CS237 Probability in Computing.
Course Syllabus, Spring 2017

1 Official Description

Introduction to basic probabilistic concepts and methods used in computer science. Develops an understanding of the crucial role played by randomness in computing, both as a powerful tool and as a challenge to confront and analyze. Emphasis on rigorous reasoning, analysis, and algorithmic thinking. (Counts as a Group B course for the CS major, and a background course for the CS minor.)

2 Elaboration

We focus on applications and uses of probability theory in computer science. This includes using probability to analyze data sets, algorithms and data structures, and well as to prove the correctness of algorithms.

Labs. Labs are an essential component of this classes. Each lab will involve a substantial programming component using python and/or R. Some of the labs focus on using Monte-Carlo Simulation to validate the correctness of pencil-and-paper (*i.e.,* "theory") analyses of algorithms and data structures. Some of the labs have a data-science focus, and will teach you to apply statistical approaches learned in class to real data sets.

Active learning: We will sometimes use an active learning approach in this class. This means that we will have mandatory interactive problem-solving sessions where you will work in teams to solve problems. The TF and other course staff will be on hand to help. Attendance at these problem solving sessions is mandatory.

3 Prerequisites

MA123 (or other elementary calculus class) AND CS131. We assume good working knowledge of elementary set theory and counting, and elementary calculus (*i.e.,* integration and differentiation). These topics will be very quickly reviewed in the first weeks of the course, and are also covered in Chapters 1-2 of the Schaum's Outline text, which you can read on your own in case you need a refresher.

4 Course Staff

Instructor:	Professor Alina Ene, `aene@bu.edu`, MCS291 (111 Cummington St.)
Teaching Fellow:	Tomislav Petrovic, `tomislav@bu.edu`, PSY225 (64-86 Cummington St.)
Undergrad Assistant:	Kai Bernadini, `kaidb@bu.edu`
	Alan Burstein, `alanbur@bu.edu`
	Kathryn Quirk, `kquirk@bu.edu`

5 Textbooks

We will use a combination of the following textbooks. The MU text is optional, which means that purchasing it will help you prepare for the course, but is not necessary. The LLM and Schaum's texts are required; note that LLM can be downloaded online at the URL provided, while the Schaum's text costs under $20 at the campus bookstore.

Schaum's: (Required.)
Lipschutz and Lipson. Schaum's Outlines, Probability, 2nd Edition. Probability in Computing. 2000.

LLM: (Required.)
Lehman, Leighton, and Meyer. Mathematics for Computer Science. There are several editions available, we will use the one marked as "Revised Tuesday 4th of April, 2017".
Free download from:
`https://courses.csail.mit.edu/6.042/spring17/mcs.pdf`

MU: (Optional.)
Mitzenmacher and Upfal. Probability in Computing. Cambridge University Press 2005.

6 Course Timing and Structure

Lecture:	Monday, Wednesday, Friday 10:10-11:00 AM in CAS 224
Discussion:	Monday 1:25-2:15 PM (CAS B27), 2:30-3:20 PM (MCS B21), 4:40-5:30 PM (MCS B23)
Instructor's Office Hours:	Usually Monday and Friday 11:00 AM - noon in MCS 291, but see the course calendar each week to be sure!
TF's Office Hours:	Usually Monday 11.45 AM - 1.15 PM (PSY225), Wednesday 1.00-2.30 PM (undergraduate lab), but see the course calendar each week to make sure!
UA's Office Hours:	Thursdays. Usually 5:00-7:00 PM in undergraduate lab, but see the course calendar each week to be sure!

The (gcal) course calendar, with dates of lectures and tests is here:
`https://calendar.google.com/calendar/embed?src=2h56ser839v2eho0k3rjbua440%40group.calendar.google.com&ctz=America/New_York`

Lecture attendance is required. You are responsible for all material covered in lecture. Some lecture material may not appear in the textbooks.

There are three discussion sections. If for some reason you cannot attend the section you have signed up for, arrange with the TF to attend another one. If you have a permanent scheduling conflict, please discuss this with the TF.

Some discussion sessions will be mandatory. These discussion sessions will be in-class problem-solving sessions. You will work in (randomly-selected) teams to solve problems that you must to turn in at the end of the discussion session.

