Module 0: Introduction to Bayesian Statistics (STA 360)

Professor Rebecca C. Steorts

What is Bayesian Statistics?

"Bayesian statistics is a theory in the field of statistics based on the Bayesian interpretation of probability where probability expresses a degree of belief in an event. The degree of belief may be based on prior knowledge about the event, such as the results of previous experiments, or on personal beliefs about the event. This differs from a number of other interpretations of probability, such as the frequentist interpretation that views probability as the limit of the relative frequency of an event after many trials."

-Wikipedia

Instructor

Prof. Rebecca Steorts (resteorts.github.io)

- Class: Tues/Th, 1:45 - 3:00 PM EDT (Old Chemistry 116)
- OH: Tues/Th, 3:00 - 4:00 PM EDT (Old Chemistry 216)
- Zoom ID: 925 6381 7937
- Course webpage: https://resteorts.github.io/teach/bayes21.html

Class will be in person unless you are notified! OH will be over zoom or 1-1 when possible.

Teaching Assistants

Kim Roche, PhD Student

- Lab-01: Friday: 10:15am - 11:30am EDT, Old Chemistry 101 (Back up: David)
- Meeting ID: 992 3819 1368
- Will hold OH

Lab will be in person unless notified. OH will be over zoom.

Teaching Assistants

David Buch, PhD Student

- Lab-02: Friday: 1:45 PM – 3:00 PM EDT, Old Chemistry 101 (Back up: Cathy)
- Meeting ID: 974 1981 3582
- Will hold OH

Lab will be in person unless notified. OH will be over zoom.

Teaching Assistants

Cathy Lee, PhD Student

- Responsible to majority of grading
- Will be helping with labs and answering questions on Piazza
- Will be holding office hours

OH will be over zoom.

Where to find information

- Course website (all major course information here): https://resteorts.github.io/teach/bayes21.html
- Piazza Forum:

https://piazza.com/duke

- Course syllabus https://github.com/resteorts/modern-bayes/blob/master/syllabus/syllabus-sta360-fall21.pdf

Remark: Videos on Github are too large to view, so you will need to download these if you wish to watch them. The fastest way, is to clone the repository and they will all download. They are stored this way in case your internet connection is slow.

Where to find information

- Sakai (upload homeworks): https://shib.oit.duke.edu/idp/authn/external?conversation=e1s1
- Duke Gradescope (upload homeworks): https://gradescope.com/auth/saml/duke

Other resources

- Review of probability material: https://github.com/resteorts/modern-bayes/blob/master/reading/statistical-inference.pdf
- Simon Mak's Quick Guide to Prob. Distributions https://github.com/resteorts/modern-bayes/blob/master/reading/distribution-quick-reference.pdf
- A One Pager on Prob Distributions https://github.com/resteorts/modern-bayes/blob/master/reading/common-distributions-one-pager.pdf

Course information

Course website (all major course information here):
https://resteorts.github.io/teach/bayes21.html

Course cheat sheet (a summary of course information):
https://github.com/resteorts/modern-bayes/blob/master/syllabus/deadlines-cheatsheat.pdf

