Lecture 01

Course and Bayesian Analysis Introduction

Jihong Zhang

1 Today's Lecture Objectives

1. Introduce myself
2. Syllabus
3. Extra Course Information
4. Introduce Bayesian Analysis

but, before we begin...

2 Introduce myself

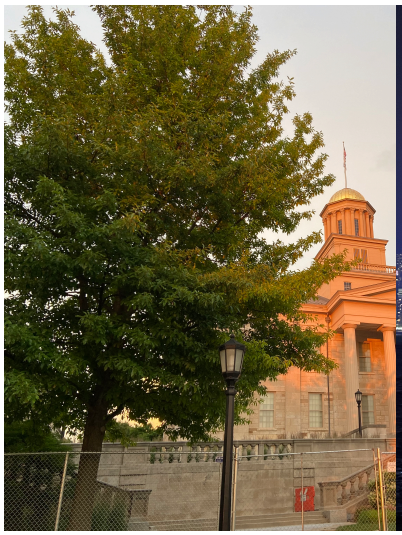
Let me introduce myself first...

3 Syllabus

Syllabus of ESRM 6553

4 Course Time

- Monday 5PM to 7:45PM:
  - 5PM to 6:15PM: First half class
  - 6:15PM to 6:30PM: 15-min break
  - 6:30PM to 7:45PM: Second half class



(a) University if Iowa

(b) Hong Kong: Victoria Harbour

(c) Chinese University of Hong Kong

5 Office Hours

- Tuesday 1:30PM to 4:30PM
- You should be able to find me in GRAD Room 109

6 Materials

- I will provide R codes and slides at the weekends before next class. You may download them on Blackboard or My website (jihongzhang.org)

7 Quiz: What is Bayesian?

1. How to Pronounce Bayesian (Real Life Examples!)
    - "B-Asian" or "Bayes-ian"?
    - What "Bayesian" mean? Assign probabilities to everything!
2. Frequentist vs Bayesian
    - Example: In U.S., more Male Asian Faculty or Female Asian Faculty?
        - Frequentist: If "Male:Female = 1:1 out of Asian Faculty" is fixed, how is the probability that the data happens?
        - Bayesian: If I believe Male:Female = 1:2 but the data says 1:1, to what degree I need to update my mind?

8 Bayesian Model Components

1. What we see: Observed Data
2. What we cannot see: Future Data, Data yet to be collected, Parameters

---

Take home Note I: Everything is random in Bayesian!

- Observed Data: Some **random** information given a unknown generation process
- Parameter: The **random** components controlling the generation process

---

Take home Note II: Every random component can be expressed as probability!

- The probability of observed data given parameters: p(Observed Data|Parameters) = **Likelihood**
- The probability of parameters: p(Parameters) = **Prior Distribution**
- The probability of parameters given observed data: p(Parameters|Observed Data) = **Posterior Distribution***

8.1 Bayesian Thinking Process

8.1.1 A toy example:

I originally thought the ratio of male Asian faculty to female is about 1:2 (**Prior**). But the data we sampled from 2000 doctoral students suggested gender ratio is 1:1 (**Data; Likelihood that the true ratio we don't know**). Based on Bayes's rule, estimate for the ratio is probably 1:1.5 (**Posterior**) after combining these two statements.

9 Bayesian Analysis: Why It Is Used?

There are at least four main reasons why people use Bayesian Analysis:

1. Missing data
    - Multiple imputation
    - More complicated model for certain types of missing data
2. Lack of software capable of handing large sized analyses
    - Have a zero-inflated Poisson model with 1000 observations and 1000 parameters? No problem in Bayesian!

