GOV 2017: Applied Bayesian Statistics for the Social Sciences

Kosuke Imai

Spring 2023

Preliminary syllabus

Abstract

This course introduces social science students to applied Bayesian statistics. We will begin by introducing Bayes’ rule, which allows us to learn from data in an intuitive and coherent way. We then cover a set of simple probabilistic models as well as powerful computational tools that will be used for the remainder of the course. Finally, we will learn about social science applications of Bayesian models including regression models, topic models, and social network models with an emphasis on foundational models. The course will build everything up from the basic principles, only requiring the knowledge of basic probability, statistics, and regression modeling along with the familiarity with R programming. The ultimate goal of this course is to teach fundamentals of Bayesian statistics that allow students to understand, implement, and even develop cutting-edge Bayesian models on their own.

1 Contact Information

Instructor
Kosuke Imai
Office: CGIS K306
Email: imai@harvard.edu
URL: https://imai.fas.harvard.edu/
Office Hours: Mondays 1:30pm – 3:00pm (sign up at http://bit.ly/ImaiOfficeHours or reach out to me via Slack for an appointment)

Teaching Fellow
Sooahn Shin
Email: sooahnshin@g.harvard.edu
URL: http://sooahnshin.com
Office Hours: Wednesdays 4:00 – 6:00pm

2 Logistics

• Class meetings: Mondays, 9:45am–11:45am, CGIS K354 (except for Mar 20 at CGIS K031)
• Section meetings: Thursdays, 5:00pm – 6:15pm, TBD

3 Prerequisites

Students should have taken Gov 2001 and Gov 2002 or the equivalent courses in basic probability, statistics, data analysis, and regression modeling (e.g., Stat 110, Stat 111, and Stat 139).

4 Questions, Announcements, and Submissions

• The Canvas site for this course is at https://canvas.harvard.edu/courses/117655.
• We will use Gradescope at https://www.gradescope.com/courses/477983 for the submission of all assignments including the review questions and class exercises. There is a student guide you can check for any questions about the workflow.

• Rather than email, please use Ed Discussion at https://edstem.org/us/courses/31608 (the link is also available at Canvas) when asking questions about lectures, problem sets, and other course materials. This allows all students to benefit from the discussion and to help each other understand the materials. Both students and instructors are encouraged to participate in discussions and answer any questions that are posted. You may find this user guide helpful to orient yourself to the platform.

• Please feel free to use the Slack workspace for this course at https://gov-2017-s23-hpj.slack.com for any other communication with instructors and other students. You can join the Slack workspace from Canvas.

• A Google calendar that contains the information about the course logistics is available at this link.

5 Class Requirements

For each requirement, no late submission is allowed without a prior approval of the instructors.

• **Review questions (20% of final grade)** For the first three modules, you will work on one review question based on the reading assignment. You will be required to submit your answers by the beginning of the class meeting of the corresponding week. We will use the two best scores for each module to compute the final grade for the review questions. Although this means that you do not need to work on review questions for every week, we encourage you to try them in order to keep up with the course materials.

• **Exercises (30% of final grade)** For the first three modules, you will work on a couple of exercise questions. You will begin working on these exercise questions after the lecture during each class meeting. You will be required to submit your completed answers by the beginning of the section meeting of the corresponding week. We will use the two best scores for each module to compute the final grade for the exercises. Although this means that you do not need to work on exercises for every week, we encourage you to try them in order to keep up with the course materials.

• **Collaboration policy:** You may collaborate with other classmates and receive help from course staff on review questions and class exercises. **However, you must not copy anybody else’s code or answers and are required to submit your own answer. Please specify the names of your collaborators at the beginning of your answer.**

• **Final project (40% of final grade):** The final project will be completed in collaboration with another student in the class. All projects must use a Bayesian methodology. Ideal projects will either (a) apply an existing technique to answer a substantive question or (b) extend such techniques in useful ways. Students are encouraged to consult the instructor and TF throughout the semester. Students should demonstrate that they can understand and implement a new Bayesian model and apply it to a data set of interest. Please note that we are not expecting you to fully execute an original research project! Rather, think of this as an opportunity to learn and apply a cutting-edge Bayesian model to a dataset of your choice. To help keep you on track, there will be multiple deliverables throughout the semester.

