Professor SU Liangjun  
Fall, 2014  

Nonparametric Econometrics- Homework 1  
Due date: 2014/09/16  

Part I: Theoretical Questions  

1. (25 points) As discussed in the text, the objective function of the leave-one-out least squares cross-validation (LSCV) for kernel density estimation can be written as

$$CV_f (h) = \frac{1}{n^2 h} \sum_{i=1}^{n} \sum_{j=1}^{n} k \left( \frac{X_i - X_j}{h} \right) - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k \left( \frac{X_i - X_j}{h} \right),$$  

(0.1)

where for simplicity we consider the estimation of the PDF $f (\cdot)$ of a univariate $X_i$ in the IID setup, and $k (\cdot)$ is the usual convolution kernel of a usual second order symmetric kernel $k (\cdot)$. Assume that $f (\cdot)$ is four times continuously differentiable.

(a) Show that $CV_f (h)$ can be written as

$$CV_f (h) = \int k (u)^2 \frac{du}{nh} + I_n (h) + O_p (n^{-1}),$$

where $I_n (h) = (n(n-1)h^{-1}) \sum_{i=1}^{n} \sum_{j \neq i}^{n} \left[ \Phi (X_i / X_j / h) - 2k ((X_i - X_j) / h) \right].$ (Hint: $\Phi (0) = \int k (u)^2 \frac{du}{nh}$.)

(b) Show that

$$E [I_n (h)] = - \int f (x)^2 \frac{dx}{nh} + \frac{h^4}{4} \left( \int u^2 k (u) \frac{du}{nh} \right)^2 \int [f^{(2)} (x)]^2 \frac{dx}{nh} + o (h^4).$$

(c) Noting that $I_n (h) = E [I_n (h)] +$ smaller order terms, minimizing $CV_f (h)$ is asymptotically equivalent to minimizing

$$I (h) = \frac{\int k (u)^2 \frac{du}{nh}}{4} \left( \int u^2 k (u) \frac{du}{nh} \right)^2 \int [f^{(2)} (x)]^2 \frac{dx}{nh}.$$

Obtain the $h$ that minimizes $I (h)$.

(d) Now, suppose that we don’t use the leave-one-out estimator for the density $f (x)$ in constructing the cross-validated objective function $CV_f (h)$ in (0.1). Then the objective function becomes

$$\bar{CV}_f (h) = \frac{1}{n^2 h} \sum_{i=1}^{n} \sum_{j=1}^{n} k \left( \frac{X_i - X_j}{h} \right) - \frac{2}{n} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k \left( \frac{X_i - X_j}{h} \right).$$

Show that

$$\bar{CV}_f (h) = \frac{\int k (u)^2 \frac{du}{nh} - 2k (0)}{nh} + I_n (h) + O_p (n^{-1}).$$

(e) Assume that $k (\cdot)$ is maximized at 0. Show that $\bar{CV}_f (h) \equiv (nh)^{-1} \left( \int k (u)^2 \frac{du}{nh} - 2k (0) \right) + E [I_n (h)]$ is minimized at $h = 0$, which obviously violates the condition of $nh \rightarrow \infty$ as $n \rightarrow \infty$.

Answer.
(a) Using the definition of $CV_f (h)$ and the fact that $\frac{1}{n^2} = \frac{1}{n(n-1)}$, we have

$$CV_f (h) = \frac{1}{n^2 h} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k \left( \frac{X_i - X_j}{h} \right) - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k_h (X_i - X_j)$$

$$= \frac{1}{n^2 h} \sum_{i=1}^{n} k (0) + \frac{1}{n^2 h} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k \left( \frac{X_i - X_j}{h} \right) - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k_h (X_i - X_j)$$

$$= \frac{k (0)}{nh} + \frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \left[ k \left( \frac{X_i - X_j}{h} \right) - 2k \left( \frac{X_i - X_i}{h} \right) \right] - \frac{1}{n^2 (n-1)} h \sum_{i=1}^{n} \sum_{j \neq i}^{n} k \left( \frac{X_i - X_i}{h} \right)$$

$$= \frac{\int k (u)^2 du}{nh} + I_n (h) + O_p (n^{-1})$$

where $O_p (n^{-1})$ results from the fact that $\frac{1}{n^2 (n-1)h} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k \left( \frac{X_i - X_j}{h} \right) = O_p (n^{-1})$. The last claim follows from Markov inequality because by the nonnegativity of $k (\cdot)$, we have

$$E \left[ \left| \frac{1}{n^2 (n-1)h} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k \left( \frac{X_i - X_j}{h} \right) \right| \right]$$

$$= \frac{1}{n^2 (n-1)h} \sum_{i=1}^{n} \sum_{j \neq i}^{n} E \left[ k \left( \frac{X_i - X_j}{h} \right) \right] = \frac{1}{nh} \sum_{i=1}^{n} \sum_{j \neq i}^{n} E \left[ k \left( \frac{X_i - X_j}{h} \right) \right]$$

$$= \frac{1}{nh} \int \int k \left( \frac{x_1 - x_2}{h} \right) f (x_1) f (x_2) dx_1 dx_2$$

$$= \frac{1}{n} \int \int k (u) f (x_1 + hu) f (x_2) du dx_2 = O \left( \frac{1}{n} \right).$$