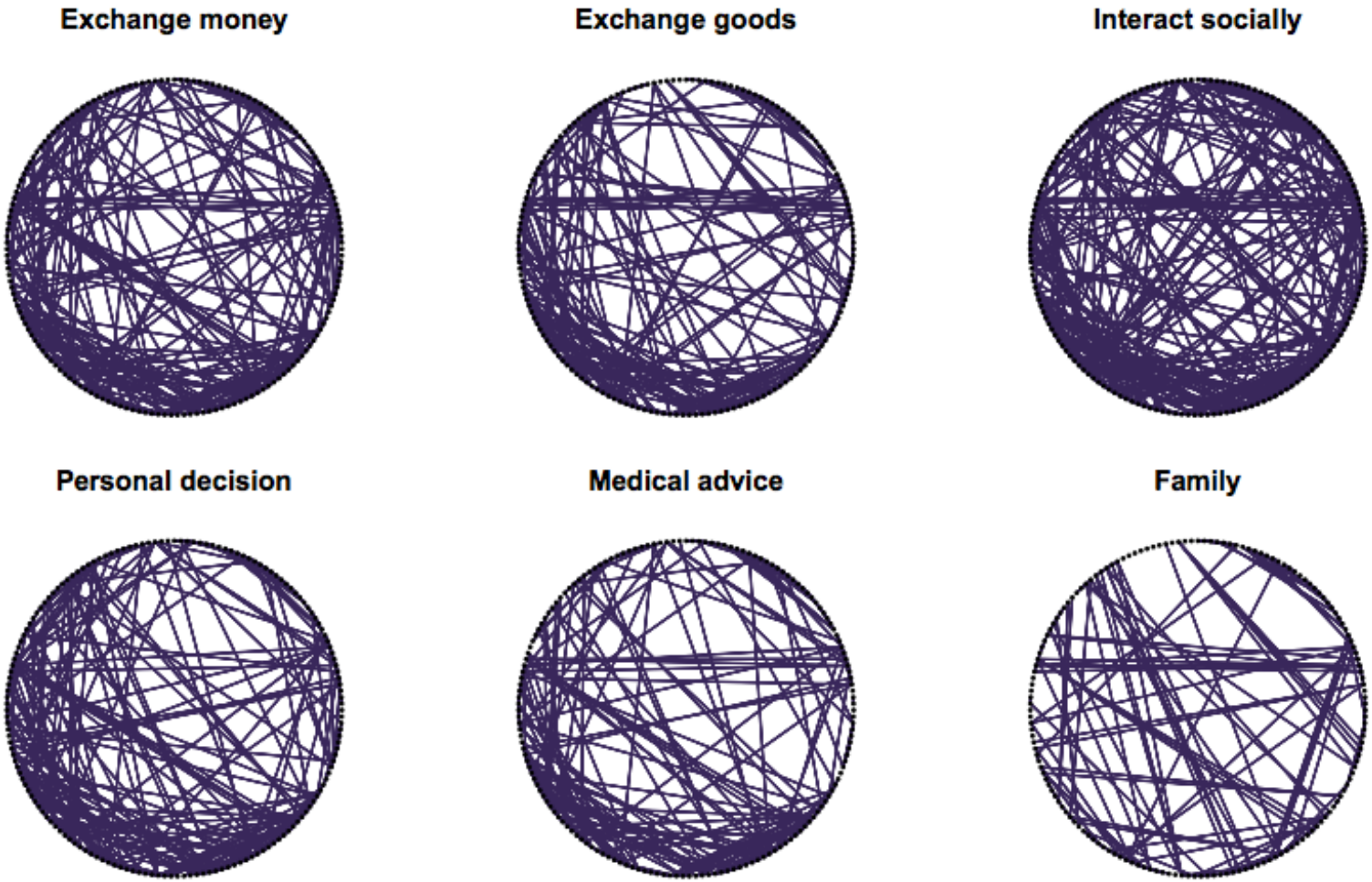
Prior Knowledge

- STA 210 https://www2.stat.duke.edu/courses/Spring19/sta210.001/
- STA 230 https://www2.stat.duke.edu/courses/Fall18/sta230/
- Linear algebra http://www.stat.columbia.edu/~fwood/Teaching/w4315/Fall2009/lecture_12
- R programming (STA 199) https://www2.stat.duke.edu/courses/Spring18/Sta199/
- github (STA 199) https://www2.stat.duke.edu/courses/Spring18/Sta199/

Course Objectives

- Provide a foundation to Bayesian statistical methods
- Explore, visualize, and analyze data in a reproducible and shareable manner using Bayesian methods from the course
- Gain experience in data wrangling and munging, building Bayesian models, visualizing them, and interpreting them
- Work on problems and case studies inspired by and based on real-world questions and data
- Learn to effectively communicate results through written assignments and exams

Social Networks



Precision Medicine

AI AND HEALTH

Editor: **Daniel B. Neill, H.J. Heinz III College,** Carnegie Mellon University, neill@cs.cmu.edu

A $3 Trillion Challenge to Computational Scientists: Transforming Healthcare Delivery

Suchi Saria, *Johns Hopkins University*

Healthcare spending in the US is nearing $3 trillion per year, but in spite of this expenditure, the US is outpaced by most developed countries in terms of health and quality of life outcomes—for example, it ranks 36th internationally in life expectancy.[1] The share of health spending in its gross domestic product has increased sharply, from 5 percent of GDP in 1960 to more than 17 percent today,[2] a rate of increase that's widely believed to be unsustainable.[3]

Policy and regulatory reform have important roles to play in addressing these challenges. Yet one of the largest underexplored avenues is the better paper records that weren't amenable to retrospective, automated analyses. The Health Information Technology for Economic and Clinical Health (HITECH) Act, a program that was part of the American Recovery and Reinvestment Act of 2009, incentivized the adoption of Electronic Health Records (EHRs) to encourage the shift from paper to digital records. That program has made more than $15.5 billion available to hospitals and healthcare professionals conditioned on their meeting certain EHR benchmarks for so-called "meaningful use." It's one of the largest investments in healthcare infrastructure ever made by the federal government.

