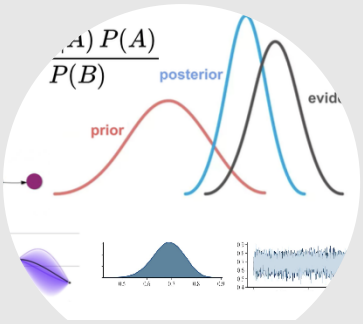


# Bayesian Reasoning in Data Science

## DATA 340 05

## Instructor Info

Cristiano Fanelli

Office Hrs: Fri 12:30-14:30p

ISC 1265

https://cristianofanelli.com

https://cfteach.github.io/brds

cfanelli@wm.edu

## Course Info

Prereq: ( CSCI 140 OR CSCI 141 OR DATA 141 OR C141 1 ) AND ( DATA 146 OR DATA 201 OR CSCI 146 ) AND ( MATH 351 OR ECON 307 )

Tues & Thurs

11a-12.20p

Boswell Hall–Room:201

## Overview

No data scientist can work without a solid grasp of conditional probability and Bayesian reasoning. Bayes'theorem allows to update our beliefs based on the occurrence of new events, steering the inference towards the truth and assessing uncertainty in predictions. This course provides an introduction to Bayesian reasoning in data science and will let you appreciate the basic building blocks of this approach through real-world examples across different areas. During the course you will learn concrete computational implementations, that will help students connect what they have read and heard with what they can program, reinforcing the material.

## Material

### Used (and recommended) texts[1]

- G. D'Agostini [GDA], *Bayesian Reasoning in Data Analysis, A Critical Introduction*, World Scientific
- O. Martin [OM], *Bayesian Analysis with Python*, Packt, Second Edition

### Other useful texts

- C. Davidson-Pilon [CDP], *Bayesian Methods for Hackers*, Addison Wesley Data & Analytics Series
- D. MacKay [DMK], Information Theory, Ingerence, and Learning Algorithms, http://www.inference.org.uk/itprnn/book.pdf
- A. B. Downey [ABD], *Think Bayes*, Second Edition, O'Reilly

### Other

Other resources will be linked in the jupyter book.

## Grading Scheme [2]

| | |
|---|---|
| 60% | Homework: 3 Assignments (lowest will be dropped - average of 2) |
| 30% | Midterm test 15%, Final project 15% |
| 10% | Participation (class or remote) |

Grades will follow the standard scale:
$\geq$ 93.00 (A); [90.00,93.00) (A-); [87.00,90.00) (B+); [83.00,87.00) (B); [80.00,83.00) (B-); [77.00,80.00) (C+); [73.00,77.00) (C); [70.00,73.00) (C-); [67.00,70.00) (D+); [63.00,67.00) (D); [60.00,63.00) (D-); $<$60.00 (F)
Curving is at the discretion of the professor.

## Midterm and Final project

Midterm: will consist of questions that would test your understanding of the concepts discussed so far in the class (due date TBD, grading period is from Oct 10-30).
Final project: will test your understanding of all the concepts discussed during the course. This project can be completed outside of class and due no later than the time assigned for our final exam by the registrar.

## Learning Objectives

- Understand the building blocks of Bayesian Reasoning and distinguish probability of causes from probability of effects
- Know modern Bayesian approaches to real-world problems
- Gain skills on probabilistic programming

[1] During the course, lectures and computational implementations will be provided as jupyter book at https://cfteach.github.io/brds

[2] Five bonus points on the homework can be obtained with a mini-project done before Thanksgiving week (due Nov 20)

# FAQs

**Can the weekly content change?**

Yes, it depends on the progress of the class. It will keep being updated on https://cristianofanelli.com

**What is a mini project?**

It's a project that should not take much of your time (a few days of work), however you can still learn very interesting real-world applications. Ask your instructor if you are interested.

## Diversity and Inclusivity Statement

This class is a place where you will be treated with respect, and welcomes individuals of all ages, backgrounds, beliefs, ethnicities, genders, gender identities, gender expressions, national origins, religious affiliations, sexual orientations, ability - and other visible and non-visible differences. All members of this class are expected to contribute to a respectful, welcoming and inclusive environment for every other member of the class.