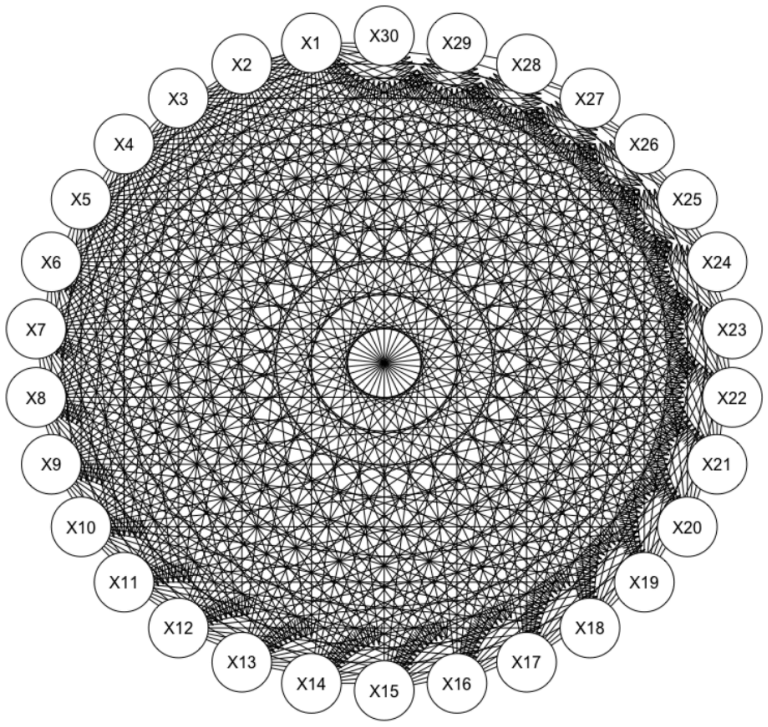
10 Bayesian Analysis: Why It Is Used? (Cont.)

3. New complex models not available in frequentist framework
    - Have a new model? (A model that estimates the probability students choose the right answers then choose the wrong answers in a multiple choice test?)
4. Enjoy the Bayesian thinking process

- It is a way of thinking that everything is random and everything can be expressed as probability. It is a way of thinking that we can update our belief as we collect more data. It is a way of thinking that we can use our prior knowledge to help us understand the data.

11 Bayesian Analysis: Why It Is Used? (Cont.)

Figure 2: A directed acyclic graph of the BayesNet model for 30-item test.



12 Bayesian Analysis: Issues

1. Subjective vs. Objective
    - Prior distribution is subjective. It is based on your prior knowledge.

- However, 1) Scientific judgement is always subjective 2) you can use objective prior distribution to avoid this issue.

2. Computationally Intensive
    - It is not a problem anymore. We have computers.
    - But we still need weeks or months to get results for some complated model and big data
3. Difficult to understand

