Problem Set 3: NP-Density estimation

Simulating the Data

- Generate a sample of n=100 points, $X_i \sim F$ for $i = 1, ..., n$ from a distribution of your choice and program (by hand) the kernel density estimator

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

using the Epanechnikov Kernel below.

$$K(x) = \frac{3}{4}\left(1 - x^2\right) \mathbf{1}(|x| \leq 1)$$

Estimation

- Plot the resulting function estimate for four different values of the bandwidth h.

Simulation Study

- Keeping the same h, repeat this process using $n = 500$ and compare (graphically) the results for all four bandwidths with $n = 100$.

*Hint: To program the estimator, use the sample of X_i and evaluate the function at each of these points **or** fix a grid on the main support of X and evaluate at each of the grid points. Make an initial guess of the bandwidth, keeping in mind the support of X.*