Unstable/slow internet?

Go to https://github.com/resteorts/modern-bayes/ and fork the repository.

- The videos are compressed, and this was done on purpose for those in the class that might have slow internet. I would suggest that everyone forks the repository to avoid any issues.
- Verify that you can play them (otherwise install something on your machine).
- Make sure to pull the repository each day. I update the repository very often as all the course resources are here (homeworks, lectures, data, videos). If you don't pull often, you might run into some issues!

Class Meetings

- Learn Bayesian statistical methods
- Lectures will consist of learning methodology and applied coding techniques
- You might find it useful to have a tablet/document camera for homeworks/OH/exams
- An alternative to this is using your phone to take pictures using Evernote (students have said this worked well in the spring). Please test things out in advance and make sure things are legible!

Labs

- Labs will reinforce concepts learned in class using R
- Apply concepts from lecture to real or synthetic data
- Many labs will be finished as part of your homework
- Bring fully-charged laptop to every lab

Required Textbooks

- [A First Course in Bayesian Statistical Methods, Peter Hoff]
  - Free PDF available online through the library. Hard copy available for purchase.
  - Assigned readings on github.
- Some of Bayesian Statistics: The Essential Parts
  - Assigned readings.

Supplemental Textbooks

- Statistical Inference
- Baby Bayes using R
- Bayesian Data Analysis
- The R Cookbook
- Github Setup Video
- Github Tutorial

Activities & Assessments

- Homework: Individual assignments combining conceptual and computational skills along with lab exercises. *Lowest score dropped.*
- Labs: **Attending lab is highly encouraged**, and, there will be a lab problem on each homework each week.
- Class/Lab: **You will be responsible for keeping up with all class and lab material on a weekly basis (even though there are recordings)!**

Lab Engagement

0. Students are expected to have attempted the lab solutions before coming to lab and have watched the lab videos provided if they had trouble.
1. Students are encouraged to post lab questions in advance to Piazza regarding lab, which should be posted no later than each Thursday at 5 PM EST.
2. Each lab TA will go over all lab tasks, providing clarifications, solutions/advice, and more advanced insights.
3. Each lab TA will take questions (in person) or go over those posted Piazza.
4. You should always indicate what lab you are in to provide help to your TA.
5. Attendance is highly encouraged in order to learn tips and tricks and in order to get hints/advice which are important for homeworks and exams.

Class Engagement

0. Students are expected to have read the assigned reading before coming to class.
1. Students are expected to have watched the pre-recorded videos.
2. Prof. Steorts will go through the concepts in class again, highlighting the most important parts, providing clarifications, solutions/advice, and more advanced insights.
3. Prof. Steorts will take questions (in person) or those that are posted in advance to Piazza by Monday at 10 AM EDT and Wed at 10 AM EDT.
4. Prof. Steorts will provide exercises for the class to work through, practice exams, and an interactive environment.
5. TAs will be present to help (as able and as available) in these activities such that we can break into small groups and students can receive more individualized attention.

Homeworks

1. All code must be written to be reproducible in Markdown.
2. All derivations can be done in any format of your choosing (word, latex, markdown, written by hand) but must be converted to a pdf document. It must be legible.
3. All files must be zipped together and submitted to Sakai as one file (including Rmd and pdf). Your pdf document must be uploaded to Duke Gradescope. Please make sure to upload early to avoid issues.
4. Ask questions early if you have a problem to a TA regarding submission issues.
5. Your lowest homework will be dropped.
6. The first two homeworks can be submitted in Sakai and on GradeScope.

Remark: Sakai is for reproducible code. Gradescope is to make grading easier for everyone. Unfortunately, there is not a platform that handles both.

Please see the syllabus for all homework guidelines.

Why Sakai + Gradescope?

1. Sakai will be used for reproducibility so this can be checked. This will account for part of your grade each assignment.
2. Gradescope will be used to:

- return homework/exams more promptly
- ensure complete fairness
- allow students to continue to learn from homeworks by reworking problems without the right answer written out to the side

Exams

The format of the exams will be discussed closer to the time of the exam. At this time, plan to be for the exams in person unless there is an announcement about this.

