

### Tutorial 3

1. Suppose that  $X$  follows a uniform distribution, then for the inner points, the local linear kernel estimator and local constant (NW) estimator have the same asymptotic MSE.
2. For model  $Y = \sin(2\pi X) + \varepsilon$  where  $X \sim \text{uniform}(0, 1)$  and  $\varepsilon$  is independent of  $X$ . If we apply local linear kernel estimation to estimate the curve, which point are expected to have the biggest absolute bias?
3. The uniform kernel function is defined as

$$K(v) = \begin{cases} 1, & \text{if } |v| \leq 1, \\ 0, & \text{if } |v| > 1. \end{cases}$$

Suppose we have the situation of equal spaced design  $(x_i, y_i)$  with  $x_i = \frac{i}{n}$ , which satisfies

$$y_i = m(x_i) + \varepsilon_i,$$

where  $\varepsilon_i, i = 1, \dots, n$  are IID  $N(0, 1)$ . Using NW kernel smoothing to estimate  $m(x), 0 < x < 1$  with  $h = n^{-0.2}$ , what is the bias and variance of the estimator? what are their limits when  $n \rightarrow \infty$ ?

4. Consider samples  $(x_i, y_i), i = 1, 2, \dots, n$ . when a single observation is distorted by a very large value, that is  $(x_k, y_k) \rightarrow (x_k, y_k + c)$  for a very large  $c$  for a fixed  $k$ . how does the estimated curve change under such a distortion? If you fit a linear regression model to the data

$$y_i = a + bx_i + \varepsilon_i$$

how does the estimated linear regression model change with such a distortion?

5. Another application of the confidence band is to check whether a pre-specified functional shape is acceptable. If the specified function is within the band, we accept the functional form. Otherwise, we reject the functional form. Based on this idea, check whether the relation between  $Y$  and  $X$  in [data 1](#) and [data 2](#) are linear regression

models? [Hint: fit a linear regression to the data. check whether the fitted regression is in the 95% pointwise confidence bands]

6. Consider samples  $(x_i, y_i), i = 1, 2, \dots, n$  and linear regression model

$$y_i = a + bx_i + \varepsilon_i$$

where  $\varepsilon_i$  are IID  $N(0, \sigma^2)$ . Suppose the least squares estimation is used. For any new point  $x$ , construct a 95% confidence interval for the regression function  $m(x) = a + bx$ .