## Accommodations

It is the policy of William & Mary to accommodate students with disabilities and qualifying diagnosed conditions in accordance with federal and state laws. Any student who feels s/he may need an accommodation based on the impact of a learning, psychiatric, physical or chronic health diagnosis should be referred to Student Accessibility Services staff at 757- 221-2509 or at sas@wm.edu. SAS staff will work with you to determine if accommodations are warranted, and if so, to help you obtain an official letter of accommodation

## Academic Integrity

Academic dishonesty is taken very seriously. Please make sure to provide citations in your work and do not turn in work that has not been done by you. Cases of academic dishonesty will be evaluated and acted upon in accordance with W&M policies. Collaboration is allowed and encouraged (*e.g.*, during lectures you may be asked to form groups with other peers to study and discuss topics covered in class). You cannot copy /paste someone else's code or answers to questions. Please learn and act in good faith at all times! If you are in doubt, you can ask me for clarification.

**Resources** For programming during this course you can use your laptop and Google COLAB; in general; if you have any issues with computing resources reach out to the instructor. For discussion and class participation we can also use Piazza (starting from the second week). Other standard platforms like Blackboard and Gradescope could be also used.

**Guidance** Students performing at a C level are required to schedule a meeting with the instructor to discuss class performance.

**Important Dates** September 12: add/drop; October 10-30: midgrade; Dec 16, Fri: Final due date.

**Expectations** You are recommended to bring your laptop to class. We will have a significant amount of active coding and discussion.

COVID: W&M is an academic community based on faculty and students convening. We will always follow W&M requirements. Testing positive for COVID is not considered a disability and is not under the purview of W&M's SAS. At present, day-to-day emergency management of COVID-19 is no longer necessary at W&M. Our course is scheduled for in-person instruction:

- ReportCOVID is no longer required, and W&M will not be operating case management or contact tracing this year.
- Students who test positive should isolate at home or off campus if that is feasible. The university does not coordinate dedicated COVID-19 isolation housing.
- Residential students who are unable to isolate off campus should plan to isolate in their residence hall following these guidelines. W&M is offering student guides and resources online.

Make-up exams will only be allowed for students who have a substantiated excuse approved by the instructor *before the due date*. Any questions should be addressed via email to the instructor. One make-up assignment is allowed for justified student absences.

# Class Schedule

## MODULE 1: Bayesian Thinking

| | | |
|---|---|---|
| Week 1<br>(8/29-9/4) | Thu, 9/1 Course Introduction<br>• course structure and grading<br>• by the end of the course: real-world<br>• causes and effects, true and measured<br>• uncertainties<br>~~Assignment: NO~~ | Lec. 1 (references therein, *e.g.*, [GDA Ch. 1 and 2]) |
| Week 2<br>(9/5-9/11) | Tue, 9/6 Bayes Theorem<br>• Derivation<br>• Prior, Likelihood, Posteriors<br>• Examples, exercises in class<br>Thu, 9/8 Thinking Probabilistically<br>Assignment: module1 - part1 (thinking) | Lec. 2, Lec. 3 (references therein); OM, Ch. 1 |
| Week 3<br>(9/12-9/18) | Tue, 9/13 Probability distributions — a reminder<br>• Different distributions<br>• Priors, Conjugate priors<br>Thu, 9/15 Programming Probabilistically<br>Assignment: module1 - part2<br>• communicating a Bayesian analysis<br>• intro to probabilistic programming | Lec. 4, Lec. 5 (references therein); OM, Ch. 1, Ch. 2 |
| Week 4<br>(9/19-9/25) | Tue, 9/20 Programming Probabilistically<br>• Central limit theorem<br>• Law of large numbers<br>Thu, 9/22 Programming Probabilistically<br>Assignment: module1 - part3<br>• pyMC3 primer<br>• Gaussian inferences | Lec. 6, Lec. 7 (references therein); OM, Ch. 2 |
| Week 5<br>(9/26-10/2) | Tue, 9/27 Bayesian Inference<br>• Continuous quantities<br>• Counting Experiments<br>• Gaussian likelihood<br>Thu, 9/29 Programming Probabilistically<br>Assignment: module1 - part4<br>• Programming Probabilistically | Lec. 8, Lec. 9 (references therein); OM, Ch. 2 |

