \leq z) = \exp\{-2 \exp(-z)\}$ for all $z \in \mathbb{R}$, and

$$a_T = (2 \log T)^{1/2},$$

$$b_T = a_T + a_T^{-1} \left[(\alpha^{-1} - 2^{-1}) \log(a_T^2/2) + \log \{C^{1/\alpha} H_\alpha(2\pi)^{-1/2} 2^{(\alpha^{-1} - 2^{-1})}\} \right]$$

with $H_1 = 1$ and $H_2 = \pi^{-1/2}$.

Proposition 4.7. *Under Assumption (A), as $n \rightarrow \infty$, if $h = h_n \rightarrow 0$, one has that*

$$a_h \left(\sup_{x \in \mathcal{I}_h} |\Xi_{n,3}(x)/\Lambda| - b_h \right) \xrightarrow{d} Z,$$

in which

$$a_h = (2 \log h^{-1})^{1/2}, \quad b_h = a_h + 2^{-1} a_h^{-1} \log \left(\frac{C}{2\pi^2} \right),$$

$$C = 2^{-1} \Lambda^{-1} \int_{-1}^{-1} \{K^{(1)}(v)\}^2 dv, \quad \Lambda = \sqrt{\int_{-1}^1 K^2(v) dv}$$

Proof. Note that

$$\begin{aligned} \mathcal{L}\{\Xi_{n,3}(x), x \in \mathcal{I}_h\} &= \mathcal{L}\left\{h^{-1/2} \int_{\mathcal{X}} K(h^{-1}(u-x)) dW_n(u), x \in \mathcal{I}_h\right\} \\ &= \mathcal{L}\left\{h^{-1/2} \int_{\mathcal{X}} K(h^{-1}u-x) dW_n(u), x \in [1, h^{-1}-1]\right\} \\ &= \mathcal{L}\left\{\int_0^{h^{-1}} K(u-x) dW_n(u), x \in [1, h^{-1}-1]\right\} \\ &= \mathcal{L}\left\{\int_0^{h^{-1}} K(u-1-x) dW_n(u), x \in [0, h^{-1}-2]\right\} \end{aligned}$$

denote $\xi(x) = \int_0^{h^{-1}} K(u-1-x) dW_n(u)$, $x \in [0, h^{-1}-2]$, one has that

$$\begin{aligned} \mathbb{E}\{\xi(x)\xi(x')\} &= \int_0^{h^{-1}} K(u-1-x)K(u-1-x') du \\ &= \int_{-1}^1 K(v)K(v+x-x') dv \end{aligned}$$

and $\mathbb{E}\{\xi^2(x)\} = \int_{-1}^1 K^2(v) dv$. Then

$$\begin{aligned} r(x-x') &= \frac{\mathbb{E}\{\xi(x)\xi(x')\}}{\sqrt{\mathbb{E}\{\xi^2(x)\}\mathbb{E}\{\xi^2(x')\}}} \\ &= \frac{\int_{-1}^1 K(v)K(v+x-x') dv}{\int_{-1}^1 K^2(v) dv} \end{aligned}$$

and

$$\begin{aligned}
 r(t) - 1 &= \frac{\int_{-1}^{-1} K(v)K(v+t)dv}{\int_{-1}^{-1} K^2(v)dv} - 1 \\
 &= \frac{\int_{-1}^{-1} K(v)\{K(v+t) - K(v)\}dv}{\int_{-1}^{-1} K^2(v)dv} \\
 &= \frac{\int_{-1}^{-1} K(v)\{K^{(1)}(v)t + 2^{-1}K^{(2)}(v)t^2 + o(t^2)\}dv}{\int_{-1}^{-1} K^2(v)dv} \\
 &= \frac{\int_{-1}^{-1} K(v)K^{(2)}(v)dv}{2 \int_{-1}^{-1} K^2(v)dv} t^2 + o(t^2) \\
 &= \frac{\int_{-1}^{-1} K(v)dK^{(1)}(v)}{2 \int_{-1}^{-1} K^2(v)dv} t^2 + o(t^2) \\
 &= -\frac{\int_{-1}^{-1} \{K^{(1)}(v)\}^2 dv}{2 \int_{-1}^{-1} K^2(v)dv} t^2 + o(t^2)
 \end{aligned}$$

Then the conditions of Lemma 4.6 holds with $\alpha = 2$, $C = 2^{-1}\Lambda^{-1} \int_{-1}^{-1} \{K^{(1)}(v)\}^2 dv$ in which $\Lambda = \sqrt{\int_{-1}^{-1} K^2(v)dv}$, $T = h^{-1} - 2$, $a_T = (2 \log T)^{1/2}$ and $b_T = a_T + 2^{-1}a_T^{-1} \log(\frac{C}{2\pi^2})$, then one has that

$$a_T \left(\sup_{x \in [0, T]} |\xi(x)/\Lambda| - b_T \right) \xrightarrow{d} Z.$$

as $T \rightarrow \infty$. Besides, one can easily obtain that

$$a_h a_T^{-1} \rightarrow 1, \quad a_h(b_T - b_h) \rightarrow 0.$$

Therefore, by using Slutsky's theorem, one can obtain that as $n \rightarrow \infty$,

$$\begin{aligned}
 a_h \left(\sup_{x \in \mathcal{I}_h} |\Xi_{n,3}(x)/\Lambda| - b_h \right) &= a_h \left(\sup_{x \in [0, T]} |\xi(x)/\Lambda| - b_h \right) \\
 &= a_h a_T^{-1} a_T \left(\sup_{x \in [0, T]} |\xi(x)/\Lambda| - b_T \right) + a_h(b_T - b_h) \xrightarrow{d} Z.
 \end{aligned}$$

