Problemset 1

Econometrics III

LAURA MAYORAL
Instituto de Análisis Económico and Barcelona School of Economics

Spring 2023

INSTRUCTIONS:

(1) You can work in groups, max. 3 people;
(2) If you work in groups, you can submit a group answer, clearly specifying the members of the group.
(3) Please submit via classroom.
(4) Deadline: Wednesday 26, 6:00 PM.

Note: you can find the data for the exercises in Hansen's book here

1. Answer the following questions in a concise and clear manner.

a) Describe the role of the bandwidth in the nonparametric estimation of the density and describe the tradeoffs implied by its choice.

b) Explain the logic behind the optimal bandwidth and the difference between the optimal estimator of the bandwidth and the plug-in estimator.

c) Describe the problem of the bias in the computation of CI for the density estimator and briefly describe potential solutions.

2. Answer the following questions in a concise and clear manner.

a) Provide intuition about why the optimal bandwidth in kernel regression is $h^* = O(N^{-0.2})$.

b) Using a) and assuming that you use the optimal bandwidth, what's the convergence rate of the kernel regression estimator?

c) Compare the rate you obtained in b) with the convergence rate of the OLS estimator. Explain the differences and their implications.

3. Download the data (qreg0902.asc or qreg0902.data) for health expenditure from Cameron's website. (see mma04p2qreg.do for a description of this dataset and the variables in it). For each question, please clearly describe what you see.

a) Plot in the same graph two histograms for the log of household total expenditure and the log of household medical expenditure. Repeat using different values for the number of bins (one large value, one small value).

b) Use a Kernel estimate to estimate the density of the log of household medical expenditure. Choose the bandwidth using the plug-in estimator.

c) Repeat the same graph, trying different bandwidths (half and double of plug-in estimator, for instance).

d) Using the plug-in bandwidth estimator, estimate the density of these variables using three different kernels.

e) Add a normal distribution to the estimated density (obtained using default values) . Do the data look (log)normal?

f) Include the CI to the previous graph.

4. Cameron and Trivedi's book, Problem 9.4.

5. Hansen's book, exercise 19.9.

6. Hansen's book, exercise 19.10.

7. Hansen's book, exercise 19.11.

Good luck!