STAT 824 Project

Each student will conduct independent research in a special topic of nonparametric statistics and submit a written report and give a 15-minute presentation to the class at the end of the semester.

Part 1: *Settle this with me by Thursday, January 19th, before class.* Identify a topic in nonparametric statistics to study and have it approved by me. There is a (far-from-exhaustive) list of suggestions at the end of this document. *You cannot choose a topic that another student has already chosen.*

Part 2 (30% of project grade) : *Submit in class on Thursday, March 2nd.* Identify 6 important papers related to your topic. Among the 6 papers, there must be 2 papers *which propose estimators of the same thing*, so that you can later compare the performance of the estimators in a simulation study. Submit a literature review based on these 6 papers; specifically, write a summary of at least 12 sentences for *each* of the 6 papers and submit these summaries in a numbered list.

Part 3 (10%): *Submit in class on Thursday, March 16th.* Based on your literature review, identify two estimation methods which can be compared to one another in a simulation study. Submit a proposal for your project in which you describe the simulation study you intend to run and identify a data set on which you will illustrate the estimation methods. The proposal should also briefly describe why this project will be interesting (should give a motivation for the project). The proposal should be about the length of an abstract: 150 to 200 words. This will serve as your roadmap to completing the rest of the project.

Part 4 (30%): *Submit by 8:00 a.m. Wednesday, April 26th in my mailbox.* Turn in a written report of at least 8 pages (in LaTeX) with these sections:

1. **Abstract**: In which you describe the setting of your project and your findings.
2. **Introduction**: In which you describe the topic, giving its background and motivation, and provide a review of relevant papers in the area. Your literature review here should be in paragraph form and have a logical flow.
3. **Methodology**: In which you describe the two estimation methods in detail (so that someone could implement them if they read your report).
4. **Data analysis**: In which you demonstrate the use of the two (or more) estimation methods on a real data set.
5. **Simulation study**: In which you compare the performances of the two (or more) estimators on synthetic data sets.
6. **Conclusions**: In which you summarize the findings of your simulation study and offer any remarks you may have about which method you think should be used and in what settings.

Part 5 (30%): *Taking place on Thursday, April 13th, Tuesday, April 18th and Thursday, April 20th.* Give a 15-minute presentation to the class, covering the content in the written report, with slides prepared in LaTeX using the Beamer class.

Some topics you might find interesting – but feel free to propose others!

1. Nonparametric regression topics:
2. Mixed model approach to nonparametric regression.
3. Estimation of a monotone regression function, i.e. $Y = m(X) + \varepsilon$ with m monotone.
4. Handling heteroscedastic variance in nonparametric regression, i.e. $Y = m(X) + \sigma(X)\varepsilon$.
5. Nonparametric regression with wavelets.
6. Nonparametric regression with orthogonal series (sums of sines and cosines).
7. Local bandwidth selection, i.e. selecting different bandwidth $h(x)$ for different x.
8. Projection pursuit regression.
9. K-nearest neighbors regression.
10. Gaussian process regression.
11. Nonparametric quantile regression.
12. Regression trees.
13. The single index model.
14. Bandwidth selection for kernel density estimation when $d \geq 2$.
15. Wavelets in nonparametric density estimation.
16. Kernel density estimation for circular data.
17. Empirical likelihood estimation.
18. Nonparametric maximum likelihood estimation.
19. The m-out-of-n bootstrap. Why is it better to resample $m < n$ out of n in some settings?
20. The jackknife estimation procedure.
21. Deconvolution of functions: $X \sim f$, $\varepsilon \sim g$, g known; estimate f having observed $X + \varepsilon$?
22. Nonparametric estimation of survival function with censored data.
23. Estimating the spectral density of a stationary time series with periodogram smoothing.