Homework 14

Liu Huihang
SA18017026
QQ: 184050473
MAIL: huihang@mail.ustc.edu.cn

Problem 1 (HW 14.1) *Let $\{X_i, Y_i\}$ be bivariate random sample, and Y_i are generated from*

$$Y_i = m(X_i) + u_i$$

where $m(\cdot)$ is a unknown univariate smooth function, μ_i satisfy $Eu_i|X_i = 0, Var(u_i|X_i) = \sigma^2(X_i), a.s.$.

(1) Get the local linear estimation of $m(x)$ and calculate the main terms of asymptotic bias and variance.

(2) Get $\hat{m}_{ll}^{(1)}(x)$, the local linear estimation of the first derivative of $m(x)$, and prove that $\hat{m}_{ll}^{(1)}(x)$ and $\hat{m}_{ll}(x)$ can be rewritten as

$$\hat{m}_{ll}^{(1)}(x) = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y}_k)(X_i - \bar{X}_k)K_{i,x}}{\sum_{i=1}^{n}(X_i - \bar{X}_k)^2 K_{i,x}}$$

and

$$\hat{m}_{ll}(x) = \bar{Y}_k - (\bar{X}_k - x)\hat{m}_{ll}^{(1)}(x)$$

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 $K_{i,x} = K_h(x - X_i)$.

Solution

(1) Linear smoother is given by

$$\hat{m}(x) = (X_x^T W_x X_x)^{-1} X_x^T W_x Y$$

where

$$X_x = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_1 - x & X_2 - x & \cdots & X_n - x \end{pmatrix}^T$$

and W_x be the $n \times n$ diagonal matrix whose (i,i) component is $K\left(\frac{X_i - x}{h}\right)$.

Because $\hat{\beta}(x) = m_{ll}^{l} l$ and we can rewrite it from matrix form to general form.

$$\begin{aligned} \hat{m}_u(x) &= \frac{\left(\sum_i K_{i,x}(X_i - x)^2, -\sum_i K_{i,x}(X_i - x)\right)}{\sum_i K_{i,x} \sum_i K_{i,x}(X_i - x)^2 - \left[\sum_i K_{i,x}(X_i - x)\right]^2} \begin{pmatrix} \sum_i K_{i,x} Y_i \\ \sum_i K_{i,x}(X_i - x) Y_i \end{pmatrix} \\ &= \frac{\sum_i K_{i,x}(X_i - x)^2 \sum_i K_{i,x} Y_i - \sum_i K_{i,x}(X_i - x) \sum_i K_{i,x}(X_i - x) Y_i}{\sum_i K_{i,x} \sum_i K_{i,x}(X_i - x)^2 - \left[\sum_i K_{i,x}(X_i - x)\right]^2} \end{aligned} \tag{1}$$

By theorem in the slide, its asymptotic bias and variance are

$$\begin{aligned} \text{bias}\left(\hat{m}_{ll}(x)\right) &= \tfrac{\kappa_{21}}{2} h^2 m''(x) \\ \text{Var}\left(\hat{m}_{ll}(x)\right) &= \tfrac{\kappa_{02}\sigma^2(x)}{nhf(x)} \end{aligned} \tag{2}$$

where $f(\cdot)$ is the density of X.

(2) From the above, we have

$$
\begin{aligned}
\hat{m}_{ll}^{(1)}(x) &= \frac{\left(-\sum_i K_{i,x}\left(X_i - x\right), \sum_i K_{i,x}\right)}{\sum_i K_{i,x} \sum_i K_{i,x}\left(X_i - x\right)^2 - \left[\sum_i K_{i,x}\left(X_i - x\right)\right]^2} \left(\sum_i K_{i,x}\left(X_i - x\right) Y_i\right) \\
&= \frac{-\sum_i K_{i,x}\left(X_i - x\right) \bar{Y}_k + \sum_i K_{i,x}\left(X_i - x\right) Y_i}{\sum_i K_{i,x}\left(X_i - x\right)^2 - \left(\bar{X}_k - x\right) \sum_i K_{i,x}\left(X_i - x\right)} \\
&= \frac{\sum_i K_{i,x}\left(X_i - x\right)\left(Y_i - \bar{Y}_k\right)}{\sum_i K_{i,x}\left(X_i - x\right)\left(X_i - \bar{X}_k\right)}
\end{aligned}
\tag{3}
$$

We can get

$$
\hat{m}_{ll}(x) = \bar{Y}_k - \left(\bar{X}_k - x\right) \hat{m}_{ll}^{(1)}(x) \tag{4}
$$

■

Problem 2 (HW 14.2) *Let's consider the following generalized linear regression model*

$$
Y|X = x \sim Exp(\lambda(x)), \lambda(x) = e^{\beta_0 + \beta_1 x}.
$$

Using local likelihood method, write an estimation function whose variables include the sample X, Y and the point x where the estimate is needed, the bandwidth h and the kernel function K.

Simulation: Generate a set of data, uses your function to estimate and uses cross-validation for optimal bandwidth selection.