MATH-517: Assignment 3

GERMI Rémi

2025-05-10

Theoretical exercise

Question 1

Let
$$K_i = K\left(\frac{X_i - x}{h}\right), \qquad u_i = X_i - x,$$
we can write the problem as a weighted least squares problem.

The minimization problem is:
$$(\widehat{\beta}_0(x), \widehat{\beta}_1(x)) = \arg\min_{\beta_0,\beta_1\in\mathbb{R}} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 u_i)^2 K_i.$$

Let
$$\mathbf{X} = \begin{pmatrix} 1 & u_1 \\ \vdots & \vdots \\ 1 & u_n \end{pmatrix}, \qquad \mathbf{W} = \operatorname{diag}(K_1, \dots, K_n), \qquad \mathbf{Y} = (Y_1, \dots, Y_n)^\top.$$

Then the weighted least squares solution is
$$\widehat{\beta}(x) = (\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{W}\mathbf{Y}.$$

The fitted value is
$$\hat{m}(x) = \widehat{\beta}_0(x) = \mathbf{e}_1^\top (\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{W}\mathbf{Y},$$
where $\mathbf{e}_1 = (1,0)^\top$. Hence,
$$\hat{m}(x) = \sum_{i=1}^{n} w_{n,i}(x) Y_i,$$
with weights
$$w_{n,i}(x) = \mathbf{e}_1^\top (\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1} \begin{pmatrix} 1 \\ u_i \end{pmatrix} K_i.$$

Question 2

Let

$$S_{n,k}(x) = \frac{1}{nh}\sum_{i=1}^{n} u_i^k K_i, \qquad k = 0, 1, 2.$$

We get

$$\mathbf{X}^\top \mathbf{W}\mathbf{X} = nh \begin{pmatrix} S_{n,0}(x) & S_{n,1}(x) \\ S_{n,1}(x) & S_{n,2}(x) \end{pmatrix} =: nh\,\mathbf{M}(x),$$

and its inverse is

$$(\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1} = \frac{1}{nh}\frac{1}{S_{n,0}(x)S_{n,2}(x) - S_{n,1}(x)^2} \begin{pmatrix} S_{n,2}(x) & -S_{n,1}(x) \\ -S_{n,1}(x) & S_{n,0}(x) \end{pmatrix}.$$

Therefore,

$$w_{n,i}(x) = \frac{1}{nh}\frac{S_{n,2}(x) - S_{n,1}(x)(X_i - x)}{S_{n,0}(x)S_{n,2}(x) - S_{n,1}(x)^2} K\Big(\frac{X_i - x}{h}\Big).$$

Question 3

Noting that

$$\mathbf{X}(1,0)^\top = \mathbf{1}_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{X}^\top \mathbf{W} = \sum_{i=1}^{n} \begin{pmatrix} 1 \\ u_i \end{pmatrix} K_i,$$

we get

$$\begin{aligned} \sum_{i=1}^{n} w_{n,i}(x) &= \sum_{i=1}^{n} \mathbf{e}_1^\top (\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1} \begin{pmatrix} 1 \\ u_i \end{pmatrix} K_i & (1) \\ &= \mathbf{e}_1^\top (\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ u_i \end{pmatrix} K_i & (2) \\ &= \mathbf{e}_1^\top (\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}\mathbf{1} & (3) \\ &= \mathbf{e}_1^\top (\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}\mathbf{X}(1,0)^\top & (4) \\ &= \mathbf{e}_1^\top \mathbf{I}_2 (1,0)^\top = 1. & (5) \end{aligned}$$

Thus, we obtain:

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

Practical exercise

- **Aim of the study :**

  The purpose of the study is to estimate and observe the optimal bandwidth $\hat{h}_{AMISE}$ for a beta distribution and to observe its behavior as a function of the block size N, the sample size n, and the parameters of the beta distribution.

- **Behavior of h when N growth:**

  From the h vs N plot (for Beta(2,2)): We observe that $\hat{h}_{AMISE}$ decreases as$N$ increases. We can explain this by the fact that increasing N divides the sample into smaller blocks. The local polynomial fits then become more localized and can better capture curvature in $m(x)$.
  This leads to a larger estimated $\hat{\theta}_{22}$ (the integrated squared second derivative), and possibly to a smaller estimated residual variance $\hat{\sigma}^2$.

  And from the definition of an increase in $\hat{\theta}_{22}$ typically implies a decrease in $\hat{h}$. Empirically, we observe exactly this behavior in the simulation.

