# STA 702: Lecture 0

## Course Overview

Merlise Clyde

# Welcome to STA 702!

# What is this course about?

Learn the foundations and theory of Bayesian inference in the context of several models.

# What is this course about?

Learn the foundations and theory of Bayesian inference in the context of several models.

Apply the models to several different problem sets.

# What is this course about?

Learn the foundations and theory of Bayesian inference in the context of several models.

Apply the models to several different problem sets.

Use Bayesian models to answer inferential questions.

# What is this course about?

Learn the foundations and theory of Bayesian inference in the context of several models.

Apply the models to several different problem sets.

Use Bayesian models to answer inferential questions.

Understand the advantages/disadvantages of Bayesian methods vs classical methods

# What is this course about?

Learn the foundations and theory of Bayesian inference in the context of several models.

Apply the models to several different problem sets.

Use Bayesian models to answer inferential questions.

Understand the advantages/disadvantages of Bayesian methods vs classical methods

We will use the Hoff book, BDA and other reference material.

# What is this course about?

Learn the foundations and theory of Bayesian inference in the context of several models.

Apply the models to several different problem sets.

Use Bayesian models to answer inferential questions.

Understand the advantages/disadvantages of Bayesian methods vs classical methods

We will use the Hoff book, BDA and other reference material.

---

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

# Instructor

Dr Merlise Clyde

- clyde@duke.edu
- https://www2.stat.duke.edu/~clyde
- 223 Old Chemistry
- https://www2.stat.duke.edu/courses/Fall22/sta702.001
- See course website for OH

# TAs

## Steven Winter

- steven.winter@duke.edu
- See course website for OH
- See course website for location

# FAQs

All materials and information will be posted on the course webpage:
https://www2.stat.duke.edu/courses/Fall22/sta702.001/

# FAQs

All materials and information will be posted on the course webpage:
https://www2.stat.duke.edu/courses/Fall22/sta702.001/

- How much theory will this class cover? A lot!

# FAQs

All materials and information will be posted on the course webpage:
https://www2.stat.duke.edu/courses/Fall22/sta702.001/

- How much theory will this class cover? A lot!
- What if I can't remember all the topics in the prerequisites?

# FAQs

All materials and information will be posted on the course webpage: https://www2.stat.duke.edu/courses/Fall22/sta702.001/

- How much theory will this class cover? A lot!
- What if I can't remember all the topics in the prerequisites?
    1. Review Chapters 1 to 5 of the Casella and Berger book You can find the solution manual here

# FAQs

All materials and information will be posted on the course webpage: https://www2.stat.duke.edu/courses/Fall22/sta702.001/

- How much theory will this class cover? A lot!
- What if I can't remember all the topics in the prerequisites?

  1. Review Chapters 1 to 5 of the Casella and Berger book You can find the solution manual here

  2. Focus on the following topics:

     - basic probability theory, random variables, transformations of random variables and change of variables, expectations of random variables, common families of probability distribution functions including multivariate distributions
     - concepts of convergence, principles of statistical inference, likelihood based inference, sampling distributions and hypothesis testing.

# FAQs

- Will we be doing "very heavy" computing? a good amount. You will be expected to be able to write your own MCMC samplers and run code long enough to show convergence

# FAQs

- Will we be doing "very heavy" computing? a good amount. You will be expected to be able to write your own MCMC samplers and run code long enough to show convergence
- What computing language will we use? R!

# FAQs

- Will we be doing "very heavy" computing? a good amount. You will be expected to be able to write your own MCMC samplers and run code long enough to show convergence
- What computing language will we use? R!
- What if I don't know R? This course assumes you already know R but you can still learn on the fly (you must be self-motivated). Here are some resources for you:
  - https://www2.stat.duke.edu/courses/Fall22/sta702.001/resources/.
  - Resources for the StaSci BootCamp 2021

# FAQs

- Will we be doing "very heavy" computing? a good amount. You will be expected to be able to write your own MCMC samplers and run code long enough to show convergence
- What computing language will we use? R!
- What if I don't know R? This course assumes you already know R but you can still learn on the fly (you must be self-motivated). Here are some resources for you:
  - https://www2.stat.duke.edu/courses/Fall22/sta702.001/resources/.
  - Resources for the StaSci BootCamp 2021
- Labs will introduce/review concepts

# FAQs

- Do I need to know a lot Bayesian statistics before taking this class?
  No

# FAQs

- Do I need to know a lot Bayesian statistics before taking this class?
  No
- What is the difference between this course and STA360 or STA602 ?

