2014 Stat 539 Biostat/Stat 572: History and Expectations

Jon Wakefield

Department of Statistics, University of Washington, WA
Department of Biostatistics, University of Washington, WA

Course Structure

- Instructors: Patrick Heagerty and Jon Wakefield.
- Stat 539 will share the first 2 lectures.
- Instructor: Marina Meila.

Historical Overview

- Biostat PhD students have theory and methods core sequences.
- Biostat PhD students: theory and applied exams, usually in 2nd year.
- Stat PhD students previously: 5 "core" sequences (stat theory/prob theory/methods/computing/stochastics) – do 2 prelims from 5, with an addtional sequence.
- Stat PhD students now: methods (570/571) and theory (581/582/583) required, plus a project course (518/539/572). (Machine Learning and Statistical Genetics slightly different).

Historical Overview

- **Methods:** historically, the 570s had a far greater emphasis on data analysis.
- During the 2000s this emphasis changed with 579 Data Analysis and Reporting now fulfiling this role.
- 572 usually covered nonparametric regression/classification, but the course (Stat/Biostat/CSSS 527) now covers these topics.
- **Stochastics:** one of 5 core sequences, previously project in Stat 518, now a research paper project requirement.
- **Statistical Learning:** one of 5 core sequences, Stat 539 recently introduced, with the same research paper project requirement.
- **Emphasize:** There are courses (518/539/572) and there are prelims (stochastics/learning/methods), and these are evaluated separately. Stat students do course and prelim, Biostat students do course (572) only.
- Prelim Dates:To be set. Exam week or the following week: Committees evaluate these.

Historical Overview

Motivation:

- How to start research?

- How to have students get into research earlier?

- Learn about faculty research interests.

- Become comtemptuous of a paper...and then move outwards.

- How to evaluate performance in methods, stochastics and learning?

Timeline:

- In 2010, 572 updated to become a research project course.

- Around the same time, 518 evolved from a research project into a research paper format. Shortly after, 539 did the same.

Biostat/Stat 572 Learning Objectives

By the end of the course the student should be able to:

- Critically evaluate a research article, and place its contents in historical context.
- Reproduce the numerical summaries of a research article.
- Reproduce the theoretical derivations of a research article.
- Clearly present the contents, and answer questions on, a research article.
- Critique a research talk.

572 Course Expectations

- Course attendance required.
- Participation in class encoucaged: including giving feeback on talks (10% of grade in 572).
- Read other students' papers: many of them are related, and it will help you give feedback.
- Introduction and update talks: not expected to have a deep understanding at this stage.
- Final paper and talk.

Stat prelim students give additional talk in June to a committee, and hand in an additional final paper.

Course Expectations

Three talks in 572:

1. Intro Talk, Start Week 2:
   - 20 minutes including 5 minutes of questions.
   - Introduce the paper and motivate the work, typically involves both scientific motivation using an example or data set, and statistical motivation
   - Basic elements of the first talk could include: Introduce the paper/Provide motivation for the work (scientific, statistical)/Review key background literature/Overview of the methods.

2. Update talk, Start Week 5:
   - 20 minutes including 5 minutes of questions.
   - Students can present on issues that have arisen, and we will allow time for discussion of potential solutions.
   - Details of the methods in terms of derivation and/or implementation Details of data simulation that could evaluate the methods Illustration of the methods using a worked example

3. Final Talk, Start Week 8:
   - 25 minutes including 5 minutes of questions.

Course Expectations

In the written report and final talk the students will be expected to:

1. Summarize the main contributions and novelty of the paper. This may include reading additional literature in order to understand the papers genesis, and to assess the papers subsequent impact.

2. Understand all of the analytical work in the paper, including filling in the details of analytic arguments.

3. Reproduce at least a subset of the simulation studies and data analyses in the paper.

4. Critique the paper.

The Write-Up for 572

The main sections of the project should be ≤ 20 pages (without appendices).
Appendices may contain longer proofs and R code. You should send an
electronic version of the latter.

The write-ups will be assessed on the following five categories:

1. *Motivation and Background Literature:* (2 points) What is the problem
   being considered, why is it important? What are

2. *Methods:* (8 points) Are the methods clearly explained? What are the
   details of the proposed method? Including relation to previous methods.

3. *Implementation:* (12 points) Simulations and data analysis. Correctly
   implemented?

4. *Conclusions and Critique:* (4 points) Did the paper stimulate other work?
   What are the main problems/assumptions of the method?

5. *Presentation:* (4 points) Is the writing style effective? Do tables, graphs
   and proofs effectively convey information?

Maximum score is 30.

The Talks for 572

The talks should be 25 minutes long, and you should expect 5 minutes of questions. Assume the audience has not seen the material previously, but is familiar with statistical methodology.

The talks will be assessed on the following three categories:

1. *Presentation:* (5 points) Was the talk clear and audible? Was the overall style effective? Were math equations, graphs and tables clear and effective? Was the pace good? Did the speaker use the time well? Did was the speaker enthusiastic? Any distracting behavior or mannerisms?

2. *Organization:* (5 points) Was the talk well organized? Good balance of introductory and more advanced material?

3. *Content:* (20 points) Was the problem addressed by the paper well motivated? Were the methods clearly explained? Was the information accurate? Did the speaker convey a deep understanding of the method?

Maximum score is 30.

Course Organization

- First three lectures will cover:
    - Reading the literature.
    - Searching the literature.
    - Preparing and delivering a presentation.
    - Writing statistical papers.
    - Designing simulation experiments.

- Then we will mostly be hearing and responding to talks.

- Students comment on talks also!

- Lots of useful materials online:
  https://courses.washington.edu/b572/lecture2013.htm