Welcome to DSCI 553: Statistical Inference and Computation II

Contents

This course will introduce Bayesian reasoning for Data Science on an inferential and predictive framework while eventually emphasizing regression analysis. We will learn how to formulate and implement inference using the prior-to-posterior paradigm.

High-Level Goals

By the end of the course, students are expected to:

- Use Bayesian reasoning when modelling data.
- Apply Bayesian statistics to regression models.
- Compare and contrast Bayesian and frequentist methods, and evaluate their relative strengths.
- Use appropriate statistical libraries and packages for performing Bayesian inference.

Teaching Team

Position	Name	Slack Handle	GHE Handle	Section
Lecture/Lab Instructor	Hedayat (Heddy) Zarkoob	`@Heddy`	`@hzarkoob`	1
Lecture/Lab Instructor	Alexi Rodríguez-Arelis	`@Alexi`	`@alexrod6`	2
Teaching Assistant	Hasti Jalali	`@Hasti (TA)`	`@hastij`	1
Teaching Assistant	Tony Liang	`@Tony Liang (TA)`	`@tliang19`	1
Teaching Assistant	Atabak Eghbal	`@Atabak`	`@ateghbal`	2
Teaching Assistant	Haley Oleynik	`@Haley Oleynik (TA)`	`@holeynik`	2
Teaching Assistant	Sky Sheng	`@Sky Sheng (TA)`	`@skysheng`	2

Lecture Schedule

This course occurs during **Block 5** in the 2024/25 school year. Course notes can be accessed here. **Typically, you should review these notes before each lecture.** There are also optional textbook readings we suggest reviewing before each lecture (if possible).

Lecture	Topic	Optional Reading Material
1	**Frequentist and Bayesian Overview, Probabilistic Generative Models, and `Stan`**	• Bayes Rules! Chapter 1: The Big (Bayesian) Picture • Getting started with `rstan`
2	**Conditional Probabilities, Bayes' Rule, and Maximum a Posteriori Estimation**	• Bayes Rules! 2.1 Building a Bayesian model for events • Bayes Rules! 2.2 Example: Pop vs soda vs coke
3	**Bayesian Statistics in Action: The Beta-Binomial Model**	• Bayes Rules! Chapter 3: The Beta-Binomial Bayesian Model • Bayes Rules! Chapter 8: Posterior Inference & Prediction (Introduction) • Bayes Rules! 8.1 Posterior estimation

Lecture	Topic	Optional Reading Material
4	**Markov Chain Monte Carlo, `Stan`, and Complex Bayesian Models**	• Bayes Rules! Chapter 6: Approximating the Posterior (Introduction) • Bayes Rules! 6.2 Markov chains via `rstan` • Bayes Rules! 6.2.1 A Beta-Binomial example • Bayes Rules! Chapter 7: MCMC under the Hood (Introduction) • Bayes Rules! 7.1 The big idea • Bayes Rules! 7.2 The Metropolis-Hastings algorithm • Bayes Rules! 7.6 Why the algorithm works
5	**Bayesian Normal Linear Regression and Hypothesis Testing**	• Bayes Rules! 8.2 Posterior hypothesis testing • Bayes Rules! Chapter 9: Simple Normal Regression (Sections 9.1 to 9.5)
6	**Bayesian Binary Logistic Regression**	• Bayes Rules! Chapter 13: Logistic Regression (Sections 13.1 to 13.3)
7	**Bayesian Hierarchical Models**	• Bayes Rules! Chapter 15: Hierarchical Models are Exciting
8	**More Hierarchical Modelling and MCMC Diagnostics**	• Bayes Rules! Chapter 6: Approximating the Posterior (Section 6.3)

See the lecture learning objectives for a detailed breakdown of lecture-by-lecture learning objectives.

Deliverables

This is an **assignment-based course**. The following deliverables will determine your course grade:

Assessment	Weight
Lab Assignment 1	12%
Lab Assignment 2	12%
Lab Assignment 3	12%
Lab Assignment 4	12%
Quiz 1	25%
Quiz 2	25%
Lecture Attendance	2%

> **Note:** A +1% final bonus mark will be granted to everybody if the class reaches a 50% response rate (or above) in the final teaching evaluations.

Lectures

Schedule

Refer to the MDS calendar.

Labs

Schedule

Refer to the MDS calendar.