- **Should N depend on n? Why?**

  Yes, the block size N (number of blocks) should be chosen in relation to the sample size n. In fact, if N is too large relative to n, blocks contain very few points, making degree-4 polynomial fits unreliable.
  Variance of the estimates increases and the denominator $(n - 5N)$ in the computation of $\hat{\sigma}^2$ can even become small or negative. If N is too small, the fits are overly global and may underestimate curvature, leading to a smaller $\hat{\theta}_{22}$ and therefore an overly large $\hat{h}$. Hence we use the C_p-like rule for try to get the best N.

  N should scale with n — larger data-sets allow more blocks, but each block must remain large enough for stable estimation.

- **What happens when we change the parameter of the Beta law?**

  When the number of observations varies a lot between different regions in the support of X e.g between Beta(5,1) or Beta(1,5) most observations concentrate near one end of the support, leaving few in the tails.

  As a result: Blocks in sparse regions have few points $\rightarrow$ unstable polynomial fits. Estimates of $\hat{\theta}_{22}$ and $\hat{\sigma}^2$ become noisier and $\hat{h}$ values fluctuate more, reflecting the heterogeneity in data density.The h vs Beta plot illustrates this.
  For instance, the U-shaped Beta(2, 2) distribution produced larger $\hat{h}$ values,
  since most observations lie near the boundaries and curvature estimates are less stable.