Office hours. We encourage you to come to our office hours. The TF, UAs, and instructor will all hold office hours. **Make sure to check the course calendar each week to check the exact time for office hours!**

The TF and instructor can also answer any questions related to personal issues during her office hour. If you need to talk to one of us in person but absolutely can't make the office hours, please send an email **with at least three options for when you are available**.

7 In-class problems, practice problems, and labs.

This course has three kinds of assignments:

1. **Labs.** There will be five lab assignments, each worth 7% of your final grade, for a total of 35% of your final grade. Labs will involve significant analytic (math) and programming (in python and R) work on an application of probability in computer science. Material for the lab, including a review/introduction to the R and python programming languages, will be covered in the discussion sessions, but the labs themselves should be completed on the student's own time. To get help with the labs, students should (1) attend discussion, or (2) attend the TF or UA office hours, or (3) post to piazza. The instructor *does not* grade, manage or answer questions on the labs. Labs are completely managed by the TF and UAs.

   Submitting labs. Labs are due at **11:59PM** on their assigned dates, submitted as a **PDF** electronically through websubmit. You may choose to hand-write your assignment and then scan it in before submitting, or you may choose to type up the assignment and then convert it to a PDF. A laTex source of the homework will be made available for students who would like to format their work as tex. Whatever format you choose to use, it is crucial that the electronic version of your assignment is legible. **Illegible assignments will not be graded kindly.**

2. **In-class problems during team-problem solving sessions.** The purpose of these problems is to help you to work through the course material in teams, with the help of your TF and UA, and prepare you for the exam and midterm. Some discussion sessions will be designated as "team problem-solving sessions", and students will work in teams to solve a set of problems during the session. Problem-solving participation counts for 7% of the final grade and will be graded mainly on degree of active, prepared participation, rather than on problem-solving success.

   Attendance will be taken at these problem solving sessions. **These sessions are mandatory.** During problem solving sessions, students will be broken up into randomly-selected teams of 7-8 students. One randomly-selected student on each team will be the *paper scribe*. All teams will be assigned the same set of questions. Teams will work through the problems together with their teammates during the session, with the UA and TF (and sometimes also instructor) on hand to help. The paper scribe writes down the team's solution on paper, and submits it to the TF at the end of the session to be graded

   Participation policy. We will have about 5 problem-solving sessions in the course, so participation a given session amounts to about 1% of your final grade. Each student may

skip *one* problem-solving session without penalty. A total of 3 points are available for each in-class problem solving session, and grades are assigned as follows:

- Each student receives 1 points for attending the session.
- If a team demonstrates active participation and submits a solution the represents an honest effort (even if it not correct), each student on the team receives 3 points.
- In the rare case that a team submits a solution that is surprisingly elegant, each student on the team receives 4 points. (This roughly represents a 0.3% bonus on the student's final grade).

3. **Practice problems.** Practice problems will assigned to help you study for the exam and midterm. It is your responsibility to solve these problems; they will be discussed during discussion sessions, but they will not be graded. If you need help with these problems, you can (1) ask for help during the discussion sessions (that are not mandatory in-class problem-solving sessions, or (2) ask the the TF, UA, or instructor during office hours.

8 Course website and communications

We will use Piazza to communicate with you and to post homeworks and labs. Please make sure you are aware of announcements on Piazza; "I didn't get the announcement" won't be an acceptable excuse.

`https://piazza.com/bu/spring2017/cascs237/home`

Feel free to post any general course- or homework-related question to Piazza. You are also welcome to post on Piazza anonymously. **Please make sure all your course-material, homework, and lab related questions on piazza are public posts. The TFs, UAs, and instructors will not answer technical questions that are made as private posts.**

If you have any personal issue (not a technical question about the course material/labs/homeworks) you can contact the TF or instructor using a private post on Piazza. Please avoid emailing the instructors, TFs and UAs.

9 Getting help

If you have questions about the practice problems, the labs, or any of the course material, there are a number of resources available to you.

1. Talk to your classmates about the problems (but, see the class collaboration policy below).
2. Post a question to the piazza board. Please realize that if you are stuck on something, it's very likely that some fraction of the class is also stuck.
3. Attend a discussion section.
4. Come to the TF's office hours. The TF can answer questions about practice problems or labs.
5. Come to the UA's office hours. The UA can answer questions about practice problems or labs.

