Practical exercise 1 : local polynomial estimation

Objective: Write a piece of R code implementing the local polynomial estimator described in Chapter 1. You can illustrate its calculation using the Yellowstone geyser data (this is obtained from `data(geyser)` in the R package `MASS`), or a data set that you have simulated yourself.

Level 1: implement the local linear estimator ($p = 1$) with the naive kernel at a given, fixed point x_0 and for a given, fixed bandwidth h.

Level 2: write a function, whose arguments should be a sample of data and a point x_0, which returns the value of the local linear estimator with the naive kernel and a given, fixed bandwidth h.

Level 3: upgrade your code in Level 2, by including as arguments in this function the order p of the polynomial, the kernel function (within a reasonable list of kernels, such as Epanechnikov, Gaussian...) and the bandwidth h.

Level 4: upgrade your code in Level 3, by implementing the cross-validation method for the choice of the bandwidth h and allowing the possibility of using this optimal choice (in addition to the possibility of inputting a specific value chosen by the user).

Level 5: speed up your code in Level 4 (eliminate "for" loops as much as possible, vectorise the function in Level 2...)

Compare your results with those of existing packages. Do you see any problem with the implementation of the estimator? (instability, unreasonably large computation time...)