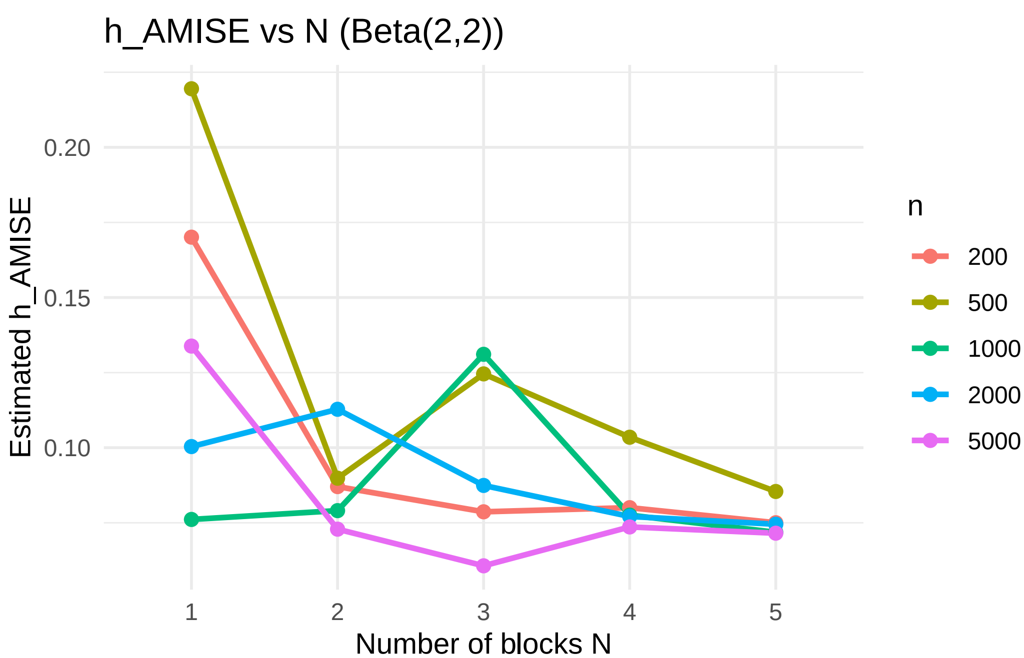
- **4. Effect of sample size n**

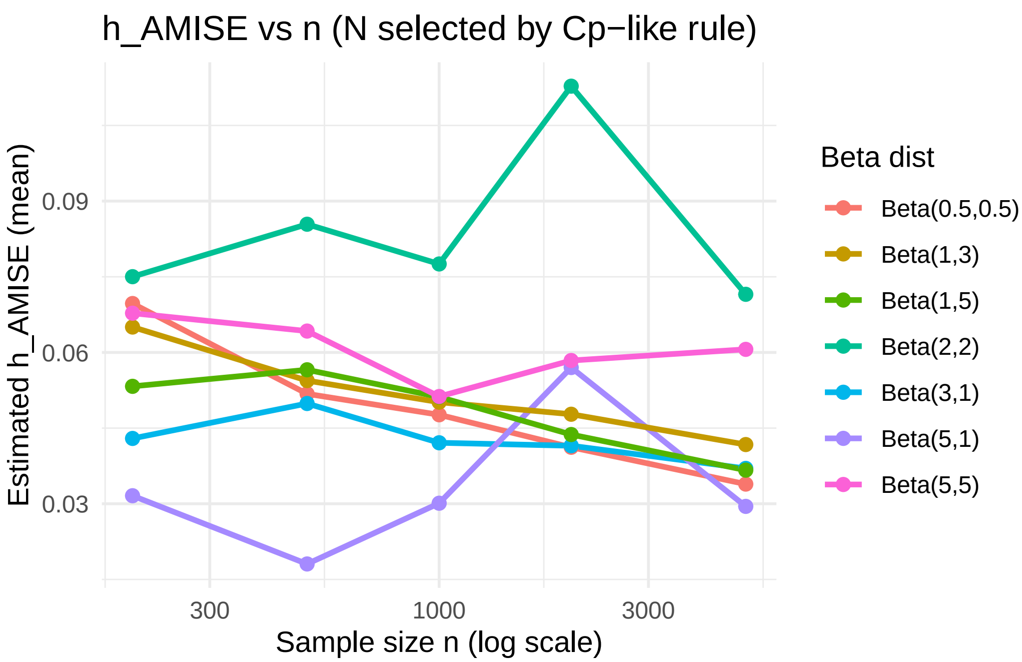
  From the h vs n plot (using C_p-selected N):

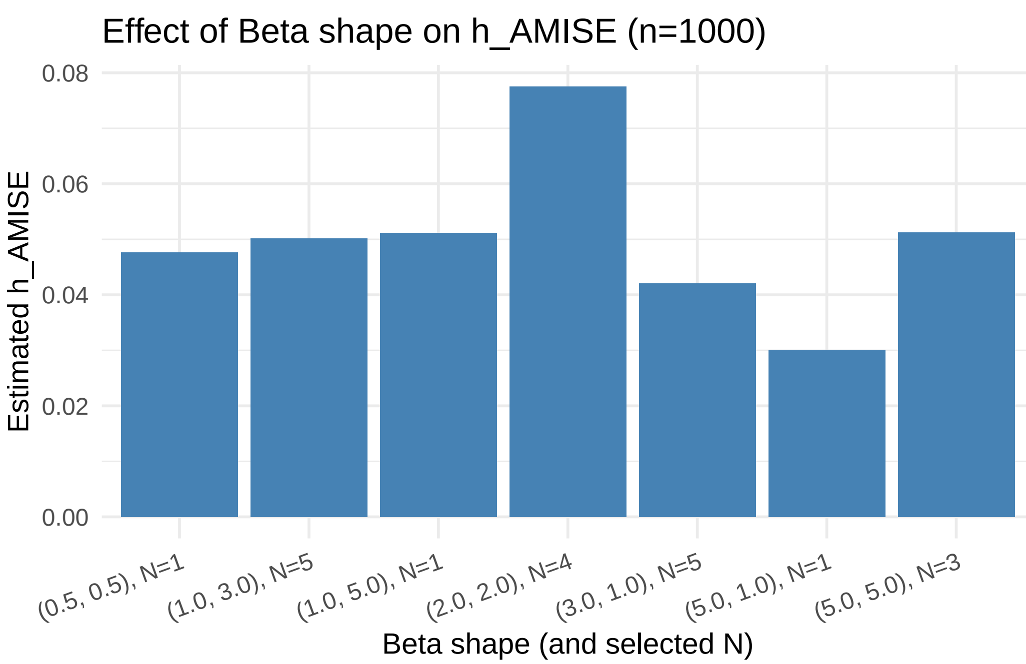
  $\hat{h}_{AMISE}$ decreases as n increases, consistent with the theoretical scaling:

  $$h_{AMISE} \propto n^{-1/5}.$$

  However, the exact slope and smoothness depend on the Beta shape and the selected N. When N is optimally reselected for each n (as done via the C_p rule),
  the dependence of $\hat{h}$ on n becomes smooth and closely follows the expected theoretical trend.








```
[1] "Simulation finished. Files produced:"

[1] "plot_h_vs_beta.png"   "plot_h_vs_n.png"      "plot_h_vs_Nblock.png"
```

```
[1] "Saved tables: simulation_results_summary.csv, selected_N_by_Cp.csv"
```