6. Come to the Instructor's office hours. The instructor can answer questions about practice problems and lecture material.

Do not send personal emails or private Piazza posts to the instructors or TF regarding the course material. These emails will not be answered.

10 Submission policy

Every submitted assignment MUST contain the following:

1. your name,
2. the name of any classmates you discussed the assignment with, or the words "no collaborators"
3. a list of sources you used (textbooks, wikipedia, research papers, etc.) to solve the lab, or the words "no sources", and
4. number of late days used for the current assignment, and total number of late days used up thus far in the semester (include on the current assignment).

We will deduct **at least 15% of the points** from submitted assignments that fail to include the four items above.

11 Late/attendance policy

Labs must be submitted on time. Students will be given a total of 3 "penalty-free" late days to use on any submitted lab at any point during the semester; once these late days have been used up, all late assignments will be given a grade of 0. All submitted labs should include a running count of the number of late days used up on that specific lab, as well as during the whole semester. (To understand how we count days, we'll use an example. Suppose that a lab is due 11:59PM on February 20. If a student submits that lab at 4:32AM on February 21, that student has used up one "full" late day. There are no fractional late days.)

Attendance at team-problem solving sessions is mandatory, but students are given the right to miss a single problem-solving session without penalty. All other missed problem-solving sessions will be assigned a grade of 0.

Please only email the instructor to ask for modifications to this policy if you are in truly extraordinary circumstances.

12 Laptop in lectures policy

You may not use your laptop during lecture unless you are actively taking notes during the lecture. If you are actively taking notes, you will be required to email them to the instructor at the end of each class.

13 Collaboration Policy

Learning probability takes practice. The purpose of the homeworks, labs, and practice problems is to help you learn. The purpose of the midterm and exam is to test this knowledge. For this reason, you are encouraged to collaborate with one another in studying the course notes, textbooks, and lecture material. Collaboration on the homework assignments is permitted and will not reduce your grade, under the following conditions:

1. You must write up your solutions completely on your own, without looking at other people's write-ups.
2. **In your solution to each lab, you must write the names of those with whom you discussed it, and all the references (textbooks, wikipedia, etc) you used to solve the problem. If you did not use any online sources or textbooks or discuss the problem with any of your classmates, please explicitly indicate that on your submitted lab.**
3. You may not work with people outside this class (but come and talk to us if you have a tutor) or get someone else to do it for you. You may not use solutions you obtained from a classmate who has taken this course in previous years.

Deviations from this policy will be taken very seriously.

Note that you are *not* permitted to collaborate on the midterm and exam. The last point is particularly important: if you don't make an honest effort but always get ideas from others, your exam score will reflect it.

It is your responsibility to know and understand the provisions of the CAS Academic Conduct Code.

14 Tests and Grading

There is a midterm and an exam. This course is cumulative; thus each test will cover all the material covered in the course, from the first day of the semester until the date of the test. The midterm will most likely be during week 9 (the week right after Spring break), we will announce the time and date soon. All tests are closed book, but you may bring you own hand-written double-sided "cheat sheet" and old-fashioned calculator (not a phone, or device that can connect to the Internet) with you to each test.

The grading formula is as follows:

Labs	35%
Midterm	20%
Exam	35%
In-class problems	7%
Participation	3%

We reserve the right to deviate from this formula.

Regrading. Grading will be managed by the TF. If you would like to request a re-grade of an exam question or an assignment, be aware that question or assignment will be completely re-graded (and potentially result in a lower grade). Prof. Ene will address re-grading issues only *after* they have been seen by the TF.

15 Topics

The following is a **tentative** list of topics. Note that Prof. Ene will write down the relevant textbook references at the beginning of each lecture and also in the practice problem handouts.

- Review: set theory, functions.
- Combinatorics: counting sequences, subsets, permutations.
- Basic probability: probability spaces, set theory and probability.
- Conditional probability.
- Independence.
- Discrete and continuous random variables and distributions.
- Expectation and variance.
- Parameter estimation and hypothesis testing.
- Concentration inequalities: Markov, Chebyshev, Chernoff.
- Randomized algorithms.