Grade Calculation

Component	Weight
Homework	30%
Exam 1	20%
Exam 2	20%
Final Exam	30%

- See the syllabus for grade breakdowns.
- Grades will **never** be curved down.
- You are expected to attend lectures and labs in order to keep up with the course material.
- There will be no attendance grade or participation grade.

Excused Absences

- Students who miss a class due to a scheduled varsity trip, religious holiday, or short-term illness should fill out the respective form.
  - These excused absences do not excuse you from assigned work.
- If you have a personal or family emergency or chronic health condition that affects your ability to participate in class, please contact your academic dean's office.
- Exam dates cannot be changed and no make-up exams will be given.

Late Work and Regrade Requests

- No late homeworks will be accepted, so please do not ask.
- No make up exams will be given.
- Regrade requests must be submitted within one week of when the assignment was returned.

Academic Honesty

All work for this class should be done in accordance with the Duke Community Standard.

To uphold the Duke Community Standard:

- I will not lie, cheat, or steal in my academic endeavors;
- I will conduct myself honorably in all my endeavors; and
- I will act if the Standard is compromised. Any violations will automatically result in a grade of 0 on the assignment and will be reported to Office of Student Conduct for further action.

Reusing Code

- Unless explicitly stated otherwise, you may make use of online resources (e.g. StackOverflow) for coding examples on assignments. If you directly use code from an outside source (or use it as inspiration), you must or explicitly cite where you obtained the code. Any recycled code that is discovered and is not explicitly cited will be treated as plagiarism.
- On individual assignments, you may discuss the assignment with one another; however, you may not directly share code or write up with other students.
- On team assignments, you may not directly share code or write up with another team. Unauthorized sharing of the code or write up will be considered a violation for all students involved.

Where to find help

- **If you have a question during lecture or lab, feel free to ask it!** There are likely other students with the same question, so by asking you will create a learning opportunity for everyone.
- **Office Hours**: A lot of questions are most effectively answered in-person, so office hours are a valuable resource. Please use them!
- Piazza: Outside of class and office hours, any general questions about course content or assignments should be posted on the forum since there are likely other students with the same questions.

Academic Resource Center

Sometimes you may need help with the class that is beyond what can be provided by the teaching team. In that instance, I encourage you to visit the Academic Resource Center.

The Academic Resource Center (ARC) offers free services to all students during their undergraduate careers at Duke. Services include Learning Consultations, Peer Tutoring and Study Groups, ADHD/LD Coaching, Outreach Workshops, and more. Because learning is a process unique to every individual, they work with each student to discover and develop their own academic strategy for success at Duke. Contact the ARC to schedule an appointment. Undergraduates in any year, studying any discipline can benefit! Contact arc@duke.edu 919-684-5917, 211 Academic Advising Center Building, East Campus – behind Marketplace.

Technology/Other

- Make sure that you have your zoom ids organized so that you're not late for class.
- Ensure the volume on all devices is set to mute.
- Refrain from engaging in activities not related to the class discussion. Browsing the web and social media, excessive messaging, playing games, etc. is not only a distraction for you but is also a distraction for everyone around you.
- If you have a question, I don't mind if you interrupt me during class.
- If you find a **typo in the slides**, please write these down and email these to myself/Cathy so they can be fixed.

Accessibility

Please contact the Student Disability Access Office (SDAO) if there is an element of the course that is not accessible to you. There you can engage in a confidential conversation about the process for requesting reasonable accommodations.

Please note that accommodations are not provided retroactively, so please contact them as soon as possible. More information can be found online at access.duke.edu.

Inclusion

In this course, we will strive to create a learning environment that is welcoming to all students and that is in alignment with Duke's Commitment to Diversity and Inclusion. If there is any aspect of the class that is not welcoming or accessible to you, please let me know immediately.

In addition, if you are experiencing something outside of class that is affecting your performance in the course, please feel free to talk with me and/or your academic dean.

Questions

Any questions regarding the format of the semester or any concerns?

Announcements

- **Please see me if you are on the waiting list**
- **Please do not come to class if you believe that you may be sick. Please reach out an let the TA team know what is going on so that we can know what is going on and can be helpful.**
- **If your situation changes during the semester for any reason, please email myself and the TAs so that we can help you.**
- **In return, I would ask that everyone be flexible and understanding of everyone else (including instructors and TAs) as this is a very difficult and trying time for everyone.**