(b) Write

$$I_n (h) = \frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \left\{ k \left( \frac{X_i - X_j}{h} \right) - 2k \left( \frac{X_i - X_i}{h} \right) \right\} = A_1 (h) - 2A_2 (h),$$

Noting that $k \left( \frac{X_i - X_j}{h} \right) = \frac{1}{h} \int k \left( \frac{X_i - x}{h} \right) k \left( \frac{X_j - x}{h} \right) dx$, we have

$$E \left[ A_1 (h) \right] = \frac{1}{n(n-1)h^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} E \left\{ \int k \left( \frac{X_i - x}{h} \right) k \left( \frac{X_j - x}{h} \right) dx \right\}$$

$$= \frac{1}{h^2} \int \left\{ E \left[ k \left( \frac{X_i - x}{h} \right) \right] \right\}^2 dx \text{ (by IID)}$$

$$= \int \left[ \frac{1}{h} \int k \left( \frac{z - x}{h} \right) f (z) dz \right]^2 dx = \int \left[ \int k (u) f (x + hu) du \right]^2 dx$$

$$= \int \left\{ \int k (u) \left[ f (x) + f^\prime (x) hu + \frac{1}{2} f^{(2)} (x) h^2 u^2 + \frac{1}{3!} f^{(3)} (x) h^3 u^3 + \frac{1}{4!} f^{(4)} (x) h^4 u^4 + o (h^4) \right] du \right\}^2 dx$$

$$= \int \left\{ f (x) + \frac{\kappa_21}{2} f^{(2)} (x) h^2 + \frac{\kappa_41}{4!} f^{(4)} (x) h^4 + o (h^4) \right\}^2 dx$$

$$= \int \left[ f (x)^2 + \kappa_21 f (x) f^{(2)} (x) h^2 + \frac{\kappa_21^2}{4} \left[ f^{(2)} (x) \right]^2 h^4 + \frac{2\kappa_41}{4!} f (x) f^{(4)} (x) h^4 + o (h^4) \right] dx$$

$$= \int f (x)^2 dx + \kappa_21 \int f (x) f^{(2)} (x) dx h^2 + \left[ \frac{\kappa_21^2}{4} \int \left[ f^{(2)} (x) \right]^2 dx + \frac{2\kappa_41}{4!} \int f (x) f^{(4)} (x) dx \right] h^4 + o (h^4).$$
where $\kappa_{ij} = \int u^i k(u)^j \, du$ for $i, j = 1, 2, 3, 4$. Similarly,

\[
E(A_2) = \frac{1}{n(n-1)}h \sum_{i=1}^{n} \sum_{j \neq i}^{n} E \left[ k \left( \frac{X_i - X_j}{h} \right) \right]
\]

\[
= \frac{1}{h} E \left[ k \left( \frac{X_1 - X_2}{h} \right) \right] \quad \text{(by IID)}
\]

\[
= \frac{1}{h} \int \int k \left( \frac{x_1 - x_2}{h} \right) f(x_1) f(x_2) \, dx_1 dx_2
\]

\[
= \int \int k(u) f(x_2 + hu) f(x_2) \, dudx_2
\]

\[
= \int \int k(u) \left[ f(x) + f'(x) hu + \frac{1}{2} f''(x) h^2 u^2 + \frac{1}{3!} f'''(x) h^3 u^3 + \frac{1}{4!} f^{(4)}(x) h^4 u^4 + o(h^4) \right] f(x) \, dudx
\]

\[
= \int \int k(u) \left[ f(x) + \frac{1}{2} f''(x) h^2 u^2 + \frac{1}{4!} f^{(4)}(x) h^4 u^4 + o(h^4) \right] f(x) \, dudx
\]

\[
= \int f(x)^2 \, dx + \kappa_{21} \int f(x) f^{(2)}(x) \, dxh^2 + \frac{\kappa_{41}}{4!} \int f(x) f^{(4)}(x) \, dxh^4 + o(h^4).
\]

It follows that

\[
E[I_n(h)] = E[A_1(h)] - 2E[A_2(h)]
\]

\[
= -\int f(x)^2 \, dx + \frac{\kappa_{21}^2}{4} \int \left[ f^{(2)}(x) \right]^2 \, dxh^4 + o(h^4)
\]

(c) Using (a) and the fact that $I_n(h) = E[I_n(h)] + \text{smaller order terms}$, we have

\[
CV_f(h) = \frac{\int k(u)^2 \, du}{nh} + I_n(h) + O_p(n^{-1})
\]

\[
= \frac{\int k(u)^2 \, du}{nh} - \int f(x)^2 \, dx + h^4 \int u^2 k(u) \, du \int \left[ f^{(2)}(x) \right]^2 \, dx + o(h^4) + \text{smaller order terms}.
\]

