CS237 Probability in Computing.
Course Syllabus, Fall 2014

September 16, 2014

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 CS Background course for the concentration.)

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.

Data-science approach: New for this year, the course will have a data focus. That is, many of the assignments will require you to apply the statistical approaches we learn in class to real data sets, and will include a python programming component. Our intention is to help you start to prepare for jobs in the "data science" area.

Active learning: We will be using an active learning approach in this class. This means that some chalkboard lectures will be replaced with an interactive problem-solving sessions where you will work in teams to solve problems. The TF and I will be on hand to help with the problems, and to give feedback. This is also a great way to meet your classmates, and form study groups.

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 Sharon Goldberg, `goldbe@cs.bu.edu`, MCS135 (111 Cummington St.)
Please make sure that all course-related email has "CS237" in the subject line.

Teaching Fellow: Davide Proserpio, `dproserp@bu.edu`, MCS136 (111 Cummington St.)

5 Textbooks.

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

M&U: Optional. Mitzenmacher and Upfal. Probability in Computing. Cambridge University Press 2005. (Available for purchase in the campus bookstore.)

Schaum's: Required. Lipschutz and Lipson. Schaum's Outlines, Probability, 2nd Edition. Probability in Computing. 2000. (Available for purchase in the campus bookstore.)

L&L: Required. Lehman and Leighton. Mathematics for Computer Science. 2004. Available for download from:
`http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/`
`/6-042j-mathematics-for-computer-science-spring-2005/lecture-notes/`

6 Course Timing and Communications.

Lecture:	Tuesday, Thursday 11:00-12:30 PM in GCB 207
Discussion:	Monday 1:00-2:00 (MCS B31), 3:00-4:00 (PSY B41)
Instructor's Office Hours:	Tuesday 1:00-4:00 PM in MCS 135
TF's Office Hours:	Monday 5:00-6:00 PM in MCS 136, Thursday 5:00-7:00 PM in undergraduate lab.

The (gcal) course calendar, with dates of lectures and tests is here:
`https://www.google.com/calendar/embed?src=ptagucofunkqutn5m653gr6rfo%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. Some lectures will also include in-class practice problem solving sessions.

There are two 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.

This course website is up on blackboard

`http://learn.bu.edu/`

We will also use Piazza to communicate with you. 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/fall2014/cascs237/home`

Feel free to post any general course- or homework-related question to piazza. Questions to piazza will be answered by the TF and instructor every weekday morning except Wednesday. Please only email the TF or Instructor directly if you have a personal issue (not a technical question about the course material) that cannot be posted to the bulletin board.

We encourage you to come to our office hours. If you need to talk to one of us in person but absolutely can't make the office hours, please send us an email with at least three options for when you are available (for Professor Goldberg, please check her calendar at

`http://www.google.com/calendar/embed?src=sharon.goldbe@gmail.com`

before proposing a time).

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

This course has three kinds of assignments:

1. **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 instructor, and prepare you for the exam and midterm. Some lectures will be designated as "team problem-solving sessions", and students will work in teams to solve a set of problems during the session. We believe that the team problem solving activity is a key learning experience; they make lectures more engaging, and students can get immediate feedback from their instructors and classmates, rather than struggling through concepts alone. Problem-solving participation counts for 15% 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 *blackboard scribe* and another randomly-selected student 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 instructor and TF on hand to help. The blackboard scribe is responsible for writing down the team's solution on the team's blackboard, while the paper scribe writes down the team's solution on paper. The paper solution is submitted to the instructor and TF at the end of the session.

   At the end of the session, the instructor and TF will explain the solutions to the in-class problems. We will do this using the solutions written on the blackboard by various teams, discussing both correct solutions to the problem, and where and why certain teams went wrong (because understanding why a solution is wrong is often crucial to understanding the concepts underlying a problem).

   Participation policy. We will have 8-10 problem-solving sessions in the course, so participation a given session amounts to about 1.25% 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 0.4% bonus on the student's final grade).

2. **Practice problem.** 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.

3. **Labs.** There will be five lab assignments, each worth 6% of your final grade, for a total of 30% of your final grade. Labs will involve significant analytic (math) and programming (in python) work on an application of probability in computer science. Material for the lab, including a review/introduction to the python programming language, will be covered in the discussion sessions, but the labs themselves should be completed on the student's own time. The TF will hold some of his office hours at the undergraduate computer lab; students are encouraged to attend these office hours to get help with the labs.

   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.**

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.
5. Come to the Instructor's office hours.

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

8 Submission policy.

Every submitted assignment MUST contain the following:

- your name,
- the name of any classmates you discussed the assignment with, or the words "no collaborators"
- a list of sources you used (textbooks, wikipedia, research papers, etc.) to solve the lab, or the words "no sources", and
- 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 20% of the points** from submitted assignments that fail to include the four items above.

9 Late/attendence 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 Sept 20. If a student submits that lab at 4:32AM on Sept 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.

10 Collaboration Policy.

Learning probability takes practice. The purpose of the homeworks, labs, and practice problems is to help you learn. The purpose of the quiz, 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 quiz, 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.

11 Tests and Grading.

There is a quiz, 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 quiz will be

in class, on Thursday October 9 11:00-12:30PM

and the midterm will be

in class, on Tuesday November 4 11:00-12:30PM

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	30%
In-class problems	10%
Quiz	10%
Midterm	20%
Exam	30%

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. Goldberg will address re-grading issues only *after* they have been seen by the TF.

12 Topics.

The following is a **tentative** list of topics. Order subject to change. Note that Prof. Goldberg will write down the relevant textbook references at the beginning of each lecture.

Week	**Topic**	**Lab**
Week 1-2	Introduction. Set theory, counting, calculus review.	
Week 3-4	Basic probability, events, random variables.	Lab 1: Intro
Week 5	Conditional probability	
Week 6-7	Quiz. Independence, universal hashing.	Lab 2: Bloom filters
Week 8-9	Distributions (Binomial, bernoulli, geometric, gaussian), expected value.	Lab 3: Distributions
Week 10	Midterm week	
Week 11	Guest lectures (instructor traveling): Variance, sampling, confidence	
Week 12-13	Gaussian distribution, statistical significance, regression	Lab 4: Regression
Week 13-15	Inequalities (Markov, Chernoff, Chebyshev), other topics.	Lab 5: TBD