# Course structure and policies

# Course structure and policies

- See: https://www2.stat.duke.edu/courses/Fall22/sta702.001/policies/

# Course structure and policies

- See: https://www2.stat.duke.edu/courses/Fall22/sta702.001/policies/
- Make use of the teaching team's office hours, we're here to help!

# Course structure and policies

- See: https://www2.stat.duke.edu/courses/Fall22/sta702.001/policies/
- 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.

# Course structure and policies

- See: https://www2.stat.duke.edu/courses/Fall22/sta702.001/policies/
- 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.
- When the teaching team has announcements for you we will send an email to your Duke email address. Please make sure to check your email daily.

# Course structure and policies

- See: https://www2.stat.duke.edu/courses/Fall22/sta702.001/policies/
- 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.
- When the teaching team has announcements for you we will send an email to your Duke email address. Please make sure to check your email daily.
- Check the dates for exams in the Course Calendar and let me know ASAP if there are issues

# 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
- Bayesian Generalized Linear Models
- Hiearchical Modeling and Random Effects
- Hamiltonian Monte Carlo
- Nonparametric Bayes Regression

# 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
- Bayesian Generalized Linear Models
- Hiearchical Modeling and Random Effects
- Hamiltonian Monte Carlo
- Nonparametric Bayes Regression

Also refer to the Class Schedule.

# Other details

- Grading
  - 5% class
  - 20% HW
  - 10% Lab
  - 20% Midterm I
  - 20% Midterm II
  - 25% Final

# Other details

- Grading
  - 5% class
  - 20% HW
  - 10% Lab
  - 20% Midterm I
  - 20% Midterm II
  - 25% Final
- No Late Submissions for HW/Lab; Drop the lowest score

# Other details

- Grading
  - 5% class
  - 20% HW
  - 10% Lab
  - 20% Midterm I
  - 20% Midterm II
  - 25% Final
- No Late Submissions for HW/Lab; Drop the lowest score
- Confirm that you have access to Sakai, Gradescope, and GitHub.

# Important Dates

Tues, Aug 30   Classes begin
Fri, Sept 9   Drop/Add ends
Fri, Oct 7   Midterm I (*tentative*)
Sat - Tues, Oct 8 - 11   No classes Fall Break |
Tue, Nov 15   Midterm II (*tentative*)
Friday Dec 2  Graduate Classes End
Mon, Dec 19   Final Exam (Perkins 060 2:00-5:00pm)

# Important Dates

Tues, Aug 30   Classes begin
Fri, Sept 9   Drop/Add ends
Fri, Oct 7   Midterm I (*tentative*)
Sat - Tues, Oct 8 - 11   No classes Fall Break |
Tue, Nov 15   Midterm II (*tentative*)
Friday Dec 2  Graduate Classes End
Mon, Dec 19   Final Exam (Perkins 060 2:00-5:00pm)

Also refer to the schedule on the website for updated breakdown of the courses. Remember to refresh the page frequently. See here: Class Schedule.

# Bayes Rules! Getting Started!

# Basics of Bayesian inference

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

# Basics of Bayesian inference

- Generally (unless otherwise stated), in this course, we will use the following notation. Let
  - $\mathcal{Y}$ be the sample space;

# Basics of Bayesian inference

- Generally (unless otherwise stated), in this course, we will use the following notation. Let
  - $\mathcal{Y}$ be the sample space;
  - $y$ be the observed data;

# Basics of Bayesian inference

- Generally (unless otherwise stated), in this course, we will use the following notation. Let
  - $\mathcal{Y}$ be the sample space;
  - $y$ be the observed data;
  - $\Theta$ be the parameter space; and

# Basics of Bayesian inference

- Generally (unless otherwise stated), in this course, we will use the following notation. Let
  - $\mathcal{Y}$ be the sample space;
  - $y$ be the observed data;
  - $\Theta$ be the parameter space; and
  - $\theta$ be the parameter of interest.

# Basics of Bayesian inference

- Generally (unless otherwise stated), in this course, we will use the following notation. Let
  - $\mathcal{Y}$ be the sample space;
  - $y$ be the observed data;
  - $\Theta$ be the parameter space; and
  - $\theta$ be the parameter of interest.
- More to come later.

# Frequentist inference

- Given data $y$, estimate the population parameter $\theta$.

# Frequentist inference

- Given data $y$, estimate the population parameter $\theta$.
- How to estimate $\theta$ under the frequentist paradigm?
  - Maximum likelihood estimate (MLE)
  - Method of moments
  - and so on...

