Problem Set: Non Parametric Statistics: Group 18

Cesar Conejo Villalobos, Xavier Bryant

3/30/2021

Question 1

Question 2

Compare the MISE and AMISE criteria in three densities in nor1mix of your choice.

1. Code (2.33) and the AMISE expression for the normal kernel, and compare the two error curves.
2. Compare them for $n = 100, 200, 500$, adding a vertical line to represent the h_{MISE} and h_{AMISE} bandwidths. Describe in detail the results and the major takeaways.

For this exercise wwe require some mathematical ideas that we will developed briefly.

We start with the KDE estimator $\hat{f}(x;h) = \sum_{i=1}^{n} K_h(x - X_i)$. The expectation and variance for this estimator are given bi the following expressions:

(1) $E[\hat{f}(x;h)] = (K_h * f)(x)$.

(2) $\text{Var}[\hat{f}(x;h)] = \frac{1}{n}((K_h^2 * f)(x) - (K_h * f)^2(x))$

Then, we develop some assymptotic expression for (1) and (2):

(3) $E[\hat{f}(x;h)] - f(x) = Bias[\hat{f}(x;h)] = \frac{1}{2}\mu_2(K)f^{''}(x)h^2 + o(h^2)$

(4) $\text{Var}[\hat{f}(x;h)] = \frac{R(K)}{nh}f(x) + o((nh)^{-1})$

Then, from whe equations (3) and (4) we obtain the following expression for the MSE:

(5) $\text{MSE}[\hat{f}(x;h)] = \frac{\mu_2^2(K)}{4}(f^{''}(x))^2 h^4 + \frac{R(K)}{nh}f(x) + o(h^4 + (nh)^{-1})$

In equations (3), (4) and (5) we have:

(6) $\mu_2(K) := \int z^2 K(z)dz$

(7) $R(K) := \int (K(x))^2 dx$

Then, we use $\text{MISE}[\hat{f}(\cdot;h)]$ as global error criteria for measuring the performance of $\hat{f}$ in relation to the target density f.

(8) $\text{MISE}[\hat{f}(\cdot;h)] = \int MSE[\hat{f}(x;h)]$

Then, we obtain the following asymptotic expansion for the MISE:

(9) $\text{MISE}[\hat{f}(\cdot;h)] = \frac{1}{4}\mu_2^2(K)R(f^{''})h^4 + \frac{R(K)}{nh} + o(h^4 + (nh)^{-1})$

We define the dominating part of equation (9) as $AMISE[\hat{f}(\cdot;h)]$ defined as:

(10) $\text{AMISE}[\hat{f}(\cdot;h)] = \frac{1}{4}\mu_2^2(K)R(f^{''})h^4 + \frac{R(K)}{nh}$

with the expression $R(f^{''})$ given by:

(11) $R(f^{''}) = \int (f^{''}(x))^2 dx$

Finally, the bandwidth the minimizes the AMISE is:

(12) $h_{AMISE} = \left[\frac{R(K)}{\mu_2^2(K)R(f^{''})n}^{1/5} \right]$

Now, we consider our particular case of study. In this case, we *reduce* our analysis considering:

a) A normal kernel $K_h(\cdot)$ with distribution $\mathcal{N}(0,1)$

b) The density function f is based on the family of normal r-mixtures:

(13) $f(x;\mu,\sigma,\boldsymbol{w}) = \sum_{j=1}^{r} w_j \phi_{\sigma_j}(x-\mu_j)$

where $w_j \geq 0$, $j = 1,...,r$ and $\sum_{j=1}^{r} w_j = 1$.

With this two expression, we can obtain a specific value for the AMISE in equation (10). With assumption a), we obating the following expressions for equations (6) and (7)

(6.1) $\mu_2(K) = 1$

(7.1) $R(K) = \frac{1}{2\sqrt{\pi}}$

Expression for equation (11) can be obtained from the following adaption of expression given in *Theorem 4.1* given by *Marron and Wand (1992)*

(11.1) $R(f^{''}) = \int (f^{''}(x))^2 dx =$

With this expression, we obtain the reduced form of the AMISE:

(10.1) $\text{AMISE}[\hat{f}(\cdot;h)] = \frac{1}{4}R(f^{''})h^4 + \frac{1}{2nh\sqrt{\pi}}$

With optimal bandwidth h_{AMISE}

(12.1) $h_{AMISE} = \left[\frac{(2\sqrt{\pi})^{-1}}{R(f^{''})n}^{1/5} \right]$

Finally, under this assumptions, we obtain a explicit and exact MISE expression of equation (8):

(14) $\text{MISE}_r[\hat{f}(\cdot;h)] = (2\sqrt{\pi}nh)^{-1} + \boldsymbol{w}^{'}\{(1-n^{-1})\Omega_2 - \Omega_1 + \Omega_0\}\boldsymbol{w}$

with $(\Omega_a)_{i,j} = \phi_{(ah^2+\sigma_i^2+\sigma_j^2)^{1/2}}(\mu_i - \mu_j)$ for $i,j = 1,...,r$

Finally, we can proceed with a numeric approach over equation (14) and obtain:

(15) $\text{arg min}_{h>0}\text{MISE}[\hat{f}(\cdot;h)]$

We start with a easy example using

Question 3