Course overview and introduction to Bayesian inference

Dr. Olanrewaju Michael Akande

Jan 10, 2020

Welcome to STA 602L!

What is this course about?

Learn the foundations of Bayesian inference.

Work through the theory of several Bayesian models.

Use Bayesian models to answer inferential questions.

Apply the models to several different problem sets.

"Prior $\rightarrow$ likelihood $\rightarrow$ posterior" over and over again!

We will follow the Hoff book closely – roughly one chapter per week.

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

Instructor

Dr. Olanrewaju Michael Akande

- olanrewaju.akande@duke.edu
- akandelanre.github.io.
- 256 Gross Hall
- Wed 9:00 - 10:00am; Thur 11:45 - 12:45pm (still subject to change!)

Lead TA

Jordan Bryan (mainly for STA 360)

jordan.bryan@duke.edu

TAs

Bai Li

bai.li@duke.edu
Wed 3:00 - 5:00pm
Old Chem 025

Zhuoqun (Carol) Wang

zhuoqun.wang@duke.edu
Tues 3:00 - 5:00pm
Old Chem 025

FAQs

All materials and information will be posted on the course webpage:

https://sta-602l-s20.github.io/Course-Website/

- How much theory will this class cover? A lot! Make sure you are especially comfortable working with probability distributions.
- Am I prepared to take this course? Yes, if you are familiar with the topics covered in the course prerequisites.
- Will we be doing "very heavy" computing? Not too heavy but yes, a good amount. You will be expected to be able to write your own MCMC sampler later on.
- 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://sta-602l-s20.github.io/Course-Website/resources/.

Course structure and policies

Course structure and policies

- All on the website (here: https://sta-602l-s20.github.io/Course-Website/policies/)
- Make use of the teaching team's office hours, we're here to help!
- Do not hesitate to come to my office during office hours or by 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.
- Please refrain from texting or using your computer for anything other than coursework during class.

Other details

- What topics will we cover? Refer to Section 9 of the syllabus (here: https://sta-602l-s20.github.io/Course-Website/syllabus_pdf/Syllabus.pdf).
- If you are auditing this course, remember to complete the audit form for the graduate school.
- Confirm that you have access to Sakai and Gradescope.

Your turn!

Introductions

- Your full name.
- The name you prefer to go by.
- One goal you hope this course would help you achieve.

Introduction to Bayesian inference

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, and validation.

Building blocks 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;
  - Θ be the parameter space; and
  - θ be the parameter of interest.
- More to come later.

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.

Building blocks of Bayesian inference

- Now, to a slightly more complicated version of Bayes' rule. First,

  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. $p(y|\theta)$ gets us the likelihood $L(y;\theta)$

  3. After observing the data y, for each $\theta \in \Theta$, update the prior distribution to a posterior distribution $p(\theta|y)$, describing our "updated" belief about θ being the true population parameter.

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

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

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;
- 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 θ.
- 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.

