The final course project is to study a research question of your choice, related to Bayesian modeling and data science. You are free to pursue any question under that broad theme. The project must be completed individually or in groups of up to 2.

> You will: Define a research question of interest, formulate a model for a data generating process, write the model in Stan/Pyro, fit it using either synthetic or real data, and write a project report.

Given that the course project is 40% of your course grade, I expect reasonable effort: ~25 hours is a reasonable benchmark, including everything (proposal, making presentation, writing report, doing the work).

Proposal

In class chat with Nikhil about your project plan 3/20

Meetings with each group like we did for the paper presentations.

Come prepared with your ideas + questions

1. Research question + method (Theory, theoretical estimand, empirical estimand, estimation strategy -- what is the estimand?
2. Data available, or will you simulate data?
3. What model you'll fit

In class project proposal presentation 3/22: ~5-6 minutes, depending on how many project teams we have

1. Add your 1-2 slides (strict limit; I will be combining pdfs for easy transitions) in pdf format by 10am 3/22 here: https://drive.google.com/drive/u/1/folders/1k8BeYvgmjX7gp7f8D3I8GRmEU-mXSj71
2. Tell us the question you'll be studying, why, and what methods you'll be using.
    a. Research question + method (Theory, theoretical estimand, empirical estimand, estimation strategy -- what is the estimand?
    b. Why is it interesting/useful/worth you spending your time on?
    c. Data available, or will you simulate data?
    d. What model you'll fit

Project proposal feedback

By 3/29, email me/presenters/Sidhika feedback about the project proposal, ~400 words

1. Do you have any questions

2. What was most interesting to you, least interesting
3. What questions do you hope they answer
4. Do you have any references/thoughts/etc that could be helpful

Final project deliverables

Final presentation: In class, last 4 class days

1. ~18 minutes total: 16 minutes presentation, leaving 2 minutes for questions
2. Powerpoint slides recommended

The presentation should be self-contained – your classmates attended the proposal review flash talks, but likely don't remember any of it.

1. It should contain everything a standard 18-20 minute conference talk contains (though related work slide can be skipped/shorter than usual): Motivation, problem setup, model, results, and discussion
    a. You don't need to walk through the stan/pyro code, but give details on the model you fit.
2. Since it is a class project, you should also spend some time discussing challenges, things that you tried that didn't work out, and whether you think there are promising research directions from the project

Written report + code submission due 5/13

Note: I have to submit final grades by Wednesday, May 24 at 3pm (ET). If you don't send me your report by at least 48 hours before then (May 22), I will have to submit a grade without your report. If you'd like further time, please talk to me and I can submit an "incomplete" for you.

Write a ~4-7 page project report, using the EC LaTeX template (https://ec22.sigecom.org/wp-content/uploads/2022/01/updated-ec22-submission-style-files.zip) or something similarly dense in terms of formatting.

You'll also submit your code, pdf, and slides via github classroom: https://classroom.github.com/a/__CowmwA

Approximate expected report content:

1. ~½ page introduction/motivation
    a. Research question + method (Theory, theoretical estimand, empirical estimand, estimation strategy -- what is the estimand, why is it interesting
2. ~½ related work. Doesn't have to be too extensive like you would have in a research paper, but discuss the few most related papers, and briefly state the most related research areas.

3. ~1-3 page model + methods
    a. Relationship to reality
    b. Data available to researchers to fit the model
    c. Synthetic data simulation model
4. ~1 pages results
5. ~½ - 1 page discussion – Reflection on the project
    a. Things you tried that didn't work out
    b. What you would try next if you were continuing the project
    c. Do you think there is enough in this direction for a research project? Do you want to continue?
    d. Any other thoughts on the project or related research areas
    e. Any other thoughts on the course
6. Appendix if relevant
    a. Bayesian model fit diagnostics – PPC, convergence, parameter forest plot, etc