  – **February 20 (Project and collaborator identification)** By this date, pairs should submit a one-page project proposal with a brief statement of the problem to be solved or the question to be answered. Before the spring break, each pair should meet with the instructor to discuss the basic plan for the project and the methodological paper for their April presentation (see below).

  – **March 20 (First deliverable)** By this date, pairs must submit a first deliverable (maximum of 3 pages) including a concise problem statement, the short presentation of a descriptive analysis of data, and a brief explanation of possible Bayesian modeling strategy to be used. Students are
encouraged to meet with the instructor to discuss the direction of their project before their class presentation.

- **April (Class presentation)** Each pair (or a group of pairs) will present a cutting-edge Bayesian model that forms the basis of their analysis. The presentation should consist of the detailed explanation of the model (15 minutes) and the results of their own implementation (5 minutes). The presentation will be followed by short Q&A. Everyone should have read the papers so that they also learn about the new Bayesian model and participate in the discussion.

- **April 26 (Preliminary result)** By this date (the final day of classes), pairs must submit a PDF slide-deck (maximum of 10 slides) with the preliminary results of your analysis. Each student will be asked to comment on the slide-deck of another student.

- **May 12 (Final project report)** By this date (last day of the exam period), pairs must submit the final report (no longer than 15 doublespaced pages) focusing on methods and results.

Although these are fixed milestones, we are happy to supervise your project throughout the semester and beyond!

- **Participation (10% of final grade):** We will assess your overall engagement with the course materials throughout the semester (class and section meetings, online discussions).

### 6 Textbook


### 7 Class Schedule

The course consists of four modules. In the first three modules, students will learn the foundation of Bayesian statistics through reading assignments, review questions, lectures, and class exercises. In the final module, students will use this foundation to learn new Bayesian models on their own. Through the final project, students will learn how to understand and implement new Bayesian models for empirical analysis.

<table>
<thead>
<tr>
<th>Modules</th>
<th>Dates</th>
<th>Week</th>
<th>Topic</th>
<th>Readings</th>
<th>Review questions</th>
<th>Project milestone</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bayesian inference</td>
<td>1/23</td>
<td>0</td>
<td>Bayes’ rule</td>
<td>BDA 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1/30</td>
<td>1</td>
<td>Single parameter models</td>
<td>BDA 2</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2/6</td>
<td>2</td>
<td>Multi-parameter models</td>
<td>BDA 3, 4</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2/13</td>
<td>3</td>
<td>Hierarchical models</td>
<td>BDA 5, 6</td>
<td>5.7</td>
<td>Proposal due on 2/20</td>
</tr>
<tr>
<td>2. Bayesian computation</td>
<td>2/27</td>
<td>4</td>
<td>Posterior simulation I</td>
<td>BDA 10, 11</td>
<td>10.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3/6</td>
<td>5</td>
<td>Posterior simulation II</td>
<td>BDA 12</td>
<td>12.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3/20</td>
<td>6</td>
<td>EM Algorithm</td>
<td>BDA 13</td>
<td>13.9, 13.10</td>
<td>First deliverable due on 3/20</td>
</tr>
<tr>
<td>3. Bayesian regression</td>
<td>3/27</td>
<td>7</td>
<td>Regression models</td>
<td>BDA 14</td>
<td>14.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4/3</td>
<td>8</td>
<td>Hierarchical regressions</td>
<td>BDA 15</td>
<td>15.6</td>
<td></td>
</tr>
<tr>
<td>4. Bayesian modeling</td>
<td>4/10</td>
<td>9</td>
<td>Student presentations</td>
<td>TBD</td>
<td></td>
<td>Preliminary results due on 4/26</td>
</tr>
<tr>
<td></td>
<td>4/17</td>
<td>10</td>
<td>Student presentations</td>
<td>TBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4/24</td>
<td>11</td>
<td>Student presentations</td>
<td>TBD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Some Potential Topics and Papers for the Final Projects

Below we list some potential topics and papers for the final projects and student presentation. Of course, students may choose other topics and papers in consultation with the instructor.

- **Ideal point estimation**
  
  
  
  
  
  

- **Topic models**
  
  
  
  
  

- **Network models**
  
  
  
  
  
• Causal inference


• Nonparametrics