□

Proposition 4.8. (*Uniform asymptotic distribution of noise term*) Under Assumption (A), as $n \rightarrow \infty$, if $h = h_n$ satisfies that $h \log n \rightarrow 0$ and $n^{-1/2}h^{-1/2} \log^{3/2} n \rightarrow 0$ and $\log h^{-1} = \mathcal{O}(\log n)$, one has that

$$a_h \left[\sqrt{nh} \Lambda^{-1} \sup_{x \in \mathcal{I}_h} f^{-1/2}(x) \left| \hat{f}(x) - \mathbb{E}\{\hat{f}(x)\} \right| - b_h \right] \xrightarrow{d} Z.$$

Proof. Note that $a_h = (2 \log h^{-1})^{1/2} = \mathcal{O}(\log^{1/2} n)$,

$$\begin{aligned}
 \left| a_h \left(\sup_{x \in \mathcal{I}_h} |\Xi_{n,2}(x)/\Lambda| - b_h \right) - a_h \left(\sup_{x \in \mathcal{I}_h} |\Xi_{n,3}(x)/\Lambda| - b_h \right) \right| &\lesssim a_h \sup_{x \in \mathcal{I}_h} |\Xi_{n,2}(x) - \Xi_{n,3}(x)| \\
 &= \mathcal{O}_p(a_h h^{1/2}) \\
 &= \mathcal{O}_p(h^{1/2} \log^{1/2} n) \\
 &= o_p(1), \\
 \left| a_h \left(\sup_{x \in \mathcal{I}_h} |\Xi_{n,0}(x)/\Lambda| - b_h \right) - a_h \left(\sup_{x \in \mathcal{I}_h} |\Xi_{n,1}(x)/\Lambda| - b_h \right) \right| &\lesssim a_h \sup_{x \in \mathcal{I}_h} |\Xi_{n,0}(x) - \Xi_{n,1}(x)| \\
 &= \mathcal{O}_p(a_h h^{1/2}) \\
 &= \mathcal{O}_p(h^{1/2} \log^{1/2} n) \\
 &= o_p(1), \\
 \left| a_h \left(\sup_{x \in \mathcal{I}_h} |\Xi_n(x)/\Lambda| - b_h \right) - a_h \left(\sup_{x \in \mathcal{I}_h} |\Xi_{n,0}(x)/\Lambda| - b_h \right) \right| &\lesssim a_h \sup_{x \in \mathcal{I}_h} |\Xi_n(x) - \Xi_{n,0}(x)| \\
 &= \mathcal{O}_{a.s.}(a_h n^{-1/2} h^{-1/2} \log n) \\
 &= \mathcal{O}_{a.s.}(n^{-1/2} h^{-1/2} \log^{3/2} n) \\
 &= o_{a.s.}(1),
 \end{aligned}$$

and

$$\mathcal{L}\{\Xi_{n,1}(x), x \in \mathcal{I}_h\} = \mathcal{L}\{\Xi_{n,2}(x), x \in \mathcal{I}_h\},$$

then one has that

$$a_h \left(\sup_{x \in \mathcal{I}_h} |\Xi_n(x)/\Lambda| - b_h \right) \xrightarrow{d} Z,$$

i.e.,

$$a_h \left[\sqrt{nh} \Lambda^{-1} \sup_{x \in \mathcal{I}_h} f^{-1/2}(x) |\hat{f}(x) - \mathbb{E}\{\hat{f}(x)\}| - b_h \right] \xrightarrow{d} Z.$$

□

Put Proposition 4.8 and Proposition 2.1 together, one can obtain the uniform asymptotic distribution of kernel density estimation.

Theorem 4.9. (*Uniform asymptotic distribution of kernel density estimation*) Under Assumptions (A), as $n \rightarrow \infty$, if $h = h_n$ satisfies that $nh^5 \log n \rightarrow 0$, $n^{-1/2} h^{-1/2} \log^{3/2} n \rightarrow 0$ and $\log h^{-1} = \mathcal{O}(\log n)$, one has that

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

Proof. Note that

$$\begin{aligned} & \left| a_h \left\{ \sqrt{nh} \Lambda^{-1} \sup_{x \in \mathcal{I}_h} f^{-1/2}(x) |\hat{f}(x) - f(x)| - b_h \right\} \right. \\ & \quad \left. - a_h \left[\sqrt{nh} \Lambda^{-1} \sup_{x \in \mathcal{I}_h} f^{-1/2}(x) |\hat{f}(x) - \mathbb{E}\{\hat{f}(x)\}| - b_h \right] \right| \\ & \lesssim a_h \sqrt{nh} \sup_{x \in \mathcal{I}_h} |\mathbb{E}\{\hat{f}(x)\} - f(x)| \\ & = \mathcal{O}(n^{1/2} h^{5/2} \log^{1/2} n) \\ & = o(1), \end{aligned}$$

then Theorem 4.9 can be obtained by using Slutsky's theorem. □

Bibliography

- Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann Stat*, 1:1071–1095.
- Bosq, D. (1998). *Nonparametric statistics for stochastic processes: estimation and prediction*. Springer, New York.
- Komlós, J., Major, P., and Tusnády, G. (1975). An approximation of partial sums of independent rv's, and the sample df. i. *Probab Theory Rel*, 32:111–131.
- Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer, New York.