13 Bayesian Analysis is popular

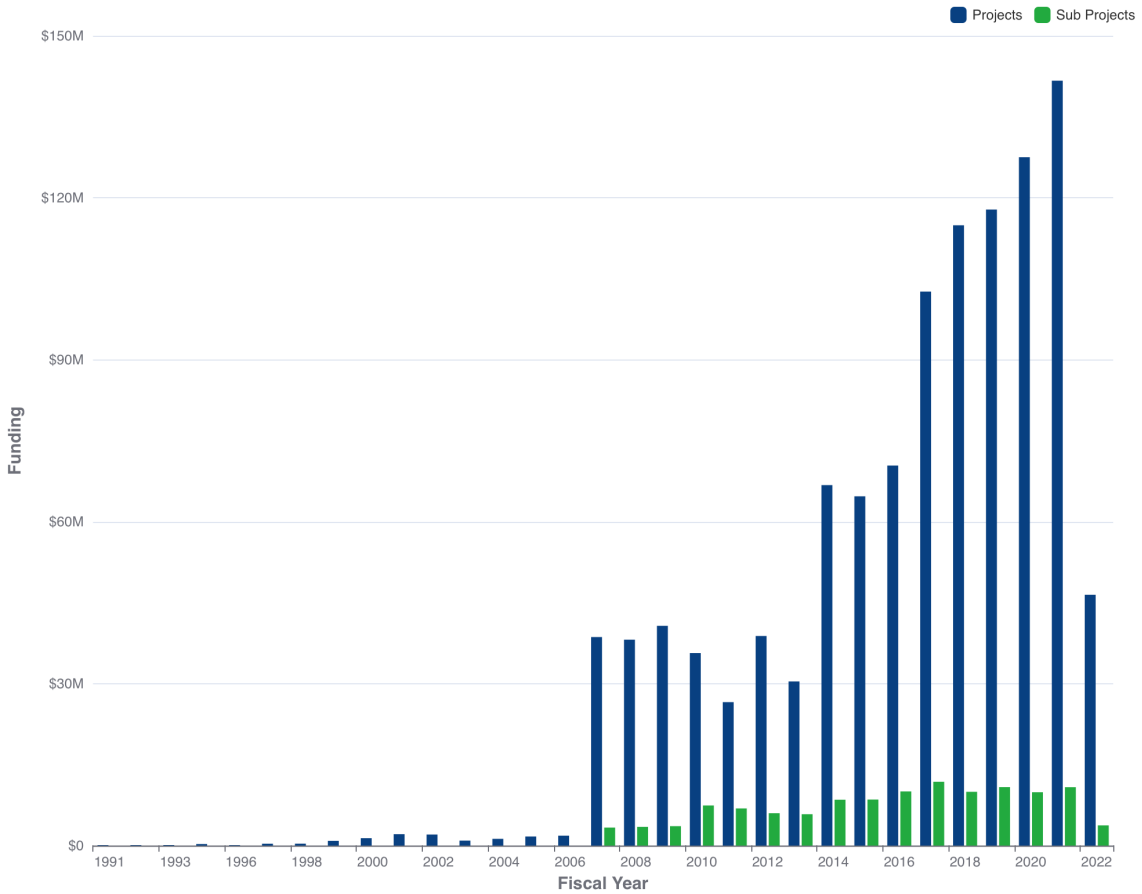


Figure 2: Funding available only for NIH, CDC, FDA, AHRQ, and ACF 2020 Spring. Source: https://report.nih.gov/

14 What topics Bayesian Analysis can cover?

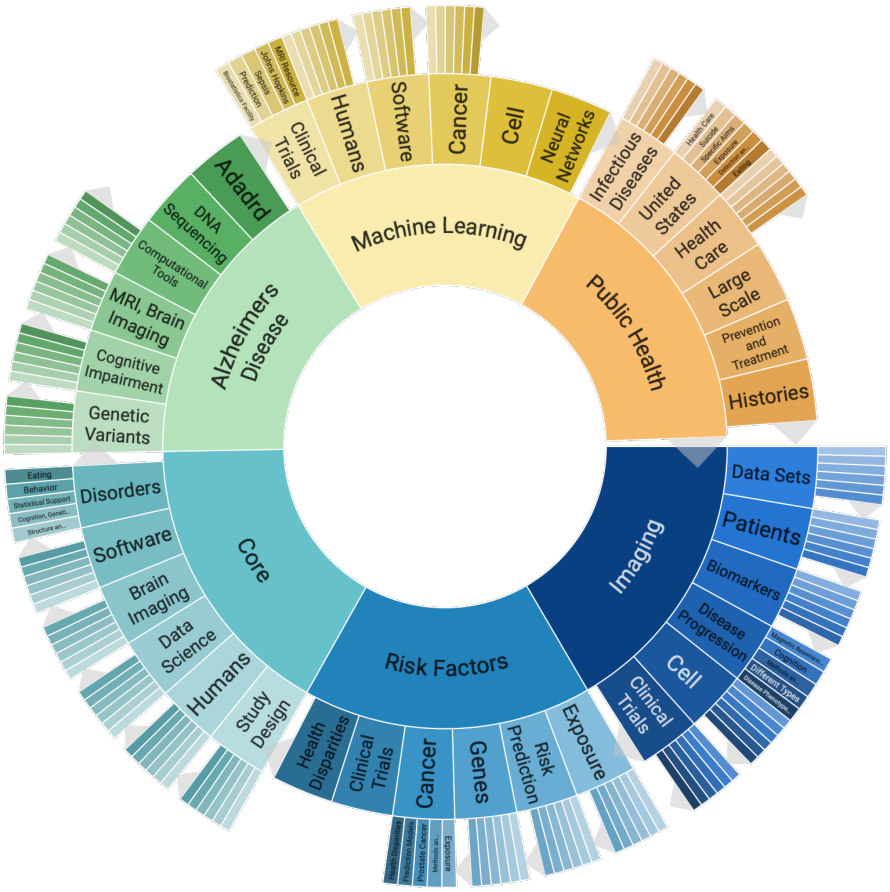


Figure 3: Funding available only for NIH, CDC, FDA, AHRQ, and ACF 2020 Spring. Source: https://report.nih.gov/

15 Next Class

We will talk about how Bayesian methods works in a little bit more technical way.

16 Suggestions

Your opinions are very important to me. Feel free to let me know if you have any suggestions on the course.