It follows that minimizing $CV_f(h)$ is asymptotically equivalent to minimizing

\[
I(h) = \frac{\int k(u)^2 \, du}{nh} + \frac{h^4}{4} \left( \int u^2 k(u) \, du \right)^2 \int \left[ f^{(2)}(x) \right]^2 \, dx
\]

\[
= \frac{\kappa_{21}}{n} + \frac{h^4 \kappa_{21}^2}{4} \int \left[ f^{(2)}(x) \right]^2 \, dx.
\]

The first order condition (FOC) is

\[
\frac{\partial I(h)}{\partial h} = -\frac{\kappa_{21}}{nh^2} + h^3 \kappa_{21}^2 \int \left[ f^{(2)}(x) \right]^2 \, dx = 0.
\]

Solving for $h$ yields

\[
h = \left\{ \frac{\kappa_{21}^2}{\kappa_{21}^4 \int \left[ f^{(2)}(x) \right]^2 \, dx} \right\}^{1/5} n^{-1/5}.
\]

We can also check for the second order condition (SOC):

\[
\frac{\partial^2 I(h)}{\partial h^2} = \frac{2\kappa_{21}^2}{nh^3} + 3h^2 \kappa_{21}^2 \int \left[ f^{(2)}(x) \right]^2 \, dx > 0.
\]
So the $h = \left\{ \frac{n_{o_2}}{\int f^{(3)}(x)^2 \, dx} \right\}^{1/5} n^{-1/5}$ minimizes $I(h)$ and thus $CV(h)$ asymptotically.

(d) Using $\frac{1}{n} = \frac{1}{n(n-1)} - \frac{1}{n^3(n-1)}$, we have

$$\hat{CV}_f(h) = \frac{1}{n^2 h} \sum_{i=1}^{n} \sum_{j=1}^{n} k\left( \frac{X_i - X_j}{h} \right) - \frac{2}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} k_h(X_i - X_j)$$

$$= \frac{1}{n^2 h} \sum_{i=1}^{n} \sum_{j=1}^{n} k\left( \frac{X_i - X_j}{h} \right) - \left\{ \frac{2}{n (n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k_h(X_i - X_j) + \frac{2}{nh} k(0) - \frac{2}{n^2 (n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k_h(X_i - X_j) \right\}$$

$$= CV_f(h) - \frac{2}{nh} k(0) + o\left( \frac{1}{n} \right)$$

where the term $o\left( \frac{1}{n} \right)$ follows from the fact that $\frac{2}{n^2 (n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k_h(X_i - X_j) = O_p \left( \frac{1}{n^2} \right)$ by Markov inequality. Then it follows from (a) that

$$\hat{CV}_f(h) = CV_f(h) - \frac{2}{nh} k(0) + o\left( \frac{1}{n} \right)$$

$$= \frac{\int k(u)^2 \, du - 2k(0)}{nh} + I_n(h) + o\left( \frac{1}{n} \right).$$

(e) By (b),

$$\hat{CV}_f(h) = \frac{1}{nh} \left( \int k(u)^2 \, du - 2k(0) \right) + E[I_n(h)]$$

$$= \frac{1}{nh} \left( \int k(u)^2 \, du - 2k(0) \right) - \int f(x)^2 \, dx + h^4 \left( \int u^2 k(u) \, du \right)^2 \int \left[ f^{(2)}(x) \right]^2 \, dx + o(h^4).$$

Minimizing $\hat{CV}_f(h)$ is asymptotically equivalent to minimizing

$$\hat{CV}_f^*(h) = \frac{1}{nh} \left( \int k(u)^2 \, du - 2k(0) \right) + h^4 \left( \int u^2 k(u) \, du \right)^2 \int \left[ f^{(2)}(x) \right]^2 \, dx.$$

Nevertheless, noting that $k(u) \leq k(0)$ for all $u \in \mathbb{R}$, we have

$$\int k(u)^2 \, du - 2k(0) \leq k(0) \int k(u) \, du - 2k(0) = k(0) - 2k(0) = -k(0) < 0.$$

That is, the first term of $\hat{CV}_f(h)$ is negative and $\hat{CV}_f^*(h)$ is strictly increasing in $h$. The minimizer can only be obtained by setting $h = 0$. This leads to $\hat{CV}_f^*(0) = -\infty$.

2. (20 points) As described in the text, a kernel density estimator for $f'$, the first derivative of the density $f$, is

$$\hat{f}'(x) = \frac{df}{dx} = \frac{1}{nh^2} \sum_{i=1}^{n} k'\left( \frac{x - X_i}{h} \right),$$

where $X_i$ are IID random variables on $\mathbb{R}$. Assume that: (1) $f$ has three continuous derivatives; (2) $k$ is a second order symmetric kernel with a square integrable derivative; and (3) $h \to 0$, $nh^3 \to \infty$ as $n \to \infty$. 


