STA237: Probability, Statistics and Data Analysis I

Shahriar Shams

Fall 2019 (last updated on Oct 02)

1 Course Description

STA237H is an introduction to probability using simulation and mathematical frameworks, with emphasis on the probability needed for more advanced study in statistical practice. Topics covered include probability spaces, random variables, discrete and continuous probability distributions, probability mass, density, and distribution functions, expectation and variance, independence, conditional probability, the law of large numbers, the central limit theorem, sampling distributions. Computer simulation will be taught and used extensively for calculations and to guide the theoretical development.

Contents, emphasis, etc. of the course is defined by means of the lecture materials - *not only the texts*. Table 1 shows the tentative lecture guide. Lecture slides will be uploaded every week. However, they are just rough, point-form notes, with no guarantee of completeness or accuracy. They should in no way be regarded as a substitute for attending the lectures or for doing the weekly non-credit assignments.

Important announcements, problem sets, additional examples, and other course info will be posted on the course web page on Quercus. Check it regularly.

2 Course Schedule

- **Lectures:**

  - L0101- Room # NF-003, Mon (9 - 11am) and Room # AH-100, Wed (9 - 11am)
  - L0201- Room # ES-1050, Mon (1 - 3pm) and Room # BA-1160, Thur (1-3pm)
  - L0301- Room # NF-003, Tue (3-5pm) and Thur (3-5pm)

- **Instructor:** Shahriar Shams, PhD in Biostatistics candidate, Dalla Lana School of Public Health, University of Toronto.

- **email:** shahriar.shams@mail.utoronto.ca (Please write "STA237" in the subject of your email. PLEASE!)

- **Office hours:** Mon 3.30pm - 5pm and Tue 1.00pm - 2.30pm, Room # HS 376

3   Textbooks

A Modern Introduction to Probability and Statistics: Understanding Why and How, 2005, by F.M Dekking, C. Kraaikamp, H.P Lopuhaa and L.E. Meester

E-book is available through the University of Toronto libraries website (https://search.library.utoronto.ca/details?7997563)

4   Computing

Statistical software **R** will be used extensively. Students will learn solving probability problems using simulations in R. *No previous exposure is expected* and R will be introduced starting from the basics.

Any code used in the lectures will be available on the course web-page for students to practice at their own time. There are two software (R and R-studio) that need to be installed. First you will have to install R. Here are the links to download this open source software:

- **for windows:** https://cran.r-project.org/bin/windows/base/

- **for mac:** https://cran.r-project.org/bin/macosx/

After you have installed R, you can install R-studio (an IDE for R) by downloading the version specific to your operating system from https://www.rstudio.com/products/rstudio/download/#download

5   Non-credit assignments

Every week after the lecture a set of exercises will be provided. These assignments are **not for credit**. They are only meant to give students opportunities to learn the materials and prepare themselves for the quizzes/tests and exam. The instructor will help you solve few of the problems from these assignments during the weekly office hours. And the TAs will also help you solve few during their office hours.

6   Assignment for credit

Sometime during the middle/end of the term, you will be doing an assignment (worth 8%) which will help you to understand few of the theories that we will be learning in this course. You are allowed to work in groups (maximum 5 members in one group). You will be asked to solve few problems using R and few analytically.

7   Online quizzes

There will be 4 online quizzes (worth 3% each) administered through Quercus. The timings are in Table 1. The quizzes will be timed and will be open on Fridays of the corresponding week.

- Quiz-1: will cover Week-1 and Week-2 materials

- Quiz-2: will cover Week-3 and Week-4 materials

- Quiz-3: will cover Week-7 and Week-8 materials

- Quiz-4: will cover Week-9 and Week-10 materials

8 Evaluation

- **Quizzes + Assignment:** $12\% + 8\% = 20\%$

- **Mid-term test:** $35\% \begin{cases} L0101 - Oct\, 23, 9 - 11am \\ L0201 - Oct\, 24, 1 - 3pm \\ L0301 - Oct\, 24, 3 - 5pm \end{cases}$

- **Final:** 45% (will cover everything taught in the course)

9 Missed test

There are **NO make-up tests**. If a test is missed for a valid reason, you must email the instructor within 24 hours, and submit appropriate documentations (doctor's notes, prescriptions, money receipts) to the course instructor within 2 business days of the test. Print your name, student number, course number, and date on the submitted documents. If documents are not received in time, your test mark will be ZERO. If a test is missed for a valid reason, its weight will be shifted to the final exam.

10 Calculators

Hand calculators will be needed for this course. **NO phone calculators** are allowed in tests/exams/quizzes. Any regular one with a square root and logarithmic function will do.

11 Communicating with your Instructor

Please do not email your instructor asking questions like "how to do problem 10.3.4?", "when is the midterm?", "how to submit the assignment?". Emails like these will be ignored. If you need help with any problem you can either talk to your instructor right after the class if you expect the answer to be brief or come during the office hour. If you have any question/concern that you don't want to ask in front of others you can email the instructor to set up a one-on-one meeting.

12   Academic Integrity

You are responsible for knowing the content of the University of Toronto's Code of Behaviour on Academic Matters (`https://www.academicintegrity.utoronto.ca/perils-and-pitfalls/`).

Table 1: STA237, Fall 2019 **tentative** lecture guide

Week	Dates	Readings	Quiz
0	Sep 05 - 06	Introduction to R	-
1	Sep 09 - 13	Ch2: Outcomes, Events and Probability	-
2	Sep 16 - 20	Ch3: Conditional Probability & Independence	-
3	Sep 23 - 27	Ch4: Discrete Random Variable	1
4	Sep 30 - Oct 04	Ch5: Continuous Random variable	-
5	Oct 07 - 11	Ch7: Expectation and variance	2
6	Oct 15 - 18	Ch6: Simulating Random variables and Ch8: Variable transformation	-
7	Oct 21 - 25	Ch9: Joint distribution and Independence	week of Midterm
8	Oct 28 - Nov 01	Ch10: Covariance and Correlation	-
-	Nov 04 - 08	Reading week	-
9	Nov 11 - 15	Ch11: More Computation with Random Variables	3
10	Nov 18 - 22	Ch13: Law of Large Number	-
11	Nov 25 - 29	Ch14: Central Limit Theorem	4
12	Dec 02 - 04	Other topics	-