Question 1

No code is required for this question.

→ KDE

Consider an i.i.d. sample $\{X_1, \ldots, X_N\}$ and a non-parametric estimate $\hat{f}$ of its probability density function f defined for any $u \in \mathbb{R}$ and some real constant $h > 0$ by

$$\hat{f}(u) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{X_i - u}{h}\right)$$

(a) What is the standard deviation of function $K(u)$? (No derivations are required to answer this question.)

(b) What is the standard deviation of function $K_h(u) = K(u/h)/h$? (No derivations are required to answer this question.)

(c) Can $K(u) = \exp(-\frac{u^2}{2})$ be used to compute this estimate? Why?

(d) In order to ensure a finite-sample estimate $\hat{f}$ of f with as small a bias as possible, and using the unbiased sample variance estimate $\hat{\sigma}^2$ of $Var(X)$, indicate which of the following values of h should be used and why:

$$h_1 = 1.06\ \hat{\sigma}\ N^{-\frac{1}{5}}$$
$$h_2 = 2.34\ \hat{\sigma}\ N^{-\frac{1}{5}}$$

- the std deviation depends on the kernel function.
  σ_K for K

Q1)

- For (b) why is it $\sigma \cdot h$ & not $\frac{\sigma}{h}$

- d) why does smaller h reduce bias.

- (c) what makes a valid pdf

Q2)

Q3) why use quantiles for C.I

Avoid assumption of a theoretical distⁿ

Bootstrap C.I doesn't assume normality