Homework 10

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

Problem 1 (HW 10.2) *Suppose $X_1, \ldots. X_n$ i.i.d obey the unary density $f(x)$. The random variable U obeys a uniform distribution over $\{1, 2, ..., n\}$, let $Y = X_U + hZ$, where Z has the density $p(x)$ And independent of $X_1, \ldots. X_n$ and U.*

(1) Prove that under the given conditions of $X_1, \ldots, X_n$, Y has a density $\hat{f}(\cdot)$, which is a kernel estimate based on the samples $X_1, \ldots, X_n$ and the kernel function $p(\cdot)$ and the bandwidth h.

(2) Find the variance of Y given the conditions of $X_1, \ldots, X_n$. How does it compare with the variance of the samples based on $X_1, \ldots, X_n$?

Solution

(1) Firstly, let's focuse on Y's density function $\hat{f}(\cdot)$.

$$\begin{aligned} P(Y < x) &= P(X_U + hZ < y) \\ &= E\left[P(X_U + hZ < x | U = i)\right] \\ &= E\left[P(Z < \frac{x - X_i}{h})\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\frac{x-X_i}{h}} p(t)dt \end{aligned}$$

This leads to

$$\hat{f}(\cdot) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}p\left(\frac{x - X_i}{h}\right)$$

Actually, this is the KDE on the samples $X_1, \ldots, X_n$ and the kernel function $p(\cdot)$ and the bandwidth h.

(2) If we assume $p(\cdot)$ astisfy $\int tp(t)dt = 0$. Then the variance of Y is

$$
\begin{aligned}
Var(Y) &= \int x^2 \hat{f}(x)dx - \left(\int x\hat{f}(x)dx\right)^2 \\
&= \int x^2 \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}p\left(\frac{x-X_i}{h}\right)dx - \left(\int x\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}p\left(\frac{x-X_i}{h}\right)dx\right)^2 \\
&= \frac{1}{nh}\sum_{i=1}^{n}\int x^2 p\left(\frac{x-X_i}{h}\right)dx - \frac{1}{n^2h^2}\left(\sum_{i=1}^{n}\int xp\left(\frac{x-X_i}{h}\right)dx\right)^2 \\
&= \frac{1}{nh}\sum_{i=1}^{n}\int h(ht+X_i)^2 p(t)dt - \frac{1}{n^2h^2}\left(\sum_{i=1}^{n}\int h(ht+X_i)p(t)dt\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[h^2EZ^2 + 2hX_iEZ + X_i^2\right] - \frac{1}{n^2}\left(\sum_{i=1}^{n} hEZ + X_i\right)^2 \\
&= h^2EZ^2 + 2h\bar{X}EZ + \overline{X^2} - \left(\bar{X}^2 + h^2(EZ)^2 + 2hEZ\bar{X}\right) \\
&= \overline{X^2} - \bar{X}^2 + h^2EZ^2
\end{aligned}
$$

While the sample variance of X is

$$
Var_S(X) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \overline{X^2} - \bar{X}^2
$$

$Var(Y)$ is a little biger than $Var_S(X)$ about $h^2EZ^2 \to 0$ if $h \to 0$. ■