(a) Show that the bias and variance of \( \hat{f} (x) \) satisfy

\[
\text{Bias}\left( \hat{f} (x) \right) = \frac{h^2}{2} \int u^2 k(u) \, df^{(3)}(x) + o(h^2),
\]

and

\[
\text{Var}\left( \hat{f} (x) \right) = \frac{1}{nh^3} \int [k'(u)]^2 \, df(x) + o \left( \frac{1}{nh^3} \right).
\]

(b) Derive the expressions for the AMISE of \( \hat{f} (x) \) and find the optimal bandwidth \( h_0 \) that minimizes the AMISE.

(c) What is the rate of convergence of the MISE of \( \hat{f} (x) \)?

(d) Derive the asymptotic normality of \( \hat{f} (x) \) by using the Liapounov central limit theorem.

**Answer.**

(a) Noting that the symmetry of \( k \) implies that \( k' \) is anti-symmetric: \( k'(-u) = -k'(u) \) for all \( u \in \mathbb{R} \), we have

\[
E \left[ \hat{f} (x) \right] = \frac{1}{h^2} E \left[ k' \left( \frac{x - X_i}{h} \right) \right] = -\frac{1}{h^2} \int_{-\infty}^{\infty} k' \left( \frac{\bar{x} - x}{h} \right) f(\bar{x}) \, d\bar{x}
\]

\[
= -\frac{1}{h} \int_{-\infty}^{\infty} k'(u) \, f(x + hu) \, du = -\frac{1}{h} \int_{-\infty}^{\infty} f(x + hu) \, dk(u)
\]

\[
= -\frac{1}{h} \left[ f(x + hu) \right]_{u = -\infty}^{\infty} + \frac{1}{h} \int_{-\infty}^{\infty} k(u) \, df(x + hu)
\]

\[
= 0 + \int_{-\infty}^{\infty} k(u) \, f'(x + hu) \, du
\]

\[
= \int_{-\infty}^{\infty} k(u) \left[ f'(x) + f''(x) hu + \frac{1}{2} f^{(3)}(x) hu^2 + o(h^2) \right] \, du
\]

\[
= f'(x) + \frac{h^2}{2} \int_{-\infty}^{\infty} w^2 k(u) \, df^{(3)}(x) + o(h^2).
\]

That is,

\[
\text{Bias}\left( \hat{f} (x) \right) = \frac{h^2}{2} \int u^2 k(u) \, df^{(3)}(x) + o(h^2).
\]

Now, we calculate \( \text{Var}\left( \hat{f} (x) \right) \). Write

\[
\text{Var}\left( \hat{f} (x) \right) = \frac{1}{n^2 h^4} \text{Var} \left( \sum_{i=1}^{n} k' \left( \frac{x - X_i}{h} \right) \right) = \frac{1}{n^2 h^4} \sum_{i=1}^{n} \text{Var} \left( k' \left( \frac{x - X_i}{h} \right) \right)
\]

\[
= \frac{1}{nh^4} \text{Var} \left( k' \left( \frac{x - X_i}{h} \right) \right) = \frac{1}{nh^4} \left( \text{Var} \left( k' \left( \frac{x - X_i}{h} \right) \right) + \text{Var} \left( \frac{h^2}{2} \int_{-\infty}^{\infty} w^2 k(u) \, df^{(3)}(x) + o(h^2) \right) \right)^2.
\]

Noting that

\[
E \left\{ \left[ k' \left( \frac{x - X_i}{h} \right) \right]^2 \right\} = \int \left[ k' \left( \frac{x - \bar{x}}{h} \right) \right]^2 f(\bar{x}) \, d\bar{x} = \int \left[ k' \left( \frac{\bar{x} - \bar{x}}{h} \right) \right]^2 f(\bar{x}) \, d\bar{x}
\]

\[
= h \int |k'(u)|^2 \, f(x + hu) \, du
\]

\[
= h \int |k'(u)|^2 \, df(x) + O(h^3)
\]

5
and
\[ \left\{ E \left[ k' \left( \frac{x - X_i}{h} \right) \right] \right\}^2 = \{ h \left[ f' (x) + O (h^2) \right] \}^2 = h^2 f' (x)^2 + O (h^4), \]
we have
\[
\text{Var} \left( \hat{f}' (x) \right) = \frac{1}{nh^3} \left[ h \int \left[ k' (u) \right]^2 du f (x) + O \left( \frac{1}{nh^2} \right) \right]
\]
\[ = \frac{1}{nh^3} \int \left[ k' (u) \right]^2 du f (x) + o \left( \frac{1}{nh^2} \right). \]

(b) Let \( \kappa_21 = \int u^2 k (u) du \) and \( \kappa = \int \left[ k' (u) \right]^2 du. \) Then
\[
\text{AMISE} \left( \hat{f}' \right) = \frac{h^4}{4 \kappa_{21}} \int \left(f^{(3)} (x) \right)^2 dx + \frac{1}{nh^3} \int \left[ k' (u) \right]^2 du f (x) dx
\]
\[ = \frac{h^4}{4 \kappa_{21}} \int \left[f^{(3)} (x) \right]^2 dx + \frac{1}{nh^3} \kappa. \]