# Frequentist inference

- Given data $y$, estimate the population parameter $\theta$.
- How to estimate $\theta$ under the frequentist paradigm?
  - Maximum likelihood estimate (MLE)
  - Method of moments
  - and so on...
- Frequentist ML estimation finds the one value of $\theta$ that maximizes the likelihood.

# Frequentist inference

- Given data $y$, estimate the population parameter $\theta$.
- How to estimate $\theta$ under the frequentist paradigm?
  - Maximum likelihood estimate (MLE)
  - Method of moments
  - and so on...
- Frequentist ML estimation finds the one value of $\theta$ 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

# 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;

# 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;

# 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; and

# 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; and
    - a computational framework for model estimation, selection, decision making and validation.

# 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; and
  - a computational framework for model estimation, selection, decision making and validation.
  - builds on likelihood inference

# Bayes' theorem - basic conditional probability

- Let's take a step back and quickly review the basic form of Bayes' theorem.

# Bayes' theorem - basic conditional probability

- 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' theorem - basic conditional probability

- 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.

# Bayes' theorem - basic conditional probability

- 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

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

# Bayes' theorem - basic conditional probability

- 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.

# Bayes' Rule more generally

1. For each $\theta \in \Theta$, specify a prior distribution $p(\theta)$ or $\pi(\theta)$, describing our beliefs about $\theta$ being the true population parameter.

# Bayes' Rule more generally

1. For each $\theta \in \Theta$, specify a prior distribution $p(\theta)$ or $\pi(\theta)$, describing our beliefs about $\theta$ 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 $\theta$. $p(y|\theta)$ proportional to the likelihood $L(\theta|y)$.

# Bayes' Rule more generally

1. For each $\theta \in \Theta$, specify a prior distribution $p(\theta)$ or $\pi(\theta)$, describing our beliefs about $\theta$ 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 $\theta$. $p(y|\theta)$ proportional to the likelihood $L(\theta|y)$.

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 $\theta$ being the true population parameter.

# Bayes' Rule more generally

1. For each $\theta \in \Theta$, specify a prior distribution $p(\theta)$ or $\pi(\theta)$, describing our beliefs about $\theta$ 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 $\theta$. $p(y|\theta)$ proportional to the likelihood $L(\theta|y)$.

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 $\theta$ being the true population parameter.

- Now, how do we get from Step 1 to 3? Bayes' rule!

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

We will use this over and over throughout the course!

# Notes on prior distributions

Many types of priors may be of interest. These may

- represent our own beliefs;

# 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

# 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.

# 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.
- and so on...

# Notes on prior distributions

- Subjective Bayes: a prior should accurately quantify some individual's beliefs about $\theta$.

# Notes on prior distributions

- Subjective Bayes: a prior should accurately quantify some individual's beliefs about $\theta$.

- Objective Bayes: the prior should be chosen to produce a procedure with "good" operating characteristics without including subjective prior knowledge.

# Notes on prior distributions

- Subjective Bayes: a prior should accurately quantify some individual's beliefs about $\theta$.

- Objective Bayes: the prior should be chosen to produce a procedure with "good" operating characteristics without including subjective prior knowledge.

- Weakly informative: prior centered in a plausible region but not overly-informative, as there is a tendency to be over confident about one's beliefs.

# Notes on prior distributions

- Subjective Bayes: a prior should accurately quantify some individual's beliefs about $\theta$.

- Objective Bayes: the prior should be chosen to produce a procedure with "good" operating characteristics without including subjective prior knowledge.

- Weakly informative: prior centered in a plausible region but not overly-informative, as there is a tendency to be over confident about one's beliefs.

-- Practical Bayes: Combines Subjective Bayes for aspects of a problem that one understands, and Objective Bayes elsewhere

# Notes on prior distributions

- The prior quantifies 'your' initial uncertainty in $\theta$ before you observe new data (new information) - this may be necessarily subjective & summarize experience in a field or prior research.

# Notes on prior distributions

- The prior quantifies 'your' initial uncertainty in $\theta$ before you observe new data (new information) - this may be necessarily subjective & summarize 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.

# Notes on prior distributions

- The prior quantifies 'your' initial uncertainty in $\theta$ before you observe new data (new information) - this may be necessarily subjective & summarize 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!)

# Notes on prior distributions

- The prior quantifies 'your' initial uncertainty in $\theta$ before you observe new data (new information) - this may be necessarily subjective & summarize 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 here: 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.