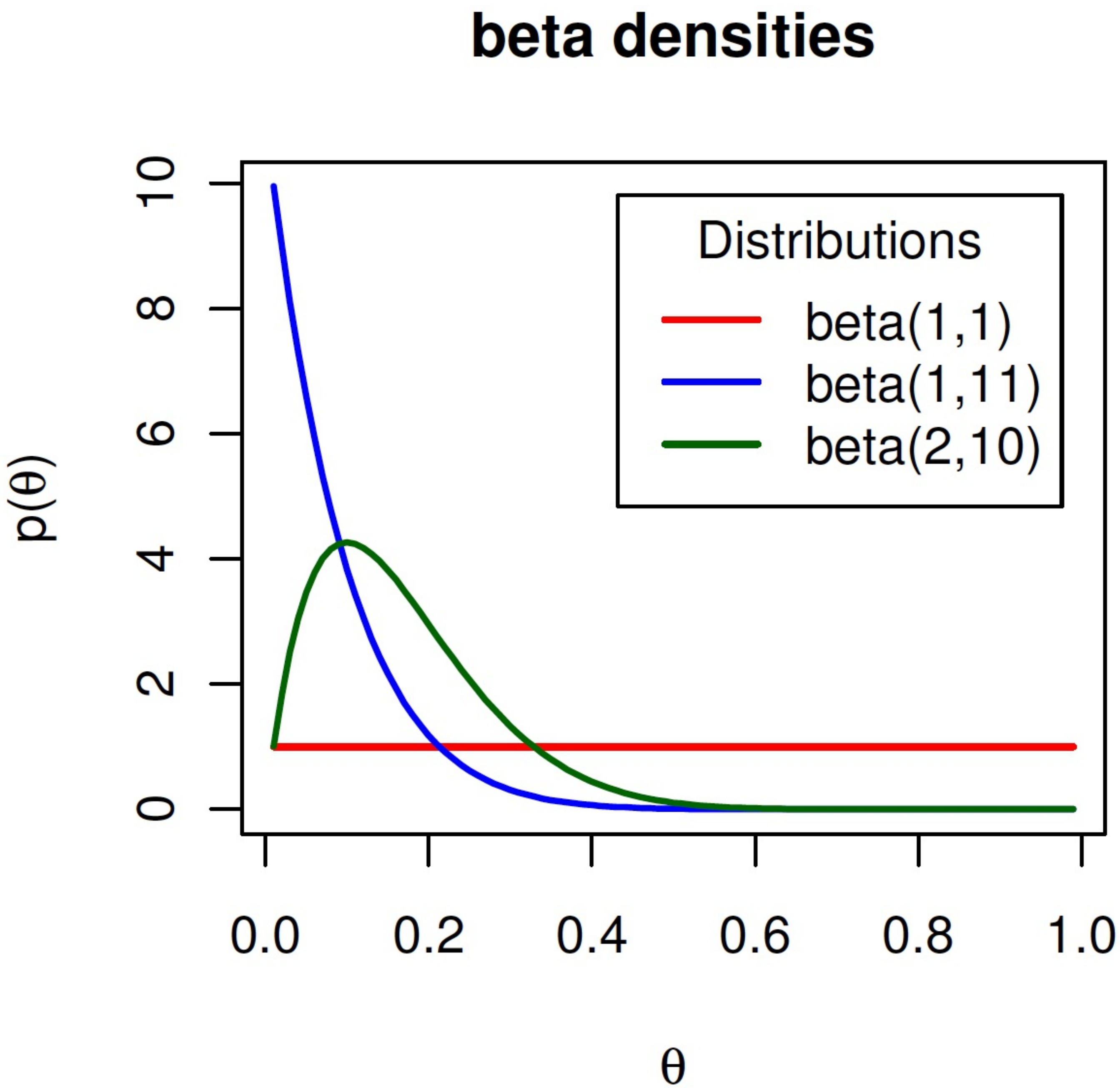
Notes on prior distributions

- The prior quantifies your initial uncertainty in θ 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.
- One (very important) role of the prior is to stabilize estimates in the presence of limited data.

Simple example - estimating a population proportion

- Suppose $\theta \in (0, 1)$ is the population proportion of individuals with diabetes in the US.
- A prior distribution for θ would correspond to some distribution that distributes probability across $(0, 1)$.
- A very precise prior corresponding to abundant prior knowledge would be concentrated tightly in a small sub-interval of $(0, 1)$.
- A vague prior may be distributed widely across $(0, 1)$ - e.g., a uniform distribution would be the common choice here.

Some possible prior densities



Beta prior densities

- These three priors correspond to Beta(1,1) [also, Unif(0,1)], Beta(1,11) and Beta(2,10) densities.
- Beta(a,b) is a probability density function (pdf) on (0,1),

$$\pi(\theta) = \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1},$$

  where $B(a,b)$ = beta function = normalizing constant ensuring the kernel integrates to one. Note: some texts write $\text{beta}(\alpha,\beta)$ instead.
- The beta(a,b) distribution has expectation $\mathbb{E} = a/(a+b)$ and the density becomes more and more concentrated as $a+b$ = prior "sample size" increases.
- The variance is $ab/[(a+b)^2(a+b+1)]$.
- We will look more carefully into the beta-binomial model next week but for now, I'll illustrate how this prior gets updated as data becomes available.