Let
\[
h_0 = \arg \min_h \left\{ \frac{h^4}{4 \kappa_{21}} \int \left[f^{(3)} (x) \right]^2 dx + \frac{1}{nh^3} \kappa \right\}. \]
The solution can be found by solving the first order condition:
\[
h_0^4 \kappa_{21} \int \left[f^{(3)} (x) \right]^2 dx - \frac{3}{nh_0^4} \kappa = 0. \]
This yields
\[
h_0 = \left( \frac{3\kappa}{\int \left[f^{(3)} (x) \right]^2 dx \kappa_{21}} \right)^{1/4} n^{-\frac{1}{4}}. \]

(c) When evaluated at the optimal bandwidth \( h_0, \)
\[
\text{AMISE} \left( \hat{f}' \right) = \frac{1}{4 \kappa_{21}} \int \left[f^{(3)} (x) \right]^2 dx h_0^4 + \frac{1}{nh_0^4} \kappa
\]
\[ = \frac{1}{4 \kappa_{21}} \int \left[f^{(3)} (x) \right]^2 dx \left( \frac{3\kappa}{\int \left[f^{(3)} (x) \right]^2 dx \kappa_{21}} \right)^{\frac{3}{4}} n^{-\frac{3}{4}} + \frac{1}{3 \kappa_{21}} \left( \frac{3\kappa}{\int \left[f^{(3)} (x) \right]^2 dx \kappa_{21}} \right)^{-\frac{3}{4}} n^{-\frac{3}{4}}
\]
\[ = \frac{7}{12} \kappa_{21} (3\kappa)^{4/7} \left( \int \left[f^{(3)} (x) \right]^2 dx \right)^{3/7} n^{-\frac{3}{4}} = O \left( n^{-\frac{3}{4}} \right). \]
It follows that \( \text{MISE} \left( \hat{f}' \right) = \text{AMISE} \left( \hat{f}' \right) + o_p \left( n^{-\frac{3}{4}} \right) = O_p \left( n^{-\frac{3}{4}} \right). \)

(d) Write
\[
\hat{f}' (x) - f' (x) = \left\{ E \left[ \hat{f}' (x) \right] - f' (x) \right\} + \left\{ \hat{f}' (x) - E \left[ \hat{f}' (x) \right] \right\} = B_n + V_n.
\]
The first term contributes to the bias of the estimator and the second term contributes to the variance. By (a),
\[
B_n = E \left[ \hat{f}' (x) \right] - f' (x) = \frac{h^2}{2} \int u^2 k (u) du f^{(3)} (x) + o \left( h^2 \right).
\]
We now demonstrate the asymptotic normality of $\sqrt{nh^3}V_n$.

Define

$$Z_{n,i} = (nh)^{-1/2} \left[ k' \left( \frac{x - X_i}{h} \right) - E k' \left( \frac{x}{h} \right) \right].$$

Then $\sqrt{nh^3}V_n = \sum_{i=1}^n Z_{n,i}$.

$$\sum_{i=1}^n \text{Var} (Z_{n,i}) = \frac{1}{nh} \sum_{i=1}^n \text{Var} \left[ k' \left( \frac{x - X_i}{h} \right) \right] = \frac{1}{h} \text{Var} \left[ k' \left( \frac{x - X_i}{h} \right) \right]$$

$$= \frac{1}{h} E \left[ \left( k' \left( \frac{x - X_i}{h} \right) \right)^2 \right] - \frac{1}{h} \left[ E \left[ k' \left( \frac{x - X_i}{h} \right) \right] \right]^2$$

$$= \int k'(u)^2 f(x + hu) du - \frac{1}{h} \left[ h \left[ f'(x) + O(h^2) \right] \right]^2$$

$$= \int k'(u)^2 du f(x) + o(1).$$

Similarly, for any $\delta > 0$, by the $C_r$ inequality (which says $E|X + Y|^r \leq 2^{r-1}E(|X|^r + |Y|^r)$ for any $r \geq 1$),

$$\sum_{i=1}^n E \left| Z_{n,i} \right|^{2+\delta} = \frac{1}{(nh)^{1+\delta/2}} \sum_{i=1}^n E \left| k' \left( \frac{x - X_i}{h} \right) - E k' \left( \frac{x - X_i}{h} \right) \right|^{2+\delta}$$

$$\leq \frac{2^{1+\delta}}{(nh)^{1+\delta/2}} \sum_{i=1}^n \left\{ E \left[ \left| k' \left( \frac{x - X_i}{h} \right) \right|^{2+\delta} \right] + \left[ E \left[ k' \left( \frac{x - X_i}{h} \right) \right] \right]^{2+\delta} \right\}$$

$$\leq \frac{2^{1+\delta}}{(nh)^{1+\delta/2}} \sum_{i=1}^n \left\{ h \int \left| k'(u) \right|^{2+\delta} f(x + hu) du + \left[ h \int k'(u) f(x + hu) du \right]^{2+\delta} \right\}$$