Lab Topics and Due Dates

	Lab Topic	Due Date
1	Introduction to `Stan` and Bayes' Rule (Lectures 1 and 2)	Refer to the MDS calendar
2	Foundations of Bayesian Inference (Lectures 3 and 4)	Refer to the MDS calendar
3	Bayesian Regression (Lectures 5 and 6)	Refer to the MDS calendar
4	Hierarchical Models and Model Diagnostics (Lectures 7 and 8)	Refer to the MDS calendar

Lab Grade Computation

Once lab grades are published on Gradescope, you will see your **raw lab mark** m. This **raw lab mark** m is the grand total of your granted marks throughout the whole lab assignment. Now, if we add up **all the marks (non-challenging and challenging)** in the handout corresponding to all `rubric={...}`, this sum is what we call the maximum raw lab mark m_{100} to get 100% as a percentage lab grade. On the other hand, if we add up **the non-challenging marks** in the handout found in `rubric={...}`, this sum is what we call the raw lab mark m_{95} to get a 95% as a percentage lab grade.

By the end of the block, **once all lab marking is finished on Gradescope**, your raw lab grades will be transferred to **Canvas**. Then, in your **Canvas gradebook**, you will see these raw lab grades (`raw lab1`, `raw lab2`, `raw lab3`, and `raw lab4`). Finally, for each of the four

labs, you will also see your final lab grades (`lab1`, `lab1`, `lab3`, and `lab4`). Let g be the final lab grade of a specific lab **as a percentage**; it will be computed as follows:

- If $m > m_{95}$, then $g = 95 + \left(\frac{m - m_{95}}{m_{100} - m_{95}} \times 5\right)$.
- If $m \leq m_{95}$, then $g = \left(\frac{m}{m_{95}}\right) \times 95$.

Quizzes

Refer to the MDS calendar.

Office hours

Refer to the MDS calendar.

Communication

We will use **Slack** as the main communication channel.

If you have any questions regarding the course content, lectures, labs, autograders, or any other course-related matters, we kindly request that you avoid direct messaging (DM) the instructor or TAs. Instead, please post your question on the DSCI 562 channel. This approach not only enables our TAs to respond promptly but also benefits other students who might have similar questions.

> **Response time:** We will try our best to reply to your inquiries as soon as possible during the normal working hours (9AM-5PM Mon-Fri). If you send us a message outside of regular working hours, please expect a response on the next working day.

Use of LLMs

LLMs, such as ChatGPT, can be helpful tools if we use them responsibly. In this course, students are permitted to use these tools to gather more information, review concepts, or brainstorm, and students must cite these tools if they use them for assignment. Having said all

this, it is **not** permitted to write any given assignment via copying and pasting AI-generated responses.

Installation

In this course, we will be using `Stan` as our inference engine along with the `R` package `rstan`. If you did not installed the software last term, **follow the installation instructions here**.

If you have installation troubles, please seek our help as soon as possible! **You can also use the #installation channel on Slack.**

Reference Material

Course textbook:

- Bayes Rules! An Introduction to Bayesian Modeling with R

Other books:

- ThinkBayes (Python)
- Doing Bayesian Data Analysis in brms and the tidyverse (R)
- Probabilistic Programming and Bayesian Methods for Hackers (Python)
- Introduction to Empirical Bayes: Examples from Baseball Statistics (R)
- Statistical Rethinking: A Bayesian Course with Examples in R and Stan (R)

Blog posts / other:

- Quora: For a non-expert, what is the difference between Bayesian and frequentist approaches?
- Probability concepts explained: Bayesian inference for parameter estimation
- MLE and MAP video from Mike Gelbart's CPSC 340.
- Web apps for visualizing probability distributions: one, another

- How Statisticians Found Air France Flight 447 Two Years After It Crashed Into Atlantic

Recommended Course Reviews

This course is taught in `R` (we will follow the `tidyverse` style guide) and `Stan` with a reasonable mathematical, statistical, and programming basis. We strongly recommend reviewing the following courses:

- *DSCI 551: Descriptive Statistics and Probability for Data Science*, for basic statistical and probabilistic concepts, and familiarity with the mathematical notation.
- *DSCI 552: Statistical Inference and Computation I*, for statistical inference concepts with a frequentist approach.
- *DSCI 561: Regression I*, for ordinary ordinary least-squares (OLS).
- *DSCI 562: Regression II*, for generalized linear models (GLMs).
- *DSCI 531: Data Visualization I*, for plotting tools using the package `ggplot2`.

Policies

See the general MDS policies.

Attribution

The course is built upon previous years' materials developed by previous instructors.

License