Welcome to STA 702

Course Overview

Merlise Clyde
Duke University

What is this course about?

- Learn the foundations and theory of Bayesian inference in the context of several models.
- Use Bayesian models to answer inferential questions.
- Apply the models to several different problems.
- Understand the advantages/disadvantages of Bayesian methods vs classical methods

> A Bayesian version will usually make things better…
>
> – Andrew Gelman.

Instructional Team

Instructor: Dr Merlise Clyde

clyde@duke.edu
223 Old Chemistry
https://www2.stat.duke.edu/~clyde

Teaching Assistant: Rick Presman

rick.presman@duke.edu

See course website for Office Hours, Policies and more!

Prerequisites

- random variables, common families of probability distribution functions and expectations
- conditional distributions
- transformations of random variables and change of variables
- principles of statistical inference (likelihoods)
- sampling distributions and hypothesis testing
- concepts of convergence

Review Chapters 1 to 5 of the Casella and Berger book

Computing

- Labs/HW will involve computing in R!
- Write your own MCMC samplers and run code long enough to show convergence
- You can learn R on the fly
  - see Resources Tab on website
  - materials from 2023 Bootcamp/Orientation

Grading Policies

- 5% class
- 20% HW
- 10% Lab
- 20% Midterm I
- 20% Midterm II
- 25% Final
- No Late Submissions for HW/Lab; Drop the lowest score
- You are encouraged to discuss assignments, but copying others work is considered a misconduct violation and will result in a 0 on the assignment
- Confirm that you have access to Sakai, Gradescope, and GitHub

Course structure and policies

- See the Syllabus
- Make use of the teaching team's office hours, we're here to help!
- Do not hesitate to come to my office hours or you can also make an appointment to discuss a homework problem or any aspect of the course.
- Please make sure to check your email daily for announcements
- Use the Reporting an issue link to report broken links or missing content

Important Dates

Tues, Aug 29	Classes begin
Fri, Sept 8	Drop/Add ends
Friday, Oct 13	Midterm I (*tentative*)
Sat - Tues, Oct 14 - 17	Fall Break
Tues, Nov 20	Midterm II (*tentative*)
Friday, Dec 1	Graduate Classes End
Dec 2 - Dec 12	Graduate Reading Period
Sat, Dec 16	**Final Exam** (Perkins 060 2:00-5:00pm)

See Class Schedule for slides, readings, HW, Labs, etc

Topics

- Basics of Bayesian Models
- Loss Functions, Inference and Decision Making
- Predictive Distributions
- Predictive Distributions and Model Checking
- Bayesian Hypothesis Testing
- Multiple Testing
- MCMC (Gibbs & Metropolis Hastings Algorithms)
- Model Uncertainty/Model Choice
- Bayesian Generalized Linear Models
- Hiearchical Modeling and Random Effects
- Hamiltonian Monte Carlo
- Nonparametric Bayes Regression

Bayes Rules! Getting Started!

Basics of Bayesian inference

Generally (unless otherwise stated), in this course, we will use the following notation. Let

- Y is a random variable from some probability distribution $p(y \mid \theta)$
- $\mathcal{Y}$ be the sample space (possible outcomes for Y)
- y is the observed data
- θ is the unknown parameter of interest
- Θ be the parameter space
- e.g. $Y \sim$ Ber(θ) where $\theta = \Pr(Y = 1)$

Frequentist inference

- Given data y, how would we estimate the population parameter θ?
  - Maximum likelihood estimate (MLE)
  - Method of moments
  - and so on…
- Frequentist MLE finds the one value of θ that maximizes the likelihood
- Typically uses large sample (asymptotic) theory to obtain confidence intervals and do hypothesis testing.

What are Bayesian methods?

- Bayesian methods are data analysis tools derived from the principles of Bayesian inference and provide
  - parameter estimates with good statistical properties;
  - parsimonious descriptions of observed data;
  - predictions for missing data and forecasts of future data with full uncertainty quantification; and
  - a computational framework for model estimation, selection, decision making and validation.
  - builds on likelihood inference

Bayes' theorem

- Let's take a step back and quickly review the basic form of Bayes' theorem.
- Suppose there are some events A and B having probabilities $\Pr(A)$ and $\Pr(B)$.
- Bayes' rule gives the relationship between the marginal probabilities of A and B and the conditional probabilities.
- In particular, the basic form of Bayes' rule or Bayes' theorem is

$$\Pr(A|B) = \frac{\Pr(A \text{ and } B)}{\Pr(B)} = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

- $\Pr(A)$ = marginal probability of event A, $\Pr(B|A)$ = conditional probability of event B given event A, and so on.
- "reverses the conditioning" e.g. Probability of Covid given a negative test versus probability of a negative test given Covid

Bayes' Rule more generally

1. For each $\theta \in \Theta$, specify a prior distribution $p(\theta)$ or $\pi(\theta)$, describing our beliefs about θ being the true population parameter.
2. For each $\theta \in \Theta$ and $y \in \mathcal{Y}$, specify a sampling distribution $p(y|\theta)$, describing our belief that the data we see y is the outcome of a study with true parameter θ. Likelihood $L(\theta|y)$ proportional to $p(y|\theta)$
3. After observing the data y, for each $\theta \in \Theta$, update the prior distribution to a posterior distribution $p(\theta|y)$ or $\pi(\theta|y)$, describing our "updated" belief about θ being the true population parameter.

Getting from Step 1 to 3? Bayes' rule!

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{\int_{\Theta} p(\tilde{\theta})p(y|\tilde{\theta})\mathrm{d}\tilde{\theta}} = \frac{p(\theta)p(y|\theta)}{p(y)}$$

where $p(y)$ obtained by Law of Total Probability

Notes on prior distributions

Many types of priors may be of interest. These may

- represent our own beliefs;
- represent beliefs of a variety of people with differing prior opinions; or
- assign probability more or less evenly over a large region of the parameter space
- designed to provide good frequentist behavior when little is known

Notes on prior distributions

Notes on prior distributions

- The prior quantifies 'your' initial uncertainty in θ before you observe new data (new information) - this may be necessarily subjective & summarizes experience in a field or prior research.
- Even if the prior is not "perfect", placing higher probability in a ballpark of the truth leads to better performance.
- Hence, it is very seldom the case that a weakly informative prior is not preferred over no prior. (Model selection is one case where one needs to be careful!)
- One (very important) role of the prior is to stabilize estimates (shrinkage) in the presence of limited data.

Next Steps

Work on Lab 0

Finally, here are some readings to entertain you. Make sure to glance through them within the next week. See Course Resources

1. Efron, B., 1986. Why isn’t everyone a Bayesian?. The American Statistician, 40(1), pp. 1-5.
2. Gelman, A., 2008. Objections to Bayesian statistics. Bayesian Analysis, 3(3), pp. 445-449.
3. Diaconis, P., 1977. Finite forms of de Finetti’s theorem on exchangeability. Synthese, 36(2), pp. 271-281.
4. Gelman, A., Meng, X. L. and Stern, H., 1996. Posterior predictive assessment of model fitness via realized discrepancies. Statistica sinica, pp. 733-760. 5. Dunson, D. B., 2018. Statistics in the big data era: Failures of the machine. Statistics & Probability Letters, 136, pp. 4-9.