

# MATH-517 Assignment 3

Dario Liuzzo

## 1 Theoretical exercise - Local Linear Regression as a Linear Smoother

### 1.1 Weight Expression Derivation

Consider the local linear regression estimator at a point  $x$ , defined by minimizing

$$\sum_{i=1}^n [Y_i - \beta_0 - \beta_1(X_i - x)]^2 K\left(\frac{X_i - x}{h}\right)$$

with respect to  $\beta_0, \beta_1$ , giving  $\hat{m}(x) = \hat{\beta}_0(x)$ .

Let  $K_i = K\left(\frac{X_i - x}{h}\right)$ . The weighted least squares problem can be written as

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^{p+1}} (Y - X\beta)^\top W (Y - X\beta)$$

where

$$X = \begin{bmatrix} 1 & X_1 - x \\ \vdots & \vdots \\ 1 & X_n - x \end{bmatrix}, \quad W = \text{diag}(K_1, \dots, K_n), \quad Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

The solution to this problem is

$$\hat{\beta} = (X^\top W X)^{-1} X^\top W Y$$

We compute

$$X^\top W X = \begin{bmatrix} \sum_{i=1}^n K_i & \sum_{i=1}^n K_i (X_i - x) \\ \sum_{i=1}^n K_i (X_i - x) & \sum_{i=1}^n K_i (X_i - x)^2 \end{bmatrix} = nh \begin{bmatrix} S_{n,0}(x) & S_{n,1}(x) \\ S_{n,1}(x) & S_{n,2}(x) \end{bmatrix},$$

where  $S_{n,k}(x) = \frac{1}{nh} \sum_{i=1}^n (X_i - x)^k K_i$ .

Also,

$$X^\top W Y = \begin{bmatrix} \sum_{i=1}^n K_i Y_i \\ \sum_{i=1}^n K_i (X_i - x) Y_i \end{bmatrix}.$$

The inverse of  $X^T W X$  is

$$\frac{1}{nh(S_{n,0}(x)S_{n,2}(x) - S_{n,1}^2(x))} \begin{bmatrix} S_{n,2}(x) & -S_{n,1}(x) \\ -S_{n,1}(x) & S_{n,0}(x) \end{bmatrix}.$$

Thus,

$$\hat{\beta}_0 = \frac{S_{n,2}(x) \sum_{i=1}^n K_i Y_i - S_{n,1}(x) \sum_{i=1}^n K_i (X_i - x) Y_i}{nh(S_{n,0}(x)S_{n,2}(x) - S_{n,1}^2(x))}.$$

Rewriting,

$$\hat{m}(x) = \hat{\beta}_0 = \sum_{i=1}^n Y_i \cdot \frac{K_i S_{n,2} - K_i (X_i - x) S_{n,1}}{nh(S_{n,0} S_{n,2} - S_{n,1}^2)}.$$

So the weight is

$$w_{ni}(x) = \frac{1}{nh} K_i \cdot \frac{S_{n,2}(x) - (X_i - x) S_{n,1}(x)}{S_{n,0}(x) S_{n,2}(x) - S_{n,1}(x)^2},$$

## 1.2 Proof that $\sum_{i=1}^n w_{ni}(x) = 1$

We have

$$\sum_{i=1}^n w_{ni}(x) = \frac{\sum_{i=1}^n K_i S_{n,2} - \sum_{i=1}^n K_i (X_i - x) S_{n,1}}{nh(S_{n,0} S_{n,2} - S_{n,1}^2)}.$$

Note that:

- $\sum_{i=1}^n K_i = nh S_{n,0}$ , so first term =  $nh S_{n,0} S_{n,2}$
- $\sum_{i=1}^n K_i (X_i - x) = nh S_{n,1}$ , so second term =  $nh S_{n,1}^2$

Thus numerator =  $nh(S_{n,0} S_{n,2} - S_{n,1}^2)$ , hence the ratio equals 1.

Therefore,  $\sum_{i=1}^n w_{ni}(x) = 1$ .

## 2 Practical Exercise - Global Bandwidth Selection

The goal of this simulation study is to analyze the impact of sample size  $n$ , block size  $N$ , and the Beta distribution parameters  $(\alpha, \beta)$  on the optimal bandwidth  $h_{\text{AMISE}}$  for local linear regression. We simulate data according to the model

$$Y_i = m(X_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2),$$

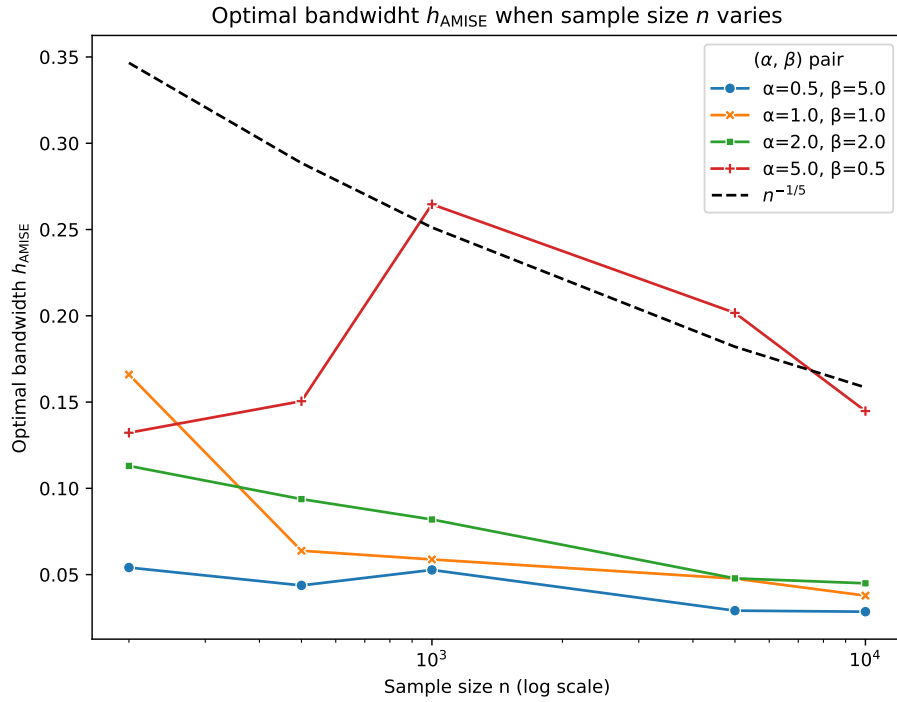
with regression function

$$m(x) = \sin \left( \left( \frac{x}{3} + 0.1 \right)^{-1} \right)$$

and covariates  $X_i \sim \text{Beta}(\alpha, \beta)$ . The variance  $\sigma^2$  is fixed, and the unknown quantities  $\sigma^2$  and  $\theta_{22}$  are estimated via local polynomial fitting in  $N$  blocks, as described in the assignment.

The analysis evaluates:

- The influence of sample size  $n$  on  $h_{\text{AMISE}}$ .
- The effect of block number  $N$  on  $\hat{\sigma}^2$  and  $\hat{\theta}_{22}$ , and consequently  $h_{\text{AMISE}}$ .
- How the Beta distribution parameters  $(\alpha, \beta)$  shape the optimal bandwidth through their effect on the distribution of  $X$ .



The figure illustrates the impact of the sample size  $n$  on the magnitude of the optimal bandwidth  $h_{\text{AMISE}}$  for various configurations of the Beta distribution parameters  $(\alpha, \beta)$ . A general decreasing trend of  $h_{\text{AMISE}}$  with increasing  $n$  is observed for symmetric Beta distributions, such as  $(\alpha, \beta) = (1, 1)$  or  $(2, 2)$ . This behavior is consistent with the intuition that, as more data become available, finer local structures of the regression function  $m(x)$  can be estimated accurately, allowing for a smaller bandwidth to balance bias and variance optimally.

However, for asymmetric distributions, particularly when the Beta distribution is strongly skewed to the right (e.g.,  $(\alpha, \beta) = (5, 0.5)$ ), the pattern of  $h_{\text{AMISE}}$  as a function of  $n$  can deviate from monotonicity. In these cases, the uneven allocation of observations across the support of  $X$  results in blocks containing

highly concentrated data regions near the mode, while other blocks may contain sparse data. This imbalance affects the stability of the blockwise polynomial estimation used to approximate the second derivative of  $m(x)$  and the local variance  $\sigma^2$ , leading to fluctuations in the computed  $h_{\text{AMISE}}$ . Consequently, the bandwidth does not decrease uniformly with increasing sample size, reflecting the interaction between sample density, distribution asymmetry, and the block-based estimation procedure.

In summary, the figure demonstrates that sample size strongly influences the magnitude of the optimal bandwidth, with smaller bandwidths generally supported by larger datasets, but the precise behavior depends on the shape of the covariate distribution. The figure was obtained by selecting the optimal number of blocks  $N$  minimizing Mallows's  $C_p$ ,

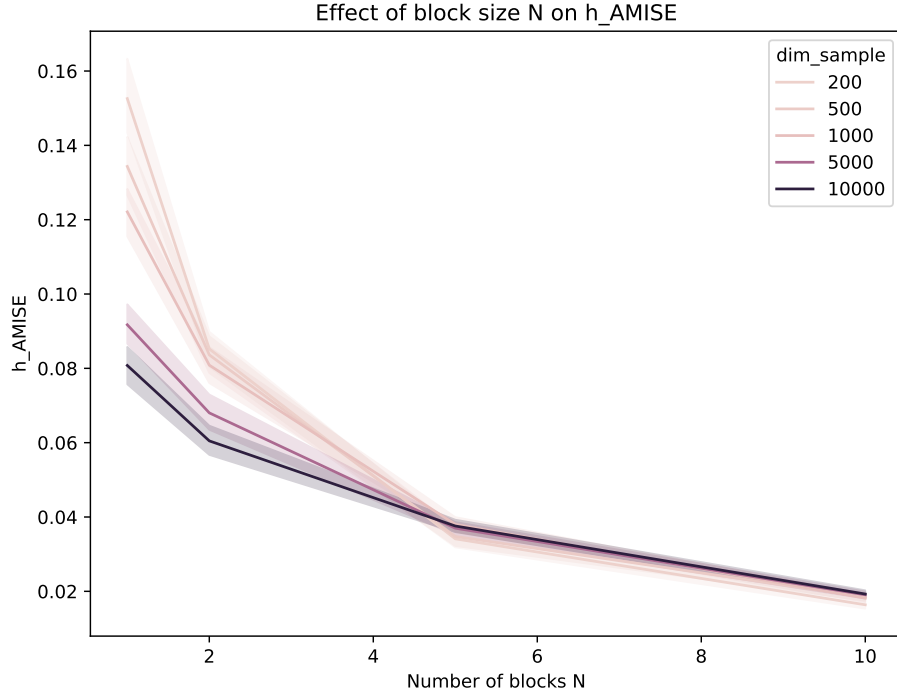
$$C_p(N) = \frac{\text{RSS}(N)}{\text{RSS}(N_{\max})/(n - 5N_{\max})} - (n - 10N),$$

where

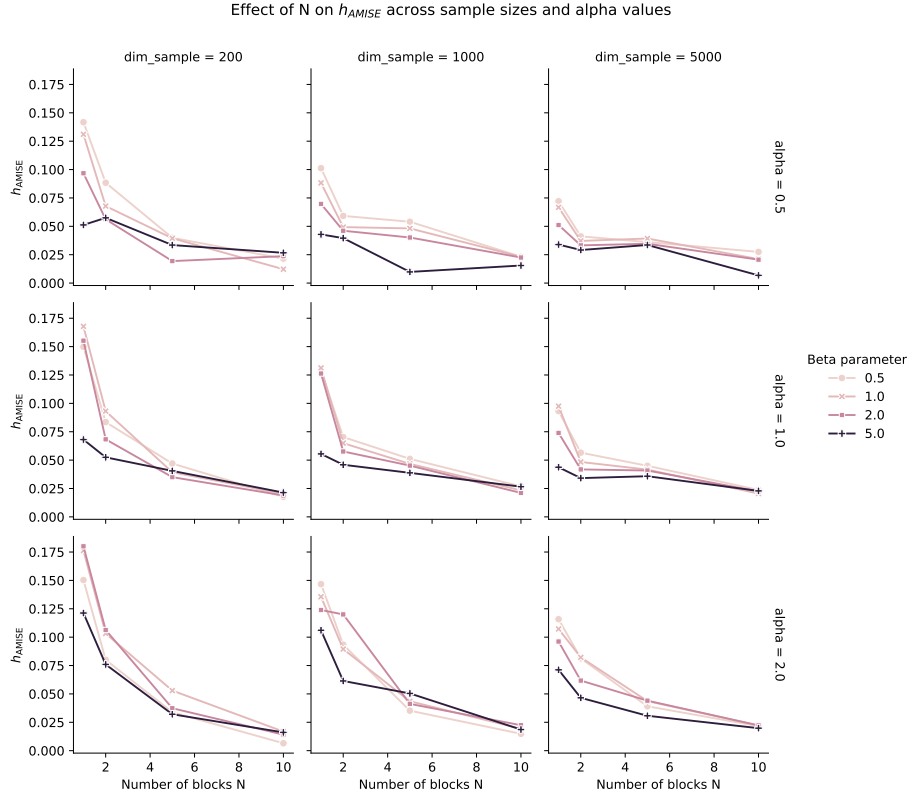
$$\text{RSS}(N) = \sum_{i=1}^n \sum_{j=1}^N \{Y_i - \hat{m}_j(X_i)\}^2 \mathbb{1}_{X_i \in x_j},$$

and

$$N_{\max} = \max \left\{ \min \left( \left\lfloor \frac{n}{20} \right\rfloor, 5 \right), 1 \right\}.$$

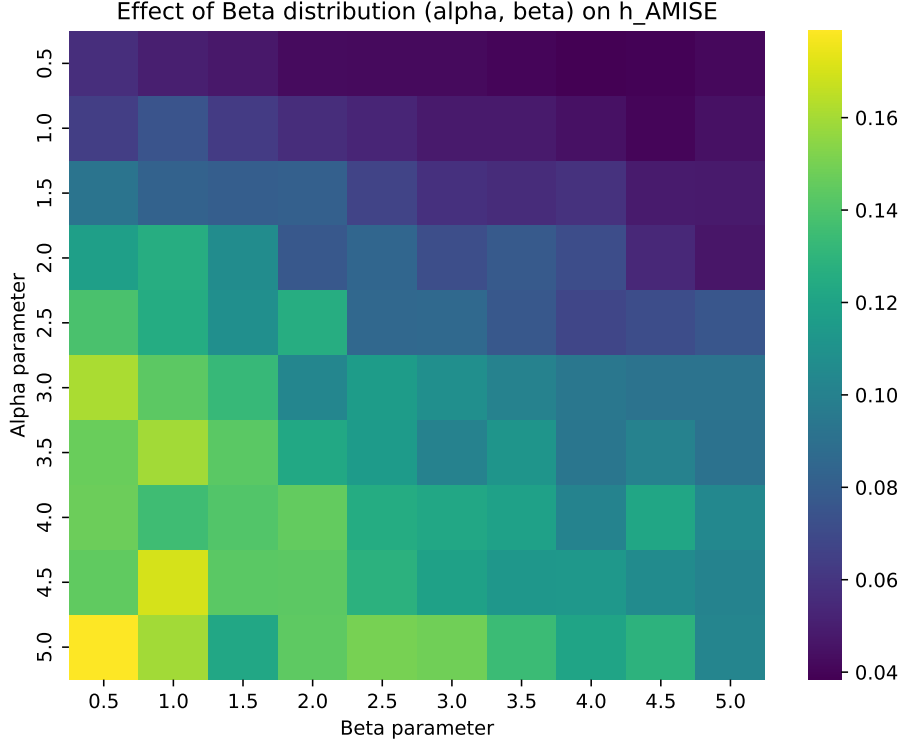


The figure illustrates the influence of the number of blocks  $N$  on the optimal bandwidth  $h_{\text{AMISE}}$ , aggregated across various Beta distribution parameter configurations. Increasing  $N$  refines the local estimation of  $\sigma^2$  and  $m''(x)$ , which reduces the bias in the plug-in estimates and can lead to smaller values of the estimated optimal bandwidth. At the same time, however, each block contains fewer observations as  $N$  grows, which increases the variance of the blockwise estimates. The observed decreasing trend of  $h_{\text{AMISE}}$  with  $N$  in this simulation suggests that the bias reduction effect dominates in the explored regime. The conclusion is that the number of blocks  $N$  should depend on the sample size  $n$ . With too few blocks, the blockwise polynomial fits suffer from bias because they average over wide regions, whereas with too many blocks, the variance of the estimates increases due to the small number of observations per block. The optimal choice of  $N$  therefore balances this bias–variance tradeoff and increases with  $n$ , since larger samples can sustain a finer partition of the covariate space. This motivates the use of data-driven selection rules such as Mallow’s  $C_p$ , which adapt  $N$  to the available sample size.



When differentiating by both  $\alpha$  and  $\beta$ , the same general trend is observed: the optimal bandwidth  $h_{\text{AMISE}}$  decreases as the number of blocks  $N$  increases. This pattern holds consistently across different sample sizes and values of  $\alpha$ , in line

with the behavior described previously.



The heatmap illustrates the dependence of the optimal bandwidth  $h_{\text{AMISE}}$  on the parameters of the Beta distribution. A clear pattern emerges: distributions that are right-skewed (high  $\alpha$ , low  $\beta$ ) correspond to larger values of  $h$ , whereas left-skewed distributions (low  $\alpha$ , high  $\beta$ ) correspond to smaller  $h$ . This behavior can be explained by the interaction between the shape of the covariate distribution and the regression function  $m(x) = \sin((x/3 + 0.1)^{-1})$ , which is very steep near  $x = 0$  and smoother elsewhere.

For right-skewed distributions, the majority of observations are located away from the steep region near  $x = 0$ . Consequently, when estimating  $m(x)$  close to 0, very few points are available locally, and the kernel must span a wider range to gather enough neighbors for a stable estimate. This necessity results in a larger optimal bandwidth. Conversely, for left-skewed distributions, most points lie near  $x = 0$ , directly in the steep region of  $m(x)$ . Even a relatively small bandwidth suffices to include a sufficient number of observations for accurate local linear estimation, resulting in a smaller optimal  $h$ .

To illustrate the performance of the local linear estimator with the optimal bandwidth  $h_{\text{AMISE}}$ , we simulated a sample of size  $n = 500$  from the model  $Y_i = m(X_i) + \epsilon_i$ , where  $X_i \sim \text{Beta}(\alpha, \beta)$  with  $(\alpha, \beta) = (2, 2)$  and  $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

with  $\sigma^2 = 0.3$ . The goal was to use the blockwise polynomial approximation to select  $h_{\text{AMISE}}$  and then construct the local linear estimator  $\hat{m}(x)$  for  $m(x) = \sin((x/3 + 0.1)^{-1})$ . This allows us to assess how well the nonparametric method captures the true regression function, particularly in regions with steep curvature and varying data density.

The figure shows the results of this procedure. The scatter points represent the simulated observations. The red line corresponds to the true regression function  $m(x)$ , while the green line corresponds to the local linear estimate  $\hat{m}(x)$  obtained using the computed optimal bandwidth.

From the visualization, we observe that the local linear estimator accurately captures the overall shape of  $m(x)$  across the support of  $X$ . In areas with a higher concentration of data points, the estimator follows the true function closely, whereas near the boundaries, particularly where the Beta density is lower, the estimator exhibits slightly higher variability, which is expected due to the limited local information.