$$= \frac{2^{1+\delta}}{(nh)^{1+\delta/2}} \int \left| k'(u) \right|^{2+\delta} du f(x) + o \left( \frac{1}{(nh)^{\delta/2}} \right) \to 0$$

where the last convergence result follows from the fact that $nh^3 \to \infty$ (in conjunction with $h \to 0$) as $n \to \infty$ implies that $nh \to \infty$ as $n \to \infty$. So we can apply the Liapounov CLT to conclude

$$\sqrt{nh^3}V_n \overset{d}{\to} N \left( 0, \int k'(u)^2 du f(x) \right).$$

Consequently,

$$\sqrt{nh^3} \left[ \hat{f}'(x) - f'(x) - \frac{h^2}{2} \int u^2 k(u) du f^{(3)}(x) - o(h^2) \right] \overset{d}{\to} N \left( 0, \int k'(u)^2 du f(x) \right).$$

The $o(h^2)$ term can be dropped if we use the bandwidth with the optimal rate: $h \propto n^{-\frac{1}{4}}$.

3. (25 points) Consider the model between a random variable $Y$ and a random variable $X$ :

$$Y = m(X) + U,$$

where $m(\cdot)$ is an unknown smooth function on $\mathbb{R}^1$. Given IID data $\{y_i, x_i\}_{i=1}^n$, please answer the following questions.
(a) Obtain the least squares estimate \( \hat{m}_{ls}(x) \) of \( m(x) = \alpha + \beta x \) and show that \( \hat{m}_{ls}(x) \) can be written in the form \( \sum_{i=1}^{n} w_i y_i \). Write down the explicit formula for \( w_i \), \( \hat{\alpha}_{ls} \), and \( \hat{\beta}_{ls} \), where \( \hat{\alpha}_{ls} \) and \( \hat{\beta}_{ls} \) are the least square estimators for \( \alpha \) and \( \beta \), respectively.

(b) Obtain the local constant (NW) estimator \( \hat{m}_{nw}(x) \) for \( m(x) \). Calculate the leading terms of its asymptotic bias and variance.

(c) Obtain the local linear (LL) estimator \( \hat{m}_{ll}(x) \) for \( m(x) \). Calculate the leading terms of its asymptotic bias and variance.

(d) Compare the asymptotic bias and variance of \( \hat{m}_{nw}(x) \) with that of \( \hat{m}_{ll}(x) \) when the same kernel \( K \) and bandwidth \( h \) are used.

(e) For the local linear estimation in (c), obtain the LL estimator \( \hat{m}_{ll}^{(1)}(x) \) for the first derivative \( m^{(1)}(x) \) of \( m(x) \). Show that \( \hat{m}_{ll}^{(1)}(x) \) and \( \hat{m}_{ll}(x) \) can be written as

\[
\hat{m}_{ll}^{(1)}(x) = \frac{\sum_{i=1}^{n} (y_i - \bar{y}_k) (x_i - \bar{x}_k) K_{i,x}}{\sum_{i=1}^{n} (x_i - \bar{x}_k)^2 K_{i,x}}, \quad \hat{m}_{ll}(x) = \bar{y}_k - (\bar{x}_k - x) \hat{m}_{ll}^{(1)}(x),
\]

where \( \bar{y}_k = \sum_{i=1}^{n} y_i K_{i,x}/\sum_{i=1}^{n} K_{i,x}, \; \bar{x}_k = \sum_{i=1}^{n} x_i K_{i,x}/\sum_{i=1}^{n} K_{i,x}, \) and \( K_{i,x} = K_h(x_i - x) \).

**Answer.**

(a) Let \( (\hat{\alpha}, \hat{\beta}) \) be the OLS estimator of \( (\alpha, \beta) \). Then

\[
\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x},
\]

\[
\hat{\beta} = \frac{\sum_{i=1}^{n} x_i - \bar{x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sum_{i=1}^{n} w_i y_i,
\]

so

\[
\hat{m}_{ls}(x) = \hat{\alpha} + \hat{\beta} x = (\bar{y} - \hat{\beta}\bar{x}) + \hat{\beta} x = \bar{y} + \hat{\beta} (x - \bar{x}) = \frac{1}{n} \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} w_i y_i (x - \bar{x}) = \sum_{i=1}^{n} w_i y_i,
\]

where

\[
w_i = \frac{1}{n} + w_0 (x - \bar{x}) = \frac{1}{n} + \frac{(x - \bar{x}) (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}.
\]

(b) The local constant (NW) estimator \( \hat{m}_{nw}(x) \) for \( m(x) \) is obtained through

\[
\hat{m}_{nw}(x) = \arg \min_{m(x)} \sum_{i=1}^{n} (y_i - m(x))^2 K \left( \frac{x - x_i}{h} \right).
\]

Explicitly,

\[
\hat{m}(x) = \frac{\sum_{i=1}^{n} K_h(x_i - x) y_i}{\sum_{i=1}^{n} K_h(x_i - x)}
\]