Assignment module 1 - parts1,2,3,4 due by 9/30 11:59pm ET

## MODULE 2: Models

| | | |
|---|---|---|
| Week 6<br>(10/3-10/9) | Tue, 10/4 Modeling linear regression<br>• Simple linear regression<br>Thu, 10/6 Programming Probabilistically<br>Assignment: module2 - part1<br>• Linear regression<br>• Robust regression | Lec. 10, Lec. 11 (references therein); OM, Ch. 3 |

| | | |
|---|---|---|
| Week 7<br>(10/10-10/16) | Tue, 10/11 Midterm test [tentative]<br>~~Thu, 10/13 NO lecture~~ [fall break 13-17] | |
| Week 8<br>(10/17-10/23) | Tue, 10/18 Modeling linear regression<br>• Hierarchical, Polynomial<br>Thu, 10/20 Generalizing linear models<br>• Logistic regression<br>Assignment: module2 - part2<br>• Logistic regression | Lec. 12, Lec. 13 (references therein); OM, Ch. 4 |
| Week 9<br>(10/24-10/30) | Tue, 10/25 Generalizing linear models<br>• Multiple logistic regression<br>Thu, 10/27 Model comparison<br>• Overfitting and underfitting<br>• Information Criteria<br>Assignment: module2 - part3<br>• MLR<br>• Overfitting/underfitting | Lec. 14, Lec. 15 (references therein); OM, Ch. 4,5 |
| Week 10<br>(10/31-11/6) | Tue, 11/1 Model comparison<br>• Bayes factors<br>Thu, 11/3 Model comparison<br>• WAIC<br>Assignment: module2 - part4<br>• BF | Lec. 16, Lec. 17 (references therein); OM, Ch. 5 |

Assignment module 2 - parts1,2,3 (4 is extra) due by 11/4 11:59pm ET

## MODULE 3: Applications

| | | |
|---|---|---|
| Week 11<br>(11/7-11/13) | ~~Tue, 11/8 NO lecture~~<br>Thu, 11/10 Gaussian processes<br>• Modeling functions<br>• Gaussian process regression<br>~~Assignment: NO~~ | Lec. 18 (references therein); OM, Ch. 7 |
| Week 12<br>(11/14-11/20) | Tue, 11/15 Applications with Gaussian processes<br>Thu, 11/17 Bayesian optimization<br>• popular libraries<br>Assignment: module 3 - part1 | Lec. 19, Lec. 20 (references therein); OM, Ch. 7 |
| Week 13<br>(11/21-11/27)<br>(Thanksgiving) | Tue, 11/22 Bayesian A/B testing<br>~~Thu, 11/24 NO lecture~~<br>Assignment: module 3 - part2 | Lec. 21 (references therein); CDP, Ch. 7 |

| | | |
|---|---|---|
| Week 14<br>(11/28-12/4) | Tue, 11/29 Inference engines<br>• Monte Carlo<br>• Markov chain<br>• MCMC<br>Thu, 12/1 Inference engines<br>• MCMC applications<br>Assignment: module 3 - part3 | Lec. 22, Lec. 23 (references therein); OM, Ch. 8 CDP, Ch. 3 |
| Week 15<br>(12/5-12/11) | Tue, 12/6 Inference engines<br>• MCMC applications<br>• Diagnosing the samples<br>Thu, 12/8 Inference engines<br>• Convergence<br>• Autocorrelation<br>~~Assignment: NO~~ | Lec. 24, Lec. 25 (references therein); OM, Ch. 8 CDP, Ch. 3 |

Assignment module 3 - parts1,2,3 due by 12/9 11:59pm ET

Final project - due by Dec 16, Fri, 2-5pm