Lab 6: Simulation Studies

2026-10-30

Table of contents

Overview **1**

Schedule **1**

Acknowledgements 2

Overview

Our lab today focuses on section **3. Simulation studies** of the paper by Li and Lin (2025), which you have read in advance.

Specifically, we'll discuss the goals of the simulation study, the choices the authors made in designing their data generating mechanisms, how those choices may affect the results of the study, and interpretation of those results as presented in Tables 2 and 3.

Schedule

1. First, to get us all refreshed on the context, I'll give a **brief overview** of their method.
2. Please form groups of 2-3 students. **Open a shared Google Doc**, giving access to everyone in your group.
3. List what were the main goals of the simulation study performed by the authors and what metrics they considered in assessing their method.
    - How do Emp SE and Est SE in Tables 2 and 3 differ? How large a discrepancy would concern you, and what additional information would you need to judge it? The average estimated SE is useful, but how is it limited in terms of assessing the procedure for estimating the population standard error? What could be a problem even if the average estimed SE were very close to the empirical standard error?
    - How many simulation replications were used? If true coverage were 95%, how much Monte Carlo variability would you expect in the reported coverage? Use this to assess a reported value. Do the authors quantify this simulation uncertainty?

4. Discuss how the authors generate their **simulated datasets**, as described in Section 3.1 of the paper. In your Google Doc, have one group member write down **pseudo-code** to generate a single dataset. Identify what is generated once and what is regenerated in each replication, and state any details you need to infer.
5. Next, **discuss the specific choices** the authors made in designing their simulated datasets. As you discuss, **write down those choices** in your Google Doc.
6. As a group, pick the **two most important choices** and justify why you picked them in the Google Doc.
7. With those choices in mind, brainstorm and write how you might simulate the data differently. Describe a data generating process that would challenge an assumption or approximation of the fitted model, and predict how it might affect coverage or standard error estimates.
8. Finally, have one group member export the Google Doc to PDF and submit it **as a group** for the "Lab 6" assignment on **Pensive**. Only one submission is needed per group.
9. Time allowing, after your group submits its assignment to Pensive, find another group that has also completed its submission and discuss your answer to question 7.

Acknowledgements

This lab was originally developed by Andrew Vaughn and James Duncan.