Nonparametric Statistics IV

Mini-project Epiphany term

Jochen Einbeck

2025-03-07

Preamble

Aim

The aim of this mini-project is to compare analytic and simulation-based approaches to uncertainty quantification for local linear regression estimation.

Methods of Analysis

The task is to study comparatively the width of two types of (pointwise) confidence intervals for the local linear regression estimator:

- analytic approaches;
- simulation-based (bootstrap-type) approaches.

Manual implementations of both approaches are encouraged.

Notes:

- For the analytic approach, use the bias and variance expressions derived in Section 4.2.5, and then apply similar reasoning as in Section 3.3.9, to derive an approximate confidence interval. *You can take asymptotic normality without proof or evidence for granted.*
- Under either approach, you may need some creativity to estimate and "plug-in" unknown quantities. Please document and justify carefully what you do. You do *not* need to resort to (or cite) literature to underpin your own developments or derivations.
- However, *if* you decide to use built-in R functions or R packages to obtain either of the intervals, please make sure to describe precisely what the used function is actually doing, which in this case *may* require you to go back to (and cite) literature in order to be able to describe this.
- You can generally refer to lecture material rather than repeating it.

Carry out such an analysis for a data set of your choice (more detail on data choice below).

The minimum required deliverables are:

- a plot showing the data with the estimated regression mean function as well as both the analytic and the bootstrapped pointwise confidence intervals;
- an assessment of the average width of the confidence intervals under each setting;

- all required code;
- an introduction (which may motivate the work from a data or methods perspective, or both), a main part which describes the code and analysis provided, including methodological developments and all decisions made, and a conclusion.

Report preparation and submission

You can work on your report directly in the .Rmd version of this document. When completed you can delete the Preamble, and just keep your own work. Produce a PDF version of your document by knitting this Markdown file when completed. **Documents which were not created using R Markdown will not be accepted**.

You may submit your report **as a PDF file** (produced using R Markdown) via Gradescope (link will be provided on Ultra), by Tuesday 29th of April 2025 (end of day).

Length

Noting that your report will include images, code, and output, the report can have up to 10 pages of length. You will only submit the compiled (knitted) Markdown file, as a PDF, but not your original Rmd file. You are not asked to submit any auxiliary files. Your document is not expected to contain an Appendix, but you may opt to include one if you run out of space. You can assume that any Appendix *will not be considered for marking.*

General marking criteria

The marking scheme is as follows:

- Appropriate structure, introduction and conclusion **10 points**
- Quality of presentation (figures/tables/equations) **10 points**
- Soundness of methodological approach **10 points**
- Correctness of implementation and analysis **10 points**
- Quality of text/writing **5 points**
- Originality/creativity **5 points**

Use of AI

It may feel tempting to use AI tools to help your work on this assignment. If you contemplate doing so, please make sure to read carefully the Department's Use of Generative AI in project work.

Datasets

To illustrate your work, you just need a data set involving at least two variables. Please use any appropriate data set of your choice for this purpose, which however may *not* include any data set (or variables of any data set) dealt with in this module so far. You can however re-use the data from your mock-project if this was a data set which was not taken from this module.

Report

Here you begin with your actual work. . . .

```
# this chunk is reserved for getting the chosen data ready.
```

Throughout your report, ensure a healthy mix of text, code, equations, images, and perhaps tables. Good visualization of results is important!

```
# from here do what you wish to do!
```