# S425/S625 syllabus (draft)

Brad Luen

Fall 2015

## 1 Instructor

Dr. Brad Luen
Room 200, Stat House, 309 N Park Ave (at 7th)
bradluen at indiana.edu (put S425/625 in the subject line)

Note that the Statistics Department will be moving at some point during the 2015–16 academic year.

## 2 Description

This course is a survey of statistical methods that do not rely on parametric assumptions. Knowledge of introductory statistics at the level of S320/S520 is assumed; this course is in some ways a sequel. As such, it will review the parametric techniques learned in that and similar introductory courses, and compare them to nonparametric alternatives to see when one technique outperforms another.

The major sections of the course are:

1. **Inference for proportions:** Key concepts of inference, including conservatism and power, and the role of assumptions. About 1 week.
2. **Classical nonparametric statistics:** One-, two-, and $k$-sample methods, including methods based on the binomial, on ranks, and on scores; paired comparisons; assocation and contingency tables. About 5 weeks.
3. **Empirical distributions and the bootstrap:** Confidence bands. Estimating the CDF from censored data. Bootstrap and resampling tests and confidence intervals. About 2 weeks.
4. **Nonparametric density estimation:** Kernel density estimates; conditional densities; comparisons to parametric density estimation. About 2 weeks.
5. **Nonparametric regression:** Local regression; splines; additive models. About 3 weeks.

The computation in the course will be performed in R.

By the course the course, you should:

- Be able to perform a variety of nonparametric inference and modeling techniques on real data.

- Know the mathematics behind basic nonparametric analyses.
- Understand the concepts motivating specific nonparametric analyses and statistical inference and modeling in general.

## 3 Prerequisites

Ideally, this course should be taken after a college level-probability course and a statistics course that uses R. However, a grade of B in S320 or a higher statistics course that involves computing should be sufficient preparation. You should already know how to calculate probabilities using software or otherwise for the fundamental probability distributions like the binomial and the normal. You should know the forms and interpretations of $t$-tests, confidence intervals, and the simple linear regression line.

You should have some experience with statistical software. It will be optimal if this software is the open-source program R. If not, download R for free from `cran.r-project.org`, install it, and start playing around with it. The clearest introductory guide to actually doing statistics in R is John Verzani's *simpleR* at `http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf`. The material up to chapter 7 (Simulations) will be immediately useful.

## 4 Lectures

Lectures will be held from 2:30 pm to 3:45 pm on Mondays and Wednesdays in Ballantine Hall 105.

## 5 Provisional office hours

S426/625-specific office hours: Tuesdays 2:30–3:30 pm
My general office hours: Mondays 1–2 pm, Tuesdays 1:30–2:30 pm

## 6 Attendance

I'll take attendance for the first few weeks of class. If you are unable to attend, please inform me before class starts.

## 7 Textbooks

James J. Higgins, *Introduction to Modern Nonparametric Statistics*
Course notes (*Some of Nonparametric Statistics*) — posted on Canvas

We'll follow Higgins up to his chapter 8 (the bootstrap) and then branch off into other things once we get to density estimation and nonparametric regression.

Addtional useful texts:

- Davison and Hinkley, *Bootstrap Methods and Their Application*
- Cosma Shalizi, *Advanced Data Analysis from an Elementary Point of View*: `http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/`

- Larry Wasserman, *All of Statistics* (for a review of parametric statistics) and *All of Nonparametric Statistics*

# 8 Grading

**Weekly homework assignments (25%):** These will consist of exercises from Higgins, plus additional problems. They will usually be announced in class and posted on Wednesdays, and will be due the following Wednesday (probably through Canvas, if I can figure out how Canvas works.) Collaboration on homework with other students is permitted and encouraged. However, you must write up your own solutions, and state who you worked with.

**Midterm tests 1 (25%) and 2 (25%)** There will be two midterms, each with takehome and in-class components. Midterm 1 will take place after we cover classical nonparameteric statistics (early October). Midterm 2 will take place after we cover density estimation (early November). Collaboration on tests with other students is not permitted under any circumstance.

**Final exam (25%)** The final exam will be takehome-only. Final grades will be submitted to the Registrar on Thursday 17th December. Final grades cannot be changed unless a mistake has been made in calculation.