STAT C180/C236.- INTRODUCTION TO BAYESIAN STATISTICS

http://www.stat.ucla.edu/~jsanchez/bayes/index.html

Course Pre-requisites: Mathematics 32B, 33B

- **Instructor:** Juana Sanchez **jsanchez@stat.ucla.edu** , **http://www.stat.ucla.edu/~jsanchez**
  Office: **Math Sciences , 8935**
- **Office hours:** *See course web*
- **Required Reading: (I) Course lecture outlines:** get the notes from the course web page. The course notes contain lecture outlines, required additional reading and some exercises that we discuss interactively in class. They also contain R programs and code (which you can also access online) **(II) Some material from Hoff, P.D. (2009) A First Course in Bayesian Statistical Methods.** Springer Texts in Statistics. (You may download this book from our library, as well as any other Springer Verlag books).
- **Recommended reading:** Gelman et al. Bayesian Data Analysis. Second Edition. Chapman and Hall. On reserve in the SEL Library. **Other:** Will be indicated in each lecture.
- **Required Software:**  http://cran.stat.ucla.edu/ Download R into your computer by clicking on the operating system that you have and then download the base and packages. The software is free. In the last 2 weeks of the quarter, we will also do a couple of lectures to learn BUGS.

WHY BAYESIAN STATISTICS

Bayesian Statistics used to be the predominant method of statistical analysis until the early thirties, when experimentation led to the development of other methods suitable mostly for experimental data. In the eighties, Bayesian statistics started gaining predominance again in the practice of statistics in many areas (medicine, computer science, physics, economics, to name a few). Not only does Bayesian statistics allows to consider experimental data, but it allows us to do inference when the notion of experimentation is unrealistic, which is the case in many sciences. Bayesian statistics also takes the con out the usual practice of data fishing, ie. incorporating subjective opinions into the analysis without a formal way of identifying that subjectivity. With Bayesian analysis, the subjective opinion of the experts involved and the data are clearly separated and explicitly indicated. Today, Bayesian statistics is so prevalent in grant panel boards, editorial boards and so many other decision making spheres that not knowing Bayesian statistics makes a student only partially prepared to face the real world.

ABOUT THE COURSE

This course is a data-driven/computationally intensive course on Bayesian Statistics with R. Each lecture originates with a data set and a reflection on the questions on populations that the data set allows us to answer. The Bayesian strategy needed to answer the questions follows. R templates that the student can change and transform for different analyses will be provided. More data analysis examples of similar nature but different context will reinforce the skills gained. Assignments will allow further practice.

Acquiring skills to be able to conduct a wide variety of Bayesian data analyses is the main goals of the course. Statistical inference will be conducted exactly and by direct simulation (if a mathematical, closed-form exists), or by grid simulation, and/or by MCMC simulation (Gibbs Sampling or Metropolis-Hasting algorithms), when a closed form does not exist. Sometimes we will use simulation when a closed form exists, so that you convince yourself that simulation works.

Data sets come from Business, Linguistics, Engineering, Social Sciences, Physical Sciences, Law, Medicine, Education, Life Sciences, Pharmacokinetics, Environmental, Vision and many other areas. This course should be of interest to any graduate or undergraduate student regardless of their major. Having seen probability theory

and some data analysis before the course, as well as R proficiency, will help, but it is not required. Undergraduates have lower requirements to pass the class than graduate students.

COURSE REQUIREMENTS

Weekly assignments. (45%) No late assignments or make ups under any circumstances.
Individual Class project: Presentation and paper due on final exams week. (40%). No late project accepted.
Participation/Attendance/Keeping up with required reading: 15% (no makeup under any circumstances)
Exercises: occasional exercises to prepare class discussion for the next lecture. Exercises will not be collected.

TOPICS COVERED IN THE COURSE

1.- The elements of a Bayesian data analysis. Conditional Probability. Law of Total Probability. Bayes' rule. Updating Probabilities. Why do we need a formal rule for learning? Sequential analysis. Likelihoods, priors and posteriors. Predictive probabilities. Assessment of priors. Posterior intervals, hypothesis testing. Exchangeability. All these elements appear in each of the topics mentioned below adapted to the particular model and problem.

2.- Inferences for proportions. Binomial and Multinomial problems, logistic regression. Laplace's rule of succession. Classification and discrimination.

3.- Inferences for means. Comparing population means. Simple and multiple regression and ANOVA.

4.- Inferences for Poisson rates. Prior to Posterior. Ratios of Poisson rates.

5.- Bayesian hierarchical/multilevel analysis. Other generalized linear models. Cross section and longitudinal models.

6.- Decision Analysis. Loss functions. Bayes risk. Value of perfect information. Value of sample information. Bayesian sample size and strategy

APPLICATIONS

Medical diagnostics. Paternity testing. Crime victimization. Second guessing undecided respondents in surveys. Linguistics, determining authorship. Mixtures. Test scores and other measures of performance in education. Randomized experiments and case control studies. Growth curves. Bioassays. Pollution. Elections. Mapping of disease. Dose-response studies. Repeated measures. Time series. Experimental design. Marketing, e-commerce.

Students are welcome to bring in areas of application that interest them and will be given a chance to incorporate them in some assignments and the course project.