The leading asymptotic bias and variance are

\[
\text{Bias}(\hat{m}(x)) \approx h^2 \left( f^{(1)}(x) m^{(1)}(x) + \frac{1}{2} f(x) m^{(2)}(x) \right) \int K(v) v^2 dv
\]

8
the NW kernel estimator, the bias of the local linear estimator is of the same order for both interior and

Solving the above two equations for \( \mu \) and \( \nu \), we have

\[
\begin{align*}
\sum_{i=1}^{n} \left( y_i - m(x) - m^{(1)}(x) (x_i - x) \right) K \left( \frac{x - x_i}{h} \right) &= 0 \\
\sum_{i=1}^{n} (x_i - x) \left( y_i - m(x) - m^{(1)}(x) (x_i - x) \right) K \left( \frac{x - x_i}{h} \right) &= 0.
\end{align*}
\]

The leading asymptotic bias and variance are

\[
\text{Bias}(\hat{m}(x)) \approx \frac{h^2}{2} f(x) m^{(2)}(x) \int K(\cdot) \nu^2 d\nu
\]

and

\[
\text{Var}(\hat{m}(x)) \approx \frac{\sigma^2(x) \int K(\cdot)^2 d\nu}{nhf(x)}
\]

For the detailed calculation, see the lecture note.

(d) We notice that the variances are the same but the biases are different. In particular, the bias of the local linear estimator does not depend on the density of the regressor. As a matter of fact, unlike the NW kernel estimator, the bias of the local linear estimator is of the same order for both interior and boundary points on the support of the regressors.

Given the expressions in part (c), it suffices to show that

\[
\sum_{i=1}^{n} (y_i - \overline{y}_k) (x_i - x) K_{i,x} = \sum_{i=1}^{n} (y_i - \overline{y}_k) (x_i - \overline{x}_k) K_{i,x}
\]

and

\[
\sum_{i=1}^{n} (x_i - \overline{x}_k) (x_i - x) K \left( \frac{x - x_i}{h} \right) = \sum_{i=1}^{n} (x_i - \overline{x}_k)^2 K_{i,x}.
\]
These hold because
\[ \sum_{i=1}^{n} (y_i - \overline{y}_k) (x_i - x) K_{i,x} - \sum_{i=1}^{n} (y_i - \overline{y}_k) (x_i - \overline{x}_k) K_{i,x} = 0 \]

and
\[ \sum_{i=1}^{n} (x_i - \overline{x}_k) (x_i - x) K \left( \frac{x - x_i}{h} \right) - \sum_{i=1}^{n} (x_i - \overline{x}_k)^2 K_{i,x} = 0 \]

**Part II: Computer and Empirical Questions**

4. (10 points) To compare the asymptotic normal approximation to the distribution of the sample mean estimator with the bootstrap approximation, one can use the following algorithm:

Step 1. Draw a random sample of \( n \) observations \( \{X_1, \ldots, X_n\} \) from the \( t(3) \) distribution (student \( t \)-distribution with 3 degrees of freedom). Calculate the sample mean estimator \( \overline{X} \) for the population mean \( \mu \).

Step 2. Draw a bootstrap resample \( \{X^*_1, \ldots, X^*_n\} \) independently from \( \{X_1, \ldots, X_n\} \) with replacement. Calculate the standardized sample mean estimator \( \sqrt{n} \overline{X} \) for the bootstrap sample.

Step 3. Repeat Step 2 \( B \) times, denote the bootstrap standardized sample mean estimators by \( \{\sqrt{n} \overline{X}_b\}_{b=1}^{B} \).

(a) For \( n = 20, 200, 2000 \), and \( B = 500 \). Obtain \( \overline{X} \) and \( \{\sqrt{n} \overline{X}_b\}_{b=1}^{B} \) using the above algorithm. Estimate the density of \( \sqrt{n} \overline{X} \) by the nonparametric kernel method and plot the density function for each sample size. In the same graph, plot the density function of the asymptotic distribution of \( \sqrt{n} \overline{X} \). Compare the four plots you have obtained. Make sure that you use the cross-validated bandwidth in obtaining the kernel density estimates.

(b) Repeat part (a) with \( \sqrt{n} \overline{X} \) replaced by \( \sqrt{n} \left( \overline{X}_b - \overline{X} \right) \) in Step 2 and part (a). Does your finding in part (a) change now? What can you learn from this exercise?  

[Hint: If \( X \) is distributed as \( t(v) \), then \( E(X) = 0 \) and \( \text{Var}(X) = v/(v-2) \) for \( v > 2 \). So the CLT says that \( \sqrt{n} \overline{X} \overset{d}{\rightarrow} N(0, 3) \) in part (a).]

**Answer.** Using the normal kernel and least squares cross-validated bandwidths, we get the following picture. From the top part of the picture, we see that the distance between the bootstrap approximation and asymptotic normal approximation does not necessarily decrease as \( n \) grows, if we don’t recenter the bootstrapped mean at the original sample mean. The bottom part of the picture suggests that the correctly re-centered bootstrap procedure produces better and better approximation to the asymptotic normal distribution as \( n \) grows.

We should keep in mind, however, both the asymptotic normal distribution and the bootstrap distribution serve as approximations to the finite sample distribution of the underlying statistic of interest.
Figure 1: Comparison of asymptotic normal and bootstrap approximations

In many cases, we can prove that the bootstrap approximation outperforms the asymptotic normal approximation and the latter may deteriorate very badly in many applications. This is why we should be in favor of the bootstrap procedure even though it may be time-consuming.

5. (20 points) The objective in this exercise is to estimate the production function for China’s non-governmental businesses for the Year 2003. The data include 2052 valid observations on 7 variables: Output, Capital, Labor, Province, Ownership, Industry, and List, where the first three variables are self-defined, and the other variables are categorical variables (for example, List = 1 if a business is a listed company, 0 otherwise). Please use the dataset (Business03.xls) to answer the following questions.

(a) Run a multiple linear regression (MLR) model by regressing ln(Output) on a constant term, ln(Capital), and ln(Labor):

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u, \]

where \( Y = \ln(\text{Output}), X_1 = \ln(\text{Capital}), \) and \( X_2 = \ln(\text{Labor}) \), and \( u \) is the disturbance term. Report the regression results in the standard format. That is, you need to report the regression model, the \( t \)-values, or the \( p \)-values or the corresponding standard errors for the coefficients in the model, \( R^2 \), \( R^2_F \), and the \( F \) test statistic or its corresponding \( p \)-value.

(b) Run a nonparametric regression model by regressing ln(Output) on ln(Capital) and ln(Labor) by using the local constant procedure

\[ Y = m(X_1, X_2) + u. \]

Denote the regression estimates by \( \tilde{m}_{lc}(x_1, x_2) \). Calculate the \( R^2 \) based upon the formula

\[ R^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y}) \left( \tilde{Y}_i - \bar{Y} \right)^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2 \sum_{i=1}^{n} (\tilde{Y}_i - \bar{Y})^2} \]
which is the square of the sample correlation between $Y_i$ and $\hat{Y}_i$, where $\hat{Y}_i = \hat{m}_{lc}(X_{i1}, X_{i2})$ is the in-sample predicted value for $Y_i$, $\bar{Y} = n^{-1} \sum_{i=1}^{n} Y_i$, and $\bar{Y} = n^{-1} \sum_{i=1}^{n} \hat{Y}_i$. Plot $\hat{Y}$ against $X_1$ and $X_2$ in a three-dimensional diagram. Does the diagram lend any support to the MLR model in part (a)?

[Hint. To get the $R^2$ in the nonparametric framework, you need to estimate the regression function $m(x_1, x_2)$ at each data point.]

(c) Repeat part (b) by using the local linear procedure.

(d) Test the correct specification of the model in part (a). That is, test the null hypothesis

$$H_0 : P \{ E(Y|X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \} = 1$$

versus

$$H_1 : P \{ E(Y|X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \} < 1.$$

Answer.

(a) The regression results are reported below

$$\ln (\text{Output}) = 4.215 + 0.5578 \log (\text{Capital}) + 0.0830 \log (\text{Labor})$$

$$\begin{array}{cc}
  (0.1207) & (0.01450) & (0.0134) \\
\end{array}$$

$$R^2 = 0.5710, \ T^2 = 0.5706, \ F = 1363.82 \ (p\text{-value} = 0.0000).$$

(b) If we use the normal kernel and the bandwidth chosen by the Silverman’s rule of thumb, $R^2 = 0.6828$. The three-dimensional plot of $\hat{Y}$ against $X_1$ and $X_2$ is as follows.

According to the plot, we can see some curve relationship between the output and inputs. So we cannot say that the plot lends strong support (if any) to the relationship in (a).

(c) If we use the normal kernel and the bandwidth chosen by the Silverman’s rule of thumb, $R^2 = 0.6891$. The three-dimensional plot of $\hat{Y}$ against $X_1$ and $X_2$ is as follows.

The plot is quite similar to that in part (b), as expected. So it does not lend much support to the relationship in (a).

(d) We can construct the following test statistic

$$T_n = \frac{n (h_1h_2)^{1/2} I_n}{\hat{\sigma}}$$

where

$$I_n = \frac{1}{n} \sum_{i=1}^{n} \hat{u}_i \left\{ \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} \hat{u}_j K_{h,j,i} \right\} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \hat{u}_i \hat{u}_j K_{h,i,j},$$

$$\hat{\sigma}^2 = 2n^{-2} h_1 h_2 \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{u}_i^2 \hat{u}_j^2 K_{h,i,j}^2$$

and the definitions for other notation is same as those in the lecture note. Here we choose the standard normal kernel and use bandwidth based on the Silverman’s rule of thumb.

Given the data, $T_n = 61.1150$. Run the test with $B = 500$ bootstrap resamples, we obtain $p\text{-value} = 0.00$. Then we can reject the null at the level of 1%. 

12
Figure 2:
Production function (local linear)

Figure 3: