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 β_0, β_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

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

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

Also,

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

1

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

Unless stated otherwise, your answer to the practical part should include the following elements:

- Description of the aim of the simulation study.
- Description of the different quantities that intervene in your simulation study. Explain how these quantities (fixed or random) are defined and the reasoning behind the choices you made.
- Description of your findings using appropriate graphics and/or tables (with a caption!) that are well commented in the text.

The code should not appear in the PDF report, unless there is a specific reason to include it.

Ciao
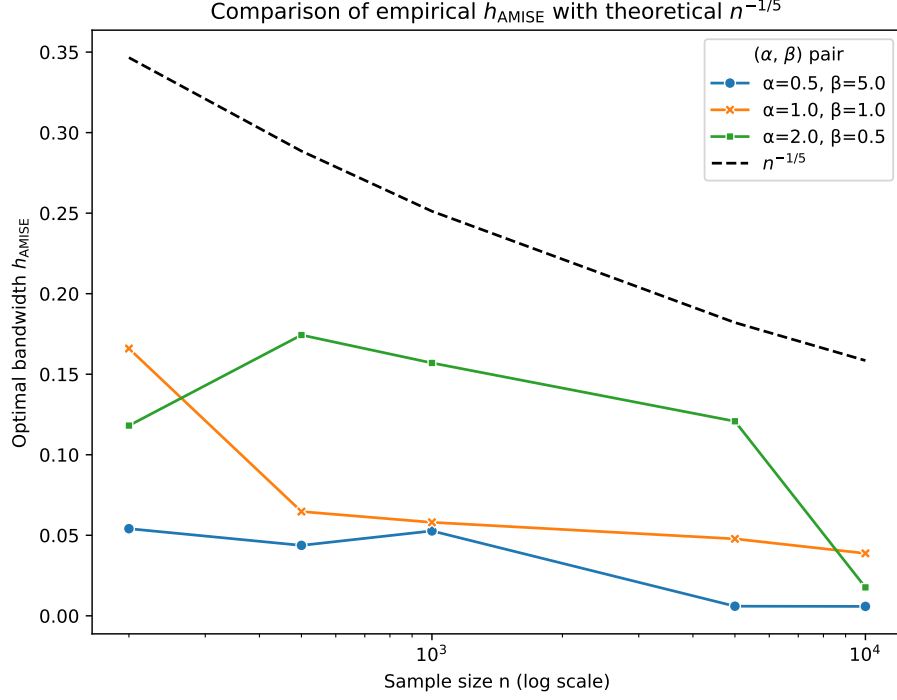
The figure compares the empirical behavior of the optimal bandwidth h_{AMISE} with the theoretical rate $n^{-1/5}$ for different Beta parameter pairs (α, β). For some distributions, such as $(1, 1)$, the empirical estimates follow the expected trend closely, confirming the theoretical prediction. However, for other parameterizations, for instance $(2, 0.5)$, certain deviations appear, with the empirical values not aligning perfectly with the $n^{-1/5}$ curve. A plausible explanation is the strong asymmetry of the Beta distribution in these cases, which leads to an uneven allocation of observations across blocks. This imbalance affects the stability of local polynomial fits, thereby producing variability in the estimation of θ_{22} and σ^2, and ultimately in the calculation of h_{AMISE}.

Effect of block size N on h_AMISE
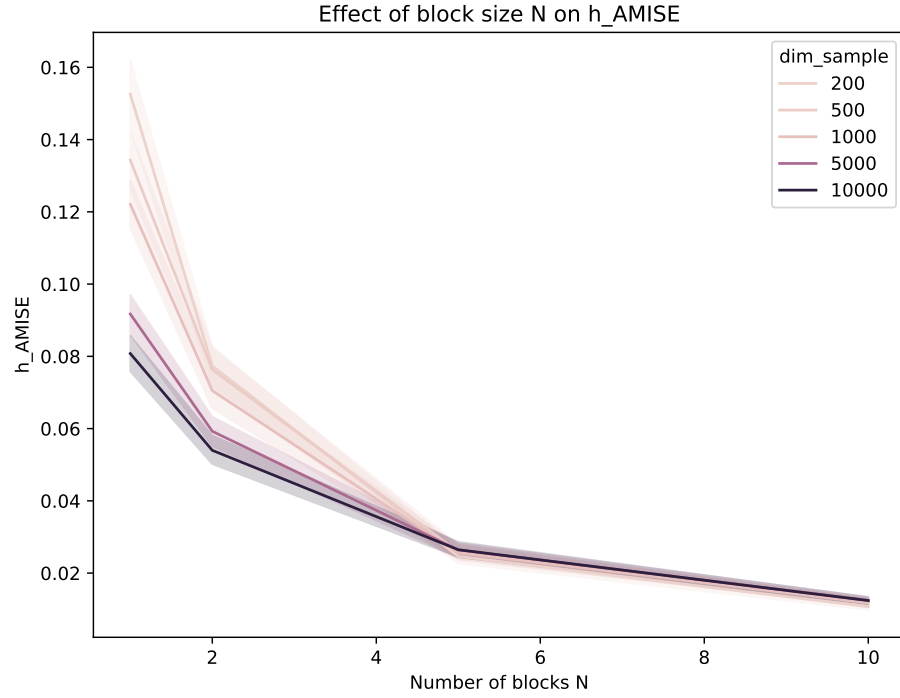
The plot shows the dependence of the optimal bandwidth h_{AMISE} on the number of blocks N, aggregated over different choices of the Beta distribution parameters. The general trend indicates that as N increases, the bandwidth decreases. This behavior is expected: partitioning the sample into a larger number of blocks effectively reduces the variance of the estimator, allowing for a finer resolution of the underlying density. Consequently, the optimal bandwidth adapts by shrinking, so as to balance bias and variance in accordance with the asymptotic mean integrated squared error (AMISE) criterion.

Effect of N on h_{AMISE} across sample sizes and alpha values
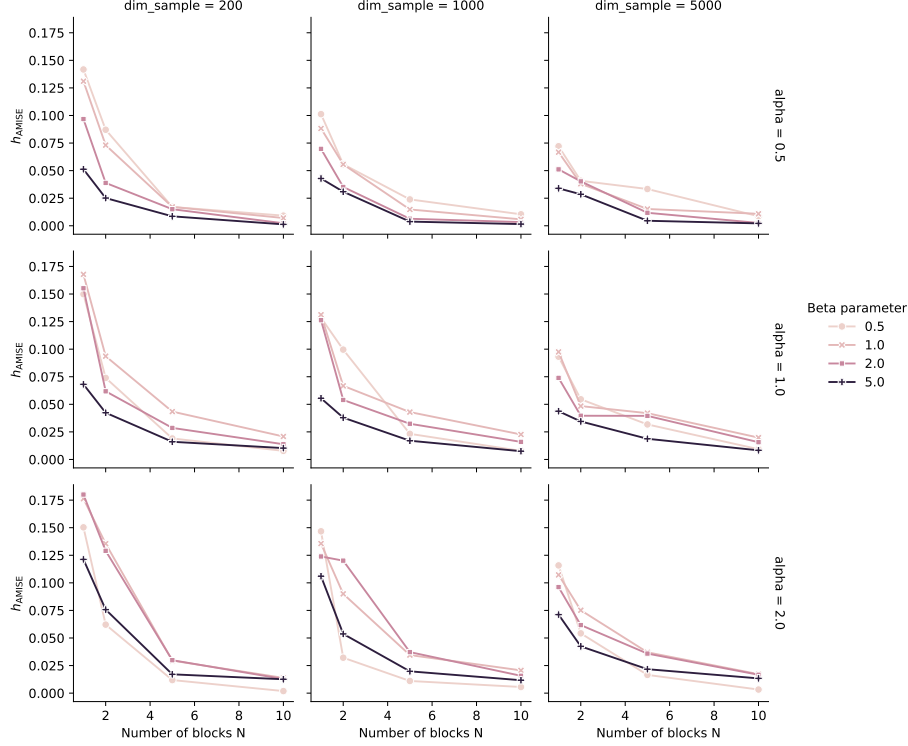
When differentiating by both α and β, the same general trend is observed: the optimal bandwidth h_{AMISE} decreases as the number of blocks N increases. This pattern holds consistently across different sample sizes and values of α, in line with the behavior described previously. Only in rare cases, typically corresponding to more asymmetric Beta distributions, slight deviations from the monotone decreasing trend appear. These discrepancies can be attributed to the uneven concentration of observations in different regions of the support, which affects the local variance estimation and therefore the stability of the optimal bandwidth.

Effect of Beta distribution (alpha, beta) on h_AMISE
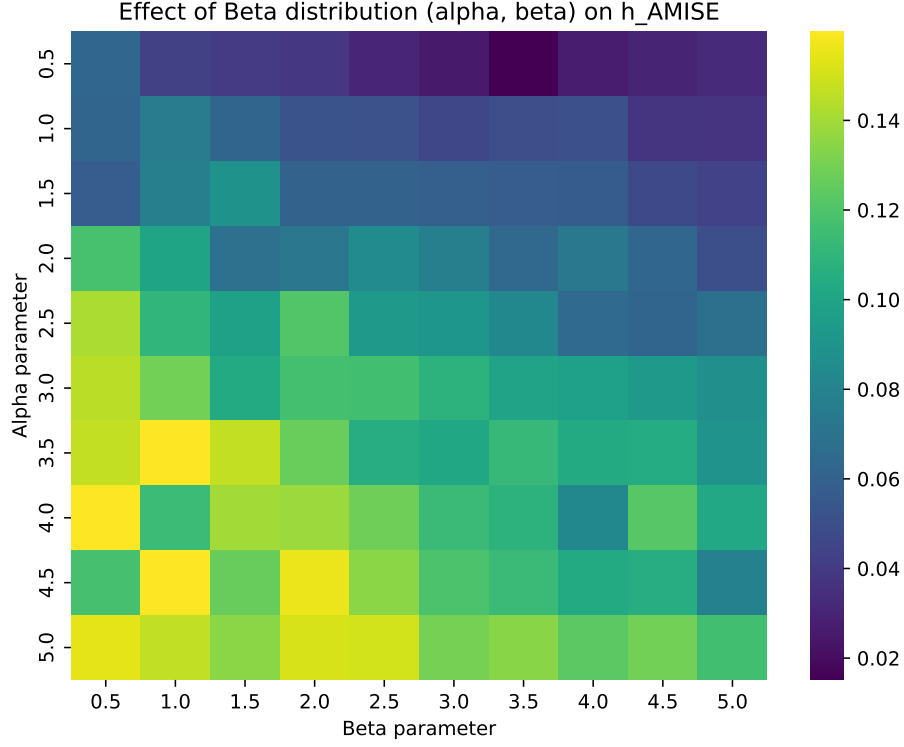
The heatmap illustrates how the optimal bandwidth h_{AMISE} depends on the parameters of the Beta distribution. A clear pattern emerges: moving towards the bottom-left corner of the heatmap (high α, low β) corresponds to larger values of h, whereas moving towards the top-right (low α, high β) leads to smaller h. This behavior is consistent with theoretical expectations. When α is large and β is small, the Beta distribution becomes heavily skewed towards the right boundary, concentrating probability mass in a narrow region. This sharp asymmetry requires a larger bandwidth to smooth effectively across the steep density gradients. Conversely, when α is small and β is large, the mass shifts towards the left boundary, producing a less concentrated distribution, for which a smaller bandwidth suffices.