Statistics 863: Advanced Statistical Computing
Fall 2018
Final Project

Overview

In this final project you will explore in more detail one of the statistical computing topics discussed in the course, implementing and applying it in the context of a specific application or methodological study. This will help you gain a deeper understanding of the topic of your choice, as well as to gain experience in translating these ideas into practice.

You should begin by choosing an application domain, data set, or methodological question of interest to you. (Theoretical topics are also fine.) This may be a topic related to your own research, or simply a topic about which you are curious. Although you are encouraged to chose problems relevant to your own research interests and activities, the project should be self-contained and the work done for Stat 863 clearly defined. Please see me with any questions. Working with a partner is possible, but the expectations will be twice as high.

The written project consists of two parts: (1) a 1-page project proposal on which I will provide comments to help you identify and scope an appropriate project for the course; and (2) a final paper, written in the style of a short journal or conference paper.

Project Proposal (due 10/15/18)

Prior to beginning work on your project you must provide me with a brief description of your proposed project for approval.

1. Find three (3) relevant papers which support the importance of your problem, discuss previous work on modeling/analysis in the area, and cover technical aspects of the methods from class you will apply. If there are more than three, select the three most relevant to cite/discuss in the proposal.

2. Prepare a brief ($\leq$1pg) description of your proposed project. Explicitly address each of the following:

   (a) Why the problem is of interest

   (b) What work has been done in this area previously

   (c) Goals of your project - what you hope to learn or demonstrate

   (d) The models or computational methods you will implement. Give models or likelihoods explicitly (but concisely).

   (e) Specify which, if any, datasets you will use.

   (f) How will you evaluate success.

   (g) Citations for the three papers you have chosen

I will return your proposal with comments and suggestions by the following class. I may also provide you with one **additional paper** reference for you to read and incorporate into your final write-up.

Individual Meetings (10/17 – 10/19)

Sign up for a brief ($\approx$ 20 min) meeting with me to discuss your proposal ideas and formulate an appropriate project. Available meeting times will be distributed by email. If you need suggestions for possible projects or help developing your idea, sign up for an early meeting before the proposal due date. If you already have a well-developed idea, it may be better to meet later in the week after you've submitted your proposal, so that I can provide feedback on it.

Progress Meetings (TBA)

You will meet with me again at least 1-2 times before submitting the final report, to discuss progress and obtain help with any problems. You should bring 3-5 brief slides to each meeting to facilitate discussion. Dates to be scheduled.

Final Paper (due 12/15/2018)

The final write-up should be in the style of a journal or conference paper. Please observe carefully the following specifications:

1. Formatting: Single space, $\geq$11pt font, $\geq$1-inch margins. LaTeX article defaults are fine.
2. Length: **7 page maximum**, *including* references and figures.
3. Your paper should contain the following sections:
   (a) *Abstract*
   (b) *Introduction* - Outline the general problem and its importance, providing any required domain-specific background. Describe the origin of any datasets used.
   (c) *Previous work* - Summarize previous approaches to this problem, their results and limitations, motivating your approach.
   (d) *Methods* - Describe the models and computational methods used in your project, emphasizing the details necessary for adapting to your problem, and any novel aspects of your approach.
   (e) *Results* - Report the results of your analysis, providing appropriate tables/plots/figures. Indicate areas where your results agree or disagree with previous work.
   (f) *Conclusions* - Interpret your results. What have you learned and contributed. Suggest topics for further work.
   (g) *References* - Bibliography of papers cited, including the papers from your proposal. Cite only critical references; should fill **no more than $\frac{1}{2}$ page.**
4. Sections (b), (c), and (d) should cite the appropriate references from your proposal, and any additional relevant references you have found.
5. Do not submit code or software output in bulk - extract relevant materials and include them as tables or figures in the body of the paper. Keep within the page limit.
6. Make sure to **spell-check** and **grammar-check** your paper before submitting.

7. Please see the following university resources on proper methods for citation of sources:

   http://www.lib.duke.edu/libguide/citing.htm

   http://www.lib.duke.edu/libguide/plagiarism.htm

Please talk to me about any questions or clarifications.

Example Projects

Examples of past class projects include:

- application of advanced statistical computing techniques to complex Bayesian models arising in your applied research
- development and testing of new algorithms
- theoretical analysis of convergence rates for specific algorithms on representative or idealized problems
- comparative empirical studies of the efficiency of different algorithms
- Many others.

These example should by no means limit your creativity - a wide range of other topics and approaches are possible. There is no requirement that you utilize any particular model or technique, rather you should choose those from class most appropriate to your problem.

A number of past student projects have led to publications. To give you an idea of the possible range from methdological to applied, some examples of these include (links can be found on my webpage):

- Ji C, Schmidler SC (2013). Adaptive Markov Chain Monte Carlo for Bayesian Variable Selection. *Journal of Computational and Graphical Statistics*, Vol 22, No 3, pp 708-728.
  (Developed new adaptive MCMC algorithms.)

- Wilson MA, Iversen ES, Clyde MA, Schmidler SC, Schildkraut JM (2010). Bayesian Model Search and Multilevel Inference for SNP Association Studies. *Annals of Applied Statistics*, Vol 4, No 3, pp 1342-64.
  (Applied enhanced MCMC sampling in an applied Bayesian context.)

- Jia B, Li J, Huang S, Schmidler SC, Wu YK (2010). Electron Beam Energy Spread Measurements Using Optical Klystron Radiation. *Physical Review Special Topics - Accelerators and Beams*, Vol 13, No 8, 080702 pp 1-14.
  (Applied Gauss-Hermite integration to improve fitting of a physics model to experimental data.)