STA237H1: Probability, Statistics and Data Analysis I

Shahriar Shams

Fall 2020 (last updated on Sept 09)

All listed times in this document are in local Toronto time. If you are in a different time zone, you are responsible for any time conversions.

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

Prerequisite: (MAT135H1, MAT136H1)/ MAT137Y1/
MAT157Y1/(MATA30H3, MATA36H3)/(MATA31H3, MATA37H3)/
MAT135Y5/MAT137Y5/MAT157Y5

Exclusion: STA247H1, STA255H1, STA257H1, ECO227Y1, STAB52H3, STA256H5, ECO227Y5

Breadth Requirements: The Physical and Mathematical Universes (5)

2   Course Schedule

- **Lectures:**

  - We will use a mix of synchronous learning and asynchronous learning.
  - Lecture slides along with pre-recorded voice overs will be uploaded to quercus (at least) 48 hours before the scheduled lecture times.
  - Lecture slides and pre-recorded voice overs will be the same for the two sections.
  - We will use the scheduled lecture times (Wednesdays 12-2pm for LEC0101 and Thursdays 8-10pm for LEC5101) for live question-answer(QA) sessions.
  - The live QA sessions will be on BBcolab and run approximately 1-1.5 hours on each day.
  - Live QA sessions are section specific, though they will cover the exact same materials and I will repeat mostly the same problem sets. So students are expected to join one of them.
  - Live QA sessions will be recorded and made available to the students.

  For example: The first recorded-lecture on "Outcomes, Events and Probability" will be uploaded by Sept 13th. On Sept 16th, at 12pm(Toronto time) we will have our live QA session based on lecture-1 for LEC0101. On Sept 17th at 8pm(Toronto time) we will have our live QA session based on lecture-1 for LEC5101.

- **Instructor: Shahriar Shams**,
  PhD in Biostatistics candidate, Dalla Lana School of Public Health,
  Assistant Professor(teaching stream), Department of Statistical Sciences,
  University of Toronto.

- **Email:** shahriar.shams@mail.utoronto.ca (Please add "STA237" at the beginning of the subject of your email. PLEASE!)

- **Office hours:** On BB colab, time to be announced later.

3   Textbooks

1. 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)

Students may use this following text book for additional practice problems:

2. Probability and Statistics: The Science of Uncertainty, Second Edition, by Michael J. Evans and Jeffrey S. Rosenthal

Available online on the web-page of Professors Evans and Rosenthal
http://www.utstat.toronto.edu/mikevans/jeffrosenthal/

4   Quizzes

There will be five quizzes in total, each covering the materials covered in the previous two lectures. Your lowest mark will be dropped and the best four will worth 10% each. Quizzes will be administered using Quercus. Crowdmark will be used for quiz submission.

Quizzes are section specific. Visit your ACORN page to know which section you are enrolled in and write your quizzes with that section only. Students writing their quiz with the wrong section will automatically get a zero in that quiz. Please note that Quercus doesn't show which section you are enrolled in. It will show LEC0101 on the top of page for all the students.

	LEC0101	LEC5101	Coverage
Quiz-1	Sep 30 (12-12.45pm)	Oct 01 (8-8.45pm)	Lectures 1 and 2
Quiz-2	Oct 14 (12-12.45pm)	Oct 15 (8-8.45pm)	Lectures 3 and 4
Quiz-3	Oct 28 (12-12.45pm)	Oct 29 (8-8.45pm)	Lectures 5 and 6
Quiz-4	Nov 25 (12-12.45pm)	Nov 26 (8-8.45pm)	Lectures 7 and 8
Quiz-5	Dec 09 (12-12.45pm)	Dec 10 (8-8.45pm)	Lectures 9 and 10

Table 1: Time table and coverage for Quizzes

5   Assignment for credit

At the middle/end of the term, you will do an assignment which will help you to understand few of the theories that we will be using in this course. The assignment will require some hand calculations and some coding in R. It will worth 15%. The assignment will be released around mid November (tentatively). Clear instructions will be given on how to complete and submit your work.

6   Mid-term

Mid-term will be held during the first week of November.

- For L0101: Nov 04, 12-2pm

- For L5101: Nov 05, 8-10pm

It will cover lectures 1 to 6. Students will be able to access the questions through Quercus. Crowdmark will be used for submission.

- Students will be required to write complete answers on paper (or using electronic devices)

- If written using pen and paper, students will be required to take pictures of their complete answers and upload them on crowdmark.

- If written using electronic devices, students can upload their answers as screenshots or saved images.

- More instructions on how to upload documents to crowdmark will be given later.

- Midterms are section specific as well. Visit your ACORN page to know which section you are enrolled in and write your midterm with that section only. Students writing their midterm with the wrong section will automatically get a zero in the midterm.

7 Evaluation

- **Quizzes:** 10% * 4 = 40%

- **Mid-term:** 20%

- **Assignment for credit:** 15%

- **Final assessment:** 25% (will cover everything taught in the course, date and time will be fixed by faculty of Arts and Science and will be announced later)

8 Missed assessment

There are NO make-up assessments of any form in this course.

- Taking the final assessment and submitting the assignment is **mandatory** for every student in order to pass this course.

- For quizzes, the lowest mark will be dropped. So missing one quiz will not affect your grade. If you miss two quizzes, your other three quiz scores will be counted and the missing 10% will be distributed to all your assessments(quiz, mid, assignment and final). If you miss more than two quizzes, you will get a zero for your third and onward(fourth or fifth) missed quiz irrespective of reason missed.

- If you miss the midterm, your midterm weight will be distributed among the other assessments.

- Students are not required to submit any doctor's note for missing any assessment.

9 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.

10   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 with questions like these will be ignored. Otherwise, students should expect a reply within 48 hours. Course related questions should be posted on the discussion board on Piazza.

11   Accessibility needs

The University of Toronto is committed to accessibility. If you require accommodations for a disability, or have any accessibility concerns about the course, the classroom, or course materials, please contact Accessibility Services as soon as possible at accessibility.services@utoronto.ca or http://www.accessibility.utoronto.ca

12   Academic Integrity

Academic integrity is fundamental to learning and scholarship at the University of Toronto. Participating honestly, respectfully, responsibly, and fairly in this academic community ensures that the University of Toronto degree that you earn will be valued as a true indication of your individual academic achievement, and will continue to receive the respect and recognition it deserves. Familiarize yourself with the University of Toronto's Code of Behaviour on Academic Matters available at (https://www.academicintegrity.utoronto.ca/perils-and-pitfalls/)

Students are not allowed to share quizzes, assignment, tests and exam questions with anyone (not even with other students taking this course) at any point in any platform. Sharing questions and submitting works completed by someone else is a huge academic offence. Please stay away from this type of behaviors.

13   Copyright

Students agree to the following terms:

- Course materials (i.e. slides, recordings, assessment questions) are your instructor's intellectual property and have been created by your instructor for students' personal use and under no circumstances should be shared, posted or distributed anywhere.

- Non-compliance with these terms violates an instructor's intellectual property rights and the Canadian Copyright Act. Students violating this agreement will be subject to disciplinary actions under the Code of Student Conduct.

Table 2: STA237, Fall 2020 **tentative** lecture guide

Lecture	Week of	Readings	Quiz
1	Sep 14	Ch2: Outcomes, Events and Probability	-
2	Sep 21	Ch3: Conditional Probability & Independence	-
3	Sep 28	Ch4: Discrete Random Variable	1
4	Oct 05	Ch5: Continuous Random variable	-
5	Oct 12	Ch7: Expectation and variance	2
6	Oct 19	Ch6: Simulating Random variables and Ch8: Variable transformation	-
7	Oct 26	Ch9: Joint distribution and Independence	3
-	Nov 02	**Midterm test**, no lecture	-
-	Nov 09	**Reading week**, no lecture, no office hours	-
8	Nov 16	Ch10: Covariance and Correlation	-
9	Nov 23	Ch11: More Computation with Random Variables	4
10	Nov 30	Ch13: Law of Large Number	-
11	Dec 07	Ch14: Central Limit Theorem	5