Lab 8: Simulation Studies

2024-11-01

Table of contents

Overview

Our lab today focuses on section **4. Numerical studies** of the paper by Cao et al., 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 Table 1 on page 765.

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.
4. Discuss how the authors generate their **simulated datasets**, as described in Section 4.1 of the paper. In your Google Doc, have one group member write down **pseudo-code** to generate a single dataset.
    - They use the multivariate normal (MVN) distribution for part of their simulation. Given a specific covariance matrix, can you show how to efficiently generate MVN data with that covariance structure using the Cholesky decomposition (described in Unit 9)?
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.
8. Finally, each group member should export the Google Doc to PDF and submit it for the "Lab 8: Simulation Studies" Gradescope assignment.
9. Time allowing, after each group member submits their assignment to Gradescope, find another group that has also completed their submission and discuss